SCMS SCHOOL OF ENGINEERING & TECHNOLOGY
PUBLICATION DETAILS 2022

Sl No:	Name	First Author	Second Author	Third Author	Fourth Author	INDEXING
1	Anandhi V	PEC2202				SCI
2	Dr.ParvathyM	PEC2201				SCOPUS
3	Praveena SKammath	PEC2203				
4	Dr.Varun Menon	PCS2204	PCS2205,PCS2206 (6TH)	PEC2202	PCS2201	SCI
		PCS2208	PCS2207(8TH), PCS2209(7TH)	PCS2212(5TH)	PCS2202(5TH), PCS2203(5th)	SCI
					PCS2210,PCS2214, PCS2211,PCS2213	SCI
5	Dr.Vidya Chandran		PME2201			SCI
						SCI
6	Dr Raghav G R			PME2203		SCI
7	DrJayanand B		PEE2201			SCI
8	Geethu R Babu	PCE2201				UGC CARE
9	Roshini K R	PCE2202				UGC CARE
10	Teenu Sycriac		PCE2203			UGC CARE
11	Dr.Santhosh Kumar	PCE2204				UGC CARE
12	Dr.Sanju Sreedharan		PCE2204			UGC CARE
13	Dr.Rahul R Pai	PCE2205				SCOPUS
14	Dr.Rathish Menon		PCE2206			SCOPUS
15	Dr.Praseeja A V	PCE2207				SCI
16	Dr.Lakshmipriya	PCE2208				SCI
17	Anoop M S	PAU2201				SCOPUS
18	Akhil Baby	PBSH2201,PBSH2202				UGC CARE
Total Publication for the calender year 2022						**30**

UAV-Enabled Non-Orthogonal Multiple Access Networks for Ground-Air-Ground Communications

Qunshu Wang, *Student Member, IEEE*, Xingwang Li, *Senior Member, IEEE*, Surbhi Bhatia, Yuanwei Liu, *Senior Member, IEEE*, Linss T. Alex, *Senior Member, IEEE*, Sunder Ali Khowaja, and Varun G. Menon, *Senior Member, IEEE*

Abstract—Both unmanned aerial vehicle (UAV) and non-orthogonal multiple access (NOMA) have gradually become promising technologies for the fifth generation (5G) driven green Internet-of-Things (IoT) networks on account of their unique advantages of massive connections, higher spectral efficiency and flexibility. Motivated by this, we propose a 3-hop NOMA UAV-aided green communication network framework, where UAVs serve as aerial relays to support two groups of ground users. A stochastic geometry approach is invoked to model the spatial positions of the two group users. Under the realistic assumption, imperfect successive interference cancelation (ipSIC) is considered. To evaluate the performance of the proposed framework, theoretical expressions are derived to facilitate the outage performance evaluation of the far user (FU) and the near user (NU). Moreover, the asymptotic behaviors for the outage probability (OP) of both the FU and the NU in the high signal-to-noise ratio (SNR) regime are explored by obtaining diversity orders. Finally, the system throughputs under the delay-limited transmission mode are investigated. Numerical results confirm that: 1) For uplink transmission, there exist outage floors for the OP of both ipSIC and perfect SIC (pSIC) due to interference from the NU; 2) For downlink transmission, an outage floor exists for the OP of the NU under the condition of ipSIC; 3) For uplink NOMA/orthogonal multiple access (OMA) transmission, the outage performances of both the FU and the NU with NOMA outperform OMA in the low SNRs, while OMA has better performance in the high SNR regime; 4) For downlink NOMA/OMA, the outage performances of both the FU and the NU under pSIC outperform OMA.

Index Terms—Non-orthogonal multiple access, unmanned aerial vehicles, stochastic geometry, imperfect successive interference cancelation.

Qunshu Wang and Xingwang Li are with the School of Physics and Electronic Information Engineering, Henan Polytechnic University, Jiaozuo 454000, China (e-mail: 15333766153@163.com; lixingwangbupt@gmail.com).

Surbhi Bhatia is with the College of Computer Science and Information Technology, King Faisal University, Al Hofuf 36362, Saudi Arabia (e-mail: sbhatia@kfu.edu.sa).

Yuanwei Liu is with the School of Electronic Engineering and Computer Science, Queen Mary University of London, London E1 4NS, U.K. (e-mail: yuanwei.liu@qmul.ac.uk).

Linss T. Alex is with the Department of Electrical and Electronics Engineering, Albertian Institute of Science and Technology, Kochi 682022, India (e-mail: linsstalex@gmail.com).

Sunder Ali Khowaja is with the Department of Telecommunication Engineering, Faculty of Engineering and Technology, University of Sindh, Jamshoro 76080, Pakistan (e-mail: sandar.ali@usindh.edu.pk).

Varun G. Menon is with the Department of Computer Science and Engineering, SCMS School of Engineering and Technology, Ernakulam 683576, India (e-mail: varunmenon@ieee.org).

Digital Object Identifier 10.1109/TGCN.2022.3152601

I. INTRODUCTION

RECENTLY, with the various applications for the green Internet-of-Things (IoT) driven by the fifth generation (5G), flexibility plays a key role for the future wireless networks [2], [3]. In this respect, unmanned aerial vehicle (UAV) communication has attracted a great number of interests in green wireless communication because of its high maneuverability, flexible deployment and cost-effective [4], [5]. In contrast to the conventional terrestrial base stations (BSs) or relays, UAVs acting as BSs or relays have the attractive advantage of rapid deployment in geographical space [6]. Therefore, employing UAVs as aerial BSs or relays is especially useful in circumstances where nodes are separated by long distance or the communication links between BSs and the access point are deteriorated by dense surroundings or buildings (e.g., large building, forests, etc.) [7]. In most of the existing works, UAVs work as aerial BSs or relays to support on-demand wireless connections during temporary events or disaster response (e.g., earthquake, fire fighting) [8]. An additional advantage of using UAVs is that their altitude enables them to establish line-of-sight (LoS) links and to overcome shadowing effect. To this end, it can improve spectrum efficiency due to LoS connections between UAVs and ground users [9]. The authors of [10] investigated the mobility and efficient deployment of multiple UAVs for supporting uplink IoT networks. In [11], the overall sum rate and average coverage probability of UAV-assisted underlaid device-to-device communication networks were investigated, in which two scenarios based the status of UAV were considered. Considering imperfect radio frequency (RF) front-ends, the authors in [12] proposed a multi-way UAV-NOMA network and derived the approximate expressions of the achievable sum-rate.

Although employing UAVs as flying BSs or relays in green communication networks can achieve promising performance, some challenges should be considered and addressed. For example, massive connectivity is expected to support immense amounts of devices with limited spectrum resources, and the

quality of service of users with worse channel condition need to be further improved [13]. To this end, non-orthogonal multiple access (NOMA) can be exploited as a promising technology to help UAV-aided networks to achieve enhanced performance. The superiority of NOMA is that it can improve spectrum efficiency which is achieved by serving multiple users in the same resources, such as frequency, time and code. In NOMA, two transmit and receive algorithms are adopted with the name of superposition coding (SC) and successive interference cancelation (SIC) [14]–[17]. So far, there exist some research contributions related to NOMA, see [18]–[23] and the references therein. In [18], the authors designed three user selection schemes for a new wireless-powered NOMA communication network and derived the OP and system throughput of the proposed schemes. The sum rate performance of the downlink NOMA networks was studied in [19]. Liu *et al.* in [20] analyzed the physical layer security of NOMA users for two different scenarios: single-antenna and multiple-antenna. In [21], Men *et al.* investigated a downlink NOMA network with a cooperative relay and derived the exact expressions and simple bounds of the outage probability (OP). To further improve spectral efficiency, authors in [22] derived the theoretical results of the OP and the ergodic sum rate of full-duplex cooperative NOMA relay networks. In [23], the authors considered a downlink multiple-input-multiple-out NOMA scenario, and proposed three sub-optimal schemes to balance the fairness and complexity of the far user (FU).

A. Existing Research Contributions on NOMA and UAV

Although the aforementioned related treatises provided significant insights of UAV and NOMA, it is not well understood for the topic of multiple UAVs-aided NOMA networks. Motivated by this, to achieve higher spectral efficiency and ubiquitous coverage, the combination of both UAV and NOMA is utmost important. Most recently, both uplink and downlink transmissions have explored the using of NOMA to support UAV-aided networks. Here we review the most recent and relevant works for the uplink and downlink UAV-aided NOMA networks. To alleviate the uplink interference of UAV-assisted cellular communication, the authors in [24] presented a cooperative NOMA design and jointly optimized the transmit power allocations and UAV's uplink rate. The authors in [25] compared outage performance of NOMA and orthogonal multiple access (OMA) with different duplex modes in the context of UAV communications and the simulated results showed that full-duplex NOMA outperforms the other two. A hybrid satellite UAV-NOMA network in [9] was presented and the accurate and approximate results for the OP considering channel estimation error were derived. In [26], the energy efficiency optimization in a multi-UAVs and multi-devices network was investigated via Lagrangian optimization and gradient-descent methods.

B. Motivation and Contributions

Most of the research works have been investigated either for the downlink or uplink scenario in the UAV-assisted NOMA systems with perfect SIC (pSIC). In particular, there is no comprehensive work to precisely analyze the system performance

in the uplink and downlink UAV-assisted NOMA systems and their respective impact on the outage performance and diversity orders. To this end, this paper considers the use of UAVs in the obstacle region, where it is non-trivial to establish a stable data link with only one deployed UAV because of obstacles. Therefore, an efficient solution is to deploy multiple UAVs to maintain uninterrupted communication between users.

This study proposes a UAV-enabled NOMA network, where two UAVs are deployed to communicate with ground users who are isolated by the obstacles. Specially, each UAV is used to establish communication between two group NOMA users. The whole communication process consists of uplink NOMA communication, point to point communication and downlink NOMA communication. We consider a more realistic assumption of imperfect SIC (ipSIC) at the receiver. As a benchmark for comparison, pSIC is also taken into account. To guarantee the coverage of all served users, we suppose that the UAVs stay at a fixed altitude h above the ground. The key contributions of our article are provided as below:

- We propose a 3-hop UAV-enabled NOMA system, where one group users aim to communicate the other group users with the aid of two decode-and-forward (DF) UAVs. Stochastic geometry approach is invoked to model the spatial positions of ground users on a 2-D plane of the two groups. To ensure the coverage, the UAVs move along a horizontal plane with a fixed height.
- We analyze the outage performance of the FU and near user (NU) for both uplink and downlink NOMA. More specifically, theoretical results for the OP of the FU and NU of the 3-hop NOMA assisted UAV networks are derived.
- We also obtain the asymptotic results for the OPs of the FU and NU in the high signal-to-noise ratio (SNR) regime and discuss the diversity order. We conclude that: 1) For uplink NOMA, OPs for the uplink converge to outage floors for the FU and NU in the high SNRs for both pSIC and ipSIC due to interference from the NU, which results in zero diversity orders; 2) For downlink NOMA, the asymptotic OP for the NU is a fixed constant and the diversity order of the NU for the ipSIC is zero.
- We finally discuss the system throughput under the delay-limited transmission mode. The results show that at high SNRs, NOMA under pSIC and ipSIC cases achieves a lower throughput than OMA due to the significant interference from the FU and NU in the uplink and vice versa.

C. Organization

The remain sections of our article are arranged as follows. In Section II, we describe the system and channel models for the considered UAV-aided NOMA networks. In Section III, we introduce and analyze the OP performance of the FU and NU in the uplink NOMA, downlink NOMA, as well as the considered system. Section IV presents the diversity orders of the FU and NU in uplink and downlink NOMA. Section V investigates the system throughput under the conditions of the ipSIC and pSIC. In Section VI, we provide the numerical simulated analyses to evaluate the performance of the proposed networks. Finally, we conclude the paper in Section VII.
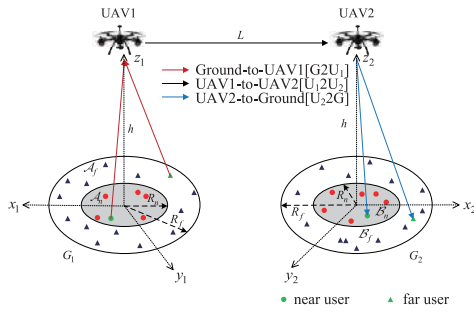
UAV1

z_1

L

UAV2

z_2

→ Ground-to-UAV1[G2U₁]
→ UAV1-to-UAV2[U₁2U₂]
→ UAV2-to-Ground[U₂2G]

h

$\mathcal{A}_f$

$\mathcal{A}_n$

R_n

R_f

x_1

G_1

y_1

h

R_f

R_n

$\mathcal{B}_n$

$\mathcal{B}_f$

x_2

G_2

y_2

• near user   ▲ far user

Fig. 1.   System model.

II. SYSTEM MODEL AND CHANNEL DESCRIPTION

A. System Model

This paper considers a UAV-enabled NOMA network with two UAVs as well as two group users G_1 and G_2, as shown in Fig. 1.[1] UAVs are deployed as DF relays that communicate with two group NOMA users G_1 and G_2. They are located on the either side of the high-rise building, and always remain inside the coverage regions of UAV1 and UAV2. We model the spatial position of the users as a homogeneous Poisson point process (PPP) with density λ. In addition, we have the assumption that all nodes are half duplex devices with the single antenna. Since it is not easy to employ SIC to all users for the NOMA system, we adopt two paired users with NOMA as in [27]. The decoding process of paired NOMA users doesn't require instantaneous channel state information (CSI), just need to grasp the distance between them and UAV. For this insight, it is assumed that UAVs coverage regions are circular and partitioned into two areas with radius R_n and R_f (assuming $R_f > R_n$), denoted by $\{\mathcal{A}_n, \mathcal{A}_f\}$ and $\{\mathcal{B}_n, \mathcal{B}_f\}$, respectively. Leveraging random user pairing principle, one user in $\mathcal{A}_n$ is randomly paired with the other user in $\mathcal{A}_f$ for the convenience of NOMA transmission. We denote that the paired NOMA users are D_{f1} and D_{n1}, which act as the FU and NU, respectively. Similarly, we denote D_{f2} and D_{n2} as the FU and NU in G_2, respectively.

The entire communication is partitioned into three phases: 1) Ground-to-UAV1 ($G2U_1$); 2) UAV1-to-UAV2 (U_12U_2); 3) UAV2-to-Ground (U_22G).

1) $G2U_1$: According to uplink NOMA protocol [28], [29], the paired NOMA users $D_{i1}(i = n, f)$ in G_1 transmit signal s_i to the UAV1 with $E\{|s_i|^2\} = 1$, where P_s is the total transmit power and a_{i1} denotes the power allocation coefficient for D_{i1} with $a_{f1} + a_{n1} = 1$.[2] Thus, the received signal at the UAV1 is given by

$$y_1 = \sum_{i=n,f} g_{i1}\sqrt{a_{i1}P_s}s_i + z_1, \qquad (1)$$

where g_{i1} represents the channel gain coefficient between UAV1 and user D_{i1}, $z_1 \sim \mathcal{CN}(0, \sigma^2)$ is the additive white Gaussian noise (AWGN). In uplink NOMA, the decoding

---

order on UAV1 is assumed to be from D_{n1} with strong channel conditions to D_{f1} with poor channel conditions (that is, D_{n1} is NU and D_{f1} is FU); otherwise, a large amount of power needs to be consumed at D_{f1} to compensate for path loss [28], [31]. Since the signal s_f is treated as a interference when UAV1 decodes s_n, D_{n1} experiences interference from D_{f1}. Once UAV1 successfully decodes the signal of D_{n1}, it will decode the signal s_f of D_{f1}. Therefore, the achievable rate of D_{n1} in uplink NOMA is given by

$$r_n^{\mathrm{UL}} = \log_2\left(1 + \frac{|g_{n1}|^2 a_{n1}\rho}{|g_{f1}|^2 a_{f1}\rho + 1}\right), \qquad (2)$$

where $\rho = P_s/\sigma^2$ is the average transmit SNR.

If $\frac{1}{3}r_n^{\mathrm{UL}} > R_2$, the signal s_n can be successfully decoded and then is subtracted at UAV1, otherwise the SIC can not be executed. It's worth noting that R_2 denotes a target date rate of D_{n1}, and multiplying 1/3 is because communication is completed in three time slots. In practice, SIC decoding is not always perfect in the first step, thus SIC error will be propagated to the next decoding. Therefore, considering ipSIC, the achievable rate of D_{f1} in uplink NOMA is given by

$$r_f^{\mathrm{UL}} = \log_2\left(1 + \frac{|g_{f1}|^2 a_{f1}\rho}{\xi_{n1}|g_{n1}|^2 a_{n1}\rho + 1}\right), \qquad (3)$$

where the parameter $\xi_{n1} \in [0, 1]$ represents the level of residual interference at UAV1. Particularly, $\xi_{n1} = 0$ and $\xi_{n1} = 1$ represent pSIC and ipSIC, respectively.

2) U_12U_2: The UAV1 sends the compound signals, $\sum_{i=n,f}\sqrt{b_i P_r}s_i$, to the UAV2, where b_i is defined as the power allocation factor at UAV1 for i-th user, P_r has been defined as the maximum transmit power from UAV1. For convenience, we assume $P_s = P_r$. Thus, the received signal at UAV2 is given as

$$y_{uu} = g_{uu}\sum_{i=n,f}\sqrt{b_i P_r}s_i + z_{uu}, \qquad (4)$$

where subscript *uu* represents the date transmission of UAV1-to-UAV2. $g_{uu} = \frac{\bar{g}_{uu}}{\sqrt{1+L^\alpha}}$ is the channel coefficient between UAV1 and UAV2, where $\bar{g}_{uu}$ is the multi-path fading, which follows Nakagami-*m* distribution, L denotes the horizontal distance from UAV1 to UAV2, and α denotes the path loss exponent. $z_{uu} \sim \mathcal{CN}(0, \sigma^2)$ denotes the AWGN. We assume $b_f > b_n$, and $b_f + b_n = 1$. According to the SIC process, UAV2 decodes the signal s_f by regarding the signal s_n as interference. Therefore, the achievable rate of s_f is given by

$$r_f = \log_2\left(1 + \frac{|g_{uu}|^2 b_f\rho}{|g_{uu}|^2 b_n\rho + 1}\right). \qquad (5)$$

If the signal s_f is successfully decoded in the first stage, UAV2 will remove s_f from the received superposed signal y_{uu}. Then, UAV2 will further decode the signal s_n. Similarly, SIC decoding may not be perfect in the first stage, so decoding error will be propagated to the next decoding process. Therefore, the achievable rate of s_n with ipSIC is given by

$$r_n = \log_2\left(1 + \frac{|g_{uu}|^2 b_n\rho}{\xi_{uu}|g_{uu}|^2 b_f\rho + 1}\right), \qquad (6)$$

---

[1]Although our work focuses on 3-hop UAV-aided NOMA network, this study can be extended to multi-hop scenario as well.

[2]Some power optimization schemes are capable of further improving system performance [30], though this is part of our ongoing work.

where the parameter $\xi_{uu} \in [0,1]$ denotes the level of residual interference at UAV2. Particularly, $\xi_{uu} = 0$ and $\xi_{uu} = 1$ represent pSIC and ipSIC, respectively.

3) $U_2 2G$: For downlink NOMA, UAV2 transmits the signals, $\sum_{i=n,f} \sqrt{a_{i2} P_r} s_i$, to D_{i2}, where a_{i2} represents the power allocation factor for D_{i2}, s_i is the desired signal of the D_{i2}.[3] Based on the protocol of downlink NOMA [19], we assume $a_{f2} > a_{n2}$ and $a_{f2} + a_{n2} = 1$. D_{i2} receives their desired signals along with the interfering signals from other user. Therefore, the received signals at D_{f2} and D_{n2} are given by

$$y_{i2} = g_{i2} \sum_{i=n,f} \sqrt{a_{i2} P_r} s_i + z_{i2}, i \in \{n,f\}, \qquad (7)$$

where g_{i2} is the channel coefficient between D_{i2} and UAV2, $z_{i2} \sim \mathcal{CN}(0, \sigma^2)$ is the AWGN at D_{i2}.

Leveraging the NOMA protocol, D_{f2} first decodes its own signal regarding D_{n2}'s signal as noise. As a result, the achievable rate of D_{f2} to decode its own signal s_f in downlink NOMA is given by

$$r_f^{\mathrm{DL}} = \log_2 \left(1 + \frac{|g_{f2}|^2 a_{f2} \rho}{|g_{f2}|^2 a_{n2} \rho + 1} \right). \qquad (8)$$

At D_{n2}, D_{n2} first decodes the signal of D_{f2}, s_f, and subtracts it from y_{n2}, and then decodes its own signal s_n. Thus, the achievable rate of D_{n2} to decode D_{f2}'s signal in downlink NOMA is given by

$$r_{n \to f}^{\mathrm{DL}} = \log_2 \left(1 + \frac{|g_{n2}|^2 a_{f2} \rho}{|g_{n2}|^2 a_{n2} \rho + 1} \right). \qquad (9)$$

Similarly, if $\frac{1}{3} r_{n \to f}^{\mathrm{DL}} > R_1$, the signal s_f can be successfully decoded and then is subtracted at D_{n2}, otherwise the SIC can not be executed. R_1 denotes a target date rate of D_{f2}. Thus, the achievable rate of D_{n2} to decode its own signal by considering ipSIC is given by

$$r_n^{\mathrm{DL}} = \log_2 \left(1 + \frac{|g_{n2}|^2 a_{n2} \rho}{\xi_{n2} |g_{n2}|^2 a_{f2} \rho + 1} \right), \qquad (10)$$

where the parameter $\xi_{n2} \in [0,1]$ denotes the level of residual interference at D_{n2}. Particularly, $\xi_{n2} = 0$ and $\xi_{n2} = 1$ represent pSIC and ipSIC, respectively.

[3]For the convenience, the transmit power for the two UAVs is assume to be the same. It is noted that our analysis is also applicable to the cases of different transmit power at UAVs.

B. Channel Model

We suppose that the wireless links between the UAV and the NOMA users are dominated by LoS. Besides path loss, the channel is susceptible to small-scale fading, therefore, we apply Nakagami-*m* distribution to model the small-scale fading of the UAV-aided NOMA systems. This is because that this fading can capture LoS communication environments between UAV and users. Thus, the channel gain from UAV*j* to its ground user D_{ij} $(i = \{n, f\}; j = \{1, 2\})$ denoted by g_{ij}, is modeled as [19]

$$g_{ij} = \frac{\bar{g}_{ij}}{\sqrt{1 + d_{ij}^\alpha}}, \qquad (11)$$

where d_{ij}^α denotes the path loss for D_{ij} located in d_{ij} away from the UAV*j*, and $\bar{g}_{ij}$ denotes the Nakagami-*m* fading channel coefficient. We consider three-dimensional cartesian coordinates (x_j, y_j, z_j), where user D_{ij} is represented by $(x_{ij}, y_{ij}, 0)$, whereas the UAV*j* is represented by $(0, 0, h)$.[4] Thus, the distance between D_{ij} and UAV*j* can be denoted as

$$d_{ij} = \left(x_{ij}^2 + y_{ij}^2 + h^2 \right)^{1/2}. \qquad (12)$$

From (12), it can be seen that the spatial distance d_{ij} is the function of the spatial position of ground user D_{ij} as well as the height of UAV*j*.

Considering the fact that FU and NU are deployed in $\mathcal{I}_n = \{\mathcal{A}_n, \mathcal{B}_n\}$ and $\mathcal{I}_f = \{\mathcal{A}_f, \mathcal{B}_f\}$, respectively. Consequently, the users follow as independently and identically distributed PPPs in the area of $\mathcal{I}_n$ and $\mathcal{I}_f$ with notations $W_i, i \in \{n, f\}$, where probability density functions (PDFs) are respectively formulated as [18]

$$f_{W_n}(w_n) = \frac{1}{\pi R_n^2}, \qquad (13)$$

$$f_{W_f}(w_f) = \frac{1}{\pi \left(R_f^2 - R_n^2 \right)}. \qquad (14)$$

Then, the cumulative density function (CDF) of the squared channel coefficient $|g_{nj}|^2$ can be expressed as (15), shown at the bottom of the page.

After some manipulations, for arbitrary α, $|g_{nj}|^2$ can be obtained as (16), shown at the bottom of the next page, where $\gamma(n, x) = (n-1)! [1 - e^{-x} \sum_{k=0}^{n-1} \frac{x^k}{k!}]$.

Similarly, for arbitrary α, the CDF of $|g_{fj}|^2$ can be formulated as (17), shown at the bottom of the next page.

After some manipulations, $|g_{fj}|^2$ can be obtained as (18), shown at the bottom of the next page.

[4]In this paper, we assume that the projections of UAV*j* are in the center of the served users. This is equivalent that the x-coordinate and y-coordinate of UAV*j* are zeros.

$$F_{|g_{nj}|^2}(x) = \int_{\mathcal{I}_n} \left(1 - e^{-\frac{m_{nj} x \left(1 + d_{nj}^\alpha \right)}{\Omega_{nj}}} \sum_{s=0}^{m_{nj}-1} \frac{1}{s!} \left(\frac{m_{nj} x \left(1 + d_{nj}^\alpha \right)}{\Omega_{nj}} \right)^s \right) f_{W_n}(w_n) dw_n$$

$$= \frac{2}{R_n^2} \int_0^{R_n} \left(1 - e^{-\frac{m_{nj} x (1 + r^\alpha)}{\Omega_{nj}}} \sum_{s=0}^{m_{nj}-1} \frac{1}{s!} \left(\frac{m_{nj} x (1 + r^\alpha)}{\Omega_{nj}} \right)^s \right) r \, dr \qquad (15)$$

III. OUTAGE PROBABILITY

In this section, we evaluate the system performance in terms of OP and diversity order. To obtain deeper insights, we first analyze individual outage performance and diversity order of the ground users for $G2U_1$ and U_22G NOMA transmission. Then, the outage performance of the overall system is analyzed.

A. $G2U_1$

Leveraging the SIC principle, the NU is in outage when $\frac{1}{3}r_n^{\text{UL}} < R_2$. Nonetheless, the UAV1 first decodes NU's signal, and then substracts it before it can detect the signal from FU. Consequently, the FU is in outage when $\frac{1}{3}r_n^{\text{UL}} < R_2$ or $\frac{1}{3}r_f^{\text{UL}} < R_1$ if $\frac{1}{3}r_n^{\text{UL}} > R_2$. The OP can be given by

$$P_{out}^{\text{UL},n} = \Pr\left(\frac{1}{3}r_n^{\text{UL}} < R_2\right), \tag{19}$$

$$P_{out}^{\text{UL},f} = 1 - \Pr\left(\frac{1}{3}r_n^{\text{UL}} > R_2, \frac{1}{3}r_f^{\text{UL}} > R_1\right). \tag{20}$$

Based on the results of (19) and (20), we will present the following theorem to discuss the OP of individual user with a given threshold of target rate.

Theorem 1: The OP of the NU D_{n1} with ipSIC in $G2U_1$ is given by (21), shown at the bottom of the page, where both K and K' are complexity-accuracy trade-off parameters. $\Theta_k = \omega_k\sqrt{1-\theta_k^2}(\theta_k+1)s_k/R_n$, $\theta_k = \cos(\frac{(2k-1)\pi}{2K})$, $s_k = \frac{R_n}{2}(\theta_k+1)$, $\omega_k = \frac{\pi}{K}$, $c_k = 1+s_k$, $\Theta_{k'} = \omega_{k'}\sqrt{1-\theta_{k'}^2}s_{k'}/.(R_f+R_n)$, $\theta_{k'} = \cos(\frac{(2k'-1)\pi}{2K'})$, $s_{k'} = \frac{R_f-R_n}{2}(\theta_{k'}+1)+R_n$, $\omega_{k'} = \frac{\pi}{K'}$, $c_{k'} = 1+s_{k'}$, $\eta_{n1} = \frac{m_{n1}\gamma_{thn}a_{f1}}{\Omega_{n1}a_{n1}}$, $\varpi_{f1} = \frac{(m_{f1})_{p_1}(\frac{m_{f1}c_{k'}}{\Omega_{f1}})^{m_{f1}}}{(\eta_{n1}c_k+\frac{m_{f1}c_{k'}}{\Omega_{f1}})^{p_1+m_{f1}}}$, $(m_{f1})_{s_1} = \frac{\Gamma(m_{f1}+s_1)}{\Gamma(m_{f1})} = m_{f1}(m_{f1}+1)\cdots(m_{f1}+s_1-1)$.

Proof: See Appendix A. ∎

Theorem 2: The OP of the FU D_{f1} with ipSIC in $G2U_1$ is given by (22), shown at the bottom of the page, where $\eta_{f1} = \frac{m_{f1}\gamma_{thf}a_{n1}\xi_{n1}}{\Omega_{f1}a_{f1}}$, $\varpi_{n1} = \frac{(m_{n1})_{p'_1}(\frac{m_{n1}c_k}{\Omega_{n1}})^{m_{n1}}}{(\eta_{f1}c_{k'}+\frac{m_{n1}c_k}{\Omega_{n1}})^{p'_1+m_{n1}}}$, $z_{n1} = \frac{(c_k\eta_{n1})^{s_1}}{s_1!}$, $z_{f1} = \frac{(c_{k'}\eta_{f1})^{s'_1}}{s'_1!}$.

Proof: See Appendix B. ∎

Corollary 1: Under the case of pSIC, upon substituting $\xi_{n1} = 0$ into (22), the approximate analytical expression for the OP of D_{f1} with pSIC is obtained as (23), shown

$$F_{|g_{nj}|^2}(x) = 1 - \frac{2e^{-\frac{m_{nj}x}{\Omega_{nj}}}}{R_n^2}\sum_{s=0}^{m_{nj}-1}\frac{1}{s!}\sum_{p=0}^{s}\binom{s}{p}\left(\frac{m_{nj}x}{\Omega_{nj}}\right)^{s-(p+\frac{\alpha}{2})}\frac{\gamma\left(p+\frac{\alpha}{2},\frac{R_n^\alpha m_{nj}x}{\Omega_{nj}}\right)}{\alpha} \tag{16}$$

$$F_{|g_{fj}|^2}(x) = \int_{\mathcal{I}_f}\left(1 - e^{-\frac{m_{fj}x(1+d_{fj}^\alpha)}{\Omega_{fj}}}\sum_{s=0}^{m_{fj}-1}\frac{1}{s!}\left(\frac{m_{fj}x\left(1+d_{fj}^\alpha\right)}{\Omega_{fj}}\right)^s\right)f_{W_f}(w_f)\,dw_f$$
$$= \frac{1}{R_f^2-R_n^2}\int_{R_n}^{R_f}\left(1 - e^{-\frac{m_{fj}x(1+r^\alpha)}{\Omega_{fj}}}\sum_{s=0}^{m_{fj}-1}\frac{1}{s!}\left(\frac{m_{fj}x(1+r^\alpha)}{\Omega_{fj}}\right)^s\right)r\,dr \tag{17}$$

$$F_{|g_{fj}|^2}(x) = 1 - \frac{2e^{-\frac{m_{fj}x}{\Omega_{fj}}}}{R_f^2-R_n^2}\sum_{s=0}^{m_{fj}-1}\frac{1}{s!}\sum_{p=0}^{s}\binom{s}{p}\left(\frac{m_{fj}x}{\Omega_{fj}}\right)^{s-(p+\frac{\alpha}{2})}$$
$$\times\frac{1}{\alpha}\left(\gamma\left(p+\frac{\alpha}{2},\frac{R_f^\alpha m_{fj}x}{\Omega_{fj}}\right) - \gamma\left(p+\frac{\alpha}{2},\frac{R_n^\alpha m_{fj}x}{\Omega_{fj}}\right)\right) \tag{18}$$

$$P_{out,ipSIC}^{\text{UL},n} = \sum_{k=1}^{K}\sum_{k'=1}^{K'}\Theta_k\Theta_{k'}\left(1 - e^{-\frac{\eta_{n1}c_k}{a_{f1}\rho}}\sum_{s_1=0}^{m_{n1}-1}\frac{(c_k\eta_{n1})^{s_1}}{s_1!}\sum_{p_1=0}^{s_1}\binom{s_1}{p_1}\frac{\varpi_{f1}}{(\rho a_{f1})^{s_1-p_1}}\right) \tag{21}$$

$$P_{out,ipSIC}^{\text{UL},f} = 1 - \left(1 - \sum_{k=1}^{K}\sum_{k'=1}^{K'}\Theta_k\Theta_{k'}\left(1 - e^{-\frac{\eta_{n1}c_k}{a_{f1}\rho}}\sum_{s_1=0}^{m_{n1}-1}\sum_{p_1=0}^{s_1}\binom{s_1}{p_1}\frac{z_{n1}\varpi_{f1}}{(\rho a_{f1})^{s_1-p_1}}\right)\right)$$
$$\times\left(1 - \sum_{k=1}^{K}\sum_{k'=1}^{K'}\Theta_k\Theta_{k'}\left(1 - e^{-\frac{\eta_{f1}c_{k'}}{\xi_{n1}a_{n1}\rho}}\sum_{s'_1=0}^{m_{f1}-1}\sum_{p'_1=0}^{s'_1}\binom{s'_1}{p'_1}\frac{z_{f1}\varpi_{n1}}{(\xi_{n1}a_{n1}\rho)^{s'_1-p'_1}}\right)\right) \tag{22}$$

at the bottom of the page, where $\varsigma_{f1} = \frac{m_{f1}\gamma_{thf}}{\Omega_{f1}a_{f1}}$, $\Upsilon_{f1} = \gamma(p'_1 + \frac{\alpha}{2}, R_f^\alpha \frac{\varsigma_{f1}}{\rho}) - \gamma(p'_1 + \frac{\alpha}{2}, R_n^\alpha \frac{\varsigma_{f1}}{\rho})$.

Proof: See Appendix C. ∎

B. $U_2 2 G$

An outage event will occur if the achievable rate is less than the threshold of target rate. Therefore, the OP of D_{f2} and D_{n2} for downlink NOMA transmission can be written as

$$P_{out}^{DL,f} = \Pr\left(r_f^{DL} < R_1\right), \tag{24}$$

$$P_{out}^{DL,n} = 1 - \Pr\left(r_{n \to f}^{DL} > R_1, r_n^{DL} > R_2\right). \tag{25}$$

respectively.

The OP of the FU D_{f2} in downlink NOMA is provided in the following theorem.

Theorem 3: The OP of the FU D_{f2} in downlink NOMA is given by

$$P_{out,ipSIC}^{DL,f} = 1 - \frac{2e^{-\tau_{f2}}}{R_f^2 - R_n^2} \sum_{s'_2=0}^{m_{f2}-1} \sum_{p'_2=0}^{s'_2} \binom{s'_2}{p'_2} \frac{\tau_{f2}^{s'_2 - p'_2 - \frac{\alpha}{2}} \Upsilon_{f2}}{\alpha s'_2!}, \tag{26}$$

where $\Upsilon_{f2} = \gamma(p'_2 + \frac{\alpha}{2}, R_f^\alpha \tau_{f2}) - \gamma(p'_2 + \frac{\alpha}{2}, R_n^\alpha \tau_{f2})$, $\tau_{f2} = \frac{m_{f2}\varsigma_{f2}}{\Omega_{f2}}$.

Proof: By plugging (8) into (24), we can obtain

$$P_{out}^{DL,f} = 1 - \underbrace{\Pr\left(|g_{f2}|^2 > \frac{\gamma_{thf}}{\rho(a_{f2} - \gamma_{thf}a_{n2})} \triangleq \varsigma_{f2}\right)}_{\Lambda}. \tag{27}$$

After some manipulations, Λ can be obtained as

$$\Lambda = 1 - F_{|g_{f2}|^2}(\varsigma_{f2})$$
$$= \frac{2e^{-\tau_{f2}}}{R_f^2 - R_n^2} \sum_{s'_2=0}^{m_{f2}-1} \frac{1}{s'_2!} \sum_{p'_2=0}^{s'_2} \binom{s'_2}{p'_2} \frac{\tau_{f2}^{s'_2 - p'_2 - \frac{\alpha}{2}}}{\alpha} \Upsilon_1, \tag{28}$$

where $\tau_{f2} = \frac{m_{f2}\varsigma_{f2}}{\Omega_{f2}}$, $\Upsilon_1 = \gamma(p'_2 + \frac{\alpha}{2}, R_f^\alpha \tau_{f2}) - \gamma(p'_2 + \frac{\alpha}{2}, R_n^\alpha \tau_{f2})$. Substituting the result in (28) into (27), Theorem 3 can be obtained. ∎

Theorem 4: The OP of the NU D_{n2} in downlink NOMA is given by

$$P_{out,ipSIC}^{DL,n} = 1 - \frac{2e^{-\tau_{n2}}}{R_n^2} \sum_{s_2=0}^{m_{n2}-1} \sum_{p_2=0}^{s_2} \binom{s_2}{p_2} \frac{\tau_{n2}^{s_2 - p_2 - \frac{\alpha}{2}}}{s_2!} \frac{\Upsilon_{n2}}{\alpha}, \tag{29}$$

where $\Upsilon_{n2} = \gamma(p_2 + \frac{\alpha}{2}, R_n^\alpha \tau_{n2})$, $\tau_{n2} = \frac{m_{n2}\vartheta}{\Omega_{n2}}$.

Proof: For the NU D_{n2}, by plugging (9) and (10) into (25), the OP of D_{n2} can be obtained as

$$P_{out,ipSIC}^{DL,n} = 1 - \Pr\left(|g_{n2}|^2 > \vartheta_{f2}, |g_{n2}|^2\right.$$

$$\left. > \frac{\gamma_{thn}}{\rho(a_{n2} - \xi_{n2}a_{f2}\gamma_{thn})} \triangleq \vartheta_{n2}\right)$$

$$= 1 - \Pr\left(|g_{n2}|^2 > \max(\vartheta_{f2}, \vartheta_{n2}) \triangleq \vartheta\right), \tag{30}$$

where $\gamma_{thf} = 2^{3R_1} - 1$, $\gamma_{thn} = 2^{3R_2} - 1$.

Similarly, using the method of Theorem 3, the results of Theorem 4 are obtained. ∎

Corollary 2: Under the condition of pSIC, the approximate analytical expression for the OP of D_{n2} with pSIC is

$$P_{out,pSIC}^{DL,n} = 1 - \frac{2e^{-\bar{\tau}_{n2}}}{R_n^2} \sum_{s_2=0}^{m_{n2}-1} \sum_{p_2=0}^{s_2} \binom{s_2}{p_2} \frac{\bar{\tau}_{n2}^{s_2 - p_2 - \frac{\alpha}{2}}}{s_2!} \frac{\bar{\Upsilon}_{n2}}{\alpha}, \tag{31}$$

where $\bar{\Upsilon}_{n2} = \gamma(p_2 + \frac{\alpha}{2}, R_n^\alpha \bar{\tau}_{n2})$, $\bar{\tau}_{n2} = \frac{m_{n2}\bar{\vartheta}}{\Omega_{n2}}$, $\bar{\vartheta} \triangleq \max(\vartheta_{f2}, \frac{\gamma_{thn}}{\rho a_{n2}})$.

Proof: According to the result of (30), the OP of the D_{n2} under the condition of pSIC can be obtained as

$$P_{out,pSIC}^{DL,n} = 1 - \Pr\left(|g_{n2}|^2 > \vartheta_{f2}, |g_{n2}|^2 > \frac{\gamma_{thn}}{\rho a_{n2}}\right)$$

$$= 1 - \Pr\left(|g_{n2}|^2 > \max\left(\vartheta_{f2}, \frac{\gamma_{thn}}{\rho a_{n2}}\right) \triangleq \bar{\vartheta}\right)$$

$$= 1 - \frac{2e^{-\bar{\tau}_{n2}}}{R_n^2} \sum_{s_2=0}^{m_{n2}-1} \sum_{p_2=0}^{s_2} \binom{s_2}{p_2} \frac{\bar{\tau}_{n2}^{s_2 - p_2 - \frac{\alpha}{2}}}{s_2!} \frac{\bar{\Upsilon}_{n2}}{\alpha}. \tag{32}$$

Then, the result of the Corollary 2 can be obtained. ∎

C. The Overall System Performance

On the basis of the analytical results on Section II, D_{f2} and D_{n2} can successfully decode s_f if satifying $r_{D_{f2}}^{s_f} = \frac{1}{3}\min(r_f^{UL}, r_f, r_f^{DL}) > R_1$, $r_{D_{n2}}^{s_f} = \frac{1}{3}\min(r_f^{UL}, r_f, r_{n \to f}^{DL}) > R_1$, respectively. Similarly, D_{n2} can correctly decode s_n if $r_{D_{n2}}^{s_n} = \frac{1}{3}\min(r_n^{UL}, r_n, r_n^{DL}) > R_2$.

1) OP of D_{f2}: The outage event occurs when s_f cannot be decoded by D_{f2} successfully, which can be expressed as

$$P_{out}^{D_{f2}} = 1 - \Pr\left(r_{D_{f2}}^{s_f} > R_1, \frac{1}{3}\min(r_{t1}^{s_n}, r_{t2}^{s_n}) > R_2\right). \tag{33}$$

The following theorem provides the OP of the FU D_{f2}.

$$P_{out,pSIC}^{UL,f} = 1 - \frac{2e^{-\frac{\varsigma_{f1}}{\rho}}}{R_f^2 - R_n^2} \sum_{s'_1=0}^{m_{f1}-1} \frac{1}{s'_1!} \sum_{p'_1=0}^{s'_1} \binom{s'_1}{p'_1} \left(\frac{\varsigma_{f1}}{\rho}\right)^{s'_1 - (p'_1 + \frac{\alpha}{2})} \frac{\Upsilon_{f1}}{\alpha}$$

$$\times \left(1 - \sum_{k=1}^{K} \sum_{k'=1}^{K'} \Theta_k \Theta_{k'} \left(1 - e^{-\frac{\eta_{n1}c_k}{a_{f1}\rho}} \sum_{s_1=0}^{m_{n1}-1} \sum_{p_1=0}^{s_1} \binom{s_1}{p_1} \frac{z_{n1}\varpi_{f1}}{(\rho a_{f1})^{s_1 - p_1}}\right)\right) \tag{23}$$

Theorem 5: The approximate analytical expression for the OP of D_{f2} with ipSIC is given by (34), shown at the bottom of the page, where $\Upsilon_1 = \gamma(p'_2 + \frac{\alpha}{2}, R_f^\alpha \tau_{f2}) - \gamma(p'_2 + \frac{\alpha}{2}, R_n^\alpha \tau_{f2})$, $\Upsilon_3 = \frac{\gamma(m_{uu}, \tau_{uu})}{\Gamma(m_{uu})}$, $\tau_{f2} = \frac{m_{f2}\varsigma_{f2}}{\Omega_{f2}}$, $\tau_{uu} = \frac{m_{uu}\varsigma_1}{\Omega_{uu}}$, $\zeta_{uu} = \frac{m_{uu}\gamma_{thn}\xi_{uu}b_f}{\Omega_{uu}b_n}$.

Proof: See Appendix D. ∎

Corollary 3: Under the condition of pSIC, the approximate analytical expression for the OP of D_{f2} with pSIC is given by (35), shown at the bottom of the page.

2) OP of D_{n2}: The OP of the event that D_{n2} cannot decode s_n correctly even though s_f can be decoded successfully is given by

$$P_{out}^{D_{n2}} = 1 - \Pr\left(r_{D_{n2}}^{s_f} > R_1, r_{D_{n2}}^{s_n} > R_2\right). \tag{36}$$

The following theorem provides the approximate expression of the OP for D_{n2}.

Theorem 6: The approximate analytical expression for the OP of D_{f2} with ipSIC is given by (37), shown at the bottom of the page.

Proof: See Appendix E. ∎

Corollary 4: Under the condition of pSIC, the approximate analytical expression for the OP of D_{f2} with pSIC is expressed as (38), shown at the bottom of the page.

Proof: See Appendix F. ∎

IV. DIVERSITY ANALYSIS

In this section, we analyze the diversity orders for uplink NOMA and downlink NOMA under the conditions of pSIC and ipSIC to acquire deeper insights in terms of the OP, which is defined as [32]

$$d = -\lim_{\rho \to \infty} \frac{\log P_{out}}{\log \rho}, \tag{39}$$

where P_{out} denotes the OP, ρ denotes the transmit average SNR.

$$P_{out,\text{pSIC}}^{D_{f2}} = 1 - \frac{4\bar{\Upsilon}e^{-\left(\frac{\varsigma_{f1}}{\rho} + \tau_{f2}\right)}}{\left(R_f^2 - R_n^2\right)^2} \sum_{s'_1=0}^{m_{f1}-1} \frac{1}{s'_1!} \sum_{p'_1=0}^{s'_1} \binom{s'_1}{p'_1}\left(\frac{\varsigma_{f1}}{\rho}\right)^{s'_1 - \left(p'_1 + \frac{\alpha}{2}\right)} \frac{\Upsilon_{f1}}{\alpha}$$

$$\times \left(1 - \sum_{k=1}^{K}\sum_{k'=1}^{K'} \Theta_k \Theta_{k'}\left(1 - \sum_{s_1=0}^{m_{n1}-1}\sum_{p_1=0}^{s_1}\binom{s_1}{p_1}\frac{z_{n1}\varpi_{f1}e^{-\frac{\eta_{n1}c_k}{a_{f1}\rho}}}{(\rho a_{f1})^{s_1-p_1}}\right)\sum_{s'_2=0}^{m_{f2}-1}\sum_{p'_2=0}^{s'_2}\binom{s'_2}{p'_2}\frac{\tau_{f2}^{s'_2-p'_2-\frac{\alpha}{2}}}{s'_2!}\frac{\Upsilon_{f2}}{\alpha}\right) \tag{34}$$

$$P_{out,\text{ipSIC}}^{D_{n2}} = 1 - \frac{2\Upsilon e^{-\tau_{n2}}}{R_n^2}\left(1 - \sum_{k=1}^{K}\sum_{k'=1}^{K'}\Theta_k\Theta_{k'}\left(1 - e^{-\frac{\eta_{n1}c_k}{a_{f1}\rho}}\sum_{s_1=0}^{m_{n1}-1}\sum_{p_1=0}^{s_1}\binom{s_1}{p_1}\frac{z_{n1}\varpi_{f1}}{(\rho a_{f1})^{s_1-p_1}}\right)\right)$$

$$\times \left(1 - \sum_{k=1}^{K}\sum_{k'=1}^{K'}\Theta_k\Theta_{k'}\left(1 - \sum_{s'_1=0}^{m_{n1}-1}\sum_{p'_1=0}^{s'_1}\binom{s'_1}{p'_1}\frac{z_{f1}\varpi_{n1}e^{-\frac{\eta_{f1}c_{k'}}{\xi_{n1}a_{n1}\rho}}}{(\xi_{n1}a_{n1}\rho)^{s'_1-p'_1}}\right)\right)\sum_{s_2=0}^{m_{n2}-1}\sum_{p_2=0}^{s_2}\binom{s_2}{p_2}\frac{\tau_{n2}^{s_2-p_2-\frac{\alpha}{2}}}{s_2!}\frac{\Upsilon_{n2}}{\alpha} \tag{35}$$

$$P_{out,\text{pSIC}}^{D_{n2}} = 1 - \frac{4\bar{\Upsilon}e^{-\left(\frac{\varsigma_{f1}}{\rho} + \bar{\tau}_{n2}\right)}}{R_n^2\left(R_f^2 - R_n^2\right)} \sum_{s'_1=0}^{m_{f1}-1} \frac{1}{s'_1!} \sum_{p'_1=0}^{s'_1} \binom{s'_1}{p'_1}\left(\frac{\varsigma_{f1}}{\rho}\right)^{s'_1 - \left(p'_1 + \frac{\alpha}{2}\right)} \frac{\Upsilon_{f1}}{\alpha}$$

$$\times \left(1 - \sum_{k=1}^{K}\sum_{k'=1}^{K'}\Theta_k\Theta_{k'}\left(1 - e^{-\frac{\eta_{n1}c_k}{a_{f1}\rho}}\sum_{s_1=0}^{m_{n1}-1}\sum_{p_1=0}^{s_1}\binom{s_1}{p_1}\frac{z_{n1}\varpi_{f1}}{(\rho a_{f1})^{s_1-p_1}}\right)\right)\sum_{s_2=0}^{m_{n2}-1}\sum_{p_2=0}^{s_2}\binom{s_2}{p_2}\frac{\bar{\tau}_{n2}^{s_2-p_2-\frac{\alpha}{2}}}{s_2!}\frac{\bar{\Upsilon}_{n2}}{\alpha} \tag{37}$$

$$P_{out,\text{ipSIC}}^{D_{f2}} = 1 - \frac{2\Upsilon e^{-\tau_{f2}}}{R_f^2 - R_n^2}\left(1 - \sum_{k=1}^{K}\sum_{k'=1}^{K'}\Theta_k\Theta_{k'}\left(1 - e^{-\frac{\eta_{n1}c_k}{a_{f1}\rho}}\sum_{s_1=0}^{m_{n1}-1}\sum_{p_1=0}^{s_1}\binom{s_1}{p_1}\frac{z_{n1}\varpi_{f1}}{(\rho a_{f1})^{s_1-p_1}}\right)\right)$$

$$\times \left(1 - \sum_{k=1}^{K}\sum_{k'=1}^{K'}\Theta_k\Theta_{k'}\left(1 - \sum_{s'_1=0}^{m_{n1}-1}\sum_{p'_1=0}^{s'_1}\binom{s'_1}{p'_1}\frac{z_{f1}\varpi_{n1}e^{-\frac{\eta_{f1}c_{k'}}{\xi_{n1}a_{n1}\rho}}}{(\xi_{n1}a_{n1}\rho)^{s'_1-p'_1}}\right)\right)\sum_{s'_2=0}^{m_{f2}-1}\sum_{p'_2=0}^{s'_2}\binom{s'_2}{p'_2}\frac{\tau_{f2}^{s'_2-p'_2-\frac{\alpha}{2}}}{s'_2!}\frac{\Upsilon_{f2}}{\alpha} \tag{38}$$

A. $G2U_1$

We first carry out high SNR approximation analysis of the corresponding OP.

Theorem 7: The asymptotic OP of the FU D_{f1} in $G2U_1$ is expressed as (40), shown at the bottom of the page, where $z_{f1} = \frac{(c_{k'}\eta_{f1})^{s'_1}}{s'_1!}$.

Proof: In the high SNRs regions, the asymptotic expression for the $f_{|\bar{g}_{ij}|^2}(x)$ can be expressed as

$$f_{|\bar{g}_{ij}|^2}(x) \approx \frac{m_{ij}{}^{m_{ij}} x^{m_{ij}-1}}{\Omega_{ij}{}^{m_{ij}} \Gamma(m_{ij})}. \tag{41}$$

Conditioned on the PPPs, the CDFs of the variables $|g_{nj}|^2$ and $|g_{fj}|^2$ can be denoted by $F_{|\bar{g}_{nj}|^2}(x)$ and $F_{|\tilde{g}_{fj}|^2}(x)$ in the high SNR region. We have

$$F_{|\tilde{g}_{nj}|^2}(x) \approx \frac{2}{R_n^2} \int_0^{R_n} \frac{\left(m_{nj}x(1+r^\alpha)\right)^{m_{nj}}}{\Omega_{nj}^{m_{nj}} m_{nj}!} r\,dr$$

$$= \frac{2}{R_n^2} \frac{(m_{nj}x)^{m_{nj}}}{\Omega_{nj}^{m_{nj}} m_{nj}!} \sum_{q=0}^{m_{nj}} \binom{m_{nj}}{q} \frac{R_n^{\alpha q+2}}{\alpha q + 2}, \tag{42}$$

and

$$F_{|\tilde{g}_{fj}|^2}(x) \approx \frac{2}{R_f^2 - R_n^2} \int_{R_n}^{R_f} \frac{\left(m_{fj}x(1+r^\alpha)\right)^{m_{fj}}}{\Omega_{fj}^{m_{nj}} m_{fj}!} r\,dr$$

$$= \frac{2}{R_f^2 - R_n^2} \frac{(m_{fj}x)^{m_{fj}}}{\Omega_{fj}^{m_{fj}} m_{fj}!} \sum_{q=0}^{m_{fj}} \binom{m_{fj}}{q}$$

$$\times \frac{R_f^{\alpha q+2} - R_n^{\alpha q+2}}{\alpha q + 2}. \tag{43}$$

Based on the result of (22), using the similar method of Theorem 2, and replacing $F_{|g_{fj}|^2}$ by $F_{|\tilde{g}_{fj}|^2}$, we can obtain the results of Theorem 7. ∎

For the NU, replacing (16) and (18) by (42) and (43), and using the corresponding proof in Section-III, the asymptotic expressions for the OP of the NU can be provided in the following theorem.

Theorem 8: The asymptotic OP of the NU D_{n1} in $G2U_1$ is given by

$$P_{\text{ipSIC},\infty}^{\text{UL},n} = \sum_{k=1}^{K} \sum_{k'=1}^{K'} \Theta_k \Theta_{k'}$$

$$\times \left(1 - \sum_{s_1=0}^{m_{n1}-1} \frac{z_{n1}(m_{f1})_{s_1} \left(\frac{m_{f1}c_{k'}}{\Omega_{f1}}\right)^{m_{f1}}}{\left(\frac{m_{f1}c_{k'}}{\Omega_{f1}} + \eta_{n1}c_k\right)^{m_{f1}+s_1}} \right) \tag{44}$$

where $(m_{f1})_{s_1}$ is the Pochhammer symbol defined as $(m_{f1})_{s_1} = \frac{\Gamma(m_{f1}+s_1)}{\Gamma(m_{f1})} = m_{f1}(m_{f1}+1)\cdots(m_{f1}+s_1-1)$, $z_{n1} = \frac{(c_k\eta_{n1})^{s_1}}{s_1!}$.

Proof: Using the similar method in Theorem 7, we can get the result of Theorem 8. ∎

Corollary 5: Under the condition of pSIC, the asymptotic OP of the FU D_{n1} in $G2U_1$ is given by (45), shown at the bottom of the page.

Corollary 6: Using the results of (40), (44) and (41), we can obtain individual user's diversity gain as

$$d^{\text{UL},D_{n1}} = 0, \tag{46}$$

$$d_{ipSIC}^{\text{UL},D_{f1}} = 0, \quad d_{pSIC}^{\text{UL},D_{f1}} = 0. \tag{47}$$

Remark 1: We can observe from Corollary 6 that the diversity orders of the FU and NU are zeros. This is because the main interferences for signals received from uplink NOMA users are composed of intra-group interference and AWGN. The intra-group interferences come from D_{f1} whose channel conditions are worse than that of D_{n1}. On the other hand, when UAV1 first decodes D_{n1}'s signal of uplink transmission, it will be interfered by D_{f1}. Additionally, the diversity order of the FU is zero as well under the condition of pSIC.

B. U_22G

For the downlink NOMA, the high SNR approximations of the corresponding OP for the FU and the NU are provided in Theorem 9 and 10.

$$P_{\text{ipSIC},\infty}^{\text{UL},f} = 1 - \left(1 - \sum_{k=1}^{K} \sum_{k'=1}^{K'} \Theta_k \Theta_{k'} \left(1 - \sum_{s_1=0}^{m_{n1}-1} \frac{z_{n1}(m_{f1})_{s_1} \left(\frac{m_{f1}c_{k'}}{\Omega_{f1}}\right)^{m_{f1}}}{\left(\frac{m_{f1}c_{k'}}{\Omega_{f1}} + \eta_{n1}c_k\right)^{m_{f1}+s_1}} \right) \right)$$

$$\times \left(1 - \sum_{k=1}^{K} \sum_{k'=1}^{K'} \Theta_k \Theta_{k'} \left(1 - \sum_{s'_1=0}^{m_{f1}-1} \frac{z_{f1}(m_{n1})_{s'_1} \left(\frac{m_{n1}c_k}{\Omega_{n1}}\right)^{m_{n1}}}{\left(\frac{m_{n1}c_k}{\Omega_{n1}} + \eta_{f1}c_{k'}\right)^{m_{n1}+s'_1}} \right) \right) \tag{40}$$

$$P_{\text{pSIC},\infty}^{\text{UL},f} = 1 - \left(1 - \sum_{k=1}^{K} \sum_{k'=1}^{K'} \Theta_k \Theta_{k'} \left(1 - \sum_{s_1=0}^{m_{n1}-1} \frac{z_{n1}(m_{f1})_{s_1} \left(\frac{m_{f1}c_{k'}}{\Omega_{f1}}\right)^{m_{f1}}}{\left(\frac{m_{f1}c_{k'}}{\Omega_{f1}} + \eta_{n1}c_k\right)^{m_{f1}+s_1}} \right) \right)$$

$$\times \left(1 - \frac{2\varsigma_{f1}^{m_{f1}}}{m_{f1}!\left(R_f^2 - R_n^2\right)} \sum_{q'_1=0}^{m_{f1}} \binom{m_{f1}}{q'_1} \frac{R_f^{\alpha q'_1+2} - R_n^{\alpha q'_1+2}}{\alpha q'_1 + 2} \right) \tag{45}$$

Theorem 9: The asymptotic OP of D_{f2} in $U_2 2G$ is given by

$$P_{\mathrm{ipSIC},\infty}^{\mathrm{DL},f} = \frac{2\tau_{f2}^{m_{f2}}}{m_{f2}!\left(R_f^2 - R_n^2\right)} \sum_{q'_1=0}^{m_{f2}} \binom{m_{f2}}{q'_1}$$
$$\times \frac{R_f^{\alpha q'_1+2} - R_n^{\alpha q'_1+2}}{\alpha q'_1 + 2}. \tag{48}$$

Proof: Using the similar method in Theorem 7, we can get the result of Theorem 9. ∎

Theorem 10: The asymptotic OP of D_{n2} in $U_2 2G$ is given by

$$P_{\mathrm{ipSIC},\infty}^{\mathrm{DL},n} = \frac{2\tau_{n2}^{m_{n2}}}{m_{n2}!R_n^2} \sum_{q_2=0}^{m_{n2}} \binom{m_{n2}}{q_2} \frac{R_n^{\alpha q_2+2}}{\alpha q_2 + 2}. \tag{49}$$

Proof: Using the similar method in Theorem 7, we can get the result of Theorem 10. ∎

Corollary 7: Under the condition of pSIC, the asymptotic OP expression is given by

$$P_{\mathrm{pSIC},\infty}^{\mathrm{DL},n} = \frac{2\bar{\tau}_{n2}^{m_{n2}}}{m_{n2}!R_n^2} \sum_{q_2=0}^{m_{n2}} \binom{m_{n2}}{q_2} \frac{R_n^{\alpha q_2+2}}{\alpha q_2 + 2}. \tag{50}$$

By plugging (48), (49) and (50) into (39), the diversity orders of the FU and NU are provided in Corollary 8.

Corollary 8: Using the results of (48), (49) and (50), we can obtain individual user's diversity gain as

$$d^{\mathrm{DL},D_{f2}} = m_{f2}, \tag{51}$$
$$d_{\mathrm{ipSIC}}^{\mathrm{DL},D_{n2}} = 0, \quad d_{\mathrm{pSIC}}^{\mathrm{DL},D_{n2}} = m_{n2}. \tag{52}$$

Remark 2: We can observe from Corollary 8 that the diversity orders of the FU and NU are m_{f2} and m_{n2} under the condition of the pSIC, respectively. For the case of ipSIC, the diversity order of the NU is zero.

V. SYSTEM THROUGHPUT

In this section, we mainly investigate the system through-put under the mode of delay-limited transmission mode, which is one of two transmission modes of operation in wireless communication. And the other is called delay-tolerant trans-mission mode, where the ergodic rate will influence the system throughput. However, for the delay-limited transmission mode, the wireless fading channels are major determinants. We con-sider delay-limited transmission mode in this paper in order to maintain a fixed rate for users. In this mode, the system throughput is only determined by OP since the sender has a fixed transmit rate.

Corollary 9: Under the condition of ipSIC, according to the analytical results of Theorem 5 and Theorem 6, the system throughput is expressed as [33]

$$\tau_{\mathrm{ipSIC}} = \left(1 - P_{out,\mathrm{ipSIC}}^{D_{f2}}\right)R_1 + \left(1 - P_{out,\mathrm{ipSIC}}^{D_{n2}}\right)R_2, \tag{53}$$

where $P_{out,\mathrm{ipSIC}}^{D_{f2}}$ and $P_{out,\mathrm{ipSIC}}^{D_{n2}}$ are obtained from (34) and (35), respectively.

Corollary 10: Under the case of pSIC, according to the analytical results of Corollary 3 and Corollary 4, the system throughput is expressed as

$$\tau_{\mathrm{pSIC}} = \left(1 - P_{out,\mathrm{pSIC}}^{D_{f2}}\right)R_1 + \left(1 - P_{out,\mathrm{pSIC}}^{D_{n2}}\right)R_2, \tag{54}$$

TABLE I
SIMULATION PARAMETERS

| Parameter	Value
Monte Carlo simulation repeated	10^5 iterations
The uplink power allocation coefficients	$a_{f1} = 0.3$, $a_{n1} = 0.7$
The downlink power allocation coefficients	$a_{f2} = 0.8$, $a_{n2} = 0.2$
The power allocation coefficients in point to point link	$b_f = 0.8$, $b_n = 0.2$
The distance between UAV1 and UAV2	$L = 100m$
The fading channel parameter	$m = 2$
Noise power	$\sigma^2 = 1$
The radius of inner circle and outer circle	$R_n = 8m$, $R_f = 10m$
The height of UAV	$h = 20m$
The path loss exponent	$v = 2$
The target rates	$R_1 = 0.01$, $R_2 = 0.1$
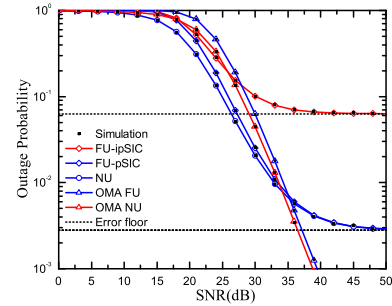
Fig. 2. OPs of users versus SNR for uplink transmission.

where $P_{out,\mathrm{pSIC}}^{D_{f2}}$ and $P_{out,\mathrm{pSIC}}^{D_{n2}}$ are obtained from (33) and (36), respectively.

VI. NUMERICAL RESULTS

In this section, some numerical results are provided to verify the correctness of our analyses. Unless otherwise specified, the simulation settings are provided in Table I. Additionally, we set $K = K' = 20$. For the sake of comparison, the case of pSIC is provided as a benchmark.

Fig. 2 depicts the OPs of D_{f1} and D_{n1} for the uplink trans-mission with both pSIC and ipSIC versus SNR. In order to facilitate comparison, the curves of OMA scheme are plotted in Fig. 2. One can observe that an outage floor exists in the high SNR regime for NOMA, which matches the conclusion of Remark 1. This is because the diversity orders of D_{f1} and D_{n1} are zeros due to the influence of residual interference using ipSIC. One can also observe that although the pSIC is carried out in the uplink transmission for NOMA, the diver-sity orders of D_{f1} and D_{n1} are still zeros. The reason is that when UAV1 decodes the stronger signal of the uplink transmis-sion, it will be interfered by the weaker signal. However, even though under the condition of pSIC, there still exist outage floors in the high SNR regions. This is due to uplink NOMA transmission turning into interference in this case. For uplink transmission the OPs of users for OMA are larger than that of NOMA in the high SNR regions, and there are no out-age floors for OMA in the high SNR regions owing to the non-interference transmission.

Fig. 3 illustrates the OPs of D_{f2} and D_{n2} for the downlink transmission with both pSIC and ipSIC versus SNR. Similarly, the curves of OMA scheme are also plotted in Fig. 3. We can observe that an outage floor exists for the OP in the high
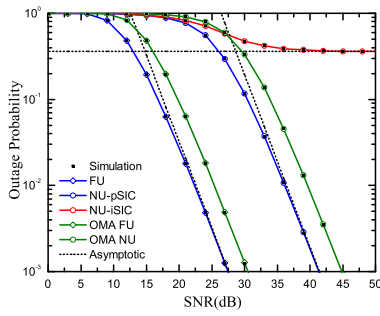
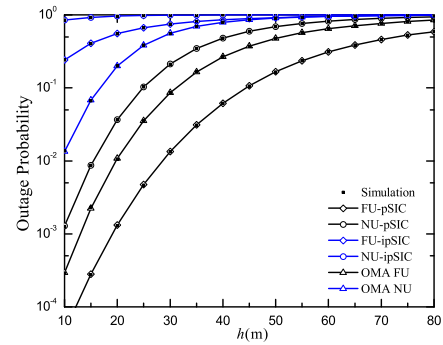Fig. 3. OPs of users versus SNR for downlink transmission.

(a) Perfect SIC

(b) Imperfect SIC
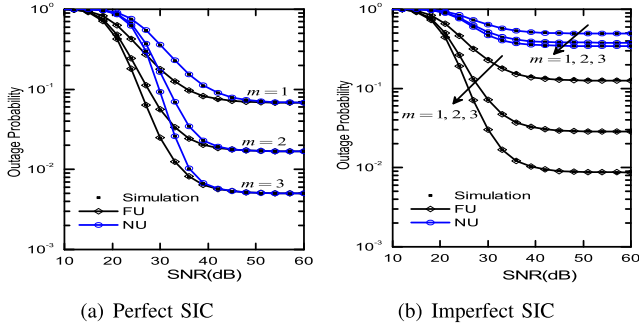
Fig. 4. OPs of users versus SNR for different m.

Fig. 5. OPs of users versus h.
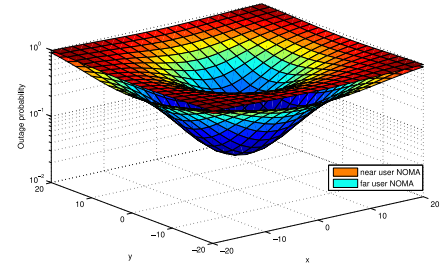

Fig. 6. OPs of users.

Fig. 7. System throughput versus SNR.

SNRs for D_{n2} under the condition of ipSIC. This is due to the residual interference resulting in zero diversity order. One can also observe that the outage performance of D_{f2} and D_{n2} increases as the SNR increases by the use of pSIC. This figure also shows that there is a non-zero diversity order for the OP under the condition of pSIC. Finally, it is obvious that whether the FU or NU, the OP for NOMA is superior to OMA, and there are no outage floors across the entire SNR region due to non-interference transmission for OMA.

Fig. 4 plots the OP versus the SNR for different fading parameters, $m = \{1, 2, 3\}$. It is shown that exact analytical results and Monte Carlo match very well. In addition, the OP decreases with the increase of the fading parameter m. When $m = 1$, the fading channel reduces to Rayleigh case. It is also observed that there exist outage floors for both pSIC and ipSIC. This is because that residual interferences are always in existence for uplink NOMA transmission. Compared with pSIC, there is a severe performance loss caused by SIC error, especially in moderate and high SNR regions. Finally, it can be observed that the OPs for the FU and NU at high SNRs have the same values under the condition of pSIC.

Fig. 5 shows theOPs of the users with both ipSIC and pSIC versus the height of the UAV h for both OMA and NOMA. From Fig. 5, we can observe that the outage performance of both D_{f2} and D_{n2} decreases as the height of UAV increases for both OMA and NOMA since the pathlosses between the ground users and UAV increase with the height of UAV. Furthermore, one can also be observe that the FU has better outage performance on account of higher received power from the UAV. This means that the fairness can be ensured if the FU is allocated more power. Finally, it is evident that the performance of users using ipSIC is poor due to the large interference and pathloss.
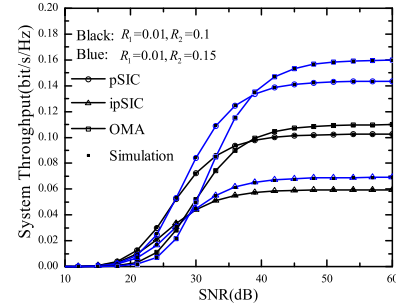
Fig. 6 plots the OPs of NOMA users versus the spatial position of users. In this case, the height of the UAV is fixed to 10m. This figure clearly depicts the influence of the distance between users and UAV in terms of the OP. We can see that the FU can achieve better outage performance, this is because that the FU is allocated higher power. Noted that the outage performance of NOMA users becomes better when $(x, y) \to (0, 0)$, as the average distance between the UAV and the NOMA users is gradually shortened. To sum up, the distances between the UAV and the served ground users limit the outage performance of the whole system.

Fig. 7 compares the throughput of the considered systems versus SNR with different targeted rates for NOMA and OMA. We set $\{R_1, R_2\} = \{0.01, 0.1; 0.01, 0.15\}$. One can observe that there exists the throughput ceiling in the high SNR region. This can be explained by the fact that the OP exists an outage floor in the high SNR region and the throughput is also determined by the targeted date rate. Note that the throughput can be improved as R_2 increasing from $R_2 = 0.1$ to $R_2 = 0.15$. Consequently, we can observe that selecting proper targeted rates is of significance when some NOMA systems are designed in practical. Finally, we can also observe that at high SNRs, NOMA achieves a small throughput than OMA

due to the large interference caused by SIC in the uplink and downlink transmissions.

VII. CONCLUSION

In this paper, a 3-hop UAV-NOMA network framework is studied. Stochastic geometry is used for modeling the spatial positions of NOMA users. More particularly, for level-performance and system-performance, new analytical expressions for the OP of the NU and FU are derived. The asymptotic expressions for the OP in the high SNRs regions and the diversity order are discussed and obtained. Also, the system throughput is evaluated and discussed under delay-limited transmission mode. The research results prove that the diversity order is zero because of residual interference in the uplink $G2 U_2$ NOMA transmission.

APPENDIX A
PROOF OF THEOREM 1

Based on (2) and (19), $P_{out}^{\mathrm{UL},n}$ can be given by (A.1), shown at the bottom of the page, where $f_{|g_{ij}|^2}(x) = \frac{\left(\frac{m_{ij}(1+d_{ij}^\alpha)}{\Omega_{ij}}\right)^{m_{ij}} x^{m_{ij}-1}}{\Gamma(m_{ij})} e^{-\frac{m_{ij}(1+d_{ij}^\alpha)x}{\Omega_{ij}}}$.

By applying [34, 3.381.4], $\Psi_1(d_{n1})$ can be calculated as (A.2), shown at the bottom of the page, where $\eta_{n1} =$

$\frac{m_{n1}\gamma_{thn} a_{f1}}{\Omega_{n1} a_{n1}}$, $(m_{f1})_{s_1} = \frac{\Gamma(m_{f1}+s_1)}{\Gamma(m_{f1})} = m_{f1}(m_{f1}+1)\cdots(m_{f1}+s_1-1)$.

Plugging (A.2) into (A.1), (A.1) can be rewritten as

$$P_{out}^{\mathrm{UL},n} = \frac{2}{R_n^2}\int_{\mathcal{I}_f}\underbrace{\int_0^{R_n} r\Psi_1(r)dr}_{\Psi_2(d_{f1})}f_{W_f}(w_f)\,dw_f. \quad \text{(A.3)}$$

Utilizing the Gaussian-Chebyshev quadrature [35], $\Psi_2(d_{f1})$ can be expressed as (A.4), shown at the bottom of the page, where $\theta_k = \cos(\frac{(2k-1)\pi}{2K})$, $\omega_k = \frac{\pi}{K}$, and $s_k = \frac{R_n}{2}(\theta_k+1)$.

Substituting (A.4), into (A.3), and utilizing the Gaussian-Chebyshev quadrature, we have (A.5), shown at the bottom of the page, where $\Theta_k = \omega_k\sqrt{1-\theta_k^2}(\theta_k+1)s_k/R_n$, $\Theta_{k'} = \omega_{k'}\sqrt{1-\theta_{k'}^2}s_{k'}/(R_f+R_n)$, $c_k = 1+s_k$, $c_{k'} = 1+s_{k'}$.

APPENDIX B
PROOF OF THEOREM 2

By substituting (2) and (3) into (20), we have (B.1), shown at the bottom of the next page. With the help of

$$P_{out}^{\mathrm{UL},n} = \Pr\left(\frac{|g_{n1}|^2 a_{n1}\rho}{|g_{f1}|^2 a_{f1}\rho+1} < 2^{3R_2}-1 \triangleq \gamma_{thn}\right)$$

$$= \int_{\mathcal{I}_f}\int_{\mathcal{I}_n}\underbrace{\int_0^\infty\left(1-e^{-\frac{m_{n1}(1+d_{n1}^\alpha)\gamma_{thn}(x a_{f1}\rho+1)}{\Omega_{n1}a_{n1}\rho}}\sum_{s_1=0}^{m_{n1}-1}\frac{1}{s_1!}\left(\frac{m_{n1}(1+d_{n1}^\alpha)\gamma_{thn}(x a_{f1}\rho+1)}{\Omega_{n1}a_{n1}\rho}\right)^{s_1}\right)f_{|g_{f1}|^2}(x)dx}_{\Psi_1(d_{n1})}$$

$$\times f_{W_n}(w_n)\,dw_n f_{W_f}(w_f)\,dw_f \quad \text{(A.1)}$$

$$\Psi_1(d_{n1}) = 1 - e^{-\frac{\eta_{n1}(1+d_{n1}^\alpha)}{a_{f1}\rho}}\sum_{s_1=0}^{m_{n1}-1}\frac{\eta_{n1}^{s_1}}{s_1!}(1+d_{n1}^\alpha)^{s_1}\sum_{p_1=0}^{s_1}\binom{s_1}{p_1}\frac{(m_{f1})_{p_1}\left(\frac{1}{a_{f1}\rho}\right)^{s_1-p_1}\left(\frac{m_{f1}\left(1+d_{f1}^\alpha\right)}{\Omega_{f1}}\right)^{m_{f1}}}{\left(\eta_{n1}\left(1+d_{n1}^\alpha\right)+\frac{m_{f1}\left(1+d_{f1}^\alpha\right)}{\Omega_{f1}}\right)^{p_1+m_{f1}}} \quad \text{(A.2)}$$

$$\Psi_2(d_{f1}) = \sum_{k=1}^K\frac{R_n w_k\sqrt{1-\theta_k^2}s_k}{2}$$

$$- \sum_{k=1}^K\sum_{s_1=0}^{m_{n1}-1}\sum_{p_1=0}^{s_1}\binom{s_1}{p_1}\frac{R_n w_k\sqrt{1-\theta_k^2}s_k e^{-\frac{\eta_{n1}(1+s_k^\alpha)}{a_{f1}\rho}}\left(\frac{1}{a_{f1}\rho}\right)^{s_1-p_1}(m_{f1})_{p_1}\left(\frac{m_{f1}\left(1+d_{f1}^\alpha\right)}{\Omega_{f1}}\right)^{m_{f1}}}{2\left(\eta_{n1}\left(1+s_k^\alpha\right)+\frac{m_{f1}\left(1+d_{f1}^\alpha\right)}{\Omega_{f1}}\right)^{p_1+m_{f1}}} \quad \text{(A.4)}$$

$$P_{out}^{\mathrm{UL},n} = \sum_{k=1}^K\sum_{k'=1}^{K'}\Theta_k\Theta_{k'}\left(1-\sum_{s_1=0}^{m_{n1}-1}\frac{(c_k\eta_{n1})^{s_1}e^{-\frac{\eta_{n1}c_k}{a_{f1}\rho}}}{s_1!}\times\sum_{p_1=0}^{s_1}\binom{s_1}{p_1}\left(\frac{1}{\rho a_{f1}}\right)^{s_1-p_1}\frac{(m_{f1})_{p_1}\left(\frac{m_{f1}c_{k'}}{\Omega_{f1}}\right)^{m_{f1}}}{\left(\eta_{n1}c_k+\frac{m_{f1}c_{k'}}{\Omega_{f1}}\right)^{p_1+m_{f1}}}\right) \quad \text{(A.5)}$$

$$I_2 = 1 - \Pr\left(|g_{f1}|^2 < \frac{\gamma_{thf}\left(\xi_{n1}|g_{n1}|^2 a_{n1} + \frac{1}{\rho}\right)}{a_{f1}}\right)$$

$$= 1 - \int_{I_n}\int_{I_f} \Psi'_1(d_{f1})f_{W_f}(\omega_f)f_{W_n}(\omega_n)\mathrm{d}\omega_f\mathrm{d}\omega_n \quad (B.2)$$

where $\Psi'_1(d_{f1}) = \int_0^\infty 1 - e^{-\frac{m_{f1}(1+d_{f1}^\alpha)\gamma_{thf}(\xi_{n1}xa_{n1}\rho+1)}{\Omega_{f1}a_{f1}\rho}}$
$\sum_{s'_1}^{m_{f1}-1}\frac{1}{s'_1!}\left(\frac{m_{f1}(1+d_{f1}^\alpha)\gamma_{thf}(\xi_{n1}xa_{n1}\rho+1)}{\Omega_{f1}a_{f1}\rho}\right)^{s'_1}f_{|g_{n1}|^2}(x)\mathrm{d}x$.

It is difficult to get the exact expression of this integral. Therefore, by applying Gaussian-Chebyshev quadrature, $F_{|g_{f_j}|^2}(x)$ can be approximated as (B.3), shown at the bottom of the page.

Based on (B.3), we calculate I_2 as follows

$$I_2 = 1 - \frac{2}{R_f^2 - R_n^2}\int_{I_n}\underbrace{\int_{R_n}^{R_f} r\Psi'_1(r)dr}_{\Psi'_2(d_{n1})}f_{W_n}(w_n)dw_n, \quad (B.4)$$

where

$$\varpi_{n1} = (m_{n1})_{p'_1}\left(\frac{m_{n1}c_k}{\Omega_{n1}}\right)^{m_{n1}}/\left(\eta_{f1}c_k' + \frac{m_{n1}c_k}{\Omega_{n1}}\right)^{p'_1+m_{n1}}.$$

Based on the result in (A.5), we have (B.7), shown at the bottom of the page.

By plugging I_1 and I_2 into (B.1), Theorem 2 can be proved.

APPENDIX C
PROOF OF COROLLARY 1

For the case of pSIC, the OP of the FU is (C.1), shown at the bottom of the next page.

$$P_{out}^{UL,f} = 1 - \Pr\left\{\log_2\left(1 + \frac{|g_{n1}|^2 a_{n1}\rho}{|g_{f1}|^2 a_{f1}\rho + 1}\right) > R_2, \log_2\left(1 + \frac{|g_{f1}|^2 a_{f1}\rho}{\xi_{n1}|g_{n1}|^2 a_{n1}\rho+1}\right) > R_1\right\}$$

$$= 1 - \underbrace{\Pr\left(|g_{n1}|^2 > \frac{\gamma_{thn}\left(|g_{f1}|^2 a_{f1}\rho + 1\right)}{a_{n1}\rho}\right)}_{I_1}\underbrace{\Pr\left(|g_{f1}|^2 > \frac{\gamma_{thf}\left(\xi_{n1}|g_{n1}|^2 a_{n1}\rho + 1\right)}{a_{f1}\rho}\right)}_{I_2} \quad (B.1)$$

$$\Psi'_1(d_{f1}) = 1 - e^{-\frac{\eta_{f1}\left(1+d_{f1}^\alpha\right)}{\xi_{n1}a_{n1}\rho}}\sum_{s'_1=0}^{m_{f1}-1}\frac{\left(\eta_{f1}\left(1+d_{f1}^\alpha\right)\right)^{s'_1}}{s'_1!}\sum_{p'_1=0}^{s'_1}\binom{s'_1}{p'_1}\left(\frac{1}{\xi_{n1}a_{n1}\rho}\right)^{s'_1-p'_1}$$

$$\times \frac{(m_{n1})_{p'_1}\left(\frac{m_{n1}(1+d_{n1}^\alpha)}{\Omega_{n1}}\right)^{m_{n1}}}{\left(\eta_{f1}\left(1+d_{f1}^\alpha\right) + \frac{m_{n1}(1+d_{n1}^\alpha)}{\Omega_{n1}}\right)^{p'_1+m_{n1}}} \quad (B.3)$$

$$\Psi'_2(d_{n1}) = \frac{R_f - R_n}{2}\sum_{k'=1}^{K'}\omega_{k'}\sqrt{1-\theta_{k'}^2}\,s_{k'}\left(1 - e^{-\frac{\eta_{f1}\left(1+s_{k'}^\alpha\right)}{\xi_{n1}a_{n1}\rho}}\sum_{s'_1=0}^{m_{f1}-1}\frac{\left(\eta_{f1}\left(1+s_{k'}^\alpha\right)\right)^{s'_1}}{s'_1!}\right.$$

$$\left.\times \sum_{p'_1=0}^{s'_1}\binom{s'_1}{p'_1}\left(\frac{1}{\xi_{n1}a_{n1}\rho}\right)^{s'_1-p'_1}\frac{(m_{n1})_{p'_1}\left(\frac{m_{n1}(1+d_{n1}^\alpha)}{\Omega_{n1}}\right)^{m_{n1}}}{\left(\eta_{f1}\left(1+s_{k'}^\alpha\right) + \frac{m_{n1}(1+d_{n1}^\alpha)}{\Omega_{n1}}\right)^{p'_1+m_{n1}}}\right) \quad (B.5)$$

$$I_2 = 1 - \sum_{k=1}^{K}\sum_{k'=1}^{K'}\Theta_k\Theta_{k'}\left(1 - e^{-\frac{\eta_{f1}c_{k'}}{\xi_{n1}a_{n1}\rho}}\sum_{s'_1=0}^{m_{f1}-1}\frac{(c_{k'}\eta_{f1})^{s'_1}}{s'_1!}\sum_{p'_1=0}^{s'_1}\binom{s'_1}{p'_1}\frac{\varpi_{n1}}{(\xi_{n1}a_{n1}\rho)^{s'_1-p'_1}}\right) \quad (B.6)$$

$$I_1 = 1 - \sum_{k=1}^{K}\sum_{k'=1}^{K'}\Theta_k\Theta_{k'}\left(1 - e^{-\frac{\eta_{n1}c_k}{a_{f1}\rho}}\sum_{s_1=0}^{m_{n1}-1}\frac{(c_k\eta_{n1})^{s_1}}{s_1!}\sum_{p_1=0}^{s_1}\binom{s_1}{p_1}\frac{\varpi_{f1}}{(\rho a_{f1})^{s_1-p_1}}\right) \quad (B.7)$$

With the help of

$$I'_2 = \frac{2e^{-\varsigma_{f1}}}{R_f^2 - R_n^2} \sum_{s'_1=0}^{m_{f1}-1} \frac{1}{s'_1!} \sum_{p'_1=0}^{s'_1} \binom{s'_1}{p'_1} \varsigma_{f1}^{s'_1-(p'_1+\frac{\alpha}{2})} \frac{\Upsilon}{\alpha},$$

(C.2)

where $\varsigma_{f1} = \frac{m_{f1}\gamma_{thf}}{\Omega_{f1}a_{f1}\rho}$, $\Upsilon = \gamma(p'_1 + \frac{\alpha}{2}, R_f^\alpha \varsigma_{f1}) - \gamma(p'_1 + \frac{\alpha}{2}, R_n^\alpha \varsigma_{f1})$.

Therefore, we obtain (C.3), shown at the bottom of the page.

APPENDIX D
PROOF OF THEOREM 5

Based on (33), we have (D.1), shown at the bottom of the page.

Then, substituting (5) and (6) into (D.1), we have (D.2), shown at the bottom of the page.

With the aid of Theorem 2 and Theorem 3, $\Pr(\frac{1}{3}r_n^{UL} > R_2, \frac{1}{3}r_f^{UL} > R_1)$ and $\Pr(\frac{1}{3}r_f^{DL} > R_1)$ can be obtained. Then plugging (D.2) into (D.1), the Theorem 5 can be obtained.

APPENDIX E
PROOF OF THEOREM 6

Similarly as (E.1), shown at the bottom of the page, substituting the results of Theorem 2 and Theorem 4 and (D.2) into (E.1), Theorem 6 can be obtained.

APPENDIX F
PROOF OF THEOREM 6

Similarly as (F.1), shown at the bottom of the page, substituting the results of Theorem 2 and (F.1), and (D.2) into (E.1), Corollary 4 can be obtained.

$$P_{out,\text{pSIC}}^{UL,D_{f1}} = 1 - \underbrace{\Pr\left(|g_{n1}|^2 > \frac{\gamma_{thn}\left(|h_{f1}|^2 a_{f1}\rho + 1\right)}{a_{n1}\rho}\right)}_{I_1} \underbrace{\Pr\left(|g_{f1}|^2 > \frac{\gamma_{thf}}{a_{f1}\rho}\right)}_{I'_2}$$

(C.1)

$$P_{out,\text{pSIC}}^{UL,D_{f1}} = 1 - \left(1 - \sum_{k=1}^{K}\sum_{k'=1}^{K'} \Theta_k\Theta_{k'} + \sum_{k=1}^{K}\sum_{k'=1}^{K'} \Theta_k\Theta_{k'} e^{-\frac{z_1}{a_{f1}\rho}} \sum_{s_1=0}^{m_{n1}-1}\sum_{p_1=0}^{s_1} \binom{s_1}{p_1}\left(\frac{1}{\rho a_{f1}}\right)^{s_1-p_1} v_1\right)$$

$$\times \frac{2e^{-\varsigma_{f1}}}{R_f^2 - R_n^2} \sum_{s'_1=0}^{m_{f1}-1} \frac{1}{s'_1!} \sum_{p'_1=0}^{s'_1} \binom{s'_1}{p'_1} \varsigma_{f1}^{s'_1-(p'_1+\frac{\alpha}{2})} \frac{\Upsilon}{\alpha}$$

(C.3)

$$P_{out}^{D_{f2}} = 1 - \Pr\left(r_{D_{f2}}^{s_f} > R_1, \frac{1}{3}\min\left(r_n^{UL}, r_n\right) > R_2\right)$$

$$= 1 - \Pr\left(\frac{1}{3}r_n^{UL} > R_2, \frac{1}{3}r_f^{UL} > R_1\right)\Pr\left(\frac{1}{3}r_n > R_2, \frac{1}{3}r_f > R_1\right)\Pr\left(\frac{1}{3}r_f^{DL} > R_1\right)$$

(D.1)

$$\Pr\left(\frac{1}{3}r_n > R_2, \frac{1}{3}r_f > R_1\right) = \Pr\left(|g_{uu}|^2 > \frac{\gamma_{thn}}{\rho\left(b_n - \gamma_{thn}\xi_{uu}b_f\right)} \triangleq \varphi_n, |g_{uu}|^2 > \frac{\gamma_{thf}}{\rho\left(b_f - \gamma_{thf}b_n\right)} \triangleq \varphi_f\right)$$

(D.2)

$$= \Pr\left(|g_{uu}|^2 > \max(\varphi_n, \varphi_f) \triangleq \varphi\right) = e^{-\frac{m_{uu}\varphi}{\Omega_{uu}}} \sum_{s=0}^{m_{uu}-1} \frac{1}{s!}\left(\frac{m_{uu}\varphi}{\Omega_{uu}}\right)^s$$

$$P_{out}^{D_{n2}} = 1 - \Pr\left(r_{D_{n2}}^{s_f} > R_1, r_{D_{n2}}^{s_n} > R_2\right)$$

$$= 1 - \Pr\left(\frac{1}{3}r_n^{UL} > R_2, \frac{1}{3}r_f^{UL} > R_1\right)\Pr\left(\frac{1}{3}r_n > R_2, \frac{1}{3}r_f > R_1\right)\Pr\left(\frac{1}{3}r_{n\to f}^{DL} > R_1, \frac{1}{3}r_n^{DL} > R_2\right)$$

(E.1)

$$\Pr\left(\frac{1}{3}r_{n\to f}^{DL} > R_1, \frac{1}{3}r_n^{DL} > R_2\right) = \frac{2e^{-\bar{\tau}_{n2}}}{R_n^2} \sum_{s_2=0}^{m_{n2}-1}\sum_{p_2=0}^{s_2} \binom{s_2}{p_2}\frac{\bar{\tau}_{n2}^{s_2-p_2-\frac{\alpha}{2}}}{s_2!} \frac{\bar{\Upsilon}_{n2}}{\alpha}.$$

(F.1)

REFERENCES

[1] J. Li, Y. Liu, X. Li, C. Shen, and Y. Chen, "Non-orthogonal multiple access in cooperative UAV networks: A stochastic geometry model," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, Honolulu, HI, USA, Sep. 2019, pp. 1–6.

[2] C. Roy, R. Saha, S. Misra, and K. Dev, "Micro-safe: Microservices- and deep learning-based safety-as-a-service architecture for 6G-enabled intelligent transportation system," *IEEE Trans. Intell. Transp. Syst.*, early access, Sep. 27, 2021, doi: 10.1109/TITS.2021.3110725.

[3] G. Raja, Y. Manaswini, G. D. Vivekanandan, H. Sampath, K. Dev, and A. K. Bashir, "AI-powered blockchain—A decentralized secure multi- party computation protocol for IoV," in *Proc. IEEE INFOCOM Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, Toronto, ON, Canada, 2020, pp. 865–870.

[4] P. K. Selvam, G. Raja, V. Rajagopal, K. Dev, and S. Knorr, "Collision- free path planning for UAVs using efficient artificial potential field algorithm," in *Proc. IEEE 93rd Veh. Technol. Conf. (VTC-Spring)*, Helsinki, Finland, 2021, pp. 1–5.

[5] A. Ranjha, G. Kaddoum, and K. Dev, "Facilitating URLLC in UAV- assisted relay systems with multiple-mobile robots for 6G networks: A prospective of agriculture 4.0," *IEEE Trans. Ind. Informat.*, early access, Nov. 30, 2021, doi: 10.1109/TII.2021.3131608.

[6] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah, "Efficient deploy- ment of multiple unmanned aerial vehicles for optimal wireless cover- age," *IEEE Commun. Lett.*, vol. 20, no. 8, pp. 1647–1650, Aug. 2016.

[7] Y. Zeng, R. Zhang, and T. J. Lim, "Wireless communications with unmanned aerial vehicles: Opportunities and challenges," *IEEE Commun. Mag.*, vol. 54, no. 5, pp. 36–42, May 2016.

[8] F. Jiang and A. L. Swindlehurst, "Optimization of UAV heading for the ground-to-air uplink," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 5, pp. 993–1005, Jun. 2012.

[9] X. Liu *et al.*, "Placement and power allocation for NOMA-UAV networks," *IEEE Wireless Commun. Lett.*, vol. 8, no. 3, pp. 965–968, Jun. 2019.

[10] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah, "Mobile unmanned aerial vehicles (UAVs) for energy-efficient Internet of Things com- munications," *IEEE Trans. Wireless Commun.*, vol. 16, no. 11, pp. 7574–7589, Nov. 2017.

[11] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah, "Unmanned aerial vehicle with underlaid device-to-device communications: Performance and tradeoffs," *IEEE Trans. Wireless Commun.*, vol. 15, no. 6, pp. 3949–3963, Jun. 2016.

[12] X. Li, Q. Wang, Y. Liu, T. A. Tsiftsis, Z. Ding, and A. Nallanathan, "UAV-aided multi-way NOMA networks with residual hardware impair- ments," *IEEE Wireless Commun. Lett.*, vol. 9, no. 9, pp. 1538–1542, Sep. 2020.

[13] Z. Qin, J. Fan, Y. Liu, Y. Gao, and G. Y. Li, "Sparse representation for wireless communications: A compressive sensing approach," *IEEE Signal Process. Mag.*, vol. 35, no. 3, pp. 40–58, May 2018.

[14] X. Li, J. Li, Y. Liu, Z. Ding, and A. Nallanathan, "Residual transceiver hardware impairments on cooperative NOMA networks," *IEEE Trans. Wireless Commun.*, vol. 19, no. 1, pp. 680–695, Jan. 2020.

[15] S. Mounchili and S. Hamouda, "Pairing distance resolution and power control for massive connectivity improvement in NOMA systems," *IEEE Trans. Veh. Technol.*, vol. 69, no. 4, pp. 4093–4103, Apr. 2020.

[16] L. Dai, B. Wang, Y. Yuan, S. Han, I. Chih-Li, and Z. Wang, "Non- orthogonal multiple access for 5G: Solutions, challenges, opportunities, and future research trends," *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 74–81, Sep. 2015.

[17] X. Li *et al.*, "Hardware impaired ambient backscatter NOMA systems: Reliability and security," *IEEE Trans. Commun.*, vol. 69, no. 4, pp. 2723–2736, Apr. 2021.

[18] Y. Liu, Z. Ding, M. Elkashlan, and H. V. Poor, "Cooperative non- orthogonal multiple access with simultaneous wireless information and power transfer," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 938–953, Apr. 2016.

[19] Z. Ding, Z. Yang, P. Fan, and H. V. Poor, "On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users," *IEEE Signal Process. Lett.*, vol. 21, no. 12, pp. 1501–1505, Dec. 2014.

[20] Y. Liu, Z. Qin, M. Elkashlan, Y. Gao, and L. Hanzo, "Enhancing the physical layer security of non-orthogonal multiple access in large- scale networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1656–1672, Mar. 2017.

[21] J. Men, J. Ge, and C. Zhang, "Performance analysis of nonorthog- onal multiple access for relaying networks over Nakagami-*m* fading channels," *IEEE Trans. Veh. Technol.*, vol. 66, no. 2, pp. 1200–1208, Feb. 2017.

[22] X. Li, M. Liu, C. Deng, P. T. Mathiopoulos, Z. Ding, and Y. Liu, "Full- duplex cooperative NOMA relaying systems with I/Q imbalance and imperfect SIC," *IEEE Wireless Commun. Lett.*, vol. 9, no. 1, pp. 17–20, Jan. 2020.

[23] Y. Liu, M. Elkashlan, Z. Ding, and G. K. Karagiannidis, "Fairness of user clustering in MIMO non-orthogonal multiple access systems," *IEEE Commun. Lett.*, vol. 20, no. 7, pp. 1465–1468, Jul. 2016.

[24] W. Mei and R. Zhang, "Uplink cooperative NOMA for cellular- connected UAV," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 3, pp. 644–656, Jun. 2019.

[25] H. E. T. Zheng, A. S. Madhukumar, R. P. Sirigina, and A. K. Krishna, "An outage probability analysis of full-duplex NOMA in UAV com- munications," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Marrakesh, Morocco, Apr. 2019, pp. 1–5.

[26] X. Xi, X. Cao, P. Yang, J. Chen, and D. Wu, "Energy-efficient resource allocation in a multi-UAV-aided NOMA network," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Nanjing, China, 2021, pp. 1–7.

[27] Z. Ding, P. Fan, and H. V. Poor, "Impact of user pairing on 5G nonorthogonal multiple-access downlink transmissions," *IEEE Trans. Veh. Technol.*, vol. 65, no. 8, pp. 6010–6023, Aug. 2016.

[28] Z. Yang, Z. Ding, P. Fan, and N. Al-Dhahir, "A general power allo- cation scheme to guarantee quality of service in downlink and uplink NOMA systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7244–7257, Nov. 2016.

[29] M. F. Kader and S. Y. Shin, "Coordinated direct and relay transmis- sion using uplink NOMA," *IEEE Wireless Commun. Lett.*, vol. 7, no. 3, pp. 400–403, Jun. 2018.

[30] G. Liu, R. Wang, H. Zhang, W. Kang, T. A. Tsiftsis, and V. C. M. Leung, "Super-modular game-based user scheduling and power allocation for energy-efficient NOMA network," *IEEE Trans. Wireless Commun.*, vol. 17, no. 6, pp. 3877–3888, Jun. 2018.

[31] R. Duan, J. Wang, C. Jiang, H. Yao, Y. Ren, and Y. Qian, "Resource allo- cation for multi-UAV aided IoT NOMA uplink transmission systems," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 7025–7037, Aug. 2019.

[32] E. Biglieri, R. Calderbank, A. Constantinides, A. Goldsmith, A. Paulraj, and H. V. Poor, *MIMO Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2007.

[33] M. Mohammadi, B. K. Chalise, H. A. Suraweera, C. Zhong, G. Zheng, and I. Krikidis, "Throughput analysis and optimization of wireless- powered multiple antenna full-duplex relay systems," *IEEE Trans. Commun.*, vol. 64, no. 4, pp. 1769–1785, Apr. 2016.

[34] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*. New York, NY, USA: Academic, 2007.

[35] Y. Ye, Y. Li, D. Wang, F. Zhou, R. Q. Hu, and H. Zhang, "Optimal transmission schemes for DF relaying networks using SWIPT," *IEEE Trans. Veh. Technol.*, vol. 67, no. 8, pp. 7062–7072, Aug. 2018.

Qunshu Wang (Student Member, IEEE) received the B.Sc. degree in electronic information engi- neering from the School of Physics and Electronic Information Engineering, Henan Polytechnic University, Jiaozuo, China, in 2019, where she is currently pursuing the M.Sc. degree in communica- tion and information systems. Her current research interests include nonorthogonal multiple access and unmanned aerial vehicles.

Xingwang Li (Senior Member, IEEE) received the B.Sc. degree from Henan Polytechnic University in 2007, the M.Sc. degree from the University of Electronic Science and Technology of China in 2010, and the Ph.D. degree from the Beijing University of Posts and Telecommunications in 2015. From 2010 to 2012, he was working with Comba Telecom Ltd., Guangzhou, China, as an Engineer. From 2017 to 2018, he was a Visiting Scholar with Queen's University Belfast, Belfast, U.K. He was also a Visiting Scholar with the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications from 2016 to 2018. He is currently an Associate Professor with the School of Physics and Electronic Information Engineering, Henan Polytechnic University, Jiaozuo, China. His research interests include MIMO communication, cooperative communication, hardware constrained communication, nonorthogonal multiple access, physical-layer security, unmanned aerial vehicles, and Internet of Things. He is currently an Editor on the Editorial Board of IEEE ACCESS, *Computer Communications*, and *KSII Transactions on Internet and Information Systems*. He has served as a TPC Member for IEEE/CIC International Conference on Communications in China 2019 and IEEE Global Communications Conference 2018.

Linss T. Alex (Senior Member, IEEE) received the B.Tech. degree in electrical and electronics engineering from Mahatma Gandhi University, India, in 2007, the master's degree in power electronics from Amrita Vishwa Vidyapeetham University, India, in 2011, and the Ph.D. degree from the Electrical and Electronics Engineering Department, SRM Institute of Science and Technology, India, in 2020. He is currently working as an Associate Professor with the Department of Electrical and Electronics Engineering, Albertian Institute of Science and Technology, A. P. J. Abdul Kalam Technological University, Thiruvananthapuram, India. He has ten years of teaching and two years of industrial experience. He has published more than 15 articles in reputed journals, namely, Springer, Taylor & Francis, IEEE TRANSACTIONS, and IEEE ACCESS. He has also reviewed more than 25 articles from the refereed journals which include *Journal of Electrical Engineering & Technology* (Springer), *Computer Communications* (Elsevier), IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS, and *The International Journal of Electrical Engineering & Education* (SAGE journals). His research interests include renewable energy systems, Internet of Things, artificial intelligence, and electric vehicles.

Surbhi Bhatia has rich ten years of teaching and academic experience. She has earned Professional Management professional certification from PMI, USA. She is currently an Assistant Professor with the Department of Information Systems, College of Computer Sciences and Information Technology, King Faisal University, Saudi Arabia. She is also the Adjunct Professor with Shoolini University, Solan, India. She has published more than 70 research papers in reputed journals and conferences in high indexing databases and has seven patents granted from the USA, Australia, and India. She has authored two books and edited nine books from Springer, Wiley, and Elsevier. She has completed five funded research projects from the Deanship of Scientific Research, King Faisal University, and the Ministry of Education, Saudi Arabia. Her research interests include machine learning, sentiment analysis, social media analtics, and information retrieval. She is an Editorial Board Member of *International Journal of Hybrid Intelligence* (Inderscience Publishers) and *SN Applied Sciences* (Springer), and an Associate Editor of *Human-Centric Computing and Information Sciences*. She is currently serving as a guest editor for special issues in reputed journals. She has delivered talks as a Keynote Speaker in IEEE conferences and faculty development programs.

Sunder Ali Khowaja received the Ph.D. degree in industrial and information systems engineering from the Hankuk University of Foreign Studies, South Korea. He has served as an Assistant Professor with the Department of Telecommunication Engineering, University of Sindh, Pakistan. He is currently associated with the Department of Mechatronics Engineering, Korea Polytechnic University, Republic of Korea, in the capacity of Postdoctoral Research Fellow. He is also serving as a Reviewer for many reputed journals, including IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, IEEE ACCESS, IEEE INTERNET OF THINGS JOURNAL, IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING, and IEEE TRANSACTIONS ON MEDICAL IMAGING. He also served as a Technical Program Committee Member of CCNC 2021, Mobicom 2021, and Globecom 2021 workshops. He is currently assisting in the capacity of Guest Editor at *Computers and Electrical Engineering*, *Human-Centric Computing and Information Sciences*, and *Sustainable Energy Assessment and Technologies*. His research interests include data analytics, deep learning, and communication systems-based applications.

Yuanwei Liu (Senior Member, IEEE) received the B.S. and M.S. degrees from the Beijing University of Posts and Telecommunications in 2011 and 2014, respectively, and the Ph.D. degree in electrical engineering from the Queen Mary University of London, U.K., in 2016. He was with the Department of Informatics, King's College London from 2016 to 2017, where he was a Postdoctoral Research Fellow. He has been a Lecturer (Assistant Professor) with the School of Electronic Engineering and Computer Science, Queen Mary University of London since 2017. His research interests include 5G and beyond wireless networks, Internet of Things, machine learning, and stochastic geometry. He received the Exemplary Reviewer Certificate of IEEE WIRELESS COMMUNICATIONS LETTERS in 2015, IEEE TRANSACTIONS ON COMMUNICATIONS in 2016 and 2017, and IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS in 2017 and 2018. He has served as a TPC Member for many IEEE conferences, such as GLOBECOM and ICC. He has served as the Publicity Co-Chair for VTC 2019-Fall. He is currently an Editor on the Editorial Board of the IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE COMMUNICATIONS LETTERS, and IEEE ACCESS. He also serves as a Guest Editor for IEEE JSTSP special issue on Signal Processing Advances for Non-Orthogonal Multiple Access in Next-Generation Wireless Network.

Varun G. Menon (Senior Member, IEEE) received the Ph.D. degree in computer science and engineering from Sathyabama University, India, in 2017. He is currently an Associate Professor with the Department of Computer Science and Engineering, SCMS School of Engineering and Technology, India. He has published more than 50 research papers in peer reviewed and highly indexed international journals and conferences. His research interests include Internet of Things, fog computing and networking, underwater acoustic sensor networks, scientometrics, educational psychology, ad-hoc networks, wireless communication, opportunistic routing, and wireless sensor networks. He is currently a Guest Editor of IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, IEEE SENSORS JOURNAL, and IEEE INTERNET OF THINGS JOURNAL. He is an Associate Editor of *IET Quantum Communications* and also an Editorial Board Member of IEEE Future Directions. He has served over 20 conferences, such as IEEE ICC, ICCCN 2020, IEEE COINS 2020, SigTelCom, ICACCI, and ICDMAI in leadership capacities, including the Program Co-Chair, the Track Chair, the Session Chair, and the Technical Program Committee Member. He is a Distinguished Speaker of ACM.

Towards Energy-Efficient and Delay-Optimized Opportunistic Routing in Underwater Acoustic Sensor Networks for IoUT Platforms: An Overview and New Suggestions

Varun G. Menon [ID],[1,2] Divya Midhunchakkaravarthy,[3] Aaromal Sujith,[2] Sonali John,[2] Xingwang Li,[4] and **Mohammad R. Khosravi**✉ [ID] [5]

[1]Department of Computer Science and Engineering, Lincoln University College, Petaling Jaya 47301, Malaysia

[2]Department of Computer Science and Engineering, SCMS School of Engineering and Technology, Ernakulam 683576, India

[3]Center of Postgraduate Studies, Lincoln University College, Petaling Jaya 47301, Malaysia

[4]School of Physics and Electronic Information Engineering, Henan Polytechnic University, Jiaozuo 454000, China

[5]Department of Computer Engineering, Persian Gulf University, Bushehr, Iran

Show less

Academic Editor: Sumarga Kumar Sah Tyagi

Published: 17 Mar 2022

Hindawi

In underwater acoustic sensor networks (UASNs), the reliable transfer of data from the source nodes located underwater to the destination nodes at the surface through the network of intermediate nodes is a significant challenge due to various unique characteristics of UASN such as continuous mobility of sensor nodes, increased propagation delay, restriction in energy, and heightened interference. Recently, the location-based opportunistic routing protocols seem to show potential by providing commendable quality of service (QoS) in the underwater environment. This study initially reviews all the latest location-based opportunistic routing protocols proposed for UASNs and discusses its possible limitations and challenges. Most of the existing works focus either on improving the QoS or on energy efficiency, and the few hybrid protocols that focus on both parameters are too complex with increased overhead and lack techniques to overcome communication voids. Further, this study proposes and discusses an easy-to-implement energy-efficient location-based opportunistic routing protocol (EELORP) that can work efficiently for various applications of UASN-assisted Internet of Underwater Things (IoUTs) platforms with reduced delay. We simulate the protocol in Aqua-Sim, and the results obtained show better performance than existing protocols in terms of QoS and energy efficiency.

1. Introduction

The genesis of life on Earth had its inception on water, from which it went on to conquer varied frontiers. With the advent of the latest technologies, today's world is more connected than ever before, but ironically the blue planet still lacks efficient underwater connectivity. Underwater acoustic sensor networks (UASNs) [1–3] made their way into the limelight of research quite recently; its pivot objectives deal with an array of versatile interests from oceanographic studies dealing with marine geology, marine ecology, and physical and chemical oceanography. Another significant application is resource extraction, which mainly concerns harnessing abundant rare-earth minerals, petroleum, and natural gas under the sea bed, calamity prevention, deep-sea climate monitoring, and protection, surveillance, and reconnaissance of strategic waters by naval forces around the globe. Conventional methods used for undertaking these tasks mentioned above require humans to physically dive into the ocean's depths or rely on remotely operated underwater vehicles (ROUVs). After the data collection process gets over, these units resurface to provide the output. The data acquired always fell short of fulfilling its objectives as there were problems like the lack of accurate real-time data, stringent storage constraints, inability in handling mobility, and capability to withstand underwater pressure.

UASNs have cell-powered sensor nodes deployed throughout the ocean bed that interact with each other and with the sonobuoys located at the surface to suffice these objectives. Their presence ensures effective communication with the sensor nodes (real time), and they are also the first responders to notify the base station if any nodes fail. Besides all these conspicuous merits, they have been used in underwater acoustic research and antisubmarine warfare for a long time, reflecting its practicality as the UASN's function under tight frequency limitations. Numerous unique features of the underwater environment make the deployment and use of UASN quite a challenging task [4–7]. Underwater conditions are different from the situations on land where the communication takes place with radio frequency (RF) aid. Unfortunately, the underwater environment consumes the energy of the RF waves and renders itself impractical. The mobility of underwater sensors with the ocean currents is another major challenge. To get better off from the challenging underwater situations, UASNs communicate using acoustic waves [8–10]. Underwater acoustic waves typically operate in the frequency range of 10 Hz to 1 MHz. The delay of propagation accompanies this slender range, but it seems to be the only viable choice forward on modicum energy store. UASNs make another edge by facilitating interfaces to communicate with autonomous underwater vehicles (AUVs) remotely. This feature will exponentially increase the range of AUV control, and this merit will also help us perceive the underwater world for research with an added advantage of burgeoning the amount of ocean monitored by human beings, which currently accounts for only 5%. Routing of data

significantly high latency, and reduced reliability [11–17]. Figure 1 presents a sample application scenario of UASNs.
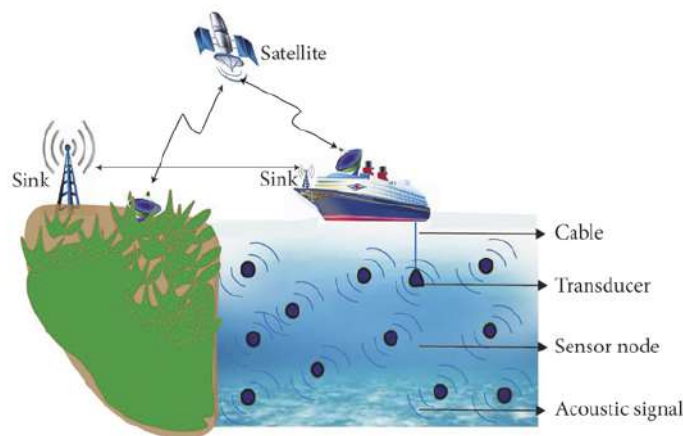

Figure 1

Underwater acoustic sensor networks (UASNs).

Terrestrial wireless sensor networks (TWSNs) have a conventional set of routing protocols that ensure good network performance. TWSNs at no point of operation face interruptions similar to ocean currents. Doppler spreading, interim path loss, and link quality loss create numerous challenges for routing in underwater environments. In a nutshell, the quality of service (QoS) and energy constraints of UASNs inextricably impede it from resorting to routing protocols of TWSNs. Majority of all the routing protocols proposed for TWSNs, thus proving to be powerless when it comes to UASNs. Numerous unconventional routing protocols were put forth in recent years for UASNs, and each focused-on energy efficiency, thus improving various QoS parameters like throughput, latency, load balancing, and robustness. Some of these protocols have already been tested for research, military applications, and catastrophe prediction. The selection of an appropriate routing protocol is significant as it is answerable for the reliable deliverance of data packets to the destination.

Table 1 presents the variations between the terrestrial wireless sensor networks (TWSNs) and underwater acoustic sensor networks (UASNs). Routing protocols in UASN face numerous design challenges. The weightage given to path selection accounts for the various problems that have to be confronted in the underwater environment, such as marine aquatic life, acoustic disturbances, propagation delay, and seismic shadow zones. Many new routing protocols are proposed to tackle these dilemmas; however, most of them lack the description of appropriate routing strategies. Routing strategies advocate the parameters, which will be extensively useful for researchers and other professionals to calibrate the effectiveness of algorithms used in UASNs to develop a strategy to tackle limitations like high propagation delay and energy usage. Picking the suitable scheme ensures engineers achieve desired productivity in applications. The routing protocols for UAWNs are mainly classified into location-based protocols and location-free protocols. The location-based protocols instrument the use of the information contained in the sensor nodes that are mostly two/three-dimensional position coordinates. In contrast, location-free/depth-based protocols depend mainly on pressure information present in sensor nodes. Most of the earlier conventional protocols proposed for UASNs selects the best path for sending data beforehand without

insurmountable dilemma, and the concept of opportunistic routing protocol (ORP) was then proposed [18 –20]. This traction appeared due to various contributing factors like the increased need for extended capacity and expectation of top-notch QoS. The basic idea behind ORP turns the table on the demerit of unreliable transmission, that is, the undesired broadcast nature exhibited by the unreliable transmission is exploited here instead of selecting the nodes beforehand. The selection of nodes in ORP happens on the go. Numerous neighboring nodes (candidate set) receive the broadcasted message. The candidates belonging to the candidate set are sorted according to the metrics and prioritized based on the probability of becoming the next-hop forwarder. The candidate with the highest priority is given the ability to forward the data packets while others discard the packets. This is known as candidate coordination. Opportunistic routing protocols have proved their robustness and adaptability to uncertain conditions by showing their significant presence in many essential fields like oil/gas pipelines, power grids, and management of metro/railroads. Currently, the advancements closely related to ORP have not yet reached their pinnacle as many problems are yet to be solved. However, the prime intention that sleeps behind it is being the ability to make a set of independently weak nodes emerge together as a virtually robust set of links. Thus, ensuring reliability which in turn plummets the retransmission rates and chop down the energy consumption of UAWNs. All of these pros and cons will be thoroughly surveyed in this study.

Table 1

Differences between the TWSN and UASN.

The major contributions of the work are highlighted as follows:

(i) We review all the major location-based opportunistic protocols proposed for routing data packets in underwater acoustic sensor networks over these years. Numerous energy-based, QoS-based, and hybrid location-based opportunistic routing protocols have been proposed in 2019, 2020, and 2021, and they promise to provide much better performance to various real-time applications deployed in UASNs. Very few works have provided reviews on these latest protocols, and we initially tried to address this research gap. We analyze and present a brief description of their working mechanism and highlight their issues and challenges. These issues can be taken up in the future for further improvement in the design of routing protocols in UASN.

(ii) From the review of the latest protocols, it is observed that increased delay and energy drainage are the two significant areas of concern that need further solutions. We then tried to model an easy-to-implement routing protocol that can guarantee energy efficiency with reduced delay to various applications deployed in UASNs.

The rest of the study is organized as follows: in Section 2, various applications using underwater acoustic sensor networks are discussed. The fundamental principles of opportunistic routing are explained in Section 3. Energy-based, QoS-based, and hybrid location-based protocols are discussed in Section 4. The proposed energy-efficient and delay optimized protocol is discussed in detail in Section 5. The future research directions are discussed in Section 6, and the study concludes in the next section.

2. Underwater Acoustic Sensor Networks

The planet we dwell upon is covered approximately with 71% of water. Under this blue element, lies countless untapped resources that will enable human society to advance in countless ways. In order to

sensor nodes can occupy different depth locations that will permit us to spread our reach even to the ocean's deepest places. The self-driven sensor nodes will collect and transfer the sensed data to the target destination using acoustic signals. The attractive applications of UASNs comprise real-time surveillance, disaster prevention, navigation assistance, water quality determination, industrial organization, fish tillage, underwater exploration, and pollution tracking. Almost each of these applications of UASN demands sensor nodes to transfer sensed data timely and precisely through the source node present underwater towards the destination node on the surface with the help of intermediate nodes in the network. However, due to the dynamic nature of UASNs, continuous node mobility, communication voids, and limited battery storage often lead to degraded network performance. In this complicated underwater environment, how to route data packets promptly and effectively even with the presence of a communication void is the most challenging research question.

2.1. Challenges in UASNs

This section presents an overview of various challenges in underwater acoustic sensor networks.

2.1.1. Acoustic Communication

The terrestrial networks depend on RF waves to execute communication. Whereas the underwater environment is highly unpredictable, and RF waves are unfortunately absorbed. Additionally, a high amount of attenuation paves the way to energy loss; hence, RF waves are not an option to be considered. Optical waves cannot be regarded as a choice because the mobility of the nodes is unable to guarantee accuracy. The only viable option in this scenario is acoustic waves. Many other flaws are associated with the usage, but a suitable routing protocol is expected to sort out the dilemma.

2.1.2. High Mobility

The underwater sensor nodes are constantly on the move. These movements are caused by ocean currents which arise due to wind, breaking waves, temperature, and salinity variations. For efficient data gathering, the movement of these nodes is vital and indispensable. However, in reality, this high mobility induces the formation of curves to the acoustic waves, which triggers the emergence of zones that makes some of the sensor nodes in the network unable to participate in data transfer, which brings forth performance issues to the network.

2.1.3. Underwater Noise and Interference

The ocean is packed with a lot of noises and interferences that arise from varied sources. Some of the underwater noises are caused by breaking waves, rain, and marine life. However, various sources are man-made like, shipping, military sonars, fishing, and research activities. These disturbances affect the quality of data packet delivery in the underwater environment.

2.1.4. Low Bandwidth

The operational frequency range of the underwater sensor nodes is primarily restricted due to the usage of acoustic waves as the medium of transmission. The bandwidth is a meager spectrum that lies between 1 kHz and 50 kHz. This zone is significant and less fragmentious to make sure we prevent any interactions

more important to have a wider bandwidth rather than a rate. Moreover, the routing protocols are forced to choose routing paths from this small frequency range for data delivery. However, due to the dynamic nature of UASNs, continuous node mobility, communication voids, and limited battery storage often lead to degraded network performance. In this complicated underwater environment, how to route data packets promptly and effectively even with the presence of a communication void is the most challenging research question.

2.1.5. Low Data Rate

Speed is a crucial factor when it comes to information exchange. The faster the data reaches the destination, the better. Unlike its counterpart (terrestrial environment), the speed at which data is transmitted in the underwater environment is influenced by numerous factors. Firstly, the propagation speed of acoustic waves is inferior to RF waves by many folds that create room for propagation delay. Secondly, there are various persuasive components like depth, temperature, and the degree of salinity of the water. The data rate is meager and accounts for approximately 100 kbps or occasionally a bit more.

2.1.6. Transmission Loss

The hurdles when it comes to underwater sensor network implementation are numerous. Acoustic waves do not guarantee any reliability for the network. On top of that, path loss, Doppler spreading, and high latency will provide a clear picture that there is a considerable amount of packet loss. Transmission loss in any network is not a desirable outcome. Interference is one of the main reasons for packet loss to occur. When the same nodes participate in data transmission continuously for an extended period, the battery can run out, resulting in a communication blackout, and the node will no longer be able to transfer data collected from some places in the network. The acoustic signals have open channels, which are more likely to be utilized by an attacker or malware and to wreak havoc in critical services like routing, localization, and synchronization. Delay variance and bit error are two constituents that can lead to a high amount of packet loss and bit error rates.

2.1.7. Error Prone

The underwater sensor nodes, unlike their conventional counterparts, are not reliable. The environment in which it has been implemented does not support its operation. The mobility, high latency, delay in propagation, high interference, noise, etc., make them highly susceptible to errors. The changes that manifest due to the variation in salinity, depth, and acoustic speed have an implicit effect on making the data transmission process error prone.

2.1.8. High Energy Consumption

Energy drainage is a significant problem in UASNs. Communication between various nodes in a network rudimentarily requires the acknowledgment of its position. The nodes are constantly swaying in harmony with the ocean currents, and it is essential to update their position consistently with their neighbors for effective participation in the data transmission process. Ironically, this position-update process drains quite some energy from these sensor nodes. Another avenue wherein the power consumption rates rocket is when packets have to be retransmitted due to high interference. The data load bestowed on end nodes that connects to the surface stations rapidly depletes the battery power, resulting in the termination of

2.1.9. Channel Attenuation

Channel attenuation is another dilemma that underwater sensor nodes have to confront. The implementation of sensor nodes in the ocean bed is beneficial only if it can collect and transfer data, but due to channel attenuation, the data collected cannot be efficiently extracted from the received signal.

2.1.10. Short Network Lifetime

The pivot grounds for the short lifetime of underwater sensor nodes are due to its source of energy, the battery storage. The nodes extensively consume energy while localization, routing, and data transfer. The hostile underwater condition makes the replacement of cells regularly a challenging task. Therefore, an efficient routing protocol has to consider energy consumption while making decisions on routing path selection. Furthermore, if the nodes run out of energy, it will result in the formation of dead nodes that can affect the network performance and data transmission to the surface stations.

2.1.11. Security and Privacy Issues

UAWNs are made and deployed to monitor places consistently that are far from the shoreline. There is a high chance that the nodes are deployed in strategic waters for specific applications. Attackers can easily manipulate UASNs to inject malicious attacks upon the network. They can also be physically destroyed by enemy divers/AUVs. In the worst-case scenario, attackers can inject fake nodes into the network to provide misguided information and use compromised nodes to extract exclusive data from the network.

2.1.12. Physical Challenges

The sensor nodes have to be fabricated so that they are compact, rigid, and waterproof and should also be able to withstand the pressure of water. Marine life is the next physical challenge that these sensor nodes have to face. In reality, it is physically impossible to protect every one of the nodes in the network.

2.2. Applications of UASNs

This section presents an overview of the major applications that use UASNs.

2.2.1. Military Applications

The military applications of UWSNs can cover a wide range of requirements from monitoring to reconnaissance. In 1982, the United Nations Convention On the Law of the Sea allowed countries to exercise jurisdiction on territorial waters up to 200 nautical miles along the baseline. The naval force guards the coastline against invaders, but the underwater regions in strategically important areas are vulnerable. This vulnerability can be defeated with the help of UASNs. It will enable the militaries to detect enemy divers, submarines, torpedoes, AUVs, and naval mines. The ability to get real-time data will enhance strategic decision-making.

2.2.2. Oceanography

in the ocean and also be able to predict or artificially simulate similar conditions that will benefit human society. Oceanography can provide efficient analysis if it can collect real-time data. UASNs can be utilized to perform experiments to unravel the mysteries of the underwater world consistently, which will indirectly help us to obtain solutions to various problems oceanographic problems.

2.2.3. Coral Conservation

Corals are one of the most beautiful living things on the planet. It takes millions of years to form barrier reefs. The Great Barrier Reef in Australia is the most extensive collection of corals on the phase of the Earth. However, it is dying due to coral bleaching due to the shift in climatic conditions throughout the world. Coral reefs around the globe are on the verge of extinction, and nations are trying to change the situation. Human interference has the likeliness to accelerate the degradation process, but data collection is seemingly impossible without human interference. UASNs are a profound solution that can provide real-time data to conservationists by limiting human interference.

2.2.4. Resource Tapping

The Earth has resources that are tucked away in the depth of the ocean. UASNs provide a way to get the know-how of these varied resources. UASNs will enable us to find out the location, approximate quantity, and dispersion pattern of resources present in the ocean bed. Petroleum and natural gases are an inevitable part of our civilization, which are unfortunately limited. New potential sources can be located for extraction using UASNs.

2.2.5. Fishing, Farming, and Recreation

The fishing industry will benefit from the use of UASNs as they will help locate groups of fish. Underwater farming has been used to cultivate seaweed, lettuce, basil, etc., in countries like Japan and Italy. Knowing the nature of the ocean is an integral part of taking out cruises for recreation purposes deep into the ocean. UASNs can tell leisure seekers about the risks of a tsunami or hurricane before they embark on a cruise.

2.2.6. Disaster Prevention/Prediction

The implementation of UASNs will enable us to detect in advance any underwater earthquakes and volcanic eruptions, which will help us prevent or predict disasters. Many aircraft have gone missing over these years in oceans, and no data about their disappearance was harnessed. One of such shocking incidents is that of Malaysia Airlines 370. UASNs, provided it is implemented correctly, will enhance us to chart the ocean, and it is possible to derive a pinpoint location of any possible crash site.

2.2.7. Climate Change

The rising sea levels and warmth have grabbed international attention. The polar ice caps are at the risk of meltdown. Accurate screening and reports of the polar meltdown can be undertaken with the help of UASNs, which will help researchers and environmentalists to find solutions to these baffling dilemmas.

Hindawi

The uncertainty of the underwater environment poses many underlying threats to establishing efficient communication strategies. The constraints on power and the constant movement of nodes due to tides make opportunistic routing (OR) a viable solution. The opportunistic routing owns different modus operandi. On receiving a data packet, the host node takes into consideration a set of eligible neighboring nodes and prioritizes them based on various parameters. These parameters are different facets like the node closest to the destination and least power draining. A suitable packet forwarder gets opted from the candidate set based on priority and availability. Thus, in the case of unreliable underwater communication, OR proves to be promising as it provides extended reliability, robustness, and QoS than other legacy routing methods.

The principle of opportunistic routing idea was initially developed in ExOR [19] in 2005. The notable advantage of this protocol is that this protocol exploits the multiplex communication opportunities in which the broadcast character belonging to the wireless network develops. The fundamental working of opportunistic routing can be implemented in this protocol, and the three main steps include the following: Initially, the sender node can broadcast the message data. Secondly, upon receiving that data, one relay node is selected as the best forwarder node. After that, the selected best relay node transports the message transmission to the next best relay, and so on. The method is continued until the data reaches the target position. Compared to traditional routing methods, the next-hop relay is selected only after it has received the data, thereby reducing the number of data retransmissions.

The UASN's operation using acoustic channels for communication has many downfalls as there is a prevalence of solid attenuation, time-varying multipath, ambient noise, and modicum propagation speed. All of these contribute to increased delay, error, limited bandwidth, high energy consumption, communication cost, and at times temporary loss of connectivity among nodes of the network. The profound influence of channel fading is crucial to confront as it can directly impact declining routing performance. The scope of application of OR is of paramount importance because it has significantly low retransmission rates, in turn reducing the power consumption. Assured packet delivery facilitation by opportunistic routing ensures no wastage of network resources. It also reduces the chance of system collapse and diminishes retransmission costs. Additionally, opportunistic routing is a versatile choice as it applies to a variety of networks.

Opportunistic routing facilitates a dynamic and instant multiple-path routing technique through opportunistic relay selection, unlike the traditional routing method. Instead of a single precomputed relay, opportunistic routing initially broadcasts a data message to a set of forwarder relay nodes. Fundamentally, these forwarders are organized according to a particular unit. The idea of this method is to select the best forwarder relay among all the nodes. This selected relay manages the process of forwarding the packets. These steps are recursively executed until the packet is transmitted to the target destination. The rudimentary opportunistic routing mechanism consists of the following four steps:

(i) Forwarder relay set choosing

(ii) Broadcasting of data to forwarder nodes

(iii) Coordination scheme is used for best relay selection

(iv) Forwarding of data by the best relay

Each node in opportunistic routing broadcasts a data packet to various adjacent hops periodically. Hence, if communication to one neighbor crashes, another nearby node that has received the data packet can transmit it. OR defines a group of various next hops as the best forwarder relay collection and indicates it as a forwarder relay set (FRS). When a message gets transmitted to FRS, numerous forwarders can obtain a similar message packet. We can avoid duplicate transmission by selecting a single candidate as the best

data packet, and so forth. The leftover candidates will discard this data packet. The FRS selection is divided into three main components as follows: (a) forwarder relay discovery, (b) prioritization variable calculation, and (c) forwarder relays selection, prioritization, and filtering.

To find out the neighbor node, periodic or nonperiodic packets are broadcasted. This neighbor node depends on the link quality, which is changeable as well as dynamic. Hence, this phase is given the charge of computing the stability and quality of the links to reach the neighborhood. Based on these values, a group of nodes is determined. Initially, every node located in the vicinity of the sender node is added to the FRS. Then, the forwarders are taken and sorted based on the chosen variable. After an FRS is elected, priorities are given to forwarder relays based on a specific value. So, the variable selection affects the network throughput significantly. The priority variable election depends on the routing application needs and targets. For some applications, such as emergency recovery, the position data is essential. Consequently, the relay nodes must know their location, and routing can be executed by selecting the relay that is locationally nearest to the target node. Controlling the number of forwarder relays can reduce the overhead and duplicate data transmissions. Moreover, since the size of FRS grows, the number of forwarders who cannot listen to one another also increases. This leads to duplicate packet transmissions. Hence, it is suitable to avoid some forwarder candidates from the FRS. This approach is termed candidate filtering or forwarder. The conventionally used filtering method avoids the forwarders, which are not suited instead of the source. But somehow, this policy cannot guarantee efficient network performance. Another method of FRS creation is discussed in some previous works. The technique is based on implementing an algorithm that can optimize correctly and compute the optimal forwarder relays set for every node, such as forwarder relay sets that are created by Dijkstra's algorithm. However, these methods do not solve the issues of duplicate data transmission.

The optimal relay selection that uses a coordination scheme is used for the coordination of packet forwarding operation between next near nodes. This scheme is responsible for selecting the most suitable forwarder relay to push forward the data packet. Coordination methods need signaling between forwarders. The basic coordination schemes are generally classified into a timer-based, contention-based method, and token-based coordination scheme. In a contention-based method, the main principle is that forwarder relays contend to transmit the data packet with the help of control messages. For example, if a sender node transmits a forward request, its near hop nodes have to compete with themselves to come to a consensus on the forwarding of the data packets. In the timer coordination method, the forwarder relays are supposed to be ranked according to a specific priority value. This rank is commonly added with the message header, which is consistent with the hierarchy in which potential forwarders are permitted to respond. Thus, the largest priority node is allowed to respond to the first slot. The next priority node responds to the upcoming time slot, etc. However, this method is straightforward and easy to carry out, timer-based coordination that incurs some delay, affecting network performance. Another method is the token-based coordination scheme, in which the transferring of data packets is only possible through a token holder. In this scheme, the duplicate message transmission is completely prevented, but it faces increased overhead control. A forwarder (relay) node contains the overhead data packets that are being sent when a token arrives. Tokens travel with connected forwarders because more miniature priority forwarder relays can listen to large priority nodes. If no token arrives, the candidates may be moved into an idle state, slowing down the network.

4. Location-Based Opportunistic Routing Protocols in UWSNs

All routing protocols for UAWNs can be classified as location-based protocols and location-free protocols. The location-based protocols instruments data contained within sensor nodes that are mainly two/three-dimensional position coordinates. At the same time, the location-free protocols/depth-based

fabricate some virtual vector pipe that exists between the source and the destination. Only the nodes in the vicinity of the "vector" right through the source and destination will have the ability to do message forwarding. Hence, routing involves only a tiny group of nodes.

Similarly, numerous location-based protocols were proposed for UASNs. Some of the protocols focused on improving energy efficiency, while others focused on improving QoS parameters like delay and delivery ratio. Recently, many hybrid protocols also have been proposed which consider both energy efficiency and QoS. This section presents a comprehensive discussion on all the latest location-based protocols proposed for UASNs.

VBF makes use of the node location information to make the routing decisions. The knowledge of the position of nodes encourages it to be faster, reliable, and scalable. With the protocol, a virtualized pipe is created from source to destination, and those nodes in the pipe possess a higher probability of becoming the forwarder nodes, while the nodes outside are disregarded. Sink-initiated query and source-initiated query are the two main ways in which VBF addresses routes to different queries. Conceptually, all nodes inside the virtual pipe have the eligibility to forward the packets, but due to limitations like energy, mobility, and propagation delay of acoustic waves, a self-adaptation algorithm was suggested. Another protocol, directional flooding-based routing (DFR) [22] defines a forwarding method formulated by the angle among the center and intermediate nodes. The nodes are responsible for forwarding the data packets through a flooding method. It also considers the quality of the link between the sender and the destination node. A significant concern with this protocol is redundant data transmission and increased energy consumption. The information-carrying routing protocol (ICRP) [23] is an influential conservative, continuous, and versatile directing protocol. The sender hub checks the current location to the final destination when it owns the data to be sent. If there is no current course or path, it starts a path development process by communicating with the information packet, conveying the route disclosure message. Every node present on the network communicates to maintain the reverse route with the information path.

Hop-by-hop vector-based forwarding (HH-VBF) [24] is another variation of VBF where each forwarder resorts to a different routing vector. HH-VBF rudimentarily is just a version of the vector-based forwarding protocol. HH-VBF is a viable option compared to VBF as it can work well with sparse networks and is not liable to the routing pipe radius threshold. However, there is an increase in the computational delay, which in turn degrades the network performance. Reliable and energy balanced routing (REBAR) protocol [25] is a routing protocol that is energy efficient that helps in varying the broadcast domain. REBAR has good reliability and increased lifespan of the network. To balance the energy consumption in the network, a flexible scheme is developed to establish the data propagation range. Here, the nodes near the destination have a modicum radius. However, the increased node movements may expend an excessive amount of energy, resulting in degradation of performance. Vector-based void avoidance (VBVA) [26] protocol is simply an extension of the VBF protocol and focuses on addressing the void problem and energy efficiency in UASNs. The protocol works similarly to VBF when there is no void but uses a revised strategy when voids appear in the network. This helps the protocol to maintain better energy efficiency even with voids in the network.

The energy-efficient and collision aware (EECA) [27] multiple-path routing methods are founded on computing two different collision-free paths using restricted-energy modified flooding. It is one of the earliest protocols that gave equal importance to the betterment of QoS and energy efficiency in the network. Here, multipath power-control transmission (MPT) allows packet data transfer within limited end-to-end data error value and reduced power of transmission. The reliable energy-efficient routing protocol [28] functions on the foundation of link quality, physical data distance, and energy available in the UASN. These three values are calculated and shared with all nodes in the network. The protocol uses a local flooding mechanism with an adaptive selection and gives good reliability and energy efficiency

Hindawi

Location-aware routing protocol (LARP) [29], the GPS is used to identify the exact area of the sink nodes. The sink nodes then broadcast the location information in the network. At the least three sink nodes are used for reference, other nodes in the network calculate their position. The sender can locate the next hop by broadcasting two things as follows: (1) location of the destination node and (2) moving direction of the packet. Packets are forwarded if the receiving node discovers that it is moving in a similar direction. The quality-of-service aware directional flooding-based routing (QoSDFR) [30] extends the DFR protocol. In this routing strategy, the sink node is responsible for sending feedback to various other nodes in the network about the channel condition, and based on the feedback, the optimal forwarder is selected. Protocol results in high throughput because of the limited energy consumption and varying channel conditions.

Scalable and efficient data gathering (SEDG) [31] protocol tries to increase the delivery ratio of the packet and also saves the modicum energy by feasible assignment of the member nodes and gateway node (GN). Here, an autonomous underwater vehicle (AUV) goes through the network area with a precomputed elliptical route and collects data from the gateway node (GN). AUV-aided efficient data gathering (AEDG) routing protocol [32] employs an AUV to gather information from gateways or intermediate nodes and use the shortest path tree (SPT) algorithm to balance the energy consumption. Besides that, AEGD designs a model that improves the result and saves energy by reducing the node members. Furthermore, the nodes live for an extended time to transfer data, thereby increasing delivery chances. The delay-aware energy-efficient routing protocol (DEEP) [33] is a delay-aware routing protocol dependent on collision rate and energy. DEEP makes use of an adaptable node aimed to minimize the collision rate. All the intermediate nodes are elected by virtue of delivery ratio and link quality. In the channel aware routing protocol (CARP) [34], the next-hop transmitter node is elected due to its distance from the previous intermediate node and available energy. Here, every intermediate node is familiar with its neighborhood between the destination node and the next hop. Sender then broadcasts a PING message to the network to compute the next forwarder. Considering a case where the hop value of a sink is lower than the sender node, it replies a PONG data. CARP uses an efficient relay selection method, which doubles the packet delivery ratio.

The novel efficiency forwarding protocol (NEFP) [35] is a proactive anycast routing protocol proposed for UWSNs. It promotes three different approaches. One defines a routing method that avoids unnecessary forwarding of packets where the collision dilemma is averted using a timer. Moreover, finally, the design uses Markov chains to calculate the probability of forwarding the data packets that encourages adaptability to constantly changing network topology. Nevertheless, the performance of the suggested protocol is decreased in the sparse region and as a result, reduces the number of forwards in the phases. Geographic and opportunistic routing protocol with depth adjustment (GEDAR) [36] is a geo-opportunistic routing protocol proposed for a minute-monitoring task. It utilizes a greedy forwarding method to advance the message towards the next hop. The source node chooses the best candidate from the forwarding set. The opportunistic routing in GEDAR reduces the number of retransmissions. GEDAR uses a recovery mode that helps to avoid the void areas. If a node is present in the void area, it will adjust its depth to overcome the void, and new messages will be queued. The greedy strategy will reschedule the node later. Markov model-based routing (MMVR) [37] selects its route from the lower surface to the top level based on changing data traffic. The routes are stable and adaptable, with fewer hops from the sender node and destination. In a localization-based dynamic routing protocol (LBDR) [38], the network is split into smaller layers, and a virtual routing vector is made within the sub-layers. Nodes will move in and out of the virtual vector based on the water current, resulting in high throughput. Void handling geo-opportunistic routing (VHGOR) [39] protocol focuses more on efficiently handling the communication holes in the network. Here, a quick hull algorithm is used to avoid a convex hull. When the node or hub reaches a convex region, rebuilding the convex void helps check an alternate and different way to resume the greedy transmission. VHGOR improves the network performance in networks with voids compared to

Geographic and opportunistic routing (GOR) [40] protocol shows efficient multi-hop data transmission in UWSNs with an upgraded strategy compared to previous protocols. Sometimes, this method gives room for the formation of the void region, and GOR tackles this issue using some void-handling algorithms. The framework considers network density, traffic load, and energy control features to bypass the empty region. The range-based low overhead localization technique (LOTUS) [41] significantly improves on earlier versions of the localization protocols. The protocol can estimate locations based on only two references, enabling this technique to work in networks with fewer nodes. The geographical duplicate reduction flooding (GDflood) [42] considers the location data regarding sensor nodes and joins it with network coding. Energy-efficient grid routing based on 3D cubes (EGRCs) [43] employs a 3D cube network that is subdivided into small cubic clusters. The cluster head is determined based on the remaining energy and position of the intermediate node. All the cluster heads then compute their intermediate node based upon the delay and location. EGRCs reduce energy consumption and end-to-end delay and increase the network performance.

Mobile energy-efficient square routing (MEES) [44] is a routing protocol focusing on energy efficiency in underwater sensor networks. The method uses a division of the network field into dense and sparse regions. A major advantage of this method is that, the mobile sink shifts in a clockwise direction that ensures the highest coverage of nodes in the network which will, in turn, result in high throughput and energy consumption. Topology control vector-based forwarding (TC-VBF) [45] is a revamped version of VBF, which tries to address the limitation of VBF in light conditions. Another protocol energy-efficient multipath grid-based geographic routing protocol (EMGGR) [46] fragments the network into 3D grids. The routing is executed in a grid-by-grid fashion with the help of gateway nodes. Disjoint paths result in high energy efficiency and a good packet delivery ratio. Balanced multiobjective optimized opportunistic routing (BMOOR) protocol [47] uses a strategy where the data from the lower surface takes the best route through the intermediate nodes to the top-level sink. Here, the nodes are located as per dynamic assessment with regards to optimal energy forwarders. The BMOOR protocol needs no spatial data, which is costly in UWSN. The protocol is developed using a generation-based bio-inspired, meta-heuristic algorithm. This helps in delay depreciation and maximization of delivery ratio, and thereby the network lifetime is enhanced. Another proposed protocol for UWSN is energy-efficient interference aware routing (EEIAR) [48] that opts for the best forwarder following the shortest distance. The shortest distance determination decreases the propagation delay. The power control-based sharp directing routing (PCR) [49] selects the most optimal transmission power level available at each submerged sensor node, which helps improve the packet delivery conveyance at each round. Also, it condemns the usage of high-power transmission and the uncontrolled consideration of neighboring hubs in the following hop candidate set, which would end up being the root cause for building the energy utilization on the network. The simulation outcomes depict that PCR diminishes the energy expenditure by adjusting the transmission power and electing the best candidates. The stateless opportunistic routing protocol (SORP) [50]uses a novel method to employ a variable forwarding area that can be reshaped and replaced according to the regional density and placement of the potential forwarding nodes to improve the energy and reliability. The protocol gives good performance compared to the previous protocols. Glider-assisted link disruption restoration mechanism (GALDRM) [51] uses a link disordering recognition with a related link rebuilding method. In the connection acknowledgment system, the group nodes gather the link data. The cluster heads gather the disruption data in link disruption and then schedules gliders as relay nodes to revive the link. Utility capacity is built up by limiting the channel. A multiplier technique illuminates the ideal area of a lightweight flyer. The simulation outputs exhibit a glider-assisted reconditioning procedure that helps to reduce energy consumption. The energy-aware void-avoidable routing protocol (EAVARP) [52] expands the network lifetime and packet delivery rate in underwater sensor networks. EAVARP includes layering and data collection phase with the help of directional forwarding strategy and uses residual energy and data transmission to avoid cyclic transmission and flooding. Fuzzy logic-based VBF

vicinity of the target node. The projection angle allows it to be selected onto the virtual routing vector pipe. The best advantage of this protocol is that it achieves better energy and throughput. However, nodes in the selected vector terminate on dealing with a high load of the message, which is similar to the conditions in VBF.

Mobility-assisted geo-opportunistic routing (MSAGOR) [54] protocol is mainly based on interference avoidance. Here, the network region is fragmented into compact cubes to diminish the interference, which helps to make additional well-informed routing strategies for better energy utilization. Moreover, an optimal number of transmitting nodes are selected from each cube based on its distance to the destination. This proximity will help to avoid void nodes. The extensive simulation results reveal that this protocol will maximize the delivery ratio and network lifetime. Totally, opportunistic routing algorithm (TORA) [55] is an anycast, geographic opportunistic routing protocol proposed for UWSN. The protocol is implemented to avert parallel transmission, bring down end-to-end delay in the network, tame the dilemma of void regions, and enhance network throughput. TORA uses time on arrival and its range-based equation to localize nodes. The energy-aware opportunistic routing (EnOR) [56] is an energy-aware opportunistic routing (EnOR) protocol that can adjust the priority level of forwarding between candidate nodes. This leads to steady energy utilization and increased network lifetime. By using the residual energy, link reliability, packet advancement ratio, and EnOR change the priority of transmission level. Adaptive hop-by-hop cone vector-based forwarding protocol [57] tries to improve the reliability of data transmissions in the sparse sensor regions by making some modifications to the base angle of the cone as per the network structure. These protocols improve the network performance by reducing the number of duplicate packets and also enable a better selection of the potential forwarder node.

Authors in reference [58] discuss implementing a modified strategy for depth-based routing that can transfer the data reliably to the surface sonobuoy. The technique mainly uses the 2-hop neighbor technique and tries to improve the delivery ratio of packets in the network. Authors in reference [59] proposed a technique combining the ant colony optimization algorithm, artificial fish swarm algorithm, and dynamic coded cooperation to improve efficiency by reducing energy consumption. Improving the flexibility of the protocol with the network was one of the major tasks of the proposed algorithm, along with finding the most optimal route. In reference [60], authors presented a Q-learning-based multi-hop cooperative routing protocol for underwater networks. Using this algorithm, the nodes with maximum Q-value were selected as the next forwarders in the network to transfer data from the source to the destination. A coding-aware strategy was proposed for efficient routing in networks with the sparse deployment of nodes [61]. The topological information was used to expand the candidate set using the protocol. An interesting approach that utilizes AUVs to carry sensor nodes to repair the routing voids when foreseeing the occurrence of voids was proposed in reference [62]. The protocol initially predicted the location for repair and then directed the AUVs to the particular location to carry out the repair process. Most of the proposed protocols are complex and incur high overhead, which degrades the performance of the network. Although many of the current protocols improve the data delivery ratio significantly, it comes at the cost of increased energy consumption. It is vital to develop a simple to implement a protocol that can take care of energy efficiency in the network while ensuring reduced delay in the network.

5. Energy-Efficient Location-Based Opportunistic Routing Protocol (EELORP)

In this section, we present the discussion on the proposed energy-efficient location-based opportunistic routing protocol (EELORP) that is designed to provide better energy efficiency and data delivery with minimum delay.

Initially, we try to provide a theoretical analysis to the proposed protocol. The focus is mainly on the delay of transmissions that can be reduced further to enhance the performance of the system. In the underwater network, a delay occurs within two different links, the wireless sensor to the wireless controller and wireless controller to the actuators. The delays are denoted by $T_{w(s-c)}$ and $T_{w(c-a)}$. Assuming the controller to be time invariant, the delays due to two sources are combined together to get total wireless sensor network delay as follows:

$$Twt = Tw\,(s-c) + Tw\,(c-a). \tag{1}$$

The computation delay of the controller can also be included in the total wireless sensor network delay. As the assumption in the wireless controller is time invariant, the decision of controller $d_{(t)}$ is independent of the time it receives the sample $S\,((\gamma b)$. So, the total wireless sensor network delay is only important for us. The analysis of UWSN stability is carried out by assuming two different scenarios as follows: (a) the continuous UWSN network delay system is considered by determining UWSN stability with constant network delay and (b) the discrete UWSN network delay system is considered by determining UWSN stability with time-varying networking delay.

For a continuous UWSN network-delayed system, the UWSN having total network delay as T_{WT} at the time $t = \gamma b$ is considered. The assumption is extended by making $T_{wt} < b$ for all values of "γ" belonging to "S" with $\gamma \in s$. The system is modeled mathematically as follows:

$$\frac{\partial}{\partial t}l\,(t) = Pl\,(t) + Qd\,(t - T_{\mathrm{WT}}), \frac{\partial}{\partial t}l\,(t) = Pl\,(t) + Qd\,(t - T_{\mathrm{WT}}), \tag{2}$$

where "t" belongs to $[\gamma b, \gamma b + b]$. Also with P, Q, R, and S as known matrices, we have the following relation:

$$M\,(t) = R\gamma\,(t) + Sd\,(t). \tag{3}$$

$d\,(t)$ is the received signal with no delay and $d(t - T_{\mathrm{WT}})$ is the received signal with delay. In the case of $d(t)$, that is, received signal with no delay $d(t) = d(\gamma b)$ for $t \in [\gamma b, (\gamma b + b)]$.

Proposition 1. *The UWSN with the delay mentioned above validates the below-mentioned difference equations. The derived equation is as follows:*

$$l\,(\gamma b + b) = \tau l\,(\gamma b) + \varepsilon_0\,(T_{\mathrm{WT}})\,d\,(\gamma b) + \varepsilon_1\,(T_{\mathrm{WT}})\,(d\,(\gamma b - b)). \tag{4}$$

For $\tau = \int_0^b e^{Pb}$, we obtain the following equation:

$$\varepsilon_0\,(T_{\mathrm{WT}}) = \int_0^{b-T_{\mathrm{WT}}} e^{Pb}d\gamma\varphi, \tag{5}$$

$$\varepsilon_1\,(T_{\mathrm{WT}}) = \int_{b-T_{\mathrm{WT}}}^{b} e^{Pb}d\gamma\varphi.$$

Now, we have the following equation:

$$l\,(t) = e^{P_{(t)}} + \int_0^t P\,(t-r)\,\varphi d\,(\gamma)\,d\gamma.\,l\,(t) = e^{P_{(t)}} + \int_0^t P\,(t-r)\,\varphi d\,(\gamma)\,d\gamma. \tag{6}$$

If the delay $t_0 > 0$, then the above equation is rewritten as follows:

$$l\,(t) = e^{P_{(t-t_0)}}l\,(t_0) \tag{7}$$

For for $t > t_0$, we have the following equation:

$$l\left(\gamma b + b\right) = \tau l\left(\gamma b\right) + \epsilon\, d\left(\gamma b\right), \tag{8}$$

where $\tau = e^{Pb}$ and. $\varepsilon = \int_0^b e^{Pr}\, d\gamma\ \varphi$.

Applying equation (7) to (4), we obtain the following equation:

$$l\left(\right)\left(\right) = e^{Pb}l\left(\gamma b\right) + \int_{\gamma b}^{\gamma b+b} e^{P(\gamma b+b-\gamma)}\mathrm{d}\gamma\varphi\mathrm{d}\left(rb - T_{\mathrm{WT}}\right)\mathrm{d}\gamma,$$

$$= \tau l\left(\gamma b\right) + \int_{\gamma b}^{\gamma b+T_{\mathrm{WT}}} e^{P(\gamma b+b-\gamma)}\mathrm{d}\gamma\varphi\mathrm{d}\left(rb\right.$$

$$-\,b) + \int_{\gamma b+T_{\mathrm{WT}}}^{\gamma b+b} e^{P(\gamma b+b-\gamma)}\mathrm{d}\gamma\varphi\mathrm{d}\left(rb\right),$$

$$\gamma^{1} = \left(\gamma b + b - \gamma\right),$$

$$\tau l\left(\gamma b\right) - \int_{b}^{b+T_{\mathrm{WT}}} e^{P\gamma^{1}}\mathrm{d}\gamma^{1}\varphi\mathrm{d}\left(\gamma b - b\right)$$

In other scenarios, multiple copies of the signal are transmitted in the time interval and they take different routes while travelling corresponding to the direct path and scattered path. The spread in the delay indicated as σ_τ of 1 to 3 .... $h(\tau)$ is the delay profile. Taking the Fourier transform, we obtain $H(f) = \int_0^\infty h(\tau)e^{-J2\pi f\tau}\,\mathrm{d}\tau$. The coherence bandwidth at which the delay profile response is almost flat, if the signal bandwidth $\beta_s < \beta_c$ is less than coherence of $\beta_c = 1/2\sigma_\tau$.

Figure 2 shows the signal for $\sigma_\tau \ll T_s$ or $\sigma_\tau \gg T_{\text{signals}}$. So, the sound signal interferes each other significantly and so on as delay spread increases to $\sigma_\tau > T_{\text{signals}}$ and $1/T_{\text{signals}} > 1/\sigma_\tau$, which implies to $\beta_s^- > 2\beta_c$ obtained as the interference. The T_x is stable and R_x is moving towards T_x, indicating the change in the frequency of sound varying due to relative motion between the T_x and R_x.

$$= \tau l\left(\gamma b\right) + \varepsilon_0\left(T_{\text{WT}}d\left(rb\right)\right) + \varepsilon_1\left(T_{\text{WT}}d\left(rb - b\right)\right).$$
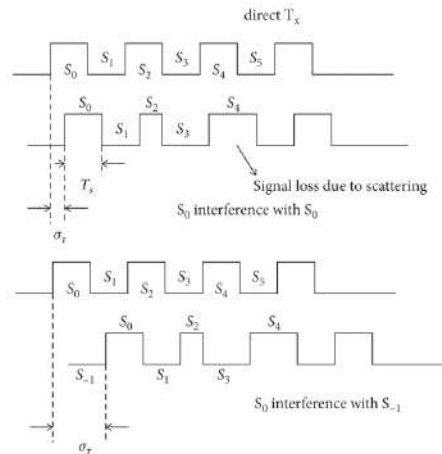
Figure 2

Physical layer-related losses and interference.

5.2. Simulation Results

In this section, we discuss the performance comparison of the proposed EELORP protocol by conducting simulations in Aqua-Sim [63–66]. Aqua-Sim is an extended version of NS-2 and offers easy implementation of underwater network scenarios. The parameters used for setting up the network are given in Table 2.

Table 2

Simulation specifications.

Using the simulations, we measure the energy consumption in nodes and the delay that occurred in the transmission of data in the UWSN. We also compare the results obtained by our proposed work with vector-based forwarding (VBF). Figure 3 shows the energy consumption by nodes in the network. From the results obtained, we can see that the nodes consume less energy using the proposed EELORP protocol

compared to the proposed scheme. This signifies the better energy efficiency offered by the proposed protocol.
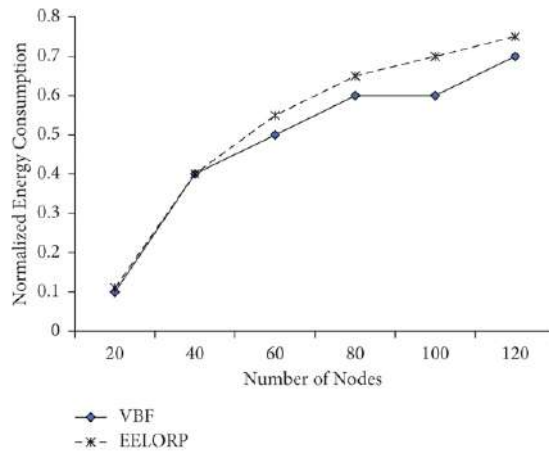
Figure 3

Normalized energy consumption versus the number of nodes.

$$l\,()\,() = e^{Pb} l\,(\gamma b)$$
$$+ \int_{\gamma b}^{b+b} e^{P(\gamma b + b - \gamma)} \mathrm{d}\gamma \varphi \mathrm{d}\,(rb$$
$$- T_{\mathrm{WT}})\,\mathrm{d}\gamma,$$

$$= \tau l\,(\gamma b)$$

Figure 4 shows the delay incurred in transmission of the data packets using the protocols in the UWSN. From the results, we can see that using the proposed method EELORP and VBF, the delay incurred remains almost similar when the number of nodes is less. But as the number of nodes increases, the EELORP has less delay compared to VBF in the network. Thus, our results show that the proposed protocol can be used efficiently for numerous possibilities in underwater acoustic sensor networks with reduced delay.

$$+ \int_{\gamma b + T_{\mathrm{WT}}}^{\gamma b + b} e^{P(\gamma b + b - \gamma)} \mathrm{d}\gamma \varphi \mathrm{d}\,(rb)\,,$$

$$\gamma^1 = (\gamma b + b -$$

$$\tau l\,(\gamma b) - \int_{b}^{b + T_{\mathrm{WT}}}$$

$$- b) - \int_{b - T_{\mathrm{WT}}}^{0}$$

$$= \tau l\,(\gamma b) + \varepsilon_0\,(T$$

$$+ \varepsilon_1\,(T_{\mathrm{WT}} d\,(rb$$
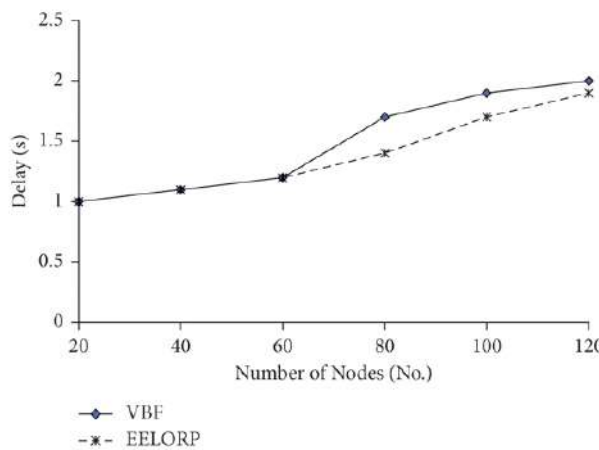
Figure 4

Delay versus number of nodes.

6. Future Research Directions

6.1. Energy Efficiency

This has emerged as one of the major research areas for opportunistic routing protocols in underwater acoustic sensor networks. With restrictions and various limitations in recharging the sensor nodes, it is very important for any routing protocol to minimize the energy usage in the nodes while ensuring that the data gets delivered to the destination. Numerous protocols have tried to improve the energy efficiency in the network, but as UASN has an unpredictable nature, we should for further improvement in this research direction.

6.2. Channel Utilization

The unique features of UASNs like high propagation delay, constant mobility of sensor nodes, high error rate, and interference lead to a major challenge in ensuring the efficient utilization of the channel. Most of the existing protocols have various limitations in channel utilization and this area would be a major area of focus.

6.3. Communication Holes

Dealing with communication holes is a major challenge in UASNs, especially in networks with sparse deployment. Frequent movement of the sensor nodes due to currents and other reasons and failure of sensor nodes due to energy drainage or damage create void areas in the network. Thus, nodes will be unable to find suitable neighbor nodes to forward the data packet to the destination.

6.4. Security

Security of the data transmitted has become one of the major requirements of the applications deploying UASNs. It is therefore vital for all the routing protocols to include a security mechanism that can secure the data from any attackers.

6.5. Reliable Delivery

Reliable delivery of data packet at the destination is a major challenge in UASNs with a dynamic environment. Due to multiple reasons like damage of nodes, lack of energy, voids, etc., the data packet might get lost in the network. It is very important for any protocol to have strategies to manage any data loss and to make sure that the data reaches the destination, also keeping the number of retransmissions to a minimum to save the energy of nodes.

7. Conclusion

This study presented a systematic survey on the location-based opportunistic routing protocols in underwater acoustic sensor networks. The study initially discussed the working of underwater sensor networks, the challenges and issues, the latest applications using UASNs, and the working of

opportunistic routing protocol (EELORP) that can be used efficiently for numerous possibilities in underwater acoustic sensor networks with reduced delay was presented. A discussion on results obtained with simulations was then presented along with comparisons with existing protocols. Finally, a brief discussion on the future research directions was presented.

Data Availability

All the output data with details can be accessed through the first author at reasonable request (email: varunmenon@ieee.org).

Conflicts of Interest

The authors declare no conflicts of interest.

References

1. T. Qiu, Z. Zhao, T. Zhang, C. Chen, and C. L. P. Chen, "Underwater Internet of things in smart ocean: system Architecture and open issues," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 7, pp. 4297–4307, 2000.
   View at: Google Scholar

2. G. Han, C. Zhang, L. Shu, and J. J. P. C. Rodrigues, "Impacts of deployment strategies on localization performance in underwater acoustic sensor networks," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 3, pp. 1725–1733, 2015.
   View at: Publisher Site | Google Scholar

3. S. Fattah, A. Gani, I. Ahmedy, M. Y. I. Idris, and I. A. Targio Hashem, "A survey on underwater wireless sensor networks: requirements, taxonomy, recent advances, and open research challenges," *Sensors*, vol. 20, no. 18, p. 5393, 2020.
   View at: Publisher Site | Google Scholar

4. G. Han, X. Long, C. Zhu, M. Guizani, and W. Zhang, "A high-availability data collection scheme based on multi-AUVs for underwater sensor networks," *IEEE Transactions on Mobile Computing*, vol. 19, no. 5, pp. 1010–1022, 2020.
   View at: Publisher Site | Google Scholar

5. A. M. Khasawneh, O. Kaiwartya, A. Khalifeh, L. M. Abualigah, and J. Lloret, "Green computing in underwater wireless sensor networks pressure centric energy modeling," *IEEE Systems Journal*, vol. 14, no. 4, pp. 4735–4745, 2020.
   View at: Publisher Site | Google Scholar

6. V. Menon, D. Midhunchakkaravarthy, S. Verma, A. Sujith, and M. Manju, "Enabling reliable communication in Internet of underwater things: applications, challenges and future directions," in *Proceedings of the 2nd International Conference on Secure Cyber Computing and*

Hindawi

7. G. Han, H. Wang, J. A. Ansere, J. Jiang, and Y. Peng, "SSLP: a stratification-based source location privacy scheme in underwater acoustic sensor networks," *IEEE Network*, vol. 34, no. 4, pp. 188–195, 2020.
   View at: Publisher Site | Google Scholar

8. V. G. Menon, "Opportunistic routing protocols in underwater acoustic sensor networks: issues, challenges, and future directions," *Magnetic Communications*, CRC Press, FL, USA, pp. 109–130, 2018.
   View at: Publisher Site | Google Scholar

9. H. Cui, Y. Zhang, X. Liu, and D. Sun, "The simulation and emulation platforms of underwater acoustic sensor networks," in *Proceedings of the IEEE/OES China Ocean Acoustics (COA)*, pp. 1–5, Harbin, China, January 2016.
   View at: Publisher Site | Google Scholar

10. Y. Chen, X. Jin, and X. Xu, "Mobile data collection paths for node cooperative underwater acoustic sensor networks," in *Proceedings of the OCEANS*, pp. 1–5, Shanghai,China, April 2016.
    View at: Publisher Site | Google Scholar

11. M. Erol-Kantarci, H. T. Mouftah, and S. Oktug, "A survey of architectures and localization techniques for underwater acoustic sensor networks," *IEEE Communications Surveys & Tutorials*, in *IEEE Communications Surveys & Tutorials*, vol. 13, no. 3, pp. 487–502, 2011.
    View at: Publisher Site | Google Scholar

12. K. S. Keerthi, B. Mahapatra, and V. G. Menon, "Into the world of underwater swarm robotics: architecture, communication, applications and challenges," *Recent Advances in Computer Science and Communications*, vol. 13, no. 2, pp. 110–119, 2020.
    View at: Publisher Site | Google Scholar

13. J. Yan, Y. Gong, C. Chen, X. Luo, and X. Guan, "AUV-aided localization for Internet of underwater things: a reinforcement-learning-based method," *IEEE Internet of Things JournalEEE Internet of Things Journal*, vol. 7, no. 10, pp. 9728–9746, 2020.
    View at: Publisher Site | Google Scholar

14. W. Zhang, G. Han, X. Wang, M. Guizani, K. Fan, and L. Shu, "A node location algorithm based on node movement prediction in underwater acoustic sensor networks," *IEEE Transactions on Vehicular Technology, IEEE Transactions on Vehicular Technology*, vol. 69, no. 3, pp. 3166–3178, 2020.
    View at: Publisher Site | Google Scholar

15. Y. Song, "Underwater acoustic sensor networks with cost efficiency for Internet of underwater things," *IEEE Transactions on Industrial Electronics, in IEEE Transactions on Industrial Electronics*, vol. 68, no. 2, pp. 1707–1716, 2021.
    View at: Publisher Site | Google Scholar

16. S. Zheng, X. Cao, F. Tong, G. Zhang, and Y. Dong, "Performance evaluation of acoustic network

International Conference on Underwater System Technology: Theory and Applications (USYS), pp. 1–4, Wuhan, China, December 2018.
View at: Publisher Site | Google Scholar

17. S. Y. Kulik, A. Y. Rodionov, F. S. Dubrovin, and P. P. Unru, "On reliability of data transmission and distance estimation using mobile underwater acoustic modems," in *Proceedings of the 25th Saint Petersburg International Conference on Integrated Navigation Systems (ICINS)*, pp. 1–4, St. Petersburg, Russia, May 2018.
View at: Publisher Site | Google Scholar

18. S. Biswas and R. Morris, "ExOR, ACM SIGCOMM - Computer Communication Review," in *Proceedings of the 2005 conference on Applications, technologies, architectures, and protocols for computer communications*, pp. 133–144, USA, October 2005.
View at: Publisher Site | Google Scholar

19. M. Ismail, M. Islam, I. Ahmad, F. Aslam, and A. Baseer, "Reliable path selection and opportunistic routing protocol for underwater wireless sensor networks," *IEEE Access*, vol. 8, pp. 100346–100364, 2020.
View at: Publisher Site | Google Scholar

20. V. G. Menon and P. M. J. Prathap, "Comparative analysis of opportunistic routing protocols for underwater acoustic sensor networks," in *Proceedings of the International Conference on Emerging Technological Trends (ICETT)*, pp. 1–5, Kollam, India, October 2016.
View at: Publisher Site | Google Scholar

21. P. Xie, J. Cui, and L. Lao, "VBF: vector-based forwarding protocol for underwater sensor networks," *NETWORKING 2006. Networking Technologies, Services, and Protocols; Performance of Computer and Communication Networks; Mobile and Wireless Communications Systems*, vol. 3976, pp. 1216–1221, 2006.
View at: Publisher Site | Google Scholar

22. H. Daeyoup and D. Kim, "DFR: directional flooding-based routing protocol for underwater sensor networks," in *Proceedings of the IEEE Oceans*, Quebec, QC, Canada, September 2007.
View at: Publisher Site | Google Scholar

23. W. Liang, H. Yu, L. Liu, B. Li, and C. Che, "Information-carrying based routing protocol for underwater acoustic sensor network," in *Proceedings of the International Conference on Mechatronics and Automation (ICMA 2007)*, pp. 729–734, Harbin, China, August 2007.
View at: Publisher Site | Google Scholar

24. N. Nicolaou, A. See, P. Xie, J. H Cui, and D. Maggiorini, "Improving the robustness of location-based routing for underwater sensor networks," in *In Proceedings of the IEEE Oceans*, Aberdeen, UK, September 2008.
View at: Google Scholar

25. J. Chen, X. Wu, and G. Rebar Chen, "A reliable and energy balanced routing algorithm for

View at: Google Scholar

26. P. Xie, Z. Zhou, Z. Peng, J. –H. Cui, and Z. Chi, "Void Avoidance in three-dimensional mobile underwater sensor networks," in *In Proceedings of the the 4th International Conference on Wireless Algorithms, Systems and Applications*, Boston, MA, USA, August 2009.
View at: Publisher Site | Google Scholar

27. M. V. PriyaA. A. Kumari, "Traffic aware multipath communication for time-critical applications in underwater acoustic sensor networks," *International Journal of Managment, IT and Engineering*, vol. 2, pp. 66–73, 2012.
View at: Google Scholar

28. P. Wang, D.-hao Fu, C.-qing Zhao, J.-chun Xing, Qi-liang Yang, and X.-fei Du, "A reliable and efficient routing protocol for Underwater Acoustic Sensor Networks," in *Proceedings of the IEEE International Conference on Cyber Technology in Automation, Control and Intelligent Systems*, Nanjing, China, May 2013.
View at: Publisher Site | Google Scholar

29. J. She, J. Wang, J. Wang, J. Zhang, and S. Wang, "Location-aware routing protocol for underwater sensor networks," in *Proceedings of the Advanced Technologies, Embedded and Multimedia for Human-Centric Computing*, Springer, New York, NY, USA, November 2014.
View at: Publisher Site | Google Scholar

30. J. Park, S. Lee, D. Kim, and Y. Hong, "QoS-aware directional flooding-based routing for underwater wireless sensor networks," in *Proceedings of the WUWNET 14th International Conference on Underwater Networks and Systems*, Rome, Italy, November 2014.
View at: Publisher Site | Google Scholar

31. I. Naveed, A. TurkiAli, N. F. Muhammad et al., "SEDG: Scalable and Efficient Data Gathering Routing Protocol for Underwater WSNs," *Procedia Computer Science*, Elsevier, Amsterdam, Netherland, vol. 2, pp. 568–575, 2015.
View at: Google Scholar

32. I. Naveed, A. Turki, N. F. Muhammad et al., "AEDG: AUV-aided Efficient Data Gathering Routing Protocol for Underwater Wireless Sensor Network," *Elsevier, Procedia Computer Science*, vol. 2, pp. 568–575, 2015.
View at: Google Scholar

33. R. W. L. Coutinho, A. Boukerche, L. F. M. Vieira, and A. A. F. Loureiro, "Modeling and analysis of opportunistic routing in low duty-cycle underwater sensor networks," in *Proceedings of the 18th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, pp. 125–132, New York, NY, USA, November 2015.
View at: Publisher Site | Google Scholar

34. S. Basagni, C. Petrioli, R. Petroccia, and D. S.. Carp, "A channel-aware routing protocol for underwater acoustic wireless networks," *Ad Hoc Networks*, vol. 34, pp. 92–104, 2015.

Hindawi

35. Q. Wang, C. Fei, L. Zhi, and Q. Qian, "A novel efficient forwarding protocol for 3-D underwater wireless sensor networks," in *In Proceedings of the IEEE 11th International Conference on Industrial Electronics and Applications*, Hefei, China, June 2016.
View at: Publisher Site | Google Scholar

36. R. W. L. Coutinho, A. Boukerche, L. F. M. Vieira, and A. A. F. Loureiro, "Geographic and opportunistic routing for under water sensor networks," *IEEE Transactions on Computers*, vol. 65, pp. 548–561, 2016.
View at: Publisher Site | Google Scholar

37. D. Li, J. Du, and L. Liu, "A data routing algorithm based on Markov model in underwater wireless sensor networks," in *Proceedings of the IEEE 16th International Conference on Ubiquitous Wireless Broadband*, Nanjing, China, December 2016.
View at: Publisher Site | Google Scholar

38. S. Han, Y. Yue, W. Meng, and X. Wu, "A localization-based routing protocol for dynamic underwater sensor networks," in *Proceedings of the IEEE Global Commuications Conference*, Washington, DC, USA, December 2016.
View at: Publisher Site | Google Scholar

39. N. Kanthimathi and Dejey, "Void handling using Geo-Opportunistic Routing in underwater wireless sensor networks," *Computers & Electrical Engineering*, vol. 64, pp. 365–379, 2017.
View at: Publisher Site | Google Scholar

40. R. W. L. Coutinho, A. Boukerche, L. F. M. Vieira, and A. A. F. Loureiro, "Performance modeling and analysis of void-handling methodologies in underwater wireless sensor networks," *Computer Networks*, vol. 126, pp. 1–14, 2017.
View at: Publisher Site | Google Scholar

41. M. Yusuf and S. Uddin, "Low-overhead range-based 3D localization technique for underwater sensor networks," in *Proceedings of the IEEE International Conference on Communications*, Kuala Lumpur, Malaysia, May 2016.
View at: Publisher Site | Google Scholar

42. E. Isufi, H. Dol, and G. Leus, "Advanced flooding-based routing protocols for underwater sensor networks," *EURASIP Journal on Applied Signal Processing*, vol. 2016, pp. 1–12, 2016.
View at: Publisher Site | Google Scholar

43. K. Wang, H. Gao, X. Xu, J. Jiag, and D. Yue, "An energy-efficient reliable data transmission scheme for complex environmental monitoring in underwater acoustic sensor networks," *IEEE Sensors Journal*, vol. 16, pp. 4051–4062, 2016.
View at: Publisher Site | Google Scholar

44. A. walayat, N. Javaid, M. Akbar, and Z. A. Khan, "MEES: mobile energy efficient square routing for underwater wireless sensor networks," in *Proceedings of the IEEE 31st International Conference on Advanced Information Networking and Applications*, Taipei, Taiwan, May 2017.

45. I. Yazgi and B. Baykal, "Topology control vector based forwarding algorithm for underwater acoustic networks," in *Proceedings of the IEEE 24th International Conference on Signal Processing and Communication Application*, Zonguldak, Turkey, May2016.
View at: Publisher Site | Google Scholar

46. F. A. Salti, N. Alzeidi, and B. R. Arafeh, "EMGGR: an energy-efficient multipath grid-based geographic routing protocol for underwater wireless sensor networks," *Wireless Networks*, vol. 23, no. 4, pp. 1301–1314, 2017.
View at: Publisher Site | Google Scholar

47. N. Kanthimathi and Dejey, "Balanced and Multi-objective Optimized Opportunistic Routing for Underwater Sensor Networks," *Wireless Personal Communications*, vol. 94, no. 4, pp. 2417–2440, 2017.
View at: Publisher Site | Google Scholar

48. A. Khan, N. Javaid, I. Ali, and H. A. Mohammad, "An energy efficient interference aware routing protocol for underwater WSNs," *KSII Transactions on Internet and Information Systems*, vol. 11, no. 11 10, 2017.
View at: Google Scholar

49. R. W. L. Coutinho, A. Boukerche, L. F. M. Vieira, and A. A. F. Coutinho, "PCR: a power control-based opportunistic routing for underwater sensor networks," in *Proceedings of the MSWiM 2018 - Proceedings of the 21st ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, pp. 173–180, New York, NY, USA, October 2018.
View at: Publisher Site | Google Scholar

50. S. M. Ghoreyshi, A. Shahrabi, and T. Boutaleb, "A stateless opportunistic routing protocol for underwater sensor networks," *Wireless Communications and Mobile Computing*, vol. 2018, Article ID 8237351, 18 pages, 2018.
View at: Publisher Site | Google Scholar

51. Z. Jin, N. Wang, Y. Su, and Q. Yang, "A glider-assisted link disruption restoration mechanism in underwater acoustic sensor networks," *Sensors*, vol. 18, no. 2, 2018.
View at: Publisher Site | Google Scholar

52. W. Zhuo, H. Guangjie, Q. Hongde, Z. Suping, and S. Yancheng, "An energy - aware and void-avoidable routing protocol for underwater sensor networks," *IEEE Access, Journals & Magazines*, vol. 6, pp. 7792–7801, 2018.
View at: Google Scholar

53. R. Bu, S. Wang, and H. Wang, "Fuzzy logic vector–based forwarding routing protocol for underwater acoustic sensor networks," *Trans. Emerg. Telecommun. Technol.*, vol. 29, pp. 1–18, 2018.
View at: Publisher Site | Google Scholar

54. F. Ahmed, Z. Wadud, N. Javaid, N. Alrajeh, M. S. Alabed, and U. Qasim, "Mobile sinks assisted

View at: Publisher Site | Google Scholar

55. Z. Rahman, F. Hashim, M. F. A. Rasid, and M. Othman, "Totally opportunistic routing algorithm (TORA) for underwater wireless sensor network," *PLoS One*, vol. 13, no. 6, 2018.
View at: Publisher Site | Google Scholar

56. R. W. L. Coutinho, A. Boukerche, L. F. M. Vieira, and A. A. F. Loureiro, "EnOR: Energy balancing routing protocol for underwater sensor networks," in *Proceedings of the IEEE International Conference on Communications*, Paris,France, May 2017.
View at: Publisher Site | Google Scholar

57. I. U. Khan, M. Islam, M. Ismail et al., "Adaptive hop-by-hop cone vector-based forwarding protocol for underwater wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 16, no. 9, Article ID 15501477209, 2020.
View at: Publisher Site | Google Scholar

58. M. Zhang and W. Cai, "Energy-efficient depth based probabilistic routing within 2-hop neighborhood for underwater sensor networks," *in IEEE Sensors Letters*, vol. 4, no. 6, pp. 1–4, 2020.
View at: Publisher Site | Google Scholar

59. Y. Chen, J. Zhu, L. Wan, S. Huang, X. Zhang, and X. Xu, "ACOA-AFSA fusion dynamic coded cooperation routing for different scale multi-hop underwater acoustic sensor networks," *in IEEE Access*, vol. 8, Article ID 186788, 2020.
View at: Publisher Site | Google Scholar

60. Y. Chen, K. Zheng, X. Fang, L. Wan, and X. Xu, "QMCR: a Q-learning-based multi-hop cooperative routing protocol for underwater acoustic sensor networks," *in China Communications*, vol. 18, no. 8, pp. 224–236, 2021.
View at: Publisher Site | Google Scholar

61. D. Zhao, G. Lun, and R. Xue, "Coding-aware opportunistic routing for sparse underwater wireless sensor networks," *IEEE Access*, vol. 9, Article ID 50187, 2021.
View at: Publisher Site | Google Scholar

62. Z. Jin, Q. Zhao, and Y. Luo, "Routing void prediction and repairing in AUV-assisted underwater acoustic sensor networks," *IEEE Access*, vol. 8, Article ID 54212, 2020.
View at: Publisher Site | Google Scholar

63. M. R. Khosravi, H. Basri, and H. Rostami, "Efficient routing for dense UWSNs with high-speed mobile nodes using spherical divisions," *The Journal of Supercomputing*, vol. 74, no. 2, pp. 696–716, 2018.
View at: Publisher Site | Google Scholar

64. M. R. Khosravi, H. Basri, H. Rostami, and S. Samadi, "Distributed random cooperation for VBF-based routing in high-speed dense underwater acoustic sensor networks," *The Journal of Supercomputing*, vol. 74, no. 11, pp. 6184–6200, 2018.

65. M. R. Khosravi, "The shortfalls of underwater sensor network simulators," *Sea Technology*, vol. 60, no. 5, p. 41, 2019.
   View at: Google Scholar

66. P. Xie, Z. Zhou, Z. Peng et al., "Aqua-Sim: an NS-2 based simulator for underwater sensor networks," in *Proceedings of the Oceans 2009*, pp. 1–7, Biloxi, MS, USA, October 2009.
   View at: Publisher Site | Google Scholar

Copyright

More related articles

A Novel Hybridized Cluster-Based Geographical Opportunistic Routing Protocol for Effective Data Routing in Underwater Wireless Sensor Networks

B. Ragavi | V. Baranidharan | K. Ramash Kumar

Delay-Driven Opportunistic Routing with Multichannel Cooperative Neighbor Discovery for Industry 4.0 Wireless Networks Based on Power and Load Awareness

Aditya Pai H. | Khalid K. Almuzaini | … | Reynah Akwafo

Multipath Route Optimization with Multiple QoS Constraints Based on Intuitionistic Fuzzy Set Theory

Peng Ren | Ran Zhang | Shuhang Luo

Energy-Efficient and Secure Opportunistic Routing Protocol for WSN: Performance Analysis with Nature-Inspired Algorithms and Its Application in Biomedical Applications

Download other formats

Order printed copies

Related articles

Hindawi

Survey

Kofi Sarpong Adu-Manu | Felicia Engmann | ... | Bernice Akusika Dulemordzi

A Directional Selective Power Routing Protocol for the Internet of Underwater Things

Manal Al-Bzoor | Ahmed Musa | ... | Taha Gharaibeh

Energy Efficiency and Reliability Considerations in Wireless Body Area Networks: A Survey

Farman Ullah | M. Zahid Khan | ... | Muhammad Fayaz

Follow us:

About Us

Contact us

Partnerships

Blog

Journals

Article Processing Charges

Print editions

Authors

Editors

Reviewers

Partnerships

Hindawi XML Corpus

Hindawi

Browse ⌄   My Settings ⌄   Help ⌄   Institutional Sign In

Institutional Sign In

All ⌄   🔍

ADVANCED SEARCH

Exploiting Benefits of IRS in Wireless Powered NOMA Networks

Publisher: IEEE   Cite This   📄 PDF

Xingwang Li ; Zhen Xie ; Zheng Chu ; Varun G. Menon ; Shahid Mumtaz ; Jianhua Zhang   **All Authors** •••

Ⓡ ⤶ © 🗁 🔔

Alerts

Manage Content Alerts

Add to Citation Alerts

Abstract

Document Sections

I. Introduction

II. System Model

III. Sum Throughput Maximization

IV. Numerical Results

V. Conclusion

Authors

Figures

References

Citations

Keywords

Metrics

More Like This

📄
Downl
PDF

Abstract: We consider an intelligent reflecting surface (IRS) wireless-powered NOMA Internet-of-Things (IoT) network, where multiple NOMA IoT devices can communicate with access po... **View more**

▸ **Metadata**

Abstract:

We consider an intelligent reflecting surface (IRS) wireless-powered NOMA Internet-of-Things (IoT) network, where multiple NOMA IoT devices can communicate with access point (AP) with the aid of IRS. Specifically, IoT devices can harvest energy from a nearby appropriative power station (PS) during the wireless energy transfer (WET) phase through the direct links as well as the reflecting links from IRS. Then, the harvested energy can achieve the wireless information transfer (WIT) during the uplink between IoT devices and AP by adopting NOMA protocol. A novel resource allocation scheme is proposed to maximized the sum throughput of the considered system by jointly optimizing the time allocation factor and phase shift matrices of WET and WIT. The optimization problem is non-convex due to multiple variables. We first transfer this problem into convex function by assuming one of the phase shift matrices fixed, and solve the transferred problem by obtaining closed-form solution. Then, the elements collaborative approximate (ECA) and the manifold space gradient descent (MSGD) algorithms are designed to optimize the phase shift matrix. ECA is iteratively optimizing one element and MSGD is focus on the derivation of a gradient descent over the manifold space. Numerical results indicate the sum throughput enhancement by IRS compared with no IRS, and highlight the advantages of IRS.

▸ **Funding Agency:**

Xingwang Li
School of Physics and Electronic Information Engineering, Henan
Polytechnic University, Jiaozuo, China

Xingwang Li (Senior Member, IEEE) received the M.Sc. degree from the University of Electronic Science and Technology of China in 2010 and the Ph.D. degree from the Beijing University of Posts and Telecommunications in 2015. From 2010 to 2012, he worked with Comba Telecom Ltd., Guangzhou China, as an Engineer. He spent one year from 2017 to 2018 as a Visiting Scholar with Queen's University Belfast, Belfast, U.K. He is cur... Show More

Xingwang Li (Senior Member, IEEE) received the M.Sc. degree from the University of Electronic Science and Technology of China in 2010 and the Ph.D. degree from the Beijing University of Posts and Telecommunications in 2015. From 2010 to 2012, he worked with Comba Telecom Ltd., Guangzhou China, as an Engineer. He spent one year from 2017 to 2018 as a Visiting Scholar with Queen's University Belfast, Belfast, U.K. He is cur... **View more**

Zhen Xie
School of Physics and Electronic Information Engineering, Henan
Polytechnic University, Jiaozuo, China

Zhen Xie received the B.Sc. degree in electronic information engineering from the School of Physics and Electronic Information Engineering, Henan Polytechnic University, Jiaozuo, China, in 2020, where she is currently pursuing the M.Sc. degree in communication and information systems. Her current research interests include nonorthogonal multiple access and intelligent reflecting surface.

Zhen Xie received the B.Sc. degree in electronic information engineering from the School of Physics and Electronic Information Engineering, Henan Polytechnic University, Jiaozuo, China, in 2020, where she is currently pursuing the M.Sc. degree in communication and information systems. Her current research interests include nonorthogonal multiple access and intelligent reflecting surface. **View more**

Zheng Chu
5GIC & 6GIC, Institute for Communication Systems, University of
Surrey, Guildford, U.K.

Zheng Chu (Member, IEEE) received the Ph.D. degree from Newcastle University, Newcastle upon Tyne, U.K, in 2016. He was with the Faculty of Science and Technology, Middlesex University, London, U.K., from 2016 to 2017. He is currently with the Institute for Communication Systems, University of Surrey, Guildford, U.K. His current research interests include smart radio environments/smart reflecting surface, 5G/6G communicat... Show More

Zheng Chu (Member, IEEE) received the Ph.D. degree from Newcastle University, Newcastle upon Tyne, U.K, in 2016. He was with the Faculty of Science and Technology, Middlesex University, London, U.K., from 2016 to 2017. He is currently with the Institute for Communication Systems, University of Surrey, Guildford, U.K. His current research interests include smart radio environments/smart reflecting surface, 5G/6G communicat... **View more**

Varun G. Menon
Department of Computer Science and Engineering, SCMS School of
Engineering and Technology, Ernakulam, India

Varun G. Menon (Senior Member, IEEE) received the Ph.D. degree in computer science and engineering from Sathyabama University, India, in 2017. He is currently an Associate Professor with the Department of Computer Science and Engineering, SCMS School of Engineering and Technology, India. He has published more than 50 research papers in peer reviewed and highly indexed International Journals and Conferences. His research i... Show More

Varun G. Menon (Senior Member, IEEE) received the Ph.D. degree in computer science and engineering from Sathyabama University, India, in 2017. He is currently an Associate Professor with the Department of Computer Science and Engineering, SCMS School of Engineering and Technology, India. He has published more than 50

research papers in peer reviewed and highly indexed International Journals and Conferences. His research i... **View more**

Shahid Mumtaz
Instituto de Telecomunicações, Aveiro, Portugal

Shahid Mumtaz (Senior Member, IEEE) is a Senior 5G Consultant with Huawei, Sweden. He is serving as a Scientific Expert and an Evaluator for various Research Funding Agencies. He has authored four technical books, 12 book chapters, more than 250 technical papers (more than 150 Journal/transaction, more than 80 conference, two IEEE best paper awards—in the area of mobile communications. Most of his publication is in the fi... Show More

Shahid Mumtaz (Senior Member, IEEE) is a Senior 5G Consultant with Huawei, Sweden. He is serving as a Scientific Expert and an Evaluator for various Research Funding Agencies. He has authored four technical books, 12 book chapters, more than 250 technical papers (more than 150 Journal/transaction, more than 80 conference, two IEEE best paper awards—in the area of mobile communications. Most of his publication is in the fi... **View more**

Jianhua Zhang
State Key Laboratory of Networking and Switching Technology,
Beijing University of Posts and Telecommunications, Beijing, China

Jianhua Zhang (Senior Member, IEEE) received the B.S. degree from the North China University of Technology in 1994, and the Ph.D. degree from the Beijing University of Posts and Telecommunications in 2003, where she is currently the Professor with Information and Engineering College. She has published more than 70 journal papers and nearly 200 conference papers. Her research interests are massive MIMO and millimeter wave ... Show More

Jianhua Zhang (Senior Member, IEEE) received the B.S. degree from the North China University of Technology in 1994, and the Ph.D. degree from the Beijing University of Posts and Telecommunications in 2003, where she is currently the Professor with Information and Engineering College. She has published more than 70 journal papers and nearly 200 conference papers. Her research interests are massive MIMO and millimeter wave ... **View more**

:≡ Contents

I. Introduction

Internet-of-Things (IoT) technology is indispensable to the beyond fifth-generation (B5G) by delivering specified functions through connecting devices or "things" over the mobile Internet, which can greatly improve the access rate of mass devices [1]–[3]. The IoT has another advantage to connect a huge number of devices for information transfer and/or exchange [4]. As one of the core components of the IoT network, machine-type communication (MTC) devices are assumed to be maintained without manual operation [5]. The sensors of IoT are designed with low power and always be powered by the batteries [6]. Therefore, this needs to periodic maintenance or change batteries to prolong their service time [7]. However, to face various applications of IoT, sensors will be deployed any scenarios, such as deployed in hospital to detect temperature and humidity [8], as well as can be worn by people to monitor their environments and physical characteristics in real time [9]. To this end, it is urgent to prolong lifetime of senors, but with limitations of battery technology, it is still an open problem [10]–[12].

Sign in to Continue Reading

Authors ⌃

Xingwang Li
School of Physics and Electronic Information Engineering, Henan Polytechnic University, Jiaozuo, China

Xingwang Li (Senior Member, IEEE) received the M.Sc. degree from the University of Electronic Science and Technology of China in 2010 and the Ph.D. degree from the Beijing University of Posts and Telecommunications in 2015. From 2010 to 2012, he worked with Comba Telecom Ltd., Guangzhou China, as an Engineer. He spent one year from 2017 to 2018 as a Visiting Scholar with Queen's University Belfast, Belfast, U.K. He is currently an Associated Professor with the School of Physics and Electronic Information Engineering, Henan Polytechnic University, Jiaozuo, China. His research interests include NOMA, ambient backscatter communication, intelligent reflecting surface, ultra reliable and low latency communication, MIMO communication, hardware constrained communication, physical-layer security, UAV communication, artificial intelligence for wireless communication, and the Internet of Things. He has served as a TPC Member, such as the IEEE Globecom, IEEE WCNC, IEEE VTC, and IEEE ICCC. He has also served as the Co-Chair for the IEEE/IET CSNDSP 2020 of the Green Communications and Networks Track. He is also the Lead/Guest Editor for the Special Issue on UAV enabled B5G/6G networks: Emerging Trends and Challenges of Physical Communication, Special Issue on Computational Intelligence and Advanced Learning for Next Generation Industrial IoT in IEEE Transactions on Network Science and Engineering. He also serves as an Editor on the Editorial Board for IEEE Access, Computer Communications, Physical Communication, EURASIP Journal on Wireless Communications and Networking, Journal of Communications and Networks, and KSII Transactions on Internet and Information Systems.

Xingwang Li (Senior Member, IEEE) received the M.Sc. degree from the University of Electronic Science and Technology of China in 2010 and the Ph.D. degree from the Beijing University of Posts and Telecommunications in 2015. From 2010 to 2012, he worked with Comba Telecom Ltd., Guangzhou China, as an Engineer. He spent one year from 2017 to 2018 as a Visiting Scholar with Queen's University Belfast, Belfast, U.K. He is currently an Associated Professor with the School of Physics and Electronic Information Engineering, Henan Polytechnic University, Jiaozuo, China. His research interests include NOMA, ambient backscatter communication, intelligent reflecting surface, ultra reliable and low latency communication, MIMO communication, hardware constrained communication, physical-layer security, UAV communication, artificial intelligence for wireless communication, and the Internet of Things. He has served as a TPC Member, such as the IEEE Globecom, IEEE WCNC, IEEE VTC, and IEEE ICCC. He has also served as the Co-Chair for the IEEE/IET CSNDSP 2020 of the Green Communications and Networks Track. He is also the Lead/Guest Editor for the Special Issue on UAV enabled B5G/6G networks: Emerging Trends and Challenges of Physical Communication, Special Issue on Computational Intelligence and Advanced Learning for Next Generation Industrial IoT in IEEE Transactions on Network Science and Engineering. He also serves as an Editor on the Editorial Board for IEEE Access, Computer Communications, Physical Communication, EURASIP Journal on Wireless Communications and Networking, Journal of Communications and Networks, and KSII Transactions on Internet and Information Systems. **View more**

Zhen Xie
School of Physics and Electronic Information Engineering, Henan Polytechnic University, Jiaozuo, China

Zhen Xie received the B.Sc. degree in electronic information engineering from the School of Physics and Electronic Information Engineering, Henan Polytechnic University, Jiaozuo, China, in 2020, where she is currently pursuing the M.Sc. degree in communication and information systems. Her current research interests include nonorthogonal multiple access and intelligent reflecting surface.

Zhen Xie received the B.Sc. degree in electronic information engineering from the School of Physics and Electronic Information Engineering, Henan Polytechnic University, Jiaozuo, China, in 2020, where she is currently pursuing the M.Sc. degree in communication and information systems. Her current research interests include nonorthogonal multiple access and intelligent reflecting surface. **View more**

Zheng Chu

5GIC & 6GIC, Institute for Communication Systems, University of Surrey, Guildford, U.K.

Zheng Chu (Member, IEEE) received the Ph.D. degree from Newcastle University, Newcastle upon Tyne, U.K, in 2016. He was with the Faculty of Science and Technology, Middlesex University, London, U.K., from 2016 to 2017. He is currently with the Institute for Communication Systems, University of Surrey, Guildford, U.K. His current research interests include smart radio environments/smart reflecting surface, 5G/6G communication networks, Internet of Things networks, artificial intelligence driven future networks, wireless security, and wireless powered networks.

Zheng Chu (Member, IEEE) received the Ph.D. degree from Newcastle University, Newcastle upon Tyne, U.K, in 2016. He was with the Faculty of Science and Technology, Middlesex University, London, U.K., from 2016 to 2017. He is currently with the Institute for Communication Systems, University of Surrey, Guildford, U.K. His current research interests include smart radio environments/smart reflecting surface, 5G/6G communication networks, Internet of Things networks, artificial intelligence driven future networks, wireless security, and wireless powered networks. **View more**

Varun G. Menon

Department of Computer Science and Engineering, SCMS School of Engineering and Technology, Ernakulam, India

Varun G. Menon (Senior Member, IEEE) received the Ph.D. degree in computer science and engineering from Sathyabama University, India, in 2017. He is currently an Associate Professor with the Department of Computer Science and Engineering, SCMS School of Engineering and Technology, India. He has published more than 50 research papers in peer reviewed and highly indexed International Journals and Conferences. His research interests include Internet of Things, fog computing and networking, underwater acoustic sensor networks, scientometrics, educational psychology, ad-hoc networks, wireless communication, opportunistic routing, and wireless sensor networks. He is currently a Guest Editor for IEEE Transactions on Industrial Informatics, IEEE Sensors Journal, and IEEE Internet of Things Journal. He is an Associate Editor of IET Quantum Communication and also an Editorial Board Member of IEEE Future Directions. He has served over 20 conferences, such as IEEE ICC, ICCCN 2020, IEEE COINS 2020, SigTelCom, ICACCI, and ICDMAI in leadership capacities, including program a co-chair, a track chair, a session chair, and a technical program committee member. He is a Distinguished Speaker of ACM.

Varun G. Menon (Senior Member, IEEE) received the Ph.D. degree in computer science and engineering from Sathyabama University, India, in 2017. He is currently an Associate Professor with the Department of Computer Science and Engineering, SCMS School of Engineering and Technology, India. He has published more than 50 research papers in peer reviewed and highly indexed International Journals and Conferences. His research interests include Internet of Things, fog computing and networking, underwater acoustic sensor networks, scientometrics, educational psychology, ad-hoc networks, wireless communication, opportunistic routing, and wireless sensor networks. He is currently a Guest Editor for IEEE Transactions on Industrial Informatics, IEEE Sensors Journal, and IEEE Internet of Things Journal. He is an Associate Editor of IET Quantum Communication and also an Editorial Board Member of IEEE Future Directions. He has served over 20 conferences, such as IEEE ICC, ICCCN 2020, IEEE COINS 2020, SigTelCom, ICACCI, and ICDMAI in leadership capacities, including program a co-chair, a track chair, a session chair, and a technical program committee member. He is a Distinguished Speaker of ACM. **View more**

Shahid Mumtaz
Instituto de Telecomunicações, Aveiro, Portugal

Shahid Mumtaz (Senior Member, IEEE) is a Senior 5G Consultant with Huawei, Sweden. He is serving as a Scientific Expert and an Evaluator for various Research Funding Agencies. He has authored four technical books, 12 book chapters, more than 250 technical papers (more than 150 Journal/transaction, more than 80 conference, two IEEE best paper awards—in the area of mobile communications. Most of his publication is in the field of Wireless Communication. He is the recipient of the NSFC Researcher Fund for Young Scientist in 2017 from China and the IEEE ComSoC Young Researcher Award in 2020. He was awarded an Alain Bensoussan Fellowship in 2012. He is the Founder and the EiC of IET Quantum Communication, a Vice-Chair: Europe/Africa Region–IEEE ComSoc: Green Communications & Computing Society, and a Vice-Chair for IEEE Standard on P1932.1: Standard for Licensed Unlicensed Spectrum Interoperability in Wireless Mobile Networks. He is an IET Fellow, IEEE ComSoc, and an ACM Distinguished Speaker.

Jianhua Zhang
State Key Laboratory of Networking and Switching Technology,
Beijing University of Posts and Telecommunications, Beijing, China

Jianhua Zhang (Senior Member, IEEE) received the B.S. degree from the North China University of Technology in 1994, and the Ph.D. degree from the Beijing University of Posts and Telecommunications in 2003, where she is currently the Professor with Information and Engineering College. She has published more than 70 journal papers and nearly 200 conference papers. Her research interests are massive MIMO and millimeter wave channel modeling and transmission techniques, channel emulator, and OTA test. She received the China Communications Best Paper Award at 2016 and shared VTC 2015 spring, and JCN2009 best paper awards. She continuously contributed to channel model standards from ITU–R M.2135 to 3GPP 36.873,900/901 and she was the Drafting Group Chairwoman of ITU–R IMT–2020 channel model.

Figures   ⌄

References   ⌄

Citations   ⌄

Keywords   ⌄

Metrics   ⌄
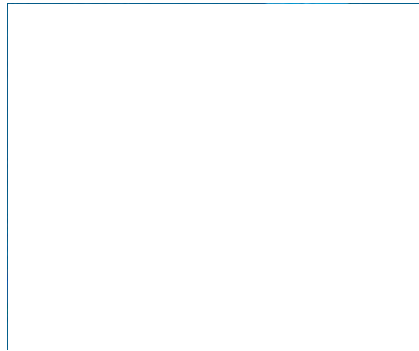
More Like This

Energy-Saving Deployment Optimization and Resource Management for UAV-Assisted Wireless Sensor Networks With NOMA

IEEE Transactions on Vehicular Technology

Published: 2022

Throughput Optimization Control Strategy for Mobile Wireless Sensor Networks Under Rician Fading Channels

2023 12th International Conference on Control, Automation and Information Sciences (ICCAIS)

Published: 2023

Show More

About IEEE *Xplore* | Contact Us | Help | Accessibility | Terms of Use | Nondiscrimination Policy | IEEE Ethics Reporting ↗ | Sitemap |
IEEE Privacy Policy

A not-for-profit organization, IEEE is the world's largest technical professional organization dedicated to advancing technology for the benefit of humanity.

IEEE Account

» Change Username/Password

» Update Address

Purchase Details

» Payment Options

» Order History

» View Purchased Documents

Profile Information

» Communications Preferences

» Profession and Education

» Technical Interests

Need Help?

» **US & Canada:** +1 800 678 4333

» **Worldwide:** +1 732 981 0060

» Contact & Support

About IEEE *Xplore* | Contact Us | Help | Accessibility | Terms of Use | Nondiscrimination Policy | Sitemap | Privacy & Opting Out of Cookies

A not-for-profit organization, IEEE is the world's largest technical professional organization dedicated to advancing technology for the benefit of humanity.

Enabling Cooperative Relay Selection by Transfer Learning for the Industrial Internet of Things

Sina Shaham, Shuping Dang, *Member, IEEE*, Miaowen Wen, *Senior Member, IEEE*, Shahid Mumtaz, *Senior Member, IEEE*, Varun G. Menon, *Senior Member, IEEE*, and Chengzhong Li

Abstract—Large manufacturing sites with movable obstacles and dynamic network topology call for reliable and efficient strategies to transmit data through the industrial Internet of Things (IoT). Cooperative communications and relay selection have shown a great potential to improve throughput and energy efficiency at the expanse of high end-to-end transmission latency. To reduce this latency, we propose to use transfer learning for relay selection in the industrial IoT. Unlike traditional approaches that are trained for a specific task, transfer learning exploits the acquired knowledge from similar tasks to assist new tasks. Transfer learning is capable of improving learning performance, reducing the need for large datasets for different setups, lowering communication overhead and computational complexity. Specifically, in this paper, we propose a generic transfer learning framework for relay selection problems in the industrial IoT. Based on the proposed framework, we design a hypothesis and test it by empirical data for convergence analysis. Also, we devise and conduct a step-by-step and rigorous hyperparameter tuning procedure for the proposed transfer learning framework. The accuracy of the proposed approach is evaluated and verified by extensive training and test datasets abiding by different statistical distributions.

Sina Shaham was with the Department of Engineering, The University of Sydney, Sydney, NSW 2006, Australia. He is now with the University of Southern California, Los Angeles, CA 90089 USA (e-mail: sshaham@usc.edu).

Shuping Dang was with the R&D Center, Guangxi Huanan Communication Company Ltd., Nanning 530007, China, and also with the Computer, Electrical and Mathematical Science and Engineering Division, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia. He is now with the Department of Electrical and Electronic Engineering, University of Bristol, Bristol BS8 1UB, U.K. (e-mail: shuping.dang@bristol.ac.uk).

Miaowen Wen is with the School of Electronic and Information Engineering, South China University of Technology, Guangzhou 510641, China (e-mail: eemwwen@scut.edu.cn).

Shahid Mumtaz is with the Instituto de Telecomunicações, 3810-193 Aveiro, Portugal (e-mail: smumtaz@av.it.pt).

Varun G. Menon is with the Department of Computer Science and Engineering, SCMS School of Engineering and Technology, Ernakulam 683576, India (e-mail: varunmenon@ieee.org).

Chengzhong Li is with the R&D Center, Guangxi Huanan Communication Company Ltd., Nanning 530007, China (e-mail: chengzhong.li@hncom.ac.cn).

Digital Object Identifier 10.1109/TCCN.2022.3147202

Index Terms—Artificial neural network (ANN), transfer learning, distributed network architecture, relay selection, industrial Internet of Things (IoT).

I. INTRODUCTION

COOPERATIVE relaying is an important technique for modern wireless communication networks to extend wireless transmission coverage and enhance reliability [1]–[4]. In most wireless communication networks, especially the dense networks in the industrial Internet of Things (IoT), there in general are a number of sensors that can serve as relays for forwarding signals from a source to the intended destination [5]. It has already been proven by many works that employing all available relays to forward information might not be an efficient strategy, but only a subset of available relays should be selected by certain criteria to carry out signal forwarding [6]. This is because multi-relay coordination is complex and requires a large amount of signaling overhead when the number of utilized relays goes large [7]–[9]. Employing a full set of available relays will also bring challenges for the destination to combine and jointly decode the signals propagating through different paths with different delays [10]. Researchers also proved that selecting a single relay to forward signals can also provide the full selection diversity as utilizing all available relays, as long as the relay selection criterion is well designed [11]. In this regard, relay selection techniques have been imperative to efficiently organize cooperative communication networks [12].

Although designing and analyzing relay selection criteria can be challenging in theory, it is also demanding to implement a certain relay selection criterion in practice. This refers to the relay selection implementation issue [13]. Conventionally, there are two ways to implement relay selection in practical wireless cooperative networks. The first one is termed the centralized implementation, which assumes a central node (e.g., a base station or an access point) that is able to get access to the global channel state information (CSI), makes decisions on which relay(s) should be selected in a centralized manner, and broadcasts decisions to all other nodes, including source, relays, and destination [14]. The centralized method requires the global CSI and reliable coordination among all nodes, which are demanding and even impossible in practice, especially when the number of relays is large and their locations are randomly distributed [15], [16]. To resolve the

drawbacks of the centralized method, the second method is termed the decentralized implementation. This decentralized method relies on a timer based decentralized algorithm, which supposes that each relay has a timer and will respond to the transmission requests via clear-to-send (CTS) and ready-to-send (RTS) messages [17]. Albeit without the requirements of global CSI and a centralized control mechanism, the decentralized method involves extra latency and could render the selection collision problem [18].

Relay selection problems are in essence resource allocation problems, in which a limited amount of hardware resource needs to be assigned so as to optimize certain system performance. Machine learning (ML) techniques, relying on both centralized or distributed architectures, are the powerful tools to deal with resource allocation and become full-fledged in recent years with the artificial intelligence (AI) enthusiasm [19]–[23]. As a result, to implement relay selection in a more efficient manner, researchers are currently resorting to ML techniques. Without prior knowledge of channels and labeled data, both reinforcement learning and unsupervised learning techniques are essentially trial-and-error methodologies, which result in inefficiency in the training phase and unnecessary latency for cooperative transmissions. Moreover, the optimization convergence can hardly be guaranteed or even assessed in trial-and-error procedures, as reference baselines are lacking. In this regard, supervised ML, as a model-driven learning method, has emerged as promising techniques to implement relay selection schemes.

Despite the advancements achieved by supervised ML models, these conventional models have shortcomings that limit their use for relay selection due to the highly dynamic nature of wireless communication networks. Some of the challenges of using conventional ML models include:

- The mismatch in statistical distributions between training and testing datasets exists due to heterogeneity of networks in the industrial IoT, deteriorating the model performance. This phenomenon could even be severer as data distribution pattern drifts in both the time and space domains;
- Conventional ML models are commonly trained for a given and specific application scenario, for which the training data is generated, sampled, and processed. However, the significant variations in modern networks, including the industrial IoT, make each application scenario quite different from the previous one [24], and therefore case-specific training by conventional ML models becomes inefficient and even impossible.
- The performance of supervised ML techniques highly depends on the availability of a large enough amount of labeled training data, which can be costly and time-consuming to collate and pre-process in practice. Moreover, even if such a huge amount of labeled training data exists, the overlong training duration could become another bottleneck to restrict performance improvement, particularly for low-latency industrial applications.

To address the aforementioned challenges for exploiting the supervised ML models for cooperative relay selection, we adopt transfer learning as an effective tool to enhance the performance of traditional models for relay selection in the industrial IoT. Our specific contributions include:

- To deal with the statistical heterogeneity of data, we construct a generic transfer learning framework to enable relay selection, which is suited for most application scenarios in the industrial IoT with a distributed network architecture.
- We design a hypothesis test for the generic framework to estimate whether the ANN after training has converged and can be used in practice. Based on this generic framework, a step-by-step example is demonstrated on how transfer learning can facilitate relay selection for a specific wireless application by using training and testing data sets generated from different distribution models.
- We investigate a series of system configurations and hyperparameters of the proposed framework as well as system imperfections and obtain insights into their effects on the learning performance.

The rest of the paper is organized as follows. In Section III, we introduce the fundamentals, which become the building blocks of the transfer learning based relay selection framework. In Section IV, we propose the generic framework and explain it in detail. Based on the generic framework, we also illustrate its realization by an exemplary case in Section V. We design and expatiate on the hypothesis test in the checking stage for determining whether the ANN after training has converged in Section VI. We further investigate the proposed framework by numerical simulations and have comprehensive discussion in Section VII. Finally, we conclude the paper in Section VIII.

II. RELATED WORK

The significance of relay selection has called for several survey papers summarizing the existing works. Distributed relay selection schemes for cooperative communications are reviewed in [25], including the classification of buffer-aided relay selection techniques in [26], the strategies focusing on capacity enhancement in [27], and cooperative relaying in industrial wireless sensor networks in [28]. Moreover, a compendium of ML applications for wireless networks is provided in [29], and the key roles of cooperative relaying and ML for sixth generation (6G) communications are recognized in [30]. In the following, we detail the existing literature for cooperative relaying and the computing techniques based on ML for performance enhancement purposes.

Broadly speaking, cooperative communications is a mechanism to distribute wireless resources among distributed nodes and improve the overall system performance [8]. The integration of cooperative relay selection techniques in the IoT infrastructure is essential due to the large number of sensors distributed in addition to large covered indoor and outdoor areas with complex obstacles and terrains [5]. The selection of relays due to highly dynamic nature of wireless environments and heterogeneous service demands is a challenging task. In the most extreme cases, the employment of either all relays or only a single relay has been shown in the literature to be inefficient [3], [6], [7]. Two traditional strategies have

been performed to address this issue, i.e., centralized [14] and decentralized implementations [17], [18]. The former requires the global CSI and the efficient coordination among all relay nodes, making it difficult to be used in practice particularly for industrial IoT. The latter does not require the global CSI; however, it could result in an intolerable latency and can also lead to collision in the selection process [18].

Because of the practical issues with the existing implementation schemes, a simplistic supervised ML based relay selection framework is constructed in [31], in which relay selection problems are treated as binary classification problems. In [32], the supervised ML aided relay selection is generalized to be maximum/minimum searching problems, and the training network architecture design issues are investigated in [33]. Multi-carrier relay selection assisted by supervised ML and sensing fusion is also studied following a similar methodology for intelligent vehicular communication networks in [34]. In [35], the physical layer security problems are incorporated in the supervised ML based relay selection framework, where there exists an external eavesdropper. The number of hidden layers is also increased to form deep artificial neural networks (ANNs) to cope with the relay selection problems in the contexts of millimeter-wave (mmWave), device-to-device (D2D), and vehicular communications [36], [37]. All above studies have confirmed that supervised ML is capable of enabling relay selection in a more efficient way that requires less computational resource and computing time than the exhaustive search. Despite being insightful and instructive, these studies only employ specific ANNs to solve specific relay selection problems in specific scenarios, which can hardly formulate a universal framework for solving a generic relay selection problem.

III. FUNDAMENTALS

A. Cooperative Communication System Model

In this paper, we consider a generic multi-relay cooperative communication system in the context of industrial IoT, where there exist a single source, a single destination, and M relays randomly distributed in a cluster (this can be an industrial workshop, an equipment room, and an assembly line). The full set of M relays is denoted as $\mathcal{M}$. The cooperative network topology is assumed to be static during the entire transmission phase, and the direct transmission between the source and the destination has been blocked due to overlong transmission distance or obstacles. These are the common case and reasonable assumptions for the industrial IoT [38]. In this regard, at least a single relay must be employed to forward the signal from the source to the destination in order to complete end-to-end cooperative transmissions. For the mth relay, its first-hop and second-hop channel coefficients at time t are denoted as $h_1(m, t)$ and $h_2(m, t)$, respectively. For simplicity, we also define the channel magnitudes for the first and second hops as $g_1(m, t) = |h_1(m, t)|$ and $g_2(m, t) = |h_2(m, t)|$, which are distributed abiding the cumulative distribution functions (CDFs) $F_1(x|m, t) = \mathbb{P}\{g_1(m, t) \leq x\}$ and $F_2(x|m, t) = \mathbb{P}\{g_2(m, t) \leq x\}$, where $\mathbb{P}\{\cdot\}$ denotes the probability of the random event enclosed. The generic multi-relay cooperative
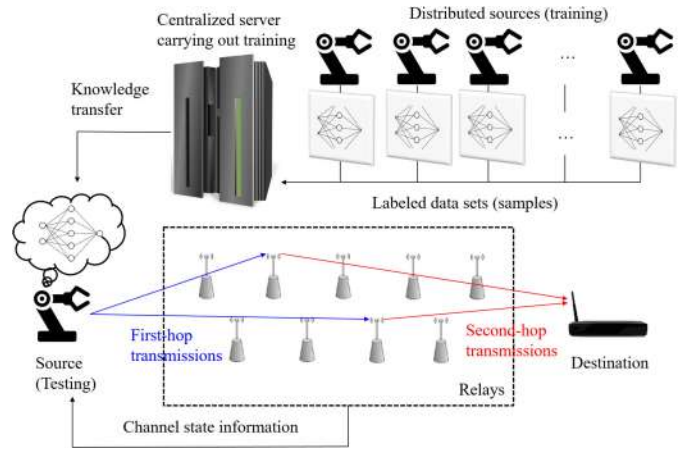

Fig. 1. A generic multi-relay cooperative communication system with $M = 9$ relays considered in this paper. $S = 2$ relays are selected from $M = 9$ relays to forward the transmitted signal from source (industrial apparatus) to destination (access point).

TABLE I
LIST OF IMPORTANT NOTATIONS USED IN THIS PAPER

Symbol	Description
M	Number of relays
$\mathcal{M}$	Set of relays
$h_i(m, t)$	ith hop channel information of mth relay at time t
$g_i(m, t)$	Magnitude of ith hop channel information of mth relay at time t
$F_i(m, t)$	CDF of $F_i(m, t)$
S	Subset of relays
$\{\mathbf{G}(t), \mathbf{S}(t)\}$	Complete labeled dataset
$\tilde{\mathbf{S}}(t)$	Output ML labels
Ω	Relay optimization objective function
Ω^*	Optimal value of objective function
$\Omega^\&$	Value estimated by TL
T	Number of optimization constraints
K_l	Number of neurons in the lth layer
$\mathbf{a}$	Architecture vector
L	Number of NN layers
$y_l^{k_l}(t)$	Output of k_lth neuron in the lth layer at the tth epoch
$w_{k_l \to k_{l+1}}(t)$	link weight between k_lth and the k_{l+1}th node
$\theta_l^{k_l}(t)$	Activation threshold at the lth layer and t epoch
φ_t	Number of training data sets
φ_c	Number of checking data sets
α	Learning rate
β	Range expansion coefficient
ϵ	Termination threshold

communication system is illustrated in Fig. 1 for clarity, and the list of important notations used in this paper is given in Table I.

B. Relay Selection Problem

Recent studies have reported that using all relays to forward source's signals is not optimal from the network's perspective. First of all, more transmit power at multiple relay nodes is consumed to process and retransmit the received signals. Also, coordinating among multiple relay modes could also result in a much higher signaling overhead and the difficulty for the destination to combine multiple received signals. To leverage the advantages of cooperative relaying, it is required
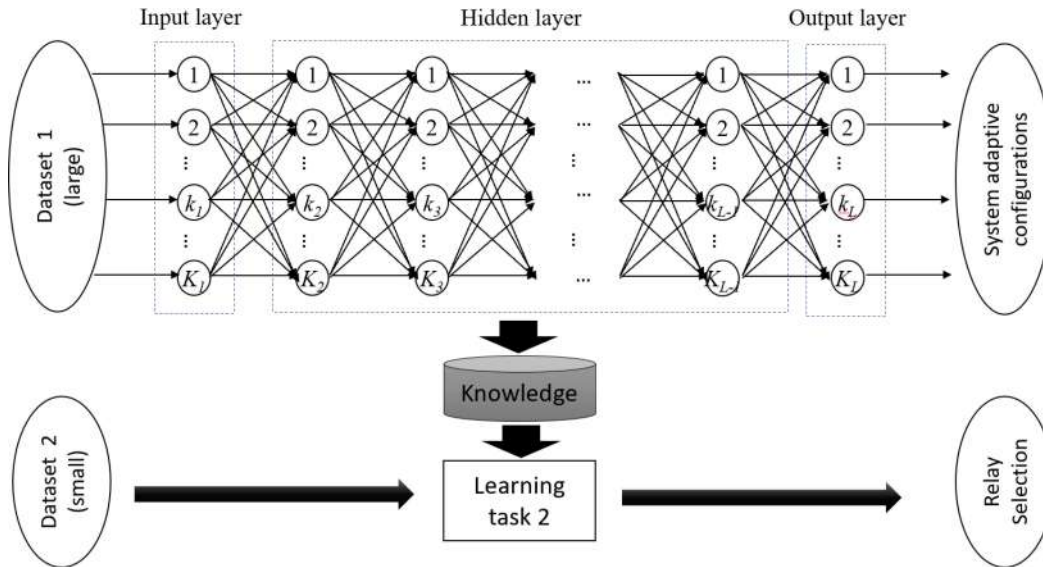
Fig. 2. System architecture of the proposed generic framework utilizing transfer learning.

to select a subset of relays $\mathcal{S} \subseteq \mathcal{M}$ to bridge the source and the destination, where we have $|\mathcal{S}| = S \in [1, M]$. For single-relay selection schemes, $S = 1$, while for multi-relay selection schemes, $S > 1$. Normally, a relay selection procedure is made based on the CSI represented by channel magnitudes and a set of ambient, protocol, and hardware constraints. In general, a typical relay selection problem can be expressed as a decision making problem aiming at maximizing a single or composite objective under certain constraints.[1] Mathematically, we can formulate a generic relay selection problem at time t as follows:

$$\mathcal{S} = \underset{\mathcal{N} \subseteq \mathcal{M}, \ |\mathcal{N}|=S}{\arg\max} \ \{\Omega\}$$
$$\text{s.t. Constraint } 1, \text{ Constraint } 2, \ldots, \text{ Constraint } T, \tag{1}$$

where Ω is the objective function, which can be expanded to be either $\Omega(\{g_1(m, t)\}_{m=1}^M, \{g_2(m, t)\}_{m=1}^M)$ depending on the *instantaneous* CSI or $\Omega(\{F_1(x|m, t)\}_{m=1}^M, \{F_2(x|m, t)\}_{m=1}^M)$ depending on the *statistical* CSI.

The research activities pertaining to relay selection problems mainly focus on

- How to model the objective function and constraints (a.k.a. relay selection criterion) so as to reflect the realistic situations and requirements of cooperative communication systems;
- How to solve the formulated relay selection problems and obtain the corresponding $\mathcal{S}$ in an efficient manner.

C. Transfer Learning

The first focus of relay selection research activities requires domain knowledge and thereby human intervention, while the second focus can be facilitated by ML and transmission

---

[1]These constraints are related to energy consumption, system complexity, user fairness, expected transmission latency, and security requirements.

environment sensing capability. Essentially, ML provides an adaptive mechanism that is able to dynamically adjust a set of system configurations, including the subset of selected relays $\mathcal{S}$, by detected information from the environment. The learning procedure relies on a special structure termed ANN, which is constructed by a vast number of connected neurons placed in several layers [39]. There are normally one input layer, one output layer, and multiple hidden layers in ANNs. Comparing the outputs to the optimal references contained in labeled training data sets, neurons in ANNs are capable of strengthening the links by increasing link weights and weakening the links by decreasing link weights. The adjustment of link weights is carried out in a dynamical manner without human intervention. Consequently, learning through a huge volume of labeled data sets, a well designed ANN can adapt to the optimal/suboptimal settings by changing the link weights and thereby produce satisfactory output [40].

Transfer learning is an emerging technique belonging to the ML family, which is proposed to deal with statistical heterogeneity of data [41]. In particular, transfer learning employs two disjoint data sets, normally following different distribution models, for training and testing purposes. In this way, the ANN trained by the training data set is able to process the testing data set with an acceptable loss of accuracy. Therefore, by transfer learning, the trained ANN is guaranteed to be general and applicable for a wide range of applications. This technique is in particular useful for coping with channel variations and enabling cooperative relay selection in complex wireless environments.

D. Neural Network Architecture and Configurations

The schematic of a fully connected ANN model is presented in Fig. 2. For a fully connected ANN, we can define an architecture vector to characterize its network architecture:

$$\mathbf{a} = [K_1, K_2, \ldots, K_L]^T, \tag{2}$$

where K_l is the number of neurons in the lth layer, and L is the number of layers. For an L-layer ANN, the first and last layers are called the input layer and the output layer, respectively. The other $L - 2$ layers are called the hidden layers. We denote the output from the k_lth neuron in the lth layer at the tth epoch as $y_l^{k_l}(t)$ and the link weight between the k_lth node and the k_{l+1}th node in the adjacent layer at the tth epoch as $w_{k_l \to k_{l+1}}(t)$. Also, the activation threshold corresponding to the lth layer at the tth epoch is denoted as $\theta_l^{k_l}(t)$.

E. Performance Metrics

To measure the training performance so as to check whether an ANN has been well trained, we can collect all outputs from the output layers of a trained ANN and reconstruct the estimated $\tilde{\mathbf{S}}(t)$ for a given labeled data set $\{\mathbf{G}(t), \mathbf{S}(t)\}$, where $\mathbf{G}(t)$ is the input vector representing the detected information from the environment, and $\mathbf{S}(t)$ is the output vector, in which the entries are unity if the corresponding relays are selected or zero, otherwise. Then, we can introduce the mean squared error (MSE) over a number of labeled data sets to measure the training performance:

$$\text{MSE}(t) = \frac{1}{M}\left\|\mathbf{S}(t) - \tilde{\mathbf{S}}(t)\right\|_F^2 \tag{3}$$

where $\left\|\cdot\right\|_F$ denotes the Frobenius norm of the enclosed matrix.

To fit the binary nature of the expected outputs, we can also utilize the mean squared quantized error (MSQE) over a number of labeled data sets to measure the training performance:

$$\text{MSQE}(t) = \frac{1}{M}\left\|\mathbf{S}(t) - \Xi\left(\tilde{\mathbf{S}}(t)\right)\right\|_F^2 \tag{4}$$

where $\Xi(\cdot)$ is the entry-wise quantization function of the matrix enclosed, which should be customized for different relay selection problems and application scenarios.

More importantly, we should also compare the value of the objective function $\Omega^{\&}$ yielded by the transfer learning based relay selection framework with the optimal value of objective function Ω^* yielded by the brute-force search. The relative error (RE) measuring the communication performance at epoch t is defined as

$$\text{RE}(t) = \left(\left|\Omega^* - \Omega^{\&}\right|\right)/\Omega^*. \tag{5}$$

IV. RELAY SELECTION BY TRANSFER LEARNING: A GENERIC FRAMEWORK

The flowchart of the generic framework implementing relay selection by transfer learning is illustrated in Fig. 3. To expatiate on the generic framework, we detail each of the functional block in this flowchart in the following subsections.

A. Set ANN Architecture and Hyperparameters

To initiate the training process and specify the data set structure, we first need to set the ANN architecture and a series of hyperparameters, which are invariant during the entire optimization process. The ANN architecture is determined by the $L \times 1$ architecture vector $\mathbf{a}$ according to (2). It should also
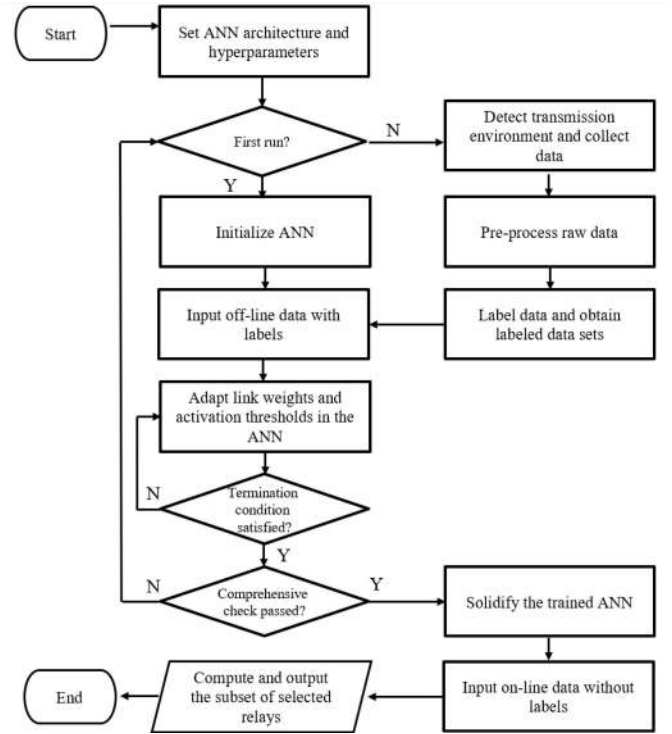
Fig. 3. Flowchart of the generic framework implementing relay selection based on transfer learning.

be noted that the number of neurons in the input layer must be aligned with the structure of data set. Similarly, the number of neurons in the output layer must be aligned with the structure of the output vector, which equals the number of all available relays,[2] i.e., $K_L = M$. The hyperparameters required to be set include: the number of data sets φ_t for training purposes, the number of data sets φ_c for checking purposes, the learning rate α, the range expansion coefficient β, and the termination threshold of the training procedure ϵ.

B. Initialize ANN

For an ANN with a given architecture, there are two sets of parameters required to be initialized before carrying out any learning algorithms. The first set contains all $K_1 K_2 + K_2 K_3 + \cdots + K_{L-1} K_L$ link weights $\{w_{k_l \to k_{l+1}}(1)\}$ and the second set incorporates $K_1 + K_2 + \cdots + K_L$ activation thresholds $\{\theta_l^{k_l}(1)\}$. In this paper, we employ a random initialization approach to set the original values of these parameters [40]. Specifically, when $l > 1$, all link weights $\{w_{k_l \to k_{l+1}}(1)\}$ and thresholds $\{\theta_d^{k_l}(1)\}$ are uniformly chosen within the range $(-\frac{\beta}{K_{l-1}}, \frac{\beta}{K_{l-1}})$, where β is the range expansion coefficient. For the special case $l = 1$, the distribution range is $(-\frac{\beta}{M}, \frac{\beta}{M})$.

---

[2]Note that, this specification is only valid for cooperative networks with static topology, where the number of relays and their locations are fixed. For dynamic cooperative networks where relay nodes are with mobility, the actual number of available relays is unknown. In this case, we can still assume the number of neurons in the output layer to be M. After network sensing and channel estimation, if the actual number of available relays is higher than M, then only M from them with stronger connectivity will be considered for selection. If the actual number of available relays is lower than M, the rest of inputs will be inactivated by setting to be nulls.

C. Detect Transmission Environment and Collect Data

The transmission environment detection is case dependent. With powerful detection capability, wireless sensors can measure the end-to-end channel quality by the instantaneous channel magnitudes in both hops by sending training sequences or pilot signals. Less capable sensors might only be able to measure instantaneous channel magnitudes in a single hop or even statistical CSI that characterizes the channel quality averaged over a period of time. Most sensors can at least detect the transmission distances through feedforward and feedback control channels or by equipped Global Positioning System (GPS). All above information can be used solely or jointly as the raw data to decide which relay(s) should be selected.

D. Pre-Process Raw Data

As there is always channel estimation errors and other noisy effects in the collected raw data, one must pre-process these raw data before using them in the initialized ANN. First, some simple noise deduction techniques can be applied to refine the information in the raw data, and then anomalies should be picked up via certain criteria and discarded. Dimension deduction techniques should also be involved to simplify the raw data and reduce the processing complexity for the ANN. More importantly, the input domain of most ANNs should be regulated so as to accelerate convergence. Therefore, all input data need to be normalized within a certain range, say $[\chi_{\min}, \chi_{\max}]$. Also, an appropriate sampling method should also be used to split φ_t training and φ_c testing data sets.

E. Label Data and Obtain Labeled Data Sets

To efficiently implement ML, labels pairing with data are suggested, which provide the optimal solutions to the optimization problems for training purposes. In essence, labels indicate the direction of adaptation for an ANN. However, generating labels is computationally costly, especially for large-scale and sophisticated problems that are not convex/concave. For most relay selection problems that are binary integer programming (BIP) problems and thereby NP-complete [42], we can employ brute-force algorithms to yield the optimal relay selection schemes for given transmission environment. For special or simple relay selection problems, in which the optimality can be guaranteed by low-complexity algorithms, we can also resort to the low-complexity algorithms to generate labels. A complete labeled data set can be expressed by $\{\mathbf{G}(t), \mathbf{S}(t)\}$. Note that, labels need to be generated for both training and testing data sets.

F. Input Off-Line Data With Labels

In order to utilize the labeled data sets in an efficient way, we must carefully design the interfaces between labeled data sets and an ANN so as to input off-line data for training purposes. Generally speaking, there are two interfaces. The first interface is used to input processed data characterizing the transmission environment (i.e., $\mathbf{G}(t)$) to the ANN. This interface regulates the input layer of ANN that the number of neurons in the input layer must equal the number of entries of $\mathbf{G}(t)$. Also, this interface should specify whether data sets are allowed to be used repeatedly for training ANN and how to repeat if allowed.[3] The second interface is used to compare the outputs of ANN to the labels and enable back-propagation mechanism to adjust ANN coefficients $\{w_{k_l \to k_{l+1}}(t)\}$ and $\{\theta_l^{k_l}(t)\}$ layer by layer. In particular, the second interface stipulates that the number of neurons in the output layer must equal the number of entries of $\mathbf{S}(t)$.

G. Adapt Link Weights and Activation Thresholds in the ANN

Comparing the outputs from the ANN $\tilde{\mathbf{S}}(t)$ to the label $\mathbf{S}(t)$, an error vector $\mathbf{E}(t) = \mathbf{S}(t) - \tilde{\mathbf{S}}(t)$ can be yielded and propagate backwards to adapt ANN coefficients $\{w_{k_l \to k_{l+1}}(t)\}$ and $\{\theta_l^{k_l}(t)\}$. A common practice is to employ gradient descent with learning rate α, i.e., the factor multiplying adaptation increments, on an iterative basis to obtain error gradients and adapt the ANN coefficients accordingly [40].

H. Check Termination Condition

Two conditions can be organized as an intersection to check whether the iterative training procedure should be terminated: *Condition 1*-The MSE is less than the termination threshold of the training procedure ϵ, and *Condition 2*-All φ_t training data sets have been exhaustively used. If both of the above conditions are satisfied, the training procedure will be terminated. However, when φ_t training data sets have been exhaustively used, but the MSE is still higher than the termination threshold of the training procedure ϵ, the training proccedure will also be stopped, but a warning message will be sent for necessary attention. It should be noted that because there are two termination conditions that must be satisfied simultaneously in order to end the iterative training procedure, we can reduce the computational complexity caused by checking the termination conditions by only inspecting *Condition 1* once after *Condition 2* has already been satisfied.

I. Check the Trained ANN Comprehensively

Once the training procedure is completed, it does not necessarily imply that the trained ANN is appropriate to be used, because of the statistical heterogeneity of data. We should employ the trained ANN to test the φ_c testing data sets with labels in terms of MSE and objective function for training and communication performance, respectively. If the results generated by the trained ANN for the testing data sets are close enough to the optima, the feasibility of the trained ANN is verified, otherwise the settings of ANN architecture and hyperparameters would be improper and require adjustments. Fixing up the problem requires human intervention and domain knowledge. Sometimes complicated hyperparameter tuning experiments are involved, which could be costly and time-consuming. To rigorously check the training convergence, various hypothesis tests can be designed to assist, which will be detailed in Section VI.

---

[3]This interface and the relevant settings are of paramount importance for avoiding overfitting.

J. Solidify the Trained ANN

Once passing the comprehensive check, the trained ANN is regarded as qualified for practical use. As a result, the trained ANN is solidified by fixing all link weights $\{w_{k_l \to k_{l+1}}(t)\}$ and activation thresholds $\{\theta_l^{k_l}(t)\}$. However, it is worth noting that solidifying an ANN does not imply that all ANN coefficients will be invariant forever. In practice, there must be a break mechanism that allows re-calculating all ANN coefficients. This break mechanism is normally activated once the statistical heterogeneity index of new $\mathbf{G}(t)$ is above certain value (this indicates that the transmission environment has considerably changed).

K. Input On-Line Data Without Labels

After solidifying the trained ANN, on-line data without labels can be imported to the input layer for computing. It is worth noting that even for on-line data, pre-processing is required so that the format of input data set can be aligned with that of training data set. In addition, a sampling mechanism or data window might be utilized to randomly or regularly store a certain volume of input on-line data for checking statistical heterogeneity index so that the break mechanism for re-calculating ANN coefficients can be activated when the index exceeds a certain value.

L. Compute and Output the Subset of Selected Relays

Processing the input on-line data produces outputs from the neurons in the output layer, which can be constructed in the format of vector, in which the entries are fractional. By quantization, the fractional entries are converted to binary values, i.e., unity or zero, which are used to indicate whether a relay should be selected or not. With the quantized output vector, the subset of selected relays corresponding to the input on-line data can be directly obtained. So far, the entire relay selection procedure is aided by transfer learning via a well designed and trained ANN.

M. Analysis of Computational Complexity

To fully understand the computational advantages brought by the proposed framework and methods, one needs to quantify and study the computational complexity in the feedforward and backward propagation processes. In the feedforward process, at every layer, the computing operations include a matrix multiplication in addition to the transformation based on the adopted activation function. The asymptotic runtime of matrix multiplication with dimensions $n \times n$ is in the order of $\mathcal{O}(n^3)$ [43]. Looking at the dimensions of the adopted neural network in this paper, the first layer involves $K_1 \times 1$ number of multiplications, and for l greater than one, the number of multiplications is $K_l \times K_{l-1}$. Therefore, the total number of multiplications denoted as $\mathrm{n_{mul}}$ can be calculated as

$$\mathrm{n_{mul}} = \sum_{l=3}^{L} K_l K_{l-1} K_{l-2} + K_2 K_1 1. \qquad (6)$$

Since the asymptotic runtime is in a cubic order, the total number of multiplications can be upper bounded by $\mathcal{O}(L \times K_{\max}^3)$,

where $K_{\max} = \max\limits_{l=1,2,\ldots,L} \{K_l\}$. Moreover, at every layer, the activation function is applied to the input vector in a linear timescale, resulting in the additional computational complexity of $\mathcal{O}(L \times K_{\max})$. Thus, the total computational complexity of the feedforward propagation process is $\mathcal{O}(L \times K_{\max}^3) + \mathcal{O}(L \times K_{\max})$ or even simpler given by $\mathcal{O}(L \times K_{\max}^3)$ as the linear term becomes negligible compared to the cubic term. With similar rationales, the computational complexity of backpropagation can be found to be $\mathcal{O}(K_{\max}^2 + K_{\max}^3 \times L)$. Considering the entire training process, one should also account for the number of iterations, in which the forward and backward propagation processes are executed, and thereby multiply the total computational complexity by φ_t.

V. RELAY SELECTION BY TRANSFER LEARNING: AN EXEMPLARY CASE

In this section, we adopt a simple and exemplary case to demonstrate how the generic transfer learning based relay selection framework constructed in the above section is materialized and utilized. In order to be instructional, we strictly follow the flowchart depicted in Fig. 3 and expatiate on each functional block following the process flow.

First of all, we need to set ANN architecture and hyperparameters. Assuming a static cooperative network where there exist $M = 4$ relays, we let $\mathbf{a} = [4, 4]^T$, which is the simplest form of the ANN fitting the four-relay selection scenario. Note that there are only input and output layers, but absence of any hidden layer. Obviously, we have $L = 2$. However, to maintain the generality of the description, we describe the application of the proposed framework in the exemplary case without specifying $L = 2$ in the following. We set $\varphi_t = 1000$ and $\varphi_c = 100$ for the numbers of training and checking data sets, respectively. For simplicity, we normalize $\alpha = 1$ and $\beta = 1$ for the learning rate and the range expansion coefficient. We also let $\epsilon = 0.05$.

Having set the range expansion coefficient $\beta = 1$, $\{w_{k_1 \to k_2}(1)\}$ and $\{\theta_1^{k_1}(1)\}$ are randomly determined by values within $(-1/4, 1/4)$. Similarly, $\{\theta_2^{k_2}(1)\}$ is also randomly determined by values within $(-1/4, 1/4)$. We hereby complete the initialization of the ANN once these initial coefficients are chosen.

To facilitate relay selection, we employ end-to-end channel magnitudes for all relays as the selection references, as it is the direct indicator of channel reliability [44]. Assuming all channel magnitudes of the first and second hops corresponding to all relays in $\varphi_t + \varphi_c = 1000 + 100 = 1100$ independent epochs have been perfectly detected without estimation error, denoted as $\{g_1(m, t)\}$ and $\{g_2(m, t)\}$, we can subsequently adopt random sampling to split these 1100 data into two groups for training and checking purposes, respectively. Unless otherwise stated, $\{g_1(m, t)\}$ and $\{g_2(m, t)\}$ are generated based on exponential distribution with the rate parameter set to one. For examining the training transferability, in Fig. 4, we have also experimented on Rician distributed training data with normalized parameters. To facilitate the training process, we first perform dimension deduction to derive the end-to-end
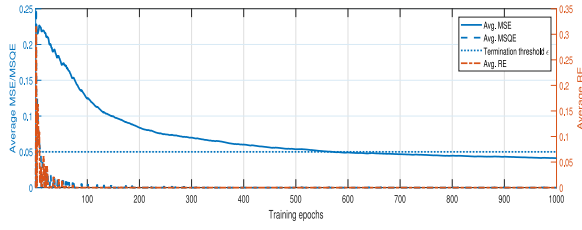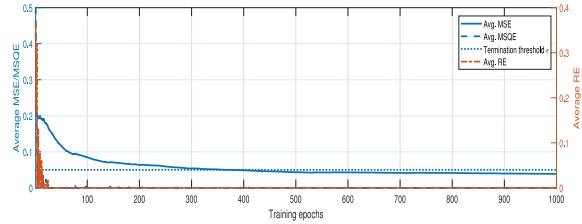
(a) Training data from exponential distribution (1st model).
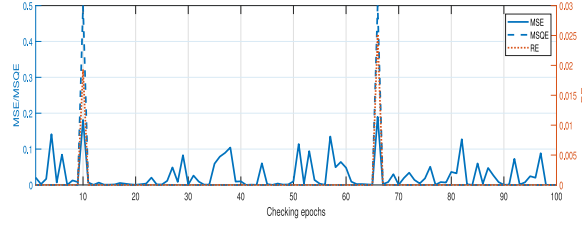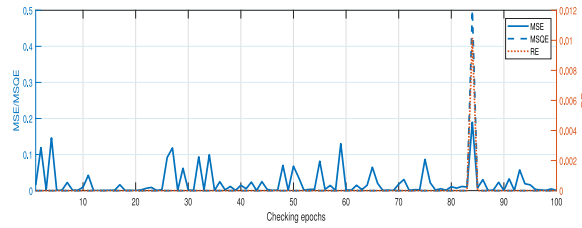

(b) Testing data from exponential distribution (1st model).
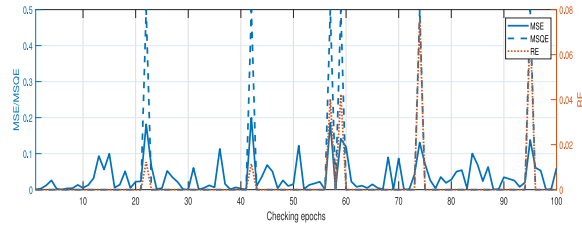
(c) Testing data from Rician distribution (first model).

(d) Training data from exponential and Rician distributions (2nd model).

(e) Testing data from exponential and Rician distributions (2nd model).

Fig. 4. Performance indicators in both of the training and testing stages for the exemplary case, given $\alpha = 1$ and $\beta = 1$.

channel magnitudes by

$$g(m, t) = \min\{g_1(m, t), g_2(m, t)\}, \tag{7}$$

which characterizes the end-to-end cooperative channel quality via the mth relay at epoch t. Then, we map the end-to-end channel magnitudes to values within the range $[\chi_{\min}, \chi_{\max}] = [-1, 1]$ for normalization purposes. The mapped value of $g(m, t)$ is determined by [45]

$$\lambda(m, t) = -1 + \frac{2\Big(g(m, t) - \inf\limits_{m}\{g(m, t)\}\Big)}{\sup\limits_{m}\{g(m, t)\} - \inf\limits_{m}\{g(m, t)\}} \in [-1, 1], \tag{8}$$

which is treated as the direct input of the ANN. In the above equation, $\inf\{\cdot\}$ and $\sup\{\cdot\}$ denote the infimum and supremum, returning the greatest lower bound and the least upper bound among all values of m.

Considering a simplistic relay selection criterion maximizing the end-to-end channel magnitude [46]

$$m^*(t) = \arg\max\limits_{m \in \mathcal{M}}\{g(m, t)\}, \tag{9}$$

we can generate the label by comparing all M end-to-end channel magnitudes at epoch t to find the maximum and the corresponding relay index, albeit with quadratic computational complexity [14]. Therefore, for epoch t, $\{\mathbf{G}(t), \mathbf{S}(t)\}$ is the labeled data set, where $\mathbf{G}(t) = [\lambda(1, t), \lambda(2, t), \ldots, \lambda(M, t)]^T$; only the $m^*(t)$th entry of $\mathbf{S}(t)$ is unity and other entries of $\mathbf{S}(t)$ are zero, which are used to compare with the outputs of the ANN to determine correction terms of ANN coefficients for the next epoch.

In the input layer, the outputs of neurons is equal to the corresponding inputs, i.e.,

$$y_1^{k_1}(t) = \lambda(k_1, t). \tag{10}$$

Without loss of generality, for the k_lth neuron in the hidden layers and the output layer, we have

$$y_l^{k_l}(t) = \mathcal{F}_A\Bigg(\sum_{k_{l-1}=1}^{K_{l-1}} w_{k_{l-1} \to k_l}(t) y_{l-1}^{k_{l-1}}(t) - \theta_l^{k_l}(t)\Bigg), \tag{11}$$

where $\mathcal{F}_A(\cdot)$ is called the activation function, which defines the way how neurons process the input signals and map them to output signals. Following most application scenarios of supervised ML, we adopt the sigmoid model for the activation function in our exemplary case and thereby let $\mathcal{F}_A(x) = 1/(1 + \exp(-x))$.

We can construct the estimated relay selection vector as

$$\tilde{\mathbf{S}}(t) = \Big[y_L^1(t), y_L^2(t), \ldots, y_M^1(t)\Big]^T, \tag{12}$$

which can be used to produce the error vector as

$$\mathbf{E}(t) = \mathbf{S}(t) - \tilde{\mathbf{S}}(t) = [e_1(t), e_2(t), \ldots, e_M(t)]^T. \tag{13}$$

According to gradient descent and the back-propagation mechanism [47], the error gradients for the neurons in the output layer can be written as

$$\delta_L^{k_L}(t) = y_L^{k_L}(t)\Big(1 - y_L^{k_L}(t)\Big)e_{k_L}(t). \tag{14}$$

Similarly, the error gradients for the neurons in the input and hidden layers can be expressed as

$$\delta_l^{k_l}(t) = y_l^{k_l}(t)\Big(1 - y_l^{k_l}(t)\Big)\sum_{k_{l+1}=1}^{K_{l+1}} \delta_{l+1}^{k_{l+1}}(t) w_{k_l \to k_{l+1}}(t). \tag{15}$$

With the error gradients, we can determine the correction terms for link weights and activation thresholds to be

$$\Delta w_{k_l \to k_{l+1}}(t) = \alpha y_l^{k_l}(t)\delta_{l+1}^{k_{l+1}}(t) \tag{16}$$

and

$$\Delta\theta_l^{k_l}(t) = -\alpha\delta_l^{k_l}(t). \tag{17}$$

Therefore, the link weights and activation thresholds are updated for the next epoch by the following relations:

$$w_{k_l \to k_{l+1}}(t+1) = w_{k_l \to k_{l+1}}(t) + \Delta w_{k_l \to k_{l+1}}(t). \tag{18}$$

and

$$\theta_l^{k_l}(t+1) = \theta_l^{k_l}(t) + \Delta\theta_l^{k_l}(t). \tag{19}$$

In this way, we train the ANN by iteratively updating both sets of ANN coefficients until the termination conditions are satisfied.

Once both termination conditions have been satisfied, we perform the comprehensive check by leveraging the checking data sets and see if the optimal sets of selected relays can be identified in most cases. For clarity, we plot the performance indicators in both of the training and checking stages for the exemplary case in Fig. 4. From this figure, we can clearly see that both average MSE and MSQE accounting previous training epochs diminish by repeated training, and average MSQE approaches zero after the initial fluctuation period within 50 training epochs. After being trained for 350 epochs, average MSE has been lower than the termination threshold $\epsilon = 0.05$. This satisfies *Condition 1* for termination. Moreover, by checking the average RE accounting previous training epochs, we can observe its convergence to zero after the dynamic fluctuation before the first 50 epochs. As for the checking stage, it is clear that for most cases the trained ANN is able to provide optimal solutions achieving the maximum channel magnitude, but a small number of overshoots over the MSQE curve exist resulting in mismatch between the yielded channel magnitude and the optimum. These overshoots are caused once more than one channel magnitudes share close values and thereby can hardly be differentiated by the trained ANN without being trained over a gigantic data space. Fortunately, because channel magnitudes are very close, the performance loss caused by the selection mismatch is trivial. With $\varphi_t = 1000$ and a sufficiently large φ_c, we numerically find the average mismatch ratio converge around 2.47%. Therefore, the trained ANN can be regarded as passing the comprehensive check and is thereby solidified.

We then test the solidified ANN by

$$[g_1(1,t), g_1(2,t), g_1(3,t), g_1(4,t)]^T$$
$$= [0.8942, 4.1632, 1.6688, 2.6932]^T \tag{20}$$

and

$$[g_2(1,t), g_2(2,t), g_2(3,t), g_2(4,t)]^T$$
$$= [2.4514, 0.6583, 4.2545, 0.7376]^T, \tag{21}$$

which give $\mathbf{G}(t) = [-0.5330, -1.0000, 1.0000, -0.8430]^T$ as the input vector by (7) and (8). The vector output after processing via the solidified ANN is $\tilde{\mathbf{S}}(t) = [0.0303, 0.0010, 0.9544, 0.0011]^T$, which can be quantized to be $\Xi(\tilde{\mathbf{S}}(t)) = [0, 0, 1, 0]^T$. Therefore, the quantized output vector indicates that the third relay should be selected, which is aligned with our expectation. The above observations on

the case study verify the effectiveness and efficiency of the proposed transfer learning based relay selection framework.

Once the generic framework is constructed, various transfer learning methods can be applied for training purposes. A common approach is to transfer the knowledge and learning weights from an original trained model for prospective learning tasks. The initial layers of the ANN tend to store more coarse information regarding the dataset, which is likely to remain the same for the other tasks sharing a similar nature. On the other hand, the last layers of the ANN tend to be more focused on detailed information of dataset. Armed with this knowledge, given a new and relatively smaller dataset, only the last one or two layers will be adapted for training, and the weights associated with the other layers are transferred from the existing model. With a similar setup as above, we trained the ANN based on the exponential distribution; however, we utilized the test dataset generated from the Rician distribution. Not surprisingly, as the distribution of checking and training datasets are different, the model's accuracy was poor, with MSE of 0.27. Then, we generated a new small training dataset from the Rician distribution with only 100 inputs and trained the last layer of the ANN model further. Such an approach resulted in the performance improvement of 40% once the Rician distributed test dataset was applied again on the model.

VI. HYPOTHESIS TEST FOR TRAINING CONVERGENCE

Although the simple checking method also works in the previous example, we design a hypothesis test to rigorously study the training convergence in this section. Because an error can be regarded as determined by many independent factors, e.g., current inputs, training data sets, and ANN initialized parameters, by the symmetry of the input vector and the conditions of the Lyapunov theorem, it is reasonable to assume that all errors in the checking stage are independent and identically distributed Gaussian random variables. Mathematically, we assume $e_m(t) \sim \mathcal{N}(\mu, \sigma^2)$, $1 \le m \le M$ and $1 \le t \le \varphi_c$, where μ and σ^2 are the mean and variance of the errors in the checking stage. Note that, both μ and σ^2 are determined in the training stage and are assumed to be fixed throughout the entire checking stage. This is because the ANN parameters and architecture will not be altered by processing the testing data sets. Therefore, this results in a stationary stochastic process capturing the system errors [48]. The above statistical assumptions can greatly facilitate the following mathematical analyses and the design of the hypothesis test for evaluating training convergence.

By the definition of MSE given in (3) and the error vector shown in (13), we can alternatively write MSE as

$$\mathrm{MSE}(t) = \frac{1}{M} \sum_{m=1}^{M} (e_m(t))^2. \tag{22}$$

Because of the assumption of $e_m(t) \sim \mathcal{N}(\mu, \sigma^2)$, $1 \le m \le M$, according to the definition and property of non-central chi-square distribution, we can $\mathrm{MSE}(t)$ as a non-central chi-square distributed variable with the following probability

density function (PDF) [49]:

$$
\hat{f}_{\mathrm{MSE}}(z) = \frac{M}{2\sigma^2}\left(\frac{z}{\mu^2}\right)^{\frac{M-2}{4}}\exp\left(-\frac{M}{2\sigma^2}\left(z+\mu^2\right)\right)
$$
$$
\times\; I_{\frac{M}{2}-1}\left(\frac{M|\mu|}{\sigma^2}\sqrt{z}\right), \quad z \geq 0, \tag{23}
$$

where $I_\nu(\cdot)$ represents the νth-order modified Bessel function of the first kind. When M is an even number, the CDF of $\mathrm{MSE}(t)$ can be written in closed form as

$$
\hat{F}_{\mathrm{MSE}}(z) = 1 - Q_{\frac{M}{2}}\left(\frac{\sqrt{M}|\mu|}{\sigma}, \frac{\sqrt{Mz}}{\sigma}\right), \quad z \geq 0, \tag{24}
$$

where $Q.(\cdot,\cdot)$ represents the generalized Marcum Q function [50], [51]. The accurate and efficient derivation of the Marcum Q function allows the derivation of useful bounds and approximations.

Therefore, based on the assumption made above, the unknown statistical parameters in (23) are the mean and derivation of the error terms, i.e., μ and σ^2. To estimate both statistical parameters, we can employ either the moment-matching estimation method or maximum likelihood estimation method. The former is easy to compute, while the later could be more accurate and has the asymptotic normality. Note that, the estimates are not unique and can be quite different by different estimation methods.

With φ_c independent observations, denoted as $\{z_i\}_{i=1}^{\varphi_c}$, we first show the estimations for μ and σ^2 by the moment-matching estimation method. By (23), it can be derived that the first and second moments of $\mathrm{MSE}(t)$ are

$$
\mathbb{E}\{\mathrm{MSE}(t)\} = \mu^2 + \sigma^2 \tag{25}
$$

and

$$
\mathbb{E}\left\{(\mathrm{MSE}(t))^2\right\} = \mu^4 + \left(\frac{2}{M}+1\right)\sigma^4 + 2\left(\frac{2}{M}+1\right)\mu^2\sigma^2, \tag{26}
$$

where $\mathbb{E}\{\cdot\}$ returns the expectation of the random variable enclosed. Denoting the first and second empirical moments as $\varrho_1 = \frac{1}{\varphi_c}\sum_{i=1}^{\varphi_c} z_i$ and $\varrho_2 = \frac{1}{\varphi_c}\sum_{i=1}^{\varphi_c} z_i^2$, both empirical moments converge in probability towards $\mathbb{E}\{\mathrm{MSE}(t)\}$ and $\mathbb{E}\{(\mathrm{MSE}(t))^2\}$, i.e., $\lim_{\varphi_c\to\infty}\mathbb{P}\{|\varrho_r - \mathbb{E}\{(\mathrm{MSE}(t))^r\}| < \varepsilon\} = 1$, where $r \in \mathbb{N}^+$ and $\varepsilon > 0$ is an arbitrarily small value. Therefore, we can construct the following equation set

$$
\begin{cases} \hat{\mu}^2 + \hat{\sigma}^2 = \varrho_1 \\ \hat{\mu}^4 + \left(\frac{2}{M}+1\right)\hat{\sigma}^4 + 2\left(\frac{2}{M}+1\right)\hat{\mu}^2\hat{\sigma}^2 = \varrho_2 \end{cases} \tag{27}
$$

to obtain the estimates of of μ and σ^2 as

$$
\begin{cases} \hat{\mu} = \pm\left[\frac{(M+2)\varrho_1^2 - M\varrho_2^2}{2}\right]^{\frac{1}{4}} \\ \hat{\sigma}^2 = \varrho_1 - \left[\frac{(M+2)\varrho_1^2 - M\varrho_2^2}{2}\right]^{\frac{1}{2}}. \end{cases} \tag{28}
$$

It should be noted from the estimates given in (28) that the estimates will be meaningful if and only if the following condition can be met: $\sqrt{\frac{M}{M+2}} < \frac{\varrho_1}{\varrho_2} < 1$. Also, the sign of $\hat{\mu}$ cannot be directly determined by solving the equation set and

should be further discussed by other means, e.g., higher-order moments and hypothesis tests.

Because of the limitations of the moment-matching estimation method, we can also resort to the maximum likelihood estimation method to estimate μ and σ^2. In particular, we can construct and utilize the log-likelihood function

$$
\mathcal{L}\left(\mu,\sigma^2\right) = \log\left(\prod_{i=1}^{\varphi_c}\hat{f}_{\mathrm{MSE}}(z_i)\right)
$$
$$
= \varphi_c\log\left(\frac{M}{2\sigma^2}\right) + \frac{(M-2)}{4}\sum_{i=1}^{\varphi_c}\log\left(\frac{z_i}{\mu^2}\right)
$$
$$
- \frac{M}{2\sigma^2}\sum_{i=1}^{\varphi_c} z_i - \frac{\varphi_c M\mu^2}{2\sigma^2}
$$
$$
+ \sum_{i=1}^{\varphi_c}\log\left(I_{\frac{M}{2}-1}\left(\frac{M|\mu|}{\sigma^2}\sqrt{z_i}\right)\right). \tag{29}
$$

Then, the maximum-likelihood estimates of μ and σ^2 can be obtained by the following criterion:

$$
\{\hat{\mu},\hat{\sigma}^2\} = \underset{\mu,\sigma^2}{\arg\max}\left\{\mathcal{L}\left(\mu,\sigma^2\right)\right\}. \tag{30}
$$

It can be proven that $\mathcal{L}(\mu,\sigma^2)$ is a quasi-concave function with respect to μ and σ^2. Therefore, we can obtain the maximum-likelihood estimates of μ and σ^2 by solving the log-likelihood equation set

$$
\begin{cases} \frac{\partial\mathcal{L}(\hat{\mu},\hat{\sigma}^2)}{\partial\hat{\mu}} = 0 \\ \frac{\partial\mathcal{L}(\hat{\mu},\hat{\sigma}^2)}{\partial\hat{\sigma}^2} = 0. \end{cases} \tag{31}
$$

However, to the best of the authors' knowledge, there is no closed-form solutions to the maximum-likelihood equation set, because of the involvement of modified Bessel functions. Various numerical computing techniques, e.g., convex/concave optimization techniques, the Newton-Raphson method, and quasi-Newton methods, can be utilized to yield numerical solutions [52]. As the computing procedures and studies of these numerical computing techniques are complex and out of the scope of this paper, we omit the detailed discussion here. In particular, if we further assume $\mu = 0$ according to the symmetry of the input vector, then the non-central chi-square distribution of MSE can be reduced to a central chi-square distribution, and the maximum-likelihood estimate of σ^2 can be derived in closed form as

$$
\hat{\sigma}^2 = \frac{1}{\varphi_c}\sum_{i=1}^{\varphi_c} z_i. \tag{32}
$$

Having obtained $\hat{\mu}$ and $\hat{\sigma}^2$, we can obtain an estimated distribution of MSE by (23) and (24) to its authentic PDF $f_{\mathrm{MSE}}(z)$ and CDF $F_{\mathrm{MSE}}(z)$.

For illustration purposes, we plot the estimated PDF and CDF of MSE yielded by the maximum likelihood estimation method in Fig. 5 for different values of M, given a sufficiently large φ_c. From this figure, it can be noticed that there still exist obvious mismatches between the empirical results and the estimated results. This is because the actual distributions of
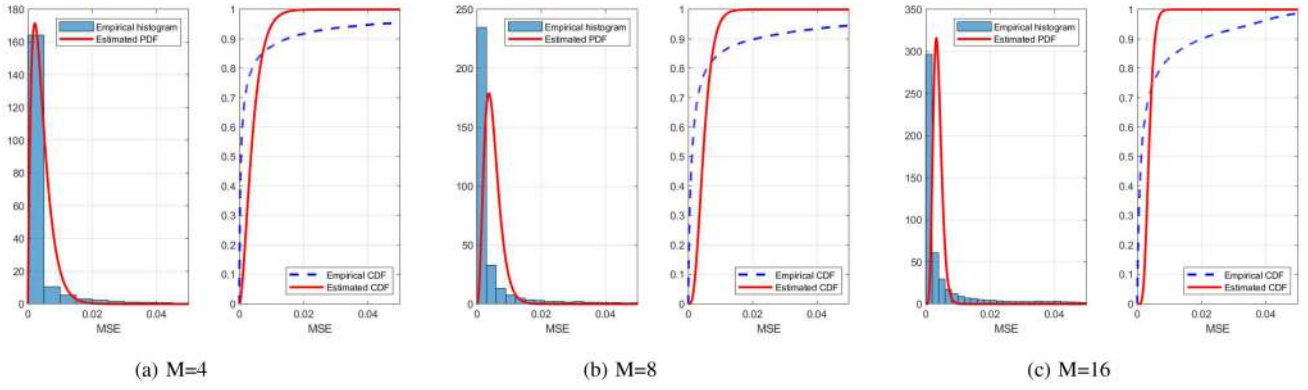
(a) M=4   (b) M=8   (c) M=16

Fig. 5. Estimated PDF and CDF of MSE yielded by the maximum likelihood estimation method, given a sufficiently large φ_c.

the error terms might not be strictly i.i.d. and Gaussian. That is, the estimation can only roughly capture the property of MSE and might not be accurate, but as our ultimate goal is to design a hypothesis test for evaluating the training convergence, this is still useful to help such a design by stipulating a proper significance level.

To rigorously examine the rationality of the aforementioned assumption and estimations leading to $e_m(t) \sim \mathcal{N}(\hat{\mu}, \hat{\sigma}^2)$, $1 \leq m \leq M$ and $1 \leq t \leq \varphi_c$, we resort to the distribution fit test:

$$H_0 : f_{\text{MSE}}(z) = \hat{f}_{\text{MSE}}(z), \quad H_1 : f_{\text{MSE}}(z) \neq \hat{f}_{\text{MSE}}(z). \quad (33)$$

To carry out the distribution fit test, we first need to split the entire domain of definition for z, i.e., $\{z \geq 0\}$, into η disjoint regions, denoted as $\{A_i\}_{i=1}^{\eta}$, and then observe the numbers of samples falling into these η disjoint regions, denoted as $\{C_i\}_{i=1}^{\eta}$. Denoting $P_i = \int_{A_i} \hat{f}_{\text{MSE}}(z)\,\mathrm{d}z$, it has been proven by Pearson that the following test statistic approximately abides the standard chi-square distribution $\chi^2(\eta - 1)$ when H_0 is true and φ_c is *sufficiently* large (empirically, $\varphi_c \geq 50$) [53]:

$$\vartheta = \sum_{i=1}^{\eta} \left[\frac{\varphi_c}{P_i} \left(\frac{C_i}{\varphi_c} - P_i \right)^2 \right] = \sum_{i=1}^{\eta} \left(\frac{C_i^2}{\varphi_c P_i} \right) - \varphi_c. \quad (34)$$

Therefore, when H_0 is true, ϑ should not be too large, otherwise the null hypothesis should be rejected. Given a level of statistical significance ω, the region of rejection can be written as

$$\vartheta \in \left[\chi_\omega^2(\eta - 1), \infty \right), \quad (35)$$

where $\chi_\omega^2(\eta - 1)$ is the upper quantile function of the standard chi-square distribution $\chi^2(\eta - 1)$, obtained by $\int_{\chi_\omega^2(\eta-1)}^{\infty} \frac{(\chi^2)^{\frac{\eta-1}{2}-1} \exp(-\frac{\chi^2}{2})}{2^{\frac{\eta-1}{2}} \Gamma(\frac{\eta-1}{2})}\,\mathrm{d}\chi^2 = \omega$, and $\Gamma(\cdot)$ is the gamma function.

Based on the assumption of the uniform Gaussian distribution of error terms, we can easily derive the variance of $\text{MSE}(t)$ by (25) and (26) as

$$\mathbb{D}\{\text{MSE}(t)\} = \mathbb{E}\left\{ (\text{MSE}(t))^2 \right\} - (\mathbb{E}\{\text{MSE}(t)\})^2$$
$$= \frac{2\sigma^2}{M} \left(2\mu^2 + \sigma^2 \right), \quad (36)$$

where $\mathbb{D}\{\cdot\}$ returns the variance of the random variable enclosed. Therefore, according to the central limit theorem, we can approximately have the distribution of $\varrho_1 = \frac{1}{\varphi_c} \sum_{i=1}^{\varphi_c} z_i$ when φ_c is large:

$$\varrho_1 \sim \mathcal{N}\left(\mu^2 + \sigma^2, \frac{2\sigma^2}{\varphi_c M} \left(2\mu^2 + \sigma^2 \right) \right). \quad (37)$$

The distribution relation given in (37) can be normalized to the following relation

$$\frac{\varrho_1 - \left(\mu^2 + \sigma^2 \right)}{\sqrt{\frac{2\sigma^2}{\varphi_c M} \left(2\mu^2 + \sigma^2 \right)}} \sim \mathcal{N}(0, 1), \quad (38)$$

which forms the basics of the hypothesis test for the training convergence.

In particular, by (25), we have $\mathbb{E}\{\text{MSE}(t)\} = \mu^2 + \sigma^2$. Therefore, we can quantify the training convergence by the average MSE yielded by φ_c testing data sets and finally design a hypothesis test to check the training convergence as follows:

$$H_0 : \mu^2 + \sigma^2 \leq \epsilon, \quad H_1 : \mu^2 + \sigma^2 > \epsilon. \quad (39)$$

Referring to (38), this is in nature a hypothesis test for the mean of a Gaussian distribution when the variance is unknown, which is also known as a Student's t-test. To construct the test statistic for a Student's t-test, we first need to construct an auxiliary statistic

$$\varrho_s = \frac{1}{\varphi_c - 1} \sum_{i=1}^{\varphi_c} (z_i - \varrho_1)^2 = \frac{1}{\varphi_c - 1} \left[\sum_{i=1}^{\varphi_c} \left(z_i^2 \right) - \varphi_c \varrho_1^2 \right], \quad (40)$$

which is an unbiased estimate of $\frac{2\sigma^2}{M}(2\mu^2 + \sigma^2)$, since $\mathbb{E}\{\varrho_s\} = \frac{2\sigma^2}{M}(2\mu^2 + \sigma^2)$. It can also be proven by linear transformation of orthogonal sample matrices that

$$\frac{M(\varphi_c - 1)}{2\sigma^2(2\mu^2 + \sigma^2)} \varrho_s \sim \chi^2(\varphi_c - 1), \quad (41)$$

and ϱ_1 and ϱ_s are statistically independent [48]. Consequently, according to the definition of the Student's t-distribution, we can have the following distribution relation:

$$\varrho_t = \frac{\varrho_1 - \left(\mu^2 + \sigma^2 \right)}{\sqrt{\varrho_s / \varphi_c}} \sim \mathcal{T}(\varphi_c - 1). \quad (42)$$
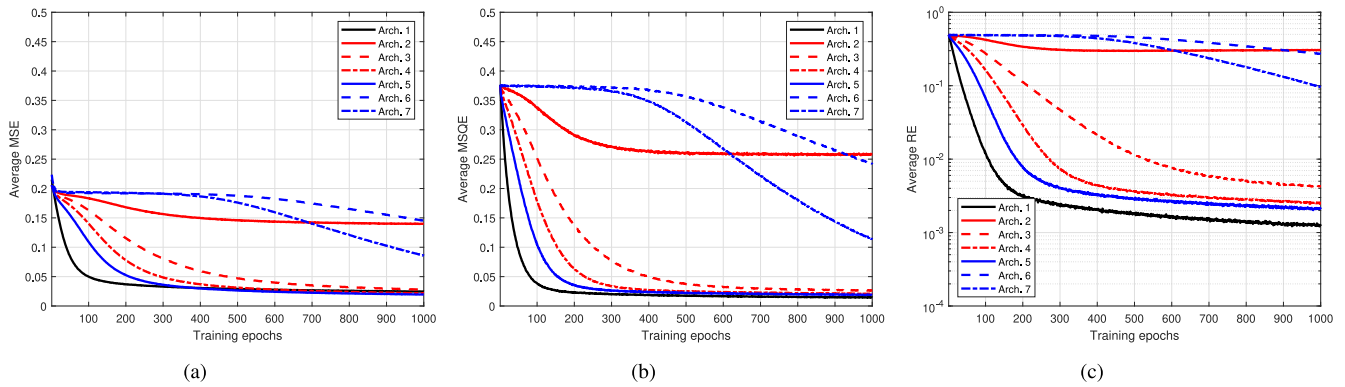
Fig. 6.   Average MSE, MSQE, RE for different ANN architectures.

Therefore, replacing $\mu^2 + \sigma^2$ by ϵ we can observe

$$\kappa = \frac{\varrho_1 - \epsilon}{\sqrt{\varrho_s/\varphi_c}} \tag{43}$$

to study the hypothesis test formulated in (39). For such a standard Student's t-test, the region of rejection can be written as

$$\kappa \in [\mathcal{T}_\omega(\varphi_c - 1), \infty), \tag{44}$$

where $\mathcal{T}_\omega(\varphi_c - 1)$ is the upper quantile function of the Student's t-distribution with $\varphi_c - 1$ degrees of freedom, which can be determined by $\int_{\mathcal{T}_\omega(\varphi_c - 1)}^{\infty} \frac{\Gamma(\frac{\varphi_c}{2})(1 + \frac{\mathcal{T}^2}{\varphi_c - 1})^{-\frac{\varphi_c}{2}}}{\sqrt{\pi(\varphi_c - 1)}\Gamma(\frac{\varphi_c - 1}{2})} \, d\mathcal{T} = \omega$, and ω is a predetermined significance level of the hypothesis test for training convergence.

VII. NUMERICAL INVESTIGATIONS AND DISCUSSION

Although effectiveness and efficiency of the proposed transfer learning based relay selection framework have been verified by the exemplary case in Section V, it is still unknown how system configurations and hyperparameters should be tuned so that the optimal training performance is achievable. To enhance the practicability of the proposed framework, we numerically investigate the effects of system configurations and hyperparameters on the training performance via Monte Carlo simulations in this section. To apply the control variates method, we adopt the same system model and training model in Section V but only change one system setting variable per subsection in following subsections. In this way, the effects of different settings can be distinguished and then studied. Meanwhile, to eliminate the randomness brought by the channel realizations and reveal the statistical effects of system configurations and hyperparameters, all presented simulation results in the following subsections are averaged over 10^4 repeated trials.

A. Artificial Neural Network Architecture

When applying transfer learning via the ANN, the most obvious question is what ANN architecture should be employed. Researchers have found that there is not a general rule or guideline to answer this question, and the ANN architecture should be case specific, depending on the optimization problem and processing requirements [54]. Without many
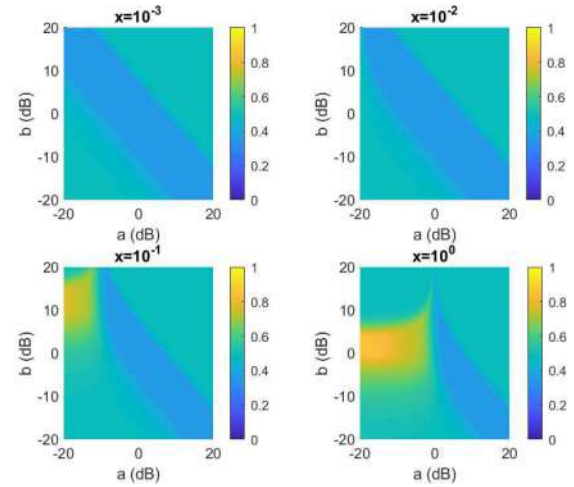
Fig. 7.   General sigmoid activation function $\mathcal{F}_A(x)$ with different coefficients a and b.

clues, we tested a series of representative ANN architectures for $M = 4$: Architecture 1: $[M, M]^T$ (the simplest architecture used as the comparison benchmark), Architecture 2: $[M, 1, M]^T$, Architecture 3: $[M, M, M]^T$, Architecture 4: $[M, 2M, M]^T$, Architecture 5: $[M, 4M, M]^T$, Architecture 6: $[M, M, M, M]^T$, and Architecture 7: $[M, 2M, 2M, M]^T$. The simulation results corresponding to different ANN architectures are presented in Fig. 6. The simulation results shown in this figure are informative and enlightening, which are inconsistent to some common verdicts and guidelines of ANN architecture design. It is surprising that the most efficient ANN architecture for such a simple relay selection problem is the simplest architecture without any hidden layer. This again implies that there does not exist the best ANN architecture for all optimization problems, but the most suited ANN architecture for a specific optimization problem. Meanwhile, increasing the number of hidden layers could lead to even a lower training efficiency. In addition, increasing the number of neurons in the hidden layers results in a higher training efficiency. Also, it is observed that there exists an evident bottleneck effect, as Architecture 2 has the worst performance.

B. Activation Function

Apart from the ANN architecture, another configuration that can significantly affect the training performance is the adopted
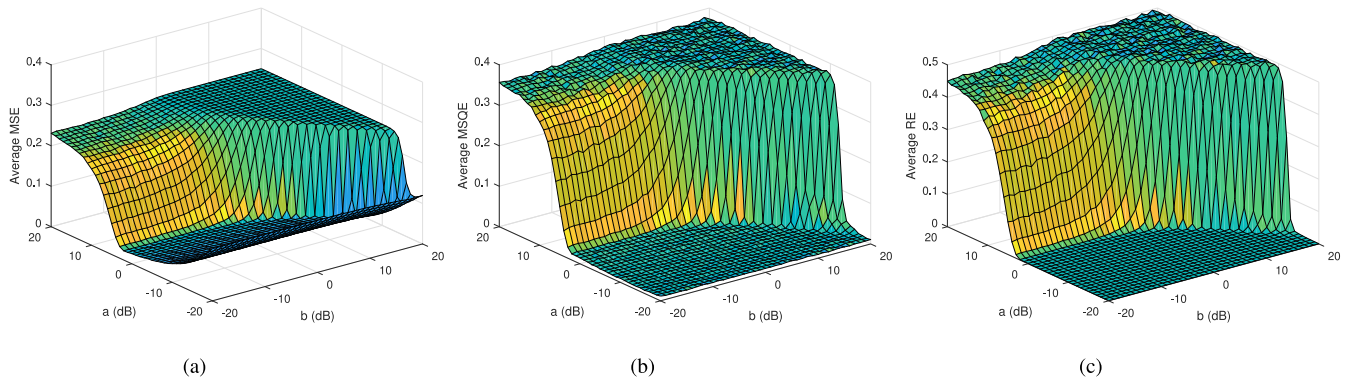
Fig. 8. Average MSE, MSQE, RE for different sigmoid activation function coefficients a and b, given $\varphi_t = 1000$.
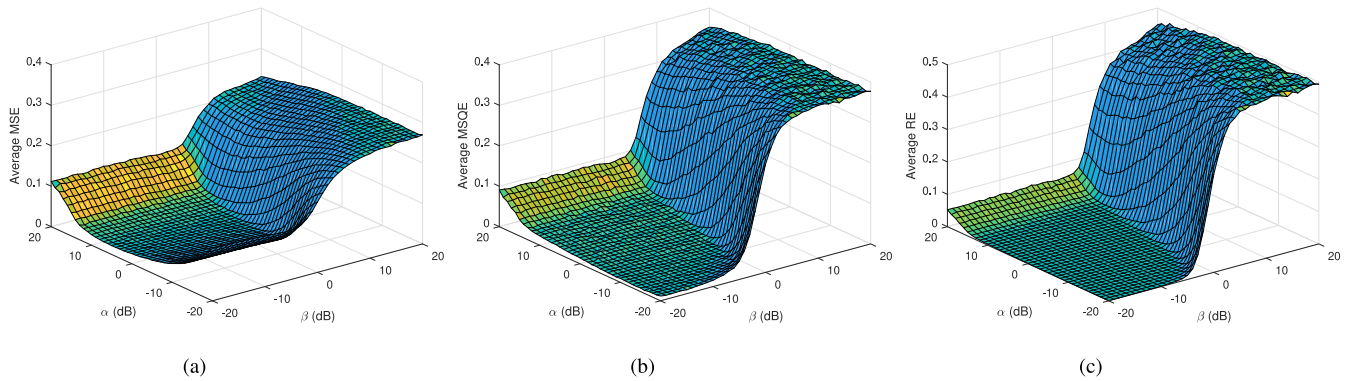

Fig. 9. Average MSE, MSQE, RE for different learning rates α and range expansion coefficients β, given $\varphi_t = 1000$.

activation function [55]. As the expected outputs of the ANN for relay selection problems are within the range [0, 1], the sigmoid activation model is more appropriate to be used than other models, e.g., step, sign, hyperbolic tangent, and linear activation models [56]. In this regard, we can express the general form of the sigmoid activation function as

$$\mathcal{F}_A(x) = \frac{1}{1 + \exp(-a(x - b))}, \qquad (45)$$

which is characterized by the coefficients a and b. We plot the sigmoid activation function $\mathcal{F}_A(x)$ with different coefficients a and b in Fig. 7.

To investigate the effects of both coefficients on the average MSE, MSQE, and RE, we illustrate the numerical results in Fig. 8. From this figure, it is clear that a should not be either too large or too small, otherwise the training performance can be degraded. The ideal value of a is around unity. However, a small a only results in a higher average MSE, but does not raise average MSQE and RE. This means that the negative impacts of small a can be effectively filtered out by quantization. On the other hand, the numerical results show that the value of b should ideally be zero.

C. Learning Rate and Range Expansion Coefficient

The learning rate α and the range expansion coefficient β have impact on the training efficiency of the ANN. The former defines the training increment for each iterative round, and the latter regulates the initialization procedure of the ANN. We

investigate the relation among the training performance, the learning rate α, and the range expansion coefficient β through numerical simulations and present the simulation results in Fig. 9. These simulation results are aligned with our expectation that the learning rate α should be kept within a proper range so as to achieve satisfactory training performance. This is because a large α might not lead to convergence forever, due to the wide oscillation alongside the optimum, while a small α leads to a small movement to the optimum per epoch and thereby requires a huge number of training data sets as well as a huge amount of training time to reach the optimum. Simulations have also corroborated that for BIP problems (the outputs of ANNs are within 0 and 1), the range expansion coefficient β should be kept small enough so as to accelerate convergence.

D. Number of Relays

The number of relays M significantly affects the computational scale and system complexity of the proposed framework, which is worth studying. We thereby study the numerical relation between the training performance and the number of relays M and depict the results in Fig. 10. From this figure, it is apparent that increasing the number of relays makes it difficult for the ANN to find the optimal subset of selected relays. However, the mismatch between the yielded subset and the optimal subset of selected relays can be reduced by a higher number of training epochs. As the average RE is sufficiently low by appropriate training with sufficient labeled data sets,
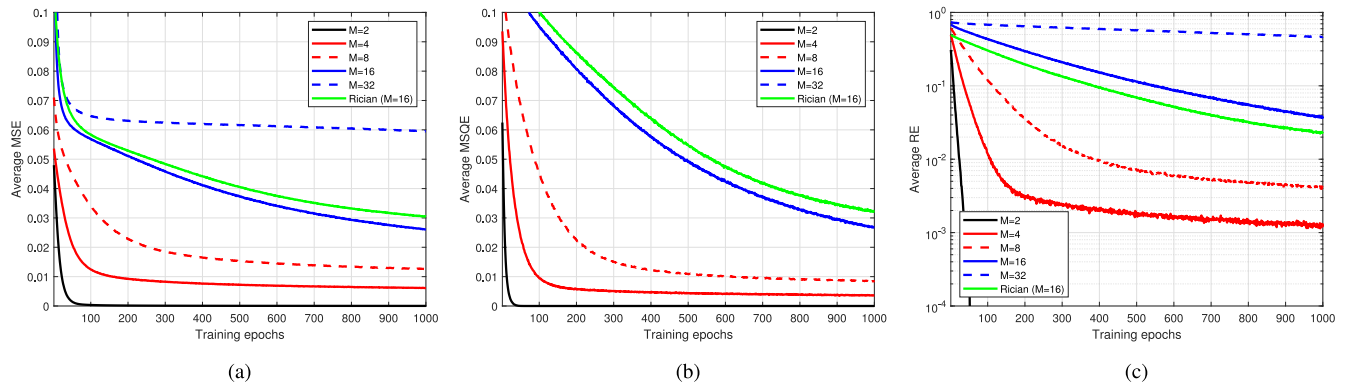
Fig. 10. Average MSE, MSQE, RE for different numbers of relays *M*.

(a) $\omega = 0.01$.

(b) $\omega = 0.20$.
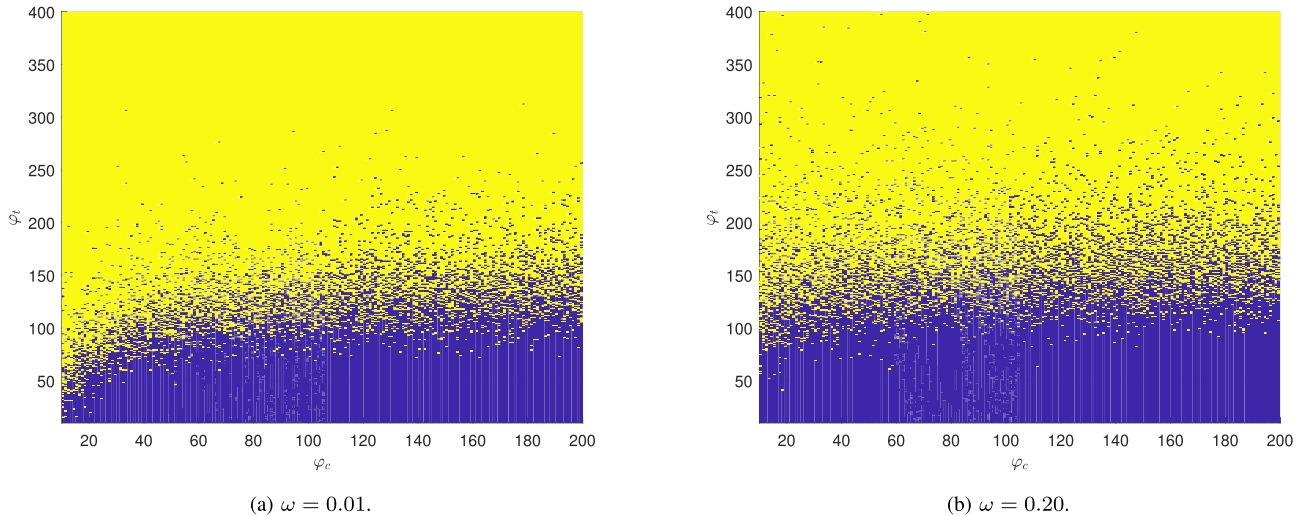
Fig. 11. Hypothesis test results represented by ρ with respect to φ_t and φ_c: the yellow and blue points correspond to the acceptance and rejection cases of the null hypothesis (i.e., $\rho = 1$ and $\rho = 0$).

the proposed framework is feasible for implementing relay selection with a large number of relays *M*.

E. Hypothesis Test for Training Convergence

To show the usefulness of the proposed hypothesis test for the training convergence, we adopt a set of toy examples with small values of φ_t and φ_c. For clarity, we plot the following statistic in Fig. 11 with respect to φ_t and φ_c, given different significance levels:

$$\rho = \mathbb{I}\{\kappa < \mathcal{T}_\omega(\varphi_c - 1)\}, \qquad (46)$$

where $\mathbb{I}\{\cdot\}$ is the Iverson bracket returning 1 if the enclosed condition is true or 0 otherwise. Specifically, $\rho = 1$ indicates that the null hypothesis $H_0 : \mu^2 + \sigma^2 \le \epsilon$ cannot be rejected, which corresponds to the belief of a converged ANN after training.

From Fig. 11, it is clear that the proposed hypothesis test in the checking stage can provide a probabilistic and significant reference for judging whether the ANN after training has converged or not. With an increasing number of training data sets φ_t, the ANN is more likely to converge in the checking stage, which is aligned with our expectation. The number of checking data sets φ_c is only related to the accessibility of statistical

regularity, while it is essentially irrelevant to the training convergence. As a result, when the number of checking data sets φ_c is sufficiently large, it is almost uncorrelated with ρ. Also, we can increase/decrease predetermined significance level ω to tighten/loosen the criterion of the proposed hypothesis test. In particular, the higher ω is set, the higher the number of training data sets φ_t would be required on average to pass the hypothesis test, and vice versa.

VIII. CONCLUSION

In this paper, we constructed a generic and comprehensive relay selection framework aided by transfer learning and ANNs. We also employed an exemplary case to demonstrate the implementation of the constructed framework and verified its effectiveness and efficiency. To be rigorous, we designed a hypothesis test based on reasonable statistical assumptions to estimate whether the ANN after training has converged and can be used in practice. We further numerical investigated the system configurations and hyperparameters of the constructed framework and obtained some insight for optimizing the framework. We have paid special attention to maintain the generality of the relay selection framework so that it can

be easily tailored to serve a variety of application scenarios of relay selections with diverse objective functions and constraints in the industrial IoT. By applying the proposed relay selection framework, it is expected that complicated relay selection scheme can be carried out in real time with lower selection complexity and processing latency. As a consequence, the coverage, connectivity, and resilience of the industrial IoT can be enhanced on an automatic and real-time basis. The proposed framework is in particular suited for large-scale wireless sensor networks in the IoT environment.

REFERENCES

[1] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.

[2] A. Nosratinia, T. E. Hunter, and A. Hedayat, "Cooperative communication in wireless networks," *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 74–80, Oct. 2004.

[3] A. Chandra, C. Bose, and M. K. Bose, "Wireless relays for next generation broadband networks," *IEEE Potentials*, vol. 30, no. 2, pp. 39–43, Mar./Apr. 2011.

[4] M. Pejanovic-Djurisic, E. Kocan, and R. Prasad, *OFDM Based Relay Systems for Future Wireless Communications* (River Publishers Series in Communications Series). Aalborg, Denmark: River Publ., 2012.

[5] E. Sisinni *et al.*, "LoRaWAN range extender for industrial IoT," *IEEE Trans. Ind. Informat.*, vol. 16, no. 8, pp. 5607–5616, Aug. 2020.

[6] W. Yang and Y. Cai, "On the performance of the block-based selective OFDM decode-and-forward relaying scheme for 4G mobile communication systems," *J. Commun. Netw.*, vol. 13, no. 1, pp. 56–62, Feb. 2011.

[7] Q. Huang, M. Ghogho, J. Wei, and P. Ciblat, "Practical timing and frequency synchronization for OFDM-based cooperative systems," *IEEE Trans. Signal Process.*, vol. 58, no. 7, pp. 3706–3716, Jul. 2010.

[8] P. Murphy and A. Sabharwal, "Design, implementation, and characterization of a cooperative communications system," *IEEE Trans. Veh. Technol.*, vol. 60, no. 6, pp. 2534–2544, Jul. 2011.

[9] Y. Li, W. Wang, and F.-C. Zheng, "Combined bulk and per-tone relay selection in cooperative OFDM systems," in *Proc. IEEE ICCC*, Beijing, China, 2012, pp. 487–491.

[10] F. Xu, F. C. M. Lau, Q. F. Zhou, and D.-W. Yue, "Outage performance of cooperative communication systems using opportunistic relaying and selection combining receiver," *IEEE Signal Process. Lett.*, vol. 16, no. 4, pp. 237–240, Apr. 2009.

[11] D. S. Michalopoulos and G. K. Karagiannidis, "Performance analysis of single relay selection in Rayleigh fading," *IEEE Trans. Wireless Commun.*, vol. 7, no. 10, pp. 3718–3724, Oct. 2008.

[12] Z. Zhang, X. Chai, K. Long, A. V. Vasilakos, and L. Hanzo, "Full duplex techniques for 5G networks: Self-interference cancellation, protocol design, and relay selection," *IEEE Commun. Mag.*, vol. 53, no. 5, pp. 128–137, May 2015.

[13] M.-H. Lu, P. Steenkiste, and T. Chen, "Design, implementation and evaluation of an efficient opportunistic retransmission protocol," in *Proc. ACM Annu. Int. Conf. Mobile Comput. Netw. (MobiCom)*, Beijing China, 2009, pp. 73–84.

[14] Y. Jing and H. Jafarkhani, "Single and multiple relay selection schemes and their achievable diversity orders," *IEEE Trans. Wireless Commun.*, vol. 8, no. 3, pp. 1414–1423, Mar. 2009.

[15] S. Dang, J. P. Coon, and G. Chen, "Outage performance of two-hop OFDM systems with spatially random decode-and-forward relays," *IEEE Access*, vol. 5, pp. 27514–27524, 2017.

[16] T. Fu, S. Dang, L. Wu, and M. Wen, "Combined relay selection in spatially random networks," *IEEE Syst. J.*, vol. 15, no. 3, pp. 3726–3729, Sep. 2021.

[17] A. Bletsas, A. Khisti, D. P. Reed, and A. Lippman, "A simple cooperative diversity method based on network path selection," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 659–672, Mar. 2006.

[18] A. Bletsas, A. Lippnian, and D. P. Reed, "A simple distributed method for relay selection in cooperative diversity wireless networks, based on reciprocity and channel measurements," in *Proc. IEEE VTC*, vol. 3. Stockholm, Sweden, 2005, pp. 1484–1488.

[19] M. Chen, U. Challita, W. Saad, C. Yin, and M. Debbah, "Artificial neural networks-based machine learning for wireless networks: A tutorial," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 4, pp. 3039–3071, 4th Quart, 2019.

[20] E. Bjornson and P. Giselsson, "Two applications of deep learning in the physical layer of communication systems [lecture notes]," *IEEE Signal Process. Mag.*, vol. 37, no. 5, pp. 134–140, Sep. 2020.

[21] M. Chen, Z. Yang, W. Saad, C. Yin, H. V. Poor, and S. Cui, "A joint learning and communications framework for federated learning over wireless networks," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 269–283, Jan. 2021.

[22] J. Kang, Z. Xiong, D. Niyato, Y. Zou, Y. Zhang, and M. Guizani, "Reliable federated learning for mobile networks," *IEEE Wireless Commun.*, vol. 27, no. 2, pp. 72–80, Apr. 2020.

[23] W. Y. B. Lim *et al.*, "Hierarchical incentive mechanism design for federated machine learning in mobile networks," *IEEE Internet Things J.*, vol. 7, no. 10, pp. 9575–9588, Oct. 2020.

[24] C. T. Nguyen *et al.*, "Transfer learning for future wireless networks: A comprehensive survey," 2021, *arXiv:2102.07572*.

[25] S. Abdulhadi, M. Jaseemuddin, and A. Anpalagan, "A survey of distributed relay selection schemes in cooperative wireless ad hoc networks," *Wireless Pers. Commun.*, vol. 63, no. 4, pp. 917–935, 2012.

[26] N. Nomikos *et al.*, "A survey on buffer-aided relay selection," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 1073–1097, 2nd Quart., 2016.

[27] Y. Bhute and A. Raut, "A survey on relay selection strategies in cooperative wireless network for capacity enhancement," *Int. J. Comput. Appl.*, vol. 65, no. 25, pp. 12–17, Mar. 2013.

[28] S. Umamaheswari and M. Sathya, "A comprehensive survey on cooperative relaying in industrial wireless sensor network," *Int. J. Eng. Res. Technol.*, vol. 6, no. 7, pp. 591–596, 2017.

[29] D. P. Kumar, T. Amgoth, and C. S. R. Annavarapu, "Machine learning algorithms for wireless sensor networks: A survey," *Inf. Fusion*, vol. 49, pp. 1–25, Sep. 2019.

[30] S. Dang, O. Amin, B. Shihada, and M.-S. Alouini, "What should 6G be?" *Nat. Electron.*, vol. 3, no. 1, pp. 20–29, 2020.

[31] K. Sankhe, C. Pradhan, S. Kumar, and G. Ramamurthy, "Machine learning based cooperative relay selection in virtual MIMO," 2015, *arXiv:1506.01910*.

[32] H. Dang, Y. Liang, L. Wei, C. Li, and S. Dang, "Enabling relay selection in cooperative networks by supervised machine learning," in *Proc. IEEE Int. Conf. Instrum. Meas. Comput. Commun. Control (IMCCC)*, Harbin, China, 2018, pp. 1459–1463.

[33] H. Dang, Y. Liang, L. Wei, C. Li, and S. Dang, "Artificial neural network design for enabling relay selection by supervised machine learning," in *Proc. 8th Int. Conf. Instrum. Meas. Comput. Commun. Control (IMCCC)*, Harbin, China, 2018, pp. 1464–1468.

[34] S. Dang, M. Wen, S. Mumtaz, J. Li, and C. Li, "Enabling multicarrier relay selection by sensing fusion and cascaded ANN for intelligent vehicular communications," *IEEE Sensors J.*, vol. 21, no. 14, pp. 15614–15625, Jul. 2021.

[35] Z. Deng, Q. Sang, Y. Gao, and C. Cai, "Optimal relay selection for wireless relay channel with external eavesdropper: A NN-based approach," in *Proc. IEEE/CIC Int. Conf. Commun. China (ICCC)*, Beijing, China, Aug. 2018, pp. 515–519.

[36] A. Abdelreheem, O. A. Omer, H. Esmaiel, and U. S. Mohamed, "Deep learning-based relay selection in D2D millimeter wave communications," in *Proc. IEEE Int. Conf. Comput. Inf. Sci. (ICCIS)*, Sakaka, Saudi Arabia, Apr. 2019, pp. 1–5.

[37] A. Mchergui, T. Moulahi, and S. Nasri, "Relay selection based on deep learning for broadcasting in VANET," in *Proc. IEEE Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Tangier, Morocco, Jun. 2019, pp. 865–870.

[38] W. Wang, S. L. Capitaneanu, D. Marinca, and E.-S. Lohan, "Comparative analysis of channel models for industrial IoT wireless communication," *IEEE Access*, vol. 7, pp. 91627–91640, 2019.

[39] C. Jiang, H. Zhang, Y. Ren, Z. Han, K.-C. Chen, and L. Hanzo, "Machine learning paradigms for next-generation wireless networks," *IEEE Wireless Commun.*, vol. 24, no. 2, pp. 98–105, Apr. 2017.

[40] M. Negnevitsky, *Artificial Intelligence: A Guide to Intelligent Systems*. Harlow, U.K.: Addison-Wesley, 2005.

[41] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[42] S. Dang, J. P. Coon, and G. Chen, "Resource allocation for full-duplex relay-assisted device-to-device multicarrier systems," *IEEE Wireless Commun. Lett.*, vol. 6, no. 2, pp. 166–169, Apr. 2017.

[43] M. B. Cohen, Y. T. Lee, and Z. Song, "Solving linear programs in the current matrix multiplication time," *J. ACM*, vol. 68, no. 1, pp. 1–39, 2021.

[44] Z. Wang, S. Dang, S. Shaham, Z. Zhang, and Z. Lv, "Basic research methodology in wireless communications: The first course for research-based graduate students," *IEEE Access*, vol. 7, pp. 86678–86696, 2019.

[45] D. Kim, "Normalization methods for input and output vectors in back-propagation neural networks," *Int. J. Comput. Math.*, vol. 71, no. 2, pp. 161–171, 1999.

[46] A. Bletsas, H. Shin, and M. Z. Win, "Cooperative communications with outage-optimal opportunistic relaying," *IEEE Trans. Wireless Commun.*, vol. 6, no. 9, pp. 3450–3460, Sep. 2007.

[47] M. Z. Rehman and N. M. Nawi, "The effect of adaptive momentum in improving the accuracy of gradient descent back propagation algorithm on classification problems," in *Proc. Int. Conf. Softw. Eng. Comput. Syst.*, 2011, pp. 380–390.

[48] A. Papoulis and S. Pillai, *Probability, Random Variables, and Stochastic Processes* (McGraw-Hill Series in Electrical Engineering: Communications and Signal Processing). New York, NY, USA: McGraw-Hill, 2002.

[49] J. Proakis and M. Salehi, *Digital Communications*. Boston, MA, USA: McGraw-Hill, 2008.

[50] S. András, A. Baricz, and Y. Sun, "The generalized marcum Q—Function: An orthogonal polynomial approach," 2010, *arXiv:1010.3348*.

[51] C. W. Helstrom, "Computing the generalized marcum Q-function," *IEEE Trans. Inf. Theory*, vol. 38, no. 4, pp. 1422–1428, Jul. 1992.

[52] R. Millar, *Maximum Likelihood Estimation and Inference: With Examples in R, SAS and ADMB* (Statistics in Practice). Chichester, U.K.: Wiley, 2011.

[53] R. C. Dahiya and J. Gurland, "Pearson chi-squared test of fit with random intervals," *Biometrika*, vol. 59, no. 1, pp. 147–153, 1972.

[54] B. M. Wilamowski, "Neural network architectures and learning algorithms," *IEEE Ind. Electron. Mag.*, vol. 3, no. 4, pp. 56–63, Dec. 2009.

[55] B. Karlik and A. V. Olgac, "Performance analysis of various activation functions in generalized MLP architectures of neural networks," *Int. J. Artif. Intell. Expert Syst.*, vol. 1, no. 4, pp. 111–122, 2011.

[56] J. Saxe and K. Berlin, "Deep neural network based malware detection using two dimensional binary program features," in *Proc. IEEE Int. Conf. Malicious Unwanted Softw. (MALWARE)*, Fajardo, PR, USA, Oct. 2015, pp. 11–20.

Miaowen Wen (Senior Member, IEEE) received the Ph.D. degree from Peking University, Beijing, China, in 2014. From 2012 to 2013, he was a Visiting Student Research Collaborator with Princeton University, Princeton, NJ, USA. He is currently a Professor with the South China University of Technology, Guangzhou, China. He has published two books and more than 130 journal papers. His research interests include a variety of topics in the areas of wireless and molecular communications. He was the recipient of the IEEE Asia–Pacific Outstanding Young Researcher Award in 2020, and four Best Paper Awards from the IEEE ITST12, the IEEE ITSC14, the IEEE ICNC16, and the IEEE ICCT19. He served as a Guest Editor for the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS (Special Issue on Spatial Modulation for Emerging Wireless Systems) and for the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING (Special Issue on Index Modulation for Future Wireless Networks: A Signal Processing Perspective). He is currently serving as an Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS, the IEEE TRANSACTIONS ON MOLECULAR, BIOLOGICAL, AND MULTISCALE COMMUNICATIONS, and the IEEE COMMUNICATIONS LETTERS, and as a Guest Editor for the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING (Special Issue on Advanced Signal Processing for Local and Private 5G Networks).

Shahid Mumtaz (Senior Member, IEEE) received the M.Sc. degree in electrical and electronic engineering from the Blekinge Institute of Technology, Karlskrona, Sweden, in 2006, and the Ph.D. degree in electrical and electronic engineering from the University of Aveiro, Aveiro, Portugal, in 2011. In 2005, he was a Research Intern with Ericsson, Stockholm, Sweden, and Huawei Research Laboratory, Karlskrona. He has more than 13 years of wireless industry experience and is currently a Senior Research Scientist and the Technical Manager with the Instituto de Telecomunicações Aveiro, Aveiro, Portugal. He has authored several conferences, journals, and books publications. His current research interests include field of architectural enhancements to 3GPP networks, such as, LTE-A user plan and control plan protocol stack, NAS, and EPC, 5G-related technologies, green communications, cognitive radio, cooperative networking, radio resource management, cross-layer design, backhaul/fronthaul, heterogeneous networks, machine-to-machine and device-to-device communication, and baseband digital signal processing.

Sina Shaham received the B.Eng. degree (Hons.) in electrical and electronic engineering from the University of Manchester, the M.Sc. degree in engineering management from the University of Warwick, and the M.Phil. degree in computer science from the University of Sydney. He is currently pursuing the Ph.D. degree with the University of Southern California. He has also worked as a Data Scientist and a Software Engineer. His research interests include privacy and machine learning.

Varun G. Menon (Senior Member, IEEE) received the Diploma degree in training and development, the M.B.A. degree in human resource management, the M.Sc. degree in applied psychology, the M.Tech. degree (with University First Rank) in computer and communication, and the Ph.D. degree in computer science and engineering. He is currently an Associate Professor and the Head of the Department of Computer Science Engineering, and International Collaborations and Corporate Relations in charge with the SCMS School of Engineering and Technology, India. His research interests include sensor technologies, Internet of Things, green IoT, wireless communication, fog computing and networking. He is a Distinguished Speaker of the Association of Computing Machinery.

Shuping Dang (Member, IEEE) received B.Eng. degree (with First Class Hons.) in electrical and electronic engineering from the University of Manchester and the B.Eng. degree in electrical engineering and automation from Beijing Jiaotong University in 2014 via a joint '2+2' dual-degree program, and the D.Phil. degree in engineering science from the University of Oxford in 2018. He joined the R&D Center, Huanan Communication Company Ltd., after graduating from the University of Oxford and worked as a Postdoctoral Fellow with the Computer, Electrical and Mathematical Science and Engineering Division, King Abdullah University of Science and Technology. He is currently a Lecturer with the Department of Electrical and Electronic Engineering, University of Bristol. His research interests include 6G communications, wireless communications, wireless security, and machine learning for communications.

Chengzhong Li is a Qualified Senior Engineer and a Certified Expert in telecommunications with the Guangxi Science & Technology Department. He is currently working as the Chief Engineer with the R&D Center, Guangxi Huanan Communication Company Ltd., Nanning, China. He has been working on the telecommunication industries for more than ten years and specializes in telecommunication network design, feasibility study, and project management. He led a number of feasibility studies of major provincial and municipal telecommunication research and construction projects. His research interests include information integration, big data, artificial intelligence, and telecommunication economics.

ScienceDirect®

Design and analysis of logistic agent-based swarm-neural network for intelligent transportation system

Monagi H. Alkinani [a] ✉, Abdulwahab Ali Almazroi [b] ✉, Mainak Adhikari [c], Varun G. Menon [d] ✉

[a] Department of Computer Science and Artificial Intelligence, College of Computer Sciences and Engineering, University of Jeddah, Saudi Arabia
[b] Department of Information Technology, College of Computing and Information Technology at Khulais, University of Jeddah, Saudi Arabia
[c] Department of Computer Science, Indian Institute of Information Technology Lucknow, India
[d] Department of Computer Science and Engineering, SCMS School of Engineering and Technology, India

🔘 Check for updates

Show less ⌃

☰ Outline | ⤴ Share 💬 Cite

Get rights and content ↗

Abstract

Driven by the massive number of connected vehicles and the stringent requirements of data-intensive applications, logistics transportation systems have evolved to fully comprehend its effectiveness and quality to meet public transportation needs. As a result, to meet public transportation requirements and analyze the data efficiently at the edge of the networks, an advanced artificial intelligent technique needs to be introduced to make the transportation system intelligent by supporting efficient decision making, intelligent traffic control, and intrusion and misuse detection.Motivated by the challenges mentioned above, in this paper, we develop a logistic agent-based model for analyzing public transports such as cars, bus or trains in the intelligent transportation system. The intelligent logistic framework is built on a parallel neural network structure, known as a Swarm-Neural Network (SWNN). The proposed SWNN model analyzes the sensory data and recognizes the public transportation at the edge of the networks. The SWNN model is constructed so that it fits within the intelligent logistic transportation framework, and the proposed model shortens the transit time of every small-scale logistics delivery to its destination. The performance of the proposed SWNN model is evaluated using a standard TMD dataset, where the SWNN model is trained using data, retrieved multiple sensors such as accelerometer, gyroscope, magnetometer, and audio sensors. The features of the sensory data are extracted based on a 5-s time interval. The performance of the proposed SWNN model is studied over various standard machine learning techniques such as Random Forest, XGBoost, and Decision Tree. As per the simulation results, the proposed technique achieves 78–98% accuracy over a real-time dataset's different sets of features.

Keywords

Intelligent transportation system; Logistic framework; Edge networks; Swarm-neural network; Agent-based model

1. Introduction

Recent advancements in technology, together with the exponential rise in the number of vehicles and rapid urbanization, have paved the way to numerous developments and increased research in Intelligent Transportation Systems (ITS) [1], [2], [3], [4]. ITS involves the interconnection of vehicles and infrastructures and provides opportunities for continuous communications, interactions, and resource sharing among vehicles with different applications for ubiquitous data analysis and processing and maximizes transportation infrastructure's efficiency with improved service quality. Recently, there has been a massive demand in supply chain management for goods to be efficiently transported from warehouses to final destinations, leading to the rapid development of ITS-enabled logistic transportation systems [5], [6].

ITS-enabled logistic transportation systems and their gradual adoption of Autonomous Vehicles (AVs) are bringing massive advancements to smart cities [7]. But this has also led to the generation of unprecedented amounts of data, posing major challenges to the applications deploying ITS. With the deployment of numerous sensors in vehicles for various applications, there is a sharp increase in the data produced in ITS, rising from the Trillion-byte level to Petabyte [8], [1]. Analyzing the sensory data and recognizing the public transportation systems with minimum delay is a significant concern that requires efficient and robust solutions to meet the ever-increasing demand for efficiency and quality of service [9], [10], [11].

1.1. Related works

Over these years, many methods have been proposed for analyzing and processing the data for intelligent transportation systems. Most of the methods have focused on the management and processing of data in cloud servers. The authors in [12] discuss a technique for real-time active, safe driving with a three-tier cloud computing framework. The technique helps predict and analyze the high threat of backward shock waves using the data that has been collected from vehicles' state information. An intelligent technique using cloud and big data for systematic traffic control is presented in [13]. The system uses artificial intelligence and deep learning to predict traffic flow and congestion. Managing different types of data, including video, is a significant challenge in ITS, and authors in [14] discuss an interesting technique for effective video management with the cloud. The authors also try to solve the challenges in balancing the load and storage issues using a novel parallel computing model. A recommendation system for vehicle speed using a cloud/fog computing service infrastructure is discussed in [15]. The system uses a game approach to satisfy the objectives, where the drivers are the players and the vehicle's speed is their strategy.

With the advent of edge/fog computing, many applications in ITS have tried to reduce the latency in data storage and retrieval and processing that was a significant issue with the cloud-based systems. A comprehensive survey on edge cloud computing for ITS and connected vehicles is presented in [16]. The paper presents exciting discussions and future research perspectives on how edge cloud computing can be efficiently used in ITS. In [17], a technique for identification of traffic flows on the edge node with the help of deep learning is discussed. The authors discuss a real-time vehicle tracking counter that combines identification and tracking algorithms for vehicles to detect the traffic flow. An edge-enabled distributed trustworthy storage framework with reinforcement learning in ITS is discussed in [18]. The system adopts an intelligent technique for dynamic allocation for storage using reinforcement learning based on trustworthiness and popularity. A technique that uses multiple fog servers to detect the identity of vehicles is proposed in [19]. The system uses a voting mechanism to detect the most suitable fog server, determining the real identity and the trajectory. Despite all the aforementioned works, analysis and management of data at the edge for an efficient public transportation system is still a significant concern.

1.2. Motivation and contributions

Due to the increasing growth of the connected vehicles and the long communication distance between the vehicles and cloud servers, analyzing the vehicular data in the centralized cloud server may increase the noise on the monitoring dataset and deteriorate the prediction accuracy. Thus, one of the critical challenges is to design an intelligent transportation framework that can analyze the monitoring vehicular data locally while meeting the resource requirements during data analytics. In this paper, to address the challenge mentioned above, we design a new edge-centric framework with a set of resource-constraint edge devices and a centralized cloud server for efficiently analyzing vehicular sensory data.

Another important research aspect in transportation systems is to process the vehicular sensory data with an advanced machine learning model that can process the data on resource-constraint edge devices and increase the prediction accuracy with minimum delay. One of the potential solutions in this aspect is to integrate a lightweight machine learning technique with a set of distributed edge devices to develop an intellectual and logistic framework for the transportation system that can analyze the data at the edge of the network with higher accuracy and lower precision.

Motivated by the challenges mentioned above, in this work, we develop a logistic agent-based intelligent transportation framework in edge networks for identifications of public transportation using an agent-based Swarm-Neural Network (SWNN) model. In the proposed edge framework, data scheduling and preprocessing have been performed on the distributed edge devices while analyzing the modified dataset on the centralized cloud servers using the proposed SWNN model with higher accuracy. Thus, the main novelty of the proposed method is to integrate edge networks in transportation systems for making correct decisions and introduce a new SWNN model for analyzing vehicular data with higher prediction accuracy.The significant contributions of the work are highlighted as follows.

- Design an agent-based intelligent transportation framework to assist the delivery person of any small size logistic to find in public transportation. Thus, if the smart logistic movement framework correctly identifies the vehicle such as bus, train, or taxi, then logistic movement is feasible for that vendor through those vehicles. This technique shortens the transit time of every small-scale logistics that delivers the products to its destination.

- Design an intelligent logistic framework using a parallel neural network structure to analyze vehicular data, known as the SWNN technique. The proposed strategy improves the performance of the discussed system. The SWNN model is used to recognize a vehicle, and the sensor attached to this vehicle sends data to the distributed edge server for detecting the vehicles. The SWNN model is constructed so that it fits within the intelligent logistic transportation framework and analyzes the sensory data with higher prediction accuracy

- The performance of the proposed SWNN model in the intelligent logistic transport framework is evaluated using the TMD dataset, on which the SWNN model is trained using data from multiple sensors. The sensors data used in this experiment include accelerometer, gyroscope, magnetometer, and audio sensors. Further, the sensor data are analyzed to choose the best characteristics of the proposed model. The same characteristics are then taken from additional sensor data and utilized to build the patterns. The features of the sensory data are extracted based on a 5-s time interval. The performance of the proposed SWNN model is studied over various simulation matrices with various standard machine learning techniques.

The remaining sections of the paper are organized as follows. The system model of the proposed logistic transportation system and the proposed SWNN model for analyzing the sensory data of the public transportation system in edge networks are described in Section 2. The empirical analysis of the work over various standard machine learning algorithms is described in Section 3. Finally, the conclusion and future direction of the work are highlighted in Section 4.

2. Swarm-neural network for intelligent transportation system

This section describes various components of the proposed logistic transportation system followed by the proposed SWNN model with a suitable algorithm.

2.1. System model of logistic transportation system

The intelligent transport system framework incorporates intelligent agent-based swarm-neural network approaches to give the capacity to recognize the logistic carrying vehicle. This method is meant to be easily compatible with the edge-enabled transport framework. This component of the system is implemented in the edge server to handle data processing and analysis. Different components of the proposed SWNN model of the logistic transportation system are shown in Fig. 1. As per Fig. 1, the sensory vehicular data is collected by the set of distributed edge devices, and the data is temporarily held in the data-gathering phase in the resource-constrained edge devices before being transmitted to the data processing phase. Data is sampled in this method, and the collected data is then provided to the feature selection algorithm, which calculates features from the sampled data. The selected feature is now ready to be sent into the SWNN for logistic type categorization.
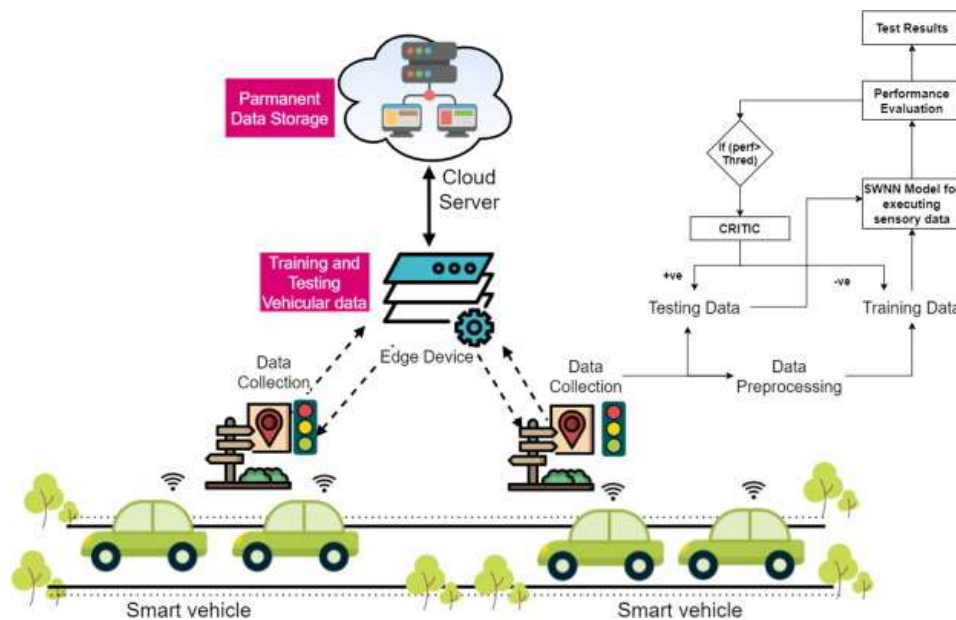
Download : Download high-res image (284KB)

Download : Download full-size image

Fig. 1. System Model of Logistic Transportation System.

The proposed technique considers the input from the vehicle's various sensors as well as a large number of vehicles submitting data to the system at the same time. This massive work of gathering data and processing it before feeding it to SWNN is broken down into four stages. The data scheduler is introduced into the system in the first phase to schedule the raw data from the sensor for data processing. The scheduler takes each sensor's data and places it in a queue to be processed by the data processing phase. The second phase is data processing, which is used to process the sensor signal using window-based sampling techniques. The features are extracted from each sample data in the third phase. The fourth phase is where the created sample is classified based on previous experience. The fifth phase's final step incorporates a rule-based decision support system to determine SWNN performance and govern overall system confidence based on that performance. The data is then stored in clod servers for any further processing and analysis in the future.

2.2. Data scheduler

In the proposed approach, the data scheduler is responsible for managing the incoming data from remote IoT devices deployed in the logistic vehicle. The scheduler is connected to several watchdogs to receive data from various IoT devices, and when the data is received, the lightweight watchdog procedures send it to the scheduler for first-come, first-served processing. In the data processing step, the data is then scheduled for processing. The data scheduler assists the system in receiving sensor data in parallel from multiple IoT devices in a single time quantum and so plays a critical role in network traffic management for the proposed method.

2.3. Data processing technique

The data processing component initially assembled the data by temporarily processing the data for each vehicle. One of the first jobs in this section is to check the correctness of each incoming data from various sensors, and only legitimate data may be sent on to the next phase. For each vehicle, sensor data is pooled over a defined time period. Following that, each vehicle's data is evaluated for feature extraction. The extraction of significant features is one of the most crucial things in sensor data processing. The data collected from numerous sensors are frequently complex and non-linear. Because of the changing traffic environment, these sensor signals with different frequencies may be seen, making them unpredictable. The sensor signals collected from various IoT devices are not stationary. As a result, the signal's enormous array is separated into N portions, with each window containing a predetermined window of size S for extracting features. The following are the features retrieved from the various windows in the proposed technique.

Correlation: This process of feature selection incorporates Correlation-based selection. The correlation between two features X_1 and X_2 is calculated

$$CO(X_1, X_2) = \frac{\sum_{r=1}^{m}(X_1-\mu_{X_1})(X_2-\mu_{X_2})}{\sqrt{\sum_{r=1}^{m}(X_1-\mu_{X_1})}\cdot\sqrt{\sum_{r=1}^{m}(X_2-\mu_{X_2})}} \tag{1}$$

where the mean values of X_1 and X_2 signal is performed as $\mu_{X_1} = \sum_{m=1}^{m} X_1$ and $\mu_{X_2} = \sum_{m=1}^{m} X_2$. The output value of $CO(X_1, X_2) \in (+1, -1)$ where the output value close to +1 indicates that X_1 and X_2 are similar and then one considered in the pattern and if it is close −1 proves the uniqueness in between X_1 and X_2. Each signal's characteristic is retrieved by examining a five-second window on it. The mean, minimum, maximum, and standard deviation values are derived from the signals of several sensors. The formula for calculating these is given below.

Mean: The means of the five seconds signal data is calculated as $Mean\,(SI)_i^t = (\frac{1}{n}\sum_{i=0}^{i} SI_i)$, where i^{th} signal value and t represent the window of the current current iteration and N is the total number of signal values in the current window.

Minimum Value: The minimum value of each windows of signal is calculated as $Minimum\,(SI)_t^i = Minimum\{SI_1^t, SI_2^t, SI_3^t, \ldots, SI_i^t\}$.

Maximum Value: The maximum value of each symptom is defined as follows. $Maximum\,(SI)_t^i = Maximum\{SI_1^t, SI_2^t, SI_3^t, \ldots, SI_i^t\}$. In the obtained features from the vehicle sensor data, the fundamental statistical characteristics are used. These are mean, standard deviation, skewness, kurtosis, and correlation computed across a small time window of signals.

2.4. Intelligent agent-based model

The intelligent critic in the suggested system maximizes the SWNN classification method's confidence by continually training the system with newly observed data. This is what the logistics system's pre-trained SWNN-based detection is built on. This intelligent critic aids SWNN classification systems in continuous assessment and switches between training and testing phases using a rule-based mechanism.

Rule 1: $SNN_{accuracy}^{best} \geqslant \theta \Rightarrow Detection_{ready}$

Rule 2: $SNN_{accuracy}^{best} < \theta \Rightarrow Learning_{ready}$ where $\forall(NN_{accuracy}^{best}) \in BoNN_{accuracy}$,

$SNN_{accuracy}^{best}$ is the best accuracy produced by neurons in the SWNN population, θ is the threshold value. Thus, when Rule 1 of the intelligent critic is met, the test data is provided to SWNN as an input. In the instance of Rule 2, the intelligent critic guarantees that the SWNN classifier is allowed to continue training. The intelligent critic notifies the gateway and edge servers if any reply is observed in the test data.

2.5. Proposed Swarm-Neural Network (SWNN) model

The SWNN classification approach is used in this proposed strategy of the intelligent logistic transportation system for the detection of different types of logistic models. The SWNN classification method is used here to detect the transport mode efficiently. This approach is empirical, making the SWNN more fit for testing different sensor data, which varies due to dynamic traffic conditions. The SWNN is designed to complete its detection task through three different phases. In the first phase, a pre-defined set of neural networks of the same architecture is generated to create the population.

The initial rounds of weights and bias matrix are also generated during this phase. In the second phase, the training is started initially, and this training phase is started for the first time; then, the switching between the training and testing cycle depends on the intelligent critic. The weight and bias matrix of the neural networks in the populations is performed two times during the training phase. In the training phase, the weight and bias change based on the backpropagation algorithm, and then lastly, it is modified based on the population's best performing neural network structure.

In the proposed method, the neural network for the population is created and stored in the queue. The creation of the neural network involves the generation of the weight and bias matrix for each layer of the neural network. The weight matrix and bias matrix generation require generating the random numbers using the following equations.

$$W_i = R_i + w_C \tag{2}$$

$$B_i = R_i + b_C \tag{3}$$

where W_i and B_i is the weight and bias of ith neural network in the population of the size m and $R_i \in (0, 1)$ and w_c and b_c is constant. Now once the weight and bias for all the neural networks are created the population, then the SWNN classifier is ready to start the training phase. The actual output of every neural network in the population is presented as follows.

$$Y_i = \sum_{l=1}^{n} W_i P_K^T + B_i \tag{4}$$

where Y_i is the actual output of the ith neuron in the neural network population. The patterns are fed into every neural network of the neural population and it is done in every iteration. Now the evaluation of neurons is performed in parallel with others, so at the end of each iteration the performance is kept in the list YO_r, and to represent the parallel the $\prod$ symbol is used.

$$YO_r = (Y_1, Y_2, \ldots \ldots Y_i)^r = \prod_{r=1}^{r} (Y_i)_r \tag{5}$$

Each transport pattern is labeled with the type of transport and the label of the pattern is the target TR_K for pattern P_k. Now, these patterns are considered to be neural network populations and feed into the population by considering one at a time. So, after every iteration, the calculation of the error EO_r for the rth iteration is performed. The calculation of the Error is performed as follows.

$$EO_r = |TR_k^i - Y_k^i|_r \tag{6}$$

where TR_K^i is the target of pattern P_k The sum of square error or SSE is calculated for i^{th} neuron in the SWNN population as follows.

$$SSE_i = EO_r^T EO_r = |TR_i - Y^i|_r^T |TR_i - Y^i|_r \tag{7}$$

The matrix format for calculating the SSE is presented in Eqn. (7). Now, the first phase of the training of SWNN is performed with the back-propagation method.In this method the error is back-propagated and the sensitivity or S_l of lth layer is calculated from the $(l + 1)^{th}$ layer. The calculation of the sensitivity is calculated for all the neural layers of a neural network and it is calculated for every neural network in the population. The calculation of the sensitivity, SE_l of the single 3-layer feed-forward neural network is calculated as follows.

$$SE_l = \frac{\delta net_{l+1}}{\delta net_l} \frac{\delta F}{\delta net_{l+1}} = F_l(net_l)(W_{l+1})^T SE_{l+1} \tag{8}$$

where $F_l(net_l)(W_{l+1})$ is the Jacobin matrix. In this method for a neural network in the SWNN population, the sensitivity is calculated over 3-layer feed-forward network structure. So, the calculation of SE_l is started from the last

layer and moves back to the first layer where the sensitivity is calculated as follows.

$$
\begin{aligned}
SE_1 &= \frac{\delta F}{\delta net_l} \\
&= \frac{\delta (TR_k - Y)^T (TR_k - Y)}{\delta net_l} \\
&= -2(TR_k - Y)F_l(net_l)
\end{aligned}
\tag{9}
$$

Now, in the back-propagation phase, the weight and bias of all the neural networks in the neural population are calculated as follows.

$$
W_{i+1}^l = W_i^l - \alpha SE_l (Y_{l-1})^T
\tag{10}
$$

$$
B(i+1)^l = B_i^l - \alpha SE_l
\tag{11}
$$

Then Eqs. (10), (11) is helped us to modify the weight and bias of the neural network of the neural population. After a single iteration the neural population $NN_p\ op$ is evaluated and the performance of the best neuron $NN_b\ est$ is selected based on the following rule.

If $(Accuracy(NN_{pop}) \geqslant \theta)$ Then.

$$
Max(Accuracy(NN_{pop})) = NN_{best}
$$

Now, when this NN_{best} is decided then the last round of the weight and bias modification is performed. This equation of weight and bias modification considers the W_{best} and B_{best} of NN_{best} as the best weight and bias of the population for an iteration. The equation is constructed as follows.

$$
W_{i+1} = W_i + R_1 e^{(DW)}(W_{best} - W_i) + R_2(1 - rand(0,1))
\tag{12}
$$

$$
B_{i+1} = B_i + R_1 e^{(DB)}(B_{best} - B_i) + R_2(1 - rand(0,1))
\tag{13}
$$

where the *DW* and *DB* are the euclidean distance between the best and current weight and bias of the neural network. R_1 and R_2 are the two constant. The rand(0,1) is the random number in between 0,1. Based on the Eqs. (12), (13) is the weight and bias of all neural networks in the neural population is updated. The algorithm of the proposed SWNN model for intelligent transportation system is presented in Algorithm 1.

Algorithm 1

Proposed SWNN Model

---

1 INPUT: Pattern P_k, Target TR_k, Neural population NN_{pop}, Population weight W_{pop}, Population Bias B_{pop}, Learning rate α

2 OUTPUT: $NN_{best}, W_{i+1}, B_{i+1}$

1: Fixed the number of iteration: EP_k

2: **for** K:1 to $Size(NN_{pop})$ **do**

3: create weight and bias matrix for every NN based on Eqs. (2), (3)

4: $W_{pop} \leftarrow W_i$

5: $B_{pop} \leftarrow B_i$

6: **end for**

7: Maximum Independent Run = M_r

8: **while** $(Counter \leqslant M_r)$ **do**

9: **for** LP_i:1 to $Size(NN_{pop})$ **do**

10: Create lightweight process: $LP_{process}^i$

11: $LP_{process}^i \leftarrow NN_{Model}(\alpha, W_i, B_i)$

12: Each neuron in NN_{pop} updates W_i and B_i using Eqs. (10), (11)

13: Error of each NN $E_i \Rightarrow E_{pop}$

14: Calculate $Accuracy(NN_{pop})$ using Eq. (14)

15: **end for**

16: **if** $Max_{accuracy}(NN_{pop}) \geqslant \theta)$ **then**

17: Update Best neuron in NN_{pop}

18: $NN_{best}^{new} \Rightarrow Max_{accuracy}(NN_{pop})$

19: **end if**

20: Find $W_{best} \Rightarrow NN_{best}$

21: Find $B_{best} \Rightarrow NN_{best}$

22: **for** p:1 to $Size(NN_{pop})$ **do**

23: $DW \leftarrow Euclidean(W_i, W_{best})$

24: $DB \leftarrow Euclidean(B_i, B_{best})$

25: Update W_i using Eq. (12)

26: Update B_i using Eq. (13)

27: **end for**

28: $Counter = Counter + 1$

29: **if** $Accuracy(NN_{best}) \geqslant \Phi$ **then**

30: Break

31: **end if**

32: **end while**

---

3. Empirical evaluation

The proposed method's success is dependent on the SWNN-based classification approaches' performance, which aids in detecting the logistic transit type. This section evaluates and improves the suggested swarm neural network approach for the TMD dataset for optimum performance by modifying different parameters over a wide range of values. The SWNN is evaluated for its performance in terms of solution quality and stability in delivering the best solution. The simulation is performed over a dataset that is relevant to the theory presented in this work, and in this dataset, data is sent via car sensors, making it relevant for the experiment. The performance of the suggested technique is compared with the standard benchmark algorithms such as Random Forest, XGBoost, and Decision Tree.

3.1. Simulation setup and dataset

The test is carried out on an Intel i5 **9 th** generation computer with 8GB RAM and a 4GB graphics card, as well as a 1TB hard drive. The Numpy, Scipy, and Panda libraries are used to create the SWNN simulation, and the Matplotlib library is used to plot the studied results. The proposed method's performance is assessed in terms of solution quality, which is measured by calculating accuracy. The typical machine algorithm's classification error, precision, recall, and accuracy are evaluated.

The TMD dataset is separated into sections based on the number of features. These features are derived from the sensor data received and processed by the data processing section. The sensor signal is separated using a set-sized window in the processing stage. The features are then gathered from this window. The experiment employs an accelerometer, gyroscope, magnetometer, and audio sensors. The dataset is now sorted into three groups depending on the available features. The first dataset contains 12 features, the second contains 32 features, and the third dataset

contains 36 features. For comparing the performance of the proposed approach and other machine algorithms across different dimensions of the transport dataset, the dataset is divided into these three groups. The performances of the proposed method and the standard machine algorithm are analyzed based on classification error, precision, recall, and accuracy as follows,

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (14)$$

$$Precision = \frac{TP}{TP+FP} \qquad (15)$$

$$Recall = \frac{TP}{TP+FN} \qquad (16)$$

where TP = true positive, TN = True negative, FP = False Positive, and FN = False Negative. In the performance measurement, the high value of precision and recall reflects the highly accurate result.

The correlation value of the characteristics in the dataset is used to analyze them and the correlation matrix is constructed based on Eqn. 1. Such matrix is shown in Fig. 2. The following rule is used during the correlation-based feature selection process.
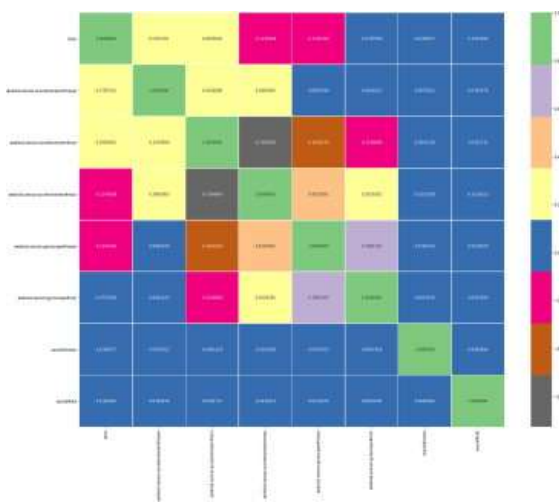
Download : Download high-res image (161KB)
Download : Download full-size image

Fig. 2. Correlation matrix of the 12 feature dataset without the implementation of the Rule.

If $CO(X_1, X_2) > 0.5$ *Then*

Remove X_2

Where X_1 and X_2 are the two different features. This rule states that the strongly correlated for the correlation matrix are eliminated. After eliminating the strongly linked features, the correlation matrix is re-plotted and shown in Fig. 3.
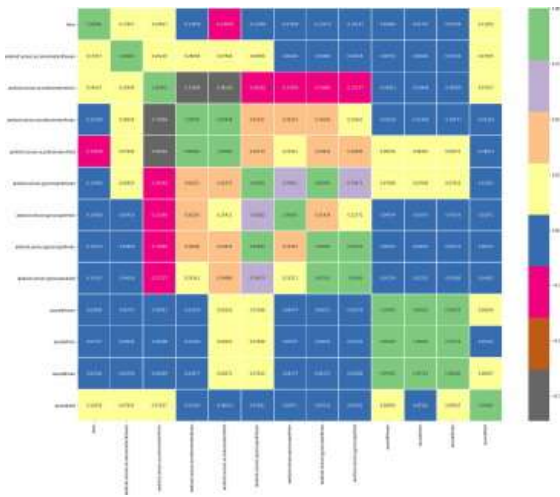
Fig. 3. Correlation matrix of the 12 feature dataset after the implementing the Rule.

The correlation matrix of 32 feature dataset without and with the implementation of the Rule are shown in Fig. 4, Fig. 5, respectively. Similarly, the correlation matrix of 36 feature dataset without and with the implementation of the Rule are shown in Fig. 6, Fig. 7, respectively.

Fig. 4. Correlation matrix of the 32 feature dataset without the implementation of the Rule.
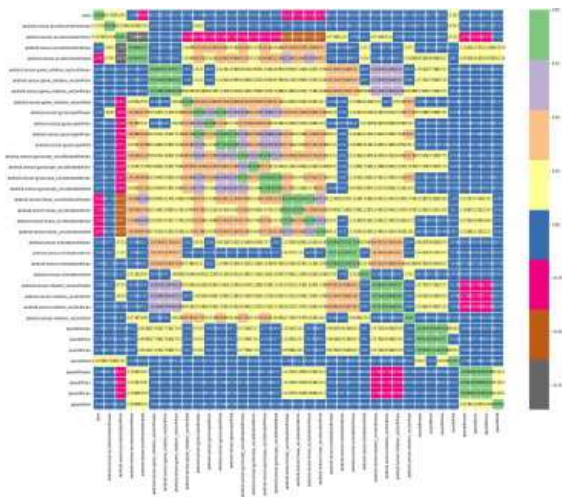
Fig. 5. Correlation matrix of the 32 feature dataset after the implementing the Rule.

Fig. 6. Correlation matrix of the 36 feature dataset without the implementation of the Rule.

Fig. 7. Correlation matrix of the 36 feature dataset after the implementing the Rule.

3.2. Parameter analysis for SWNN model

The parameter analysis for the proposed method is performed in an empirical way to adjust the parameters for the SWNN method over the transport mode dataset. The adjustment of these parameters will remain the same for any number of features used for the dataset. The parameter analysis for every parameter is initiated with a wide range of values, but only a few of them are selected for the presentation purpose. The parameters that impact the performance of the proposed SWNN model are neural population size and learning rate. The population size is tested with the different values, but the 5, 10, and 20 population size is used for the presentation purpose. (See Table 1, Table 2, Table 3).

Table 1. Performance analysis of SWNN by varying the learning rate with population size 5.

	Population size =5		
	Accuracy	Precision	Recall
Range1 (0.0001, 0.009)	0.781	0.701	0.771
Range2 (0.01, 0.06)	0.801	0.799	0.787
Range3 (0.06 0.125)	0.883	0.853	0.853

Table 2. Performance analysis of SWNN by varying the learning rate with population size 10.

	Population size =10		
	Accuracy	Precision	Recall
Range1 (0.0001, 0.009)	0.867	0.857	0.857
Range2 (0.01, 0.06)	0.901	0.901	0.901
Range3 (0.06 0.125)	0.976	0.966	0.966

Table 3. Performance analysis of SWNN by varying the learning rate with population size 20.

	Population size =20		
	Accuracy	Precision	Recall
Range1 (0.0001, 0.009)	0.971	0.970	0.970
Range2 (0.01, 0.06)	0.988	0.988	0.988
Range3 (0.06 0.125)	0.994	0.989	0.989

Considering the higher population size becomes unnecessary as the SWNN achieves its maximum accuracy with a population size of 20. The error produced with population size 10 and 20 is 0.024 and 0.006. Hence, population size 20 is good with the learning rate range 0.06 to 0.125. The Learning rate ranges are used instead of the single learning rate for neural networks of the neural population. The use of different learning rates creates a variation in the performances in the neural networks of the neural population. The learning rate in which the best performance is observed is 0.06 to 0.125. The higher learning rate is not used here because the proposed learning rate produces the best accuracy over the transport mode dataset. The overall fixed values of the different parameters are described in Table 4.

Table 4. Parameters of Swarm-NN.

Sl. No.	Parameter Names	Value
1	NN layer size	3
2	Noneuron/layer	1
3	No of NN in population	20
4	Learning range for neural population	0.6 to 0.125
5	ϕ (Eq. 16, Eq.17)	0.25

3.3. Comparative result analysis and discussion

Proposed swarm neural network-based techniques are applied over the dataset consisting of different numbers of features. This variation is analyzed to understand the stability of the proposed SWNN model over different dimensions of the dataset. The solution quality is also analyzed by comparing the performance of the proposed solution with another standard benchmark algorithm. The algorithms are Random Forest [20], Decision Tree [21], and K-Nearest Neighbor [22]. From the past few decades, Random Forest, Decision Tree, and K-Nearest Neighbour algorithms have achieved higher prediction accuracy, precision, and recall over the other machine learning models in various fields related to the Internet of Things, such as smart transportation systems, smart healthcare, smart agriculture, etc. Thus, to show the outperformance of the proposed SWNN model in the agent-based logistic transportation framework, we have introduced Random Forest, Decision Tree, and K-Nearest Neighbour algorithms for comparative analysis over a benchmark dataset. The performance of these algorithms is shown in Table 8.

From Table 5, Table 6, Table 7, it is observed that the performance of the algorithm increases with the number of features. Now, if the minimum number of features are available then the performance of the SWNN is also better than that of the standard algorithms. The performance of the SWNN over the dataset with 12 features is 24.3% more than that of the average accuracy of the other three algorithms, Table 5, Table 8. In the case of the dataset with 32 and 36 features, the performance increases by 26.3% and 14% respectively.

Table 5. Comparative analysis of proposed method (SWNN) with standard algorithms over 12 features.

Algorithms	Precision	Recall	Accuracy
Random Forest	0.71	0.71	0.77
Decision Tree	0.67	0.68	0.69
K-nearest Neighbour	0.51	0.49	0.51
Proposed Method	0.88	0.89	0.9

Table 6. Comparative analysis of proposed method (SWNN) with standard algorithms over 32 features.

Algorithms	Precision	Recall	Accuracy
Random Forest	0.75	0.76	0.8
Decision Tree	0.67	0.67	0.68
K-nearest Neighbour	0.54	0.53	0.54
Proposed Method	0.89	0.9	0.94

Table 7. Comparative analysis of proposed method (SWNN) with standard algorithms over 36 features.

Algorithms	Precision	Recall	Accuracy
Random Forest	0.83	0.83	0.84

Algorithms	Precision	Recall	Accuracy
Decision Tree	0.85	0.84	0.88
K-nearest Neighbour	0.78	0.79	0.8
Proposed Method	0.97	0.97	0.98

Table 8. Comparative analysis of the proposed method over different machine learning algorithms.

	12 Features	32 Features	36 Features
Random Forest	0.77	0.80	0.84
Decision Tree	0.69	0.68	0.88
K-nearest Neighbour	0.51	0.54	0.8
Proposed Method	0.90	0.94	0.98

The performance of the standard machine learning algorithms and the proposed SWNN approach is compared in Table 5, where the random forest achieves the second-highest performance and the SWNN outperforms all other methods on the dataset with 12 features. The random forest yielded a 23% inaccuracy, while the proposed SWNN approach produced 10% inaccuracy. Within the limited simulated environment, there is a significant 13% improvement in performance. The random forest has the second-best performance over the 32 feature dataset with a reported error rate of 20%. The performance of the proposed SWNN approach has improved by 0.04% for this 32-feature dataset compared to the previous one. This significant improvement in the SWNN method claimed that the performance is attributable to the increased availability of features in the dataset. Now, a similar pattern of performance improvement can be seen with the dataset with 36 features, which produces only 0.02% error for logistic type prediction. The performance of the decision tree is the second-best in the list for the dataset with 36 features. However, the performance of the decision tree is 10% less than the proposed SWNN approach, which makes the SWNN approach the best-suited method for predicting this kind of logistic transportation.

This performance is an enhancement in the proposed algorithm is only possible for the parallel implementation of the neural networks from the neural population and its weight adjustment strategy. The learning of the proposed method is dynamic because the selection of the best neural network is performed in each iteration and then the learning of the best neural network of the population is shared with the other neural network in the population. This performance of the proposed SWNN strategy also reflects its stability in the performance over the different datasets. The performance quality of the proposed method is represented graphically in Fig. 8, Fig. 9 and this shows that the quality of the solution is also outperforming the other strategies of machine algorithm presented here with a limited simulation environment.
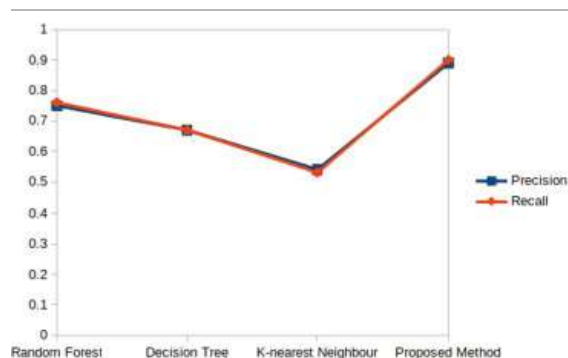
Download : Download high-res image (70KB)

Download : Download full-size image

Fig. 8. Comparative analysis of the proposed method with other standard machine learning algorithm based on precision and recall value.
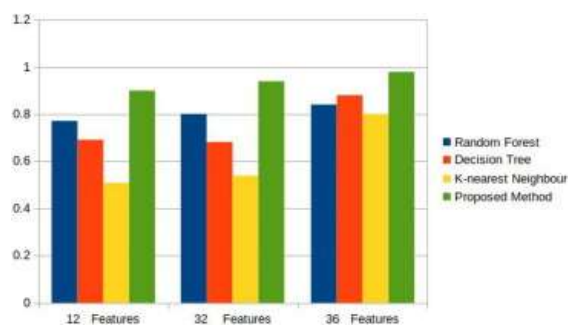
Download : Download high-res image (109KB)

Download : Download full-size image

Fig. 9. Comparative analysis of the proposed method with other standard <u>machine learning algorithm</u> over different numbers of features.

4. Conclusion

This paper has designed a new logistic agent-based model for analyzing public transports at the edge of the networks and making decisions intelligently for public transportation. For analyzing real-time sensory data of public transportation, retrieved from multiple vehicle sensors, we have introduced a new Swarm-Neural Network (SWNN) model. The SWNN model is constructed with a set of parallel neural networks, and it fits within the intelligent logistic transportation framework for data analytics. The proposed model shortens the transit time of every small-scale logistics delivery to its destination. The main goal of the SWNN model is to detect the transportation model efficiently by training the weight and bias matrix of the neural networks. The performance of the proposed SWNN model is evaluated over a standard dataset, namely the TMD dataset that consists of multiple sensory data of public transports such as accelerometer, gyroscope, magnetometer, and audio sensors. During the evaluation, the features of the sensory data are extracted based on a 5-s time interval before analyzing the dataset using the proposed SWNN model. A set of simulation analyses with multiple features have been performed to show the outperformance of the proposed SWNN model over the standard <u>machine learning techniques</u>. The comparative analysis represents that the proposed SWNN model achieves 78–98% accuracy over the standard machine learning algorithms. In the future, we will develop a logistic agent-based intelligent framework at edge networks with advanced communication technology such as 6G networks to reduce network congestion and analyze the vehicular sensory data with higher accuracy.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

Recommended articles

References

[1]   L. Zhu, F.R. Yu, Y. Wang, B. Ning, T. Tang

Big data analytics in intelligent transportation systems: A survey

IEEE Trans. Intell. Transp. Syst., 20 (1) (2019), pp. 383-398

CrossRef ↗   Google Scholar ↗

[2]   M. Won
        Intelligent traffic monitoring systems for vehicle classification: A survey
        IEEE Access, 8 (2020), pp. 73340-73358

        CrossRef ↗   View in Scopus ↗   Google Scholar ↗

[3]   M. Abbasi, A. Najafi, M. Rafiee, M.R. Khosravi, V.G. Menon, G. Muhammad
        Efficient flow processing in 5g-envisioned sdn-based internet of vehicles using gpus
        IEEE Trans. Intell. Transp. Syst., 22 (8) (2021), pp. 5283-5292

        CrossRef ↗   View in Scopus ↗   Google Scholar ↗

[4]   B. Ji, X. Zhang, S. Mumtaz, C. Han, C. Li, H. Wen, D. Wang
        Survey on the internet of vehicles: Network architectures and applications
        IEEE Commun. Stand. Mag., 4 (1) (2020), pp. 34-41

        CrossRef ↗   View in Scopus ↗   Google Scholar ↗

[5]   J.J.Q. Yu, A.Y.S. Lam
        Autonomous vehicle logistic system: Joint routing and charging strategy
        IEEE Trans. Intell. Transp. Syst., 19 (7) (2018), pp. 2175-2187

        CrossRef ↗   View in Scopus ↗   Google Scholar ↗

[6]   M. Humayun, N. Jhanjhi, B. Hamid, G. Ahmed
        Emerging smart logistics and transportation using iot and blockchain
        IEEE Internet Things Mag., 3 (2) (2020), pp. 58-62

        CrossRef ↗   Google Scholar ↗

[7]   Y. Zhang, G. Zhang, R. Fierro, Y. Yang
        Force-driven traffic simulation for a future connected autonomous vehicle-enabled smart
        transportation system
        IEEE Trans. Intell. Transp. Syst., 19 (7) (2018), pp. 2221-2233

        CrossRef ↗   View in Scopus ↗   Google Scholar ↗

[8]   A. Munusamy, M. Adhikari, M.A. Khan, V.G. Menon, S.N. Srirama, L.T. Alex, M.R. Khosravi
        Edge-centric secure service provisioning in iot-enabled maritime transportation systems
        IEEE Trans. Intell. Transp. Syst. (2021), pp. 1-10

        CrossRef ↗   Google Scholar ↗

[9]   X. Zheng, W. Chen, P. Wang, D. Shen, S. Chen, X. Wang, Q. Zhang, L. Yang
        Big data for social transportation
        IEEE Trans. Intell. Transp. Syst., 17 (3) (2016), pp. 620-630

        View in Scopus ↗   Google Scholar ↗

[10]   Y. Liu, Big data technology and its analysis of application in urban intelligent transportation system, in: 2018
        International Conference on Intelligent Transportation, Big Data Smart City (ICITBS), 2018, pp. 17–19.

        Google Scholar ↗

[11]   S. Goudarzi, N. Kama, M.H. Anisi, S. Zeadally, S. Mumtaz
        Data collection using unmanned aerial vehicles for internet of things platforms
        Comput. Electr. Eng., 75 (2019), pp. 1-15

        🗎 View PDF   View article   View in Scopus ↗   Google Scholar ↗

[12]   B.-J. Chang, J.-M. Chiou
        Cloud computing-based analyses to predict vehicle driving shockwave for active safe driving in
        intelligent transportation system

IEEE Trans. Intell. Transp. Syst., 21 (2) (2020), pp. 852-866

CrossRef ↗   View in Scopus ↗   Google Scholar ↗

[13]   M. Shengdong, X. Zhengxian, T. Yixiang

Intelligent traffic control system based on cloud computing and big data mining

IEEE Trans. Industr. Inf., 15 (12) (2019), pp. 6583-6592

CrossRef ↗   View in Scopus ↗   Google Scholar ↗

[14]   Q. Hao, L. Qin

The design of intelligent transportation video processing system in big data environment

IEEE Access, 8 (2020), pp. 13769-13780

CrossRef ↗   View in Scopus ↗   Google Scholar ↗

[15]   M. Asadi, M. Fathy, H. Mahini, A.M. Rahmani

An evolutionary game approach to safety-aware speed recommendation in fog/cloud-based intelligent transportation systems

IEEE Trans. Intell. Transp. Syst. (2021), pp. 1-10

View in Scopus ↗   Google Scholar ↗

[16]   P. Arthurs, L. Gillam, P. Krause, N. Wang, K. Halder, A. Mouzakitis

A taxonomy and survey of edge cloud computing for intelligent transportation systems and connected vehicles

IEEE Trans. Intell. Transp. Syst. (2021), pp. 1-16

CrossRef ↗   Google Scholar ↗

[17]   C. Chen, B. Liu, S. Wan, P. Qiao, Q. Pei

An edge traffic flow detection scheme based on deep learning in an intelligent transportation system

IEEE Trans. Intell. Transp. Syst., 22 (3) (2021), pp. 1840-1852

CrossRef ↗   View in Scopus ↗   Google Scholar ↗

[18]   F. Qiao, J. Wu, J. Li, A.K. Bashir, S. Mumtaz, U. Tariq

Trustworthy edge storage orchestration in intelligent transportation systems using reinforcement learning

IEEE Trans. Intell. Transp. Syst., 22 (7) (2021), pp. 4443-4456

CrossRef ↗   View in Scopus ↗   Google Scholar ↗

[19]   K. Gu, K. Wang, X. Li, W. Jia

Multi-fogs-based traceable privacy-preserving scheme for vehicular identity in internet of vehicles

IEEE Trans. Intell. Transp. Syst. (2021), pp. 1-18

Google Scholar ↗

[20]   L. Breiman

Random forests

Mach. Learn., 45 (1) (2001), pp. 5-32

Google Scholar ↗

[21]   L. Breiman, J.H. Friedman, R.A. Olshen, C.J. Stone

Classification and regression trees

Routledge (2017)

Google Scholar ↗

[22]   N.S. Altman

An introduction to kernel and nearest-neighbor nonparametric regression

Am. Stat., 46 (3) (1992), pp. 175-185

[Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/00031305.1992.10475879

View in Scopus ↗   Google Scholar ↗

Cited by (12)

Transformer for Skeleton-based action recognition: A review of recent advances

2023, Neurocomputing

Show abstract ⌄

Improved Artificial Rabbits Optimization with Ensemble Learning-Based Traffic Flow Monitoring on Intelligent Transportation System ↗

2023, Sustainability (Switzerland)

Optiflow: Optimizing Traffic Flow in its with Improved Cluster Routing ↗

2023, SSRN

Action recognition techniques ↗

2023, Intelligent Multimedia Processing and Computer Vision: Techniques and applications

Hybrid Algorithm-Driven Smart Logistics Optimization in IoT-Based Cyber-Physical Systems ↗

2023, Computers, Materials and Continua

An environmental-based perspective framework: integrating IoT technology into a sustainable automotive supply chain ↗

2023, Benchmarking

> ❯   View all citing articles on Scopus ↗

Peer review under responsibility of Faculty of Engineering, Alexandria University.

ELSEVIER

ℛ RELX™

ScienceDirect

Depress-DCNF: A deep convolutional neuro-fuzzy model for detection of depression episodes using IoMT

Akshi Kumar [a], Saurabh Raj Sangwan [b], Anshika Arora [b] 👤 ✉ , Varun G. Menon [c]

[a] Department of Computing and Mathematics, Faculty of Science and Engineering, Manchester Metropolitan University, Manchester, United Kingdom

[b] Department of Computer Science and Engineering, Netaji Subhas University of Technology, New Delhi, India

[c] Department of Computer Science and Engineering, SCMS School of Engineering and Technology, Kerala, India

🔴 Check for updates

Show less ∧

⟨ₒ Share   💬 Cite

Abstract

Discernible patterns of a person's daily activities can be utilized to detect behavioral symptomatology of mental illness at early stages. Wearable Internet of Medical Things (IoMT) devices with sensors that collect motion data and provide objective measures of physical activity can help to better monitor and detect potential episodes related to the mental health conditions at earlier, more treatable stages. This research puts forward a neuro-symbolic model which uses learnable parameters with integrated knowledge for detection of depression episodes using IoMT based actigraphic input. A novel deep fuzzy model, Depress-DCNF is a hybrid of convolutional neural network (CNN) and an adaptive neuro fuzzy inference system (ANFIS) where CNN is used to extract high-level features from the motor activity recordings which are eventually combined with the discriminative statistical features to produce an optimized feature map. This optimized feature map is finally used to train the ANFIS model which accurately performs the depression classification task. The model is validated on the Depresjon benchmark dataset and compares favorably to state-of-the-art approach giving a superior performance accuracy of 85.10%.

Introduction

An individual's personality is a relatively distinct & consistent pattern of thinking, feeling, and behaving. It entails how an individual affects others and how he/she understands and views himself/herself. This general style of interacting with the world which is primarily shaped by life experiences and genetics is often prone to psychological disorders like depression and anxiety. Mental, neurological and substance use disorders make up to 13% of the total global burden of disease, leading to shortened life expectancy and dramatic quality of life differences. Depression is a mood disorder that causes distressing symptoms and affects how one feels, thinks, and handles daily activities, such as sleeping,

eating, or working. It alone affects 3.8 percent of the world's population, with 5.0 percent of adults suffering from it. It makes an individual cynical and is deeply self-damaging.

Recently, various modeling tools and natural language processing capabilities have been used to analyze the instantaneous behavior, traits, moods and mental disorders such as stress and depression[1], [2]. The Internet of Medical Things (IoMT) is poised to become a transformational force in healthcare. It has emerged as the miracle solution to deal with an overloaded health system offering a myriad of benefits such as remote patient monitoring, improved preventive care, workflow optimization and stock management[3], [4], [5], [6]. Certainly, tangible cues from psycholinguistic markers on social media and the IoMT based multimodal signals (audio, video) combined with the sensor-based psychophysiological and brain signals can help to comprehend the affective states, personality traits, emotional experiences and psychological disorders in individuals.

Wearable IoMT devices record user's movements in changing environments and can facilitate crisis intervention for the depressed, distressed and suicidal individuals[7], [8]. The curated flow of information delivered using the wearable technology in a digestible, easy-to-understand format, can deliver interventions for depression[9], [10]. Actigraph sensor is one such widely used wearable sensor that is used to evaluate the physical activity and sleep of an individual[11]. It is a non-invasive method of monitoring the human rest/activity cycles. The motion patterns are displayed as actograms that show the daily activity and rest periods and this data can be analyzed to provide a variety of objective endpoints about the circadian patterns, the level of activity, and the nighttime movements. Eventually, this actigraphic pattern analysis allows depression screening of individuals and serves as a diagnostic parameter for identifying depression episodes.

Artificial intelligence (AI) and natural language processing (NLP) capabilities have reportedly been used to analyze the instantaneous behavior, traits, moods and mental disorders such as depression, schizophrenia, and dementia[1], [2]. Learning algorithms have become more precise and accurate as these interact with training data, allowing humans to gain unprecedented insights into diagnostics, care processes, treatment variability, and patient outcomes. Deep learning models are the state-of-the-art machine learning models that provide an effective way for various medical image & signal analysis as these methods are able to extract high-level features from sequential data discarding the need of manual feature engineering[12], [13], [14]. However, stand-alone conventional statistical models, neural networks, and machine learning methods often fail to consider each input's fuzzy uncertainty factors[15]. That is, though the neural models have strong non-linearity and self-learning capabilities, but it has issues like requirement of large training data, high training times, overfitting and parameter tuning. Moreover, for AI to advance, it must understand not only the 'what' but also the 'why'. Deep neural models give black-box predictions and are flawed in their lack of model interpretability which creates trust barriers with human users as cause–effect cannot be explained. In entirety, neural models struggle at capturing compositional and causal structure from data. Alternatively, symbolic models are good at capturing the compositional and causal structure but suffer due to its cumbersome rule-based explicit embedding of human knowledge. A viable unison, a neuro-symbolic system that combines the advantages of both neural networks and symbolic AI techniques is thus the need of the hour. Neuro-symbolic AI uses both logic and language processing to accomplish complex tasks. It primarily incorporates common sense reasoning and domain knowledge into deep learning. Adaptive Neuro-Fuzzy Inference System (ANFIS) is one of the popular combinations that has been extensively used. It can solve the nonlinear "input–output" relationship between models and help eliminate the uncertainty in the feature values of the prediction model. At the same time, these models help in creating networks that are interpretable by design.

Motivated by this, we propose a novel neuro-symbolic model that uses the learning ability of deep neural networks and transparency of the fuzzy system. The proposed Depress-DCNF model combines the capabilities of Convolutional Neural Networks (CNN) and ANFIS for detection of depression episodes from an IoMT based motor activity data. The raw activity measurements are resampled and sequences of 48h recordings are used to train the CNN. The high-level features extracted from CNN are then combined with the discriminative statistical features of the activity recordings which are further used to train an ANFIS model. Information gain is used for feature reduction. Thus, the primary contributions of this research are:

- Design a neuro-symbolic classification model for depression detection task based on actigraphic inputs from IoMT for real-time intervention.

- Application of deep learning to achieve generalizability

- Use of ANFIS to better classify the data in extracted feature space than original feature space thereby enabling interpretability under uncertainty

- Evaluation and validation on the benchmark Depresjon dataset for depression screening.

The paper is organized into 5 sections. The following Section 2 discusses the related work conducted in the domain. Section 3 discusses the details of the proposed Depress-DCNF model followed by the description of experimental results obtained in Section 4 and the conclusion of the research in Section 5.

Section snippets

Related work

The prevalence & high risk associated with depression in individuals, early screening efforts and preventive interventions have been studied widely in literature. Recent studies have leveraged the data from wearable devices and soft computing techniques to assess various psychological disorders such as sleep quality[11], [16], depression[17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33], behavioral disorders[16], anxiety[17], [19], and…

Methodology

The proposed Depress-DCNF model integrates deep learning and ANFIS to implement a deep fuzzy classifier for depression detection using motor activity data. The classifier might prove useful as a first-step to identify depression episodes in the general population, to select patients who might need intervention for confirmed clinical diagnosis and treatment of depressive disorder. The proposed Depress-DCNF model is presented in Fig. 1.

It comprises five key steps namely, data preprocessing…

Results

This study proposes a deep convolutional neuro fuzzy model for depression detection where the capabilities of convolutional neural network to learn features are combined with the abilities of fuzzy inference system for classification and handling imprecision. The resampling and framing are done using python. Also, CNN is implemented in python using the Keras library. The information gain for feature reduction is implemented in the Waikato Environment for Knowledge Analysis (WEKA). WEKA has also …

Conclusion

This study proposed a hybrid deep fuzzy model for depression classification using motor activity recordings of 23 unipolar & bipolar depression patients with 32 controls in the "The Depression Dataset: Depresjon". The novel Depress-DCNF combines the capabilities of convolutional neural network to learn high-level features with that of fuzzy logic for improved classification and handling imprecision. The initial layers of CNN are used to extract the high-level features from 48 h sequences of…

CRediT authorship contribution statement

Akshi Kumar: Conceptualization, Methodology, Validation. **Saurabh Raj Sangwan:** Writing – reviewing and editing. **Anshika Arora:** Data curation, Software, Writing – original draft. **Varun G. Menon:** Visualization, Investigation, Supervision.…

Declaration of Competing Interest

Special issue articles   Recommended articles

References (36)

Kumar
Using cognition to resolve duplicacy issues in socially connected healthcare for smart cities
Comput. Commun. (2020)

Kumar *et al.*
Hierarchical deep neural network for mental stress state detection using IoT based biomarkers
Pattern Recognit. Lett. (2021)

Kumar *et al.*
Genetically optimized Fuzzy C-means data clustering of IoMT-based biomarkers for fast affective state recognition in intelligent edge analytics
Appl. Soft Comput. (2021)

JacobsonN.C. *et al.*
Deep learning paired with wearable passive sensing data predicts deterioration in anxiety disorder symptoms across 17–18 years
J. Affect. Disord. (2021)

RaglanG.B. *et al.*
Snoring and depression symptoms in pregnant women
Sleep Health (2021)

GuntukuS.C. *et al.*
Detecting depression and mental illness on social media: an integrative review
Curr. Opin. Behav. Sci. (2017)

KumarM. *et al.*
Fuzzy evaluation of heart rate signals for mental stress assessment
IEEE Trans. Fuzzy Syst. (2007)

KumarM. *et al.*
Stress monitoring based on stochastic fuzzy analysis of heartbeat intervals
IEEE Trans. Fuzzy Syst. (2012)

PhilipV. *et al.*
A review on latest internet of things based healthcare applications
Int. J. Comput. Sci. Inf. Secur. (2017)

VinojP.G. *et al.*
IoT-powered deep learning brain network for assisting quadriplegic people
Comput. Electr. Eng. (2021)

∨　View more references

Cited by (13)

[IIFDD: Intra and inter-modal fusion for depression detection with multi-modal information from Internet of Medical Things](#)

2024, Information Fusion

Show abstract ∨

[Special issue on fuzzy systems for biomedical science in healthcare](#)

2023, Applied Soft Computing

[Medical Image Classification Using Transfer Learning and Chaos Game Optimization on the Internet of Medical Things ↗](#)

2023, arXiv

[Systematic review and meta-analysis of performance of wearable artificial intelligence in detecting and predicting depression ↗](#)

2023, npj Digital Medicine

[Fuzzy Cognitive Maps for Algorithmic Decision Making in eMental Health ↗](#)

2023, Research Square

[Optimizing Smartphone Addiction Questionnaires with Smartphone Application and Soft Computing: An Intelligent Smartphone Usage Behavior Assessment Model ↗](#)

2023, Soft Computing Techniques in Connected Healthcare Systems

> 　View all citing articles on Scopus ↗

View full text

IEEE.org   IEEE *Xplore*   IEEE SA   IEEE Spectrum   More Sites   Subscribe

Subscribe   Cart   Create   Pers
Account   Sign

Browse ⌄   My Settings ⌄   Help ⌄   Institutional Sign In

Institutional Sign In

All   ▼

🔍
ADVANCED SEARCH

Deep Reinforcement Learning for Intelligent Service Provisioning in Software-Defined Industrial Fog Networks

Publisher: IEEE | Cite This | 📄 **PDF**

Indranil Sarkar   ; Mainak Adhikari   ; Sanjay Kumar   ; Varun G. Menon   **All Authors** •••

Alerts

Manage Content Alerts

Add to Citation Alerts

Abstract

Document Sections

I. Introduction

II. System Model and
Problem Formulation

III. Intelligent Service
Provisioning Strategy

IV. Performance Evaluation

V. Conclusion

Authors

Figures

References

Citations

Keywords

Metrics

More Like This

📄
Downl
PDF

Abstract: Fog computing has become a promising technology to improve the performance of low-powered Industrial Internet of Things (IIoT) devices by providing flexible and convenien... **View more**

▶ **Metadata**
Abstract:
Fog computing has become a promising technology to improve the performance of low-powered Industrial Internet of Things (IIoT) devices by providing flexible and convenient computing services at the edge of the network with minimum delay. However, due to the ever-increasing traffic load in the network, the traditional service provisioning strategies can pose high complexity as well as network congestion, resulting in high energy consumption. Owing to this issue, in this article, we propose a deep reinforcement learning (DRL)-based service provisioning strategy in a software-defined industrial fog network to minimize the energy consumption of the network. The service provisioning strategy is performed in the network data plane, whereas the DRL is deployed in the control plane to enhance network efficiency. The service provisioning problem is formulated as the Markov decision process (MDP) and further solved by adopting the concept of a deep Q network (DQN). Further, we propose a task migration policy to ensure the high availability of computing devices while meeting a single point of failure (SPOF) issue. Finally, to show the effectiveness of the proposed method, it is compared with the traditional baseline algorithms over various performance metrics.

Indranil Sarkar
Department of Information Technology, National Institute of Technology Raipur, Raipur, India

PDF

Help

Mainak Adhikari

Department of Computer Science, Indian Institute of Information Technology Lucknow, Lucknow, India

Sanjay Kumar

Department of Information Technology, National Institute of Technology Raipur, Raipur, India

Varun G. Menon

Department of Computer Science and Engineering, SCMS School of Engineering and Technology, Ernakulam, India

☰ Contents

I. Introduction

A variety of delay-sensitive tasks have emerged with the exponential growth of the Internet of Things (IoT) applications, such as smart city, smart transportation, intelligent healthcare, etc. These tasks require significant computing resources for real-time task processing. Most of the IoT devices are embedded with sensors and are involved in applications that require either local data processing or offloading data to a suitable computing device. Conventionally, the tasks are offloaded to the centralized cloud server having adequate resources for task processing and storage [1]. However, the physical distance between the IoT devices and the cloud servers increases network latency which can affect the overall performance of the network and cause network congestion. To overcome such pitfalls, CISCO has proposed a computing paradigm, termed as fog computing that can process the delay-sensitive tasks at the edge of the network with minimum delay [2]. Fog computing enables the geographical distribution of virtualized resources at the network edges, which offers the potentiality of providing features like mobility and geodistribution of IoT services.

Authors ⌃

Indranil Sarkar

Department of Information Technology, National Institute of Technology Raipur, Raipur, India

Mainak Adhikari

Department of Computer Science, Indian Institute of Information Technology Lucknow, Lucknow, India

Sanjay Kumar

Department of Information Technology, National Institute of Technology Raipur, Raipur, India

Varun G. Menon

Department of Computer Science and Engineering, SCMS School of Engineering and Technology, Ernakulam, India

Figures ⌄

References ⌄

Citations ⌄

Keywords ⌄

Metrics ⌄

PDF

Help

More Like This

A Delay and Energy Consumption Efficient Offloading Algorithm in Mobile Edge Computing System
2019 IEEE 11th International Conference on Communication Software and Networks (ICCSN)
Published: 2019

Resource Scheduling for Delay Minimization in Multi-Server Cellular Edge Computing Systems
IEEE Access
Published: 2019

Show More

PDF

Help

IEEE Account

» Change Username/Password

» Update Address

Purchase Details

» Payment Options

» Order History

» View Purchased Documents

Profile Information

» Communications Preferences

» Profession and Education

» Technical Interests

Need Help?

» **US & Canada:** +1 800 678 4333

» **Worldwide:** +1 732 981 0060

» Contact & Support

About IEEE *Xplore* | Contact Us | Help | Accessibility | Terms of Use | Nondiscrimination Policy | Sitemap | Privacy & Opting Out of Cookies

PDF

Help

ScienceDirect®

Full length article

Covert non-orthogonal multiple access communication assisted by multi-antenna jamming

Hongxing Peng [a] ✉, Wenjing He [a] ✉, Yanliang Zhang [a] ✉, Xingwang Li [a] ✉, Yuan Ding [b] ✉, Varun G. Menon [c] ✉, Sandeep Verma [d] ✉

[a] School of Physics and Electronic Information Engineering, Henan Polytechnic University, Jiaozuo 454000, China
[b] School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh EH14 4AS, Scotland, UK
[c] Department of Computer Science and Engineering, SCMS School of Engineering and Technology, Ernakulam, India
[d] Department of Electronics and Communication Engineering, Dr. B. R. Ambedkar National Institute of Technology, Jalandhar 144011, India

Check for updates

Show less ∧

ᗜ Share 🙶 Cite

Get rights and content ↗

Abstract

As the Internet of Things (IoT) becomes increasingly popular, the amount of information transmitted through the IoT network has increased significantly. Therefore, the privacy and security problem of the transmitted information has become a major area of focus. Motivated by this, this paper considers the covert communication based on non-orthogonal multiple access (NOMA), which consists of a transmitter, a legal user, a warden with power detection function and a multi-antenna jammer. To realize the covert communication between the transmitter and the legitimate user, the detection error probability of the warden is firstly derived, and then the optimal detection threshold and the minimum detection error probability (MDEP) are obtained. In addition, with the aim of designing this system, the average MDEP of the warden is calculated, and the closed form solution for the outage probability (OP) of the communication link is obtained. Then, a scheme is proposed to optimize the covertness of this system under the covertness constraint and interruption constraint, through which the maximum covert throughput of the system can be obtained. The simulated numerical results validate the theoretical analysis, and testify that: (*i*) the detection performance of the warden can be reduced by increasing the maximum jamming power of the jammer or reducing the transmitting power of the transmitter; (*ii*) by optimizing the power allocation factor, the maximum covert throughput of the system can be obtained under the premise of satisfying the covertness constraint and interruption condition; (*iii*) the proposed optimization scheme can enhance the covertness performance of this system.

Introduction

With the rapid development of mobile communication technology, data traffic is growing exponentially[1]. In the meantime, wireless communication operators are faced with a huge challenge of managing the limited spectrum resources to meet the ubiquitous connection demand. Therefore, improving the utilization of spectrum resources has become a key research direction in academia and industry[2]. In orthogonal multiple access (OMA) technologies[3], a user can only be allocated a single wireless resource, such as dividing the wireless resources according to frequency or time. Consequently, due to the shortage of spectrum resources, OMA technologies are unable to meet the future communication requirements of high rate, diverse quality of services, low-latency and massive connections[4]. Therefore, non-orthogonal multiple access (NOMA) has been proposed as a promising technology for the fifth generation (5G) wireless network[5], [6], [7], [8], [9], [10]. At present, the mainstream scheme of NOMA is to achieve signal multiplexing in power domain (PD) or code domain (CD)[11], [12].

In NOMA, the transmitters send the superposition signal, and the receivers recover their own signal by decoding the superposition signal using successive interference cancellation (SIC) [13], [14], [15], [16], [17]. Moreover, NOMA technologies allow all service users enjoy the same time, frequency and code resource block through power multiplexing. In addition, NOMA can also allocate different power to users with different channel conditions, which ensures fairness between users. Given this fact, NOMA has been extensively studied in wireless communication. In[18], Huang etal. discussed a cooperative relay network based on dual-hop NOMA, and derived the form of closed solutions of the outage probability (OP) for decode-and-forward and amplify-and-forward protocols. Considering the use of instantaneous channel state information (CSI) in the user ordering, Li etal.[19] proposed a new two-stage relay selection scheme, obtaining the analytical expressions for the OP and the diversity order of the proposed scheme. The authors of[20] investigated the cooperative NOMA system with energy collection function in the multi-cell network, and analyzed its coverage probability, ergodic rate and energy efficiency. From the viewpoint of a unified NOMA framework, the authors in[21] studied the secrecy behavior in different eavesdropping scenarios, while derived the exact and asymptotic expressions of the secrecy OP of CD/PD-NOMA and imperfect/perfect SIC to characterize the secrecy performance. A new co-operative simultaneous wireless information and power transfer NOMA protocol was proposed in[22], where the near user acted as an energy collecting relay to assist the far user. Yin etal.[23] proposed a dynamic user grouping algorithm to classify users, and compared the performance of NOMA-2000 and PD-NOMA with Rayleigh fading channels.

Due to the broadcast nature of wireless transmission, the security problem has been a great challenge, and no exception for the NOMA system. For the purpose of ensuring the integrity of information, traditional security technology adopts encryption to prevent eavesdropping[24]. However, studies in recent years have shown that many traditional encryption methods are no longer reliable with the constant improvement of computation capability[25]. In view of this fact, some research works have investigated such security problems from the aspect of physical layer security (PLS)[26], [27], [28], [29], [30], [31], [32]. However, the whole research of PLS is mainly paid attention to protecting the content of communication. Furthermore, in addition to the security of information, users also aim to hide their communication behaviors from being detected by the monitors in certain communication scenarios (such as military activities, remote health monitoring, etc.). Based on this background, a new security mode has emerged, which is called covert communication[33], [34]. Covert communication not only prevents the content of the communication from being eavesdropped, but also ensures the transmission between the two parties is non-detected by eavesdroppers.

Covert communication is designed to ensure that when wireless signals are transmitted and/or communication behaviors are performed, the signals and/or behaviors have a higher detection error probability at the warden. In recent years, covert communication has aroused extensive research interest and become a cutting-edge technology in the field of secure communication[35], [36], [37], [38], [39], [40], [41], [42]. In[35], Liu etal. proved that wireless communication can be hidden in the interference of noisy wireless networks, and discussed some new results on the active wiretapping effect. The authors of[36] studied the performance of covert communication in multiple relays-assisted IoT systems and proposed two relay selection schemes based on random selection and superior-link selection. Shmuel etal.[37] investigated the situation in which the jammer is configured with multi-antenna, and analyzed the transmission strategy at the jammer that affects the transmission rate in the case of full CSI and partial CSI. In[38], Hu etal. proposed truncated channel inversion power control (CIPC) scheme to realize covert communication, analyzed

the Willie's detection performance limits of the proposed scheme and traditional CIPC scheme, and obtained the effective covert throughputs. Shahzad et al.[39] proposed to adopt the full-duplex (FD) receiver to realize covert communication, and derived the form of closed solution of the average optimal detection probability at Willie. In[40], Xiong et al. studied covert communication on additive white Gaussian noise (AWGN) channels with the help of cognitive jammer, and showed that the cognitive jammer brought covert rate gain over the non-informed jammer. The authors in[41], [42] studied a covert communication system over a packet fading channel where the transceiver is uncertain about the relevant CSI.

The above describes some related researches on covert communication. However, to the best of our knowledge, there are few articles about covert communication combined with NOMA. The only related research works can be found in[43], [44], [45]. In uplink NOMA systems, the authors of[43] proposed the adoption of CIPC in the public communication link to make full use of the channel uncertainty to realize covert communication. In[44], Jiang et al. designed a cooperative covert communication scheme using cellular and non-cellular signals, and used the PD-NOMA to decode the covert message. In downlink NOMA systems, Tao et al.[45] used open legal communication to provide cover for covert communication, derived detection error probability and connect outage probabilities of Willie, and optimized the PAF ratio to maximize the effective covert rate.

Driven by the above discussion, we find that the existing literature often only focuses on the transmission of a single covert message in terms of covert communication. However, with the rapid development of communication technology, multi-task communication scenarios will gradually become the mainstream. Secondly, the combination of covert communication and NOMA can not only meet the covertness of the system and realize covert communication, but also improve the transmission rate of information, so as to further improve the system performance. Regarding to the combination of NOMA and covert communication, signals sent by single antenna are often used as interference signals, but most antennas in real life are multi-antenna. Therefore, we study the influence of jamming signals emitted by multi-antenna jammer on the covertness performance of the NOMA-based system. First, the detection performance of the warden is analyzed, and the optimal detection threshold and the MDEP of the warden are obtained. Then, with the aim of analyzing the covertness performance of the system, an optimization scheme is proposed, which maximizes the covert throughput by optimizing the PAF under the conditions of satisfying the covertness constraint and interruption constraint of the system. The key contributions of this work can be summarized as follows:

- We propose a scheme of covert communication which adopts a covert transmission strategy assisted by the multi-antenna jamming. The jamming adopts transmitting antenna selection (TAS) and random power with uniform distribution, which plays an uncertain role in the detection of the warden, so as to meet the requirements of covertness.

- We calculate the MDEP of the communication link between the transmitter and the covert user, obtain its closed-form expression which reveals the effects of the transmitting power and the jamming power on MDEP, and deduce the range of its optimal detection threshold.

- The covert throughput is transformed into an optimization problem, which can maximize the covert throughput under the conditions of ensuring the covertness constraint and interruption constraint, so as to achieve a balance between the dependability and covertness of the communication system. By solving this problem, the optimal PAF is finally obtained.

The specific arrangements of each part are as follows. In Section 2, the model and the jamming design of covert communication system based on NOMA are introduced. Section 3 first derives the detection error probability at Willie, and then discusses the optimal detection threshold and MDEP in the two cases. In Section 4, a scheme for optimizing the covertness of the system is provided. Section 5 presents the analysis of experimental numerical results, which verifies the correctness of theoretical analysis. The paper is summarized in Section 6.

Notations: $\mathscr{CN}\left(0, \sigma^2\right)$ is represented as a complex Gaussian random variable with a expectation of zero and a variance of σ^2. $\mathbb{E}(\cdot)$ denotes the expectation operation of random variables. Use $|\cdot|$ to represent the absolute value of a scalar. $\mathbf{Pr}(x)$ denotes the probability of a random variable x. In addition, $f_x(\cdot)$ and $F_x(\cdot)$ are the probability density function

(PDF) and cumulative distribution function (CDF) of random variables, respectively. Finally, $\exp(\cdot)$ is the exponential function.

Section snippets

System model

Firstly, the system model and the associated assumptions are introduced in detail. Secondly, with the aim of improving the system covertness, the jamming strategy is designed. Finally, the information transmission is described....

Detection indicators at Willie

In this section, we analysis the detection performance of Willie in detail. Firstly, Willie uses a power detector to perform binary detection on the received power. Secondly, the detection error probability of Willie is calculated. Willie's optimal detection threshold is solved analytically in the end....

The optimization of covertness performance

In this section, the covert communication is firstly designed from the aspect of the covert user Bob. Secondly, the covert throughput is maximized by optimizing the PAF in conditions of covertness constraint and reliability constraint....

Numerical analysis

In this section, the theoretical analysis obtained from the simulation directly verifies the correctness of the conclusions in Section 3 and IV. We analyze and investigate this system performance in different system parameters. For simplicity, the channel parameters are set as $\lambda_{AB} = \lambda_{AW} = \lambda_{CB} = \lambda_{CW} = 1$, the antenna number at Carl is set to $N_t = 3$, and the noise variance of Bob is $\sigma_B^2 = 1$.

Fig.2 shows the variation curves of the average MDEP $\bar{\xi}^*$ and $P_C^{\max}$ for different P_A. It is clear that the average MDEP ...

Conclusion

In this paper, a covert communication assisted by multi-antenna jamming scheme based on NOMA is proposed. Specifically, the multi-antenna jammer Carl transmits a random jamming signal to provide cover for the communication between the transmitter Alice and the legitimate user Bob. Based on the proposed scheme, the optimal detection threshold and its corresponding average MDEP are derived. With the aim of balancing the reliability and covertness of this system, we propose an optimization...

CRediT authorship contribution statement

Hongxing Peng: Conceptualization, Methodology, Supervision. **Wenjing He:** Investigation, Software, Data curation, Writing – original draft, Address comments. **Yanliang Zhang:** Conceptualization, Methodology, Writing – review & editing. **Xingwang Li:** Investigation, Software, Data curation, Writing – original draft, Supervision, Address comments. **Yuan Ding:** Conceptualization, Methodology, Writing – review & editing, Address comments. **Varun G. Menon:** Visualization, Writing – review & editing. **Sandeep**...

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper....

Acknowledgments

Hongxing Peng received his B. Sc. and M. Sc. degrees in industrial automation from and control theory and control engineering from Henan Polytechnic University, China in 1999 and 2003, respectively. He then received Ph. D. degree in detection technology and automatic equipment from Beijing Institute of Technology in 2009. He is currently an Associated Professor with the School of Physics and Electronic Information Engineering. His research interests include wireless communication and...

Recommended articles

References (45)

LiX. *et al.*

I/Q imbalance aware nonlinear wireless-powered relaying of B5G networks: Security and reliability analysis

IEEE Trans. Netw. Sci. Eng. (2020)

TaoY. *et al.*

A survey: Several technologies of non-orthogonal transmission for 5G

China Commun. (2015)

QianL.P. *et al.*

NOMA-enabled mobile edge computing for internet of things via joint communication and computation resource allocations

IEEE Internet Things J. (2020)

DingZ. *et al.*

A survey on non-orthogonal multiple access for 5G networks: Research challenges and future trends

IEEE J. Sel. Areas Commun. (2017)

DingZ. *et al.*

Application of non-orthogonal multiple access in LTE and 5G networks

IEEE Commun. Mag. (2017)

WangY. *et al.*

Analysis of non-orthogonal multiple access for 5G

China Commun. (2016)

LiX. *et al.*

Residual transceiver hardware impairments on cooperative NOMA networks

IEEE Trans. Wirel. Commun. (2020)

LiuY. *et al.*

Multiple-antenna-assisted non-orthogonal multiple access

IEEE Wirel. Commun. (2018)

PengW. *et al.*

A novel perspective on multiple access in 5G network: Framework and solutions

IEEE Wirel. Commun. (2019)

ZhangL. *et al.*
A survey of advanced techniques for spectrum sharing in 5G networks
IEEE Wirel. Commun. (2017)

⌄   View more references

Cited by (8)

[Covert communication for cooperative NOMA with two phases detection](#)
2023, Alexandria Engineering Journal

> *Citation Excerpt :*
>
> …Tao et al. designed a covert communication scheme with the assist of energy harvesting jammers in [28]. With this idea, Peng et al. in [29] proposed a covert communication scheme for multi-antenna jamming, in which the optimal antenna is selected to transmit the jamming information. From the perspective of detection, the authors of [30] made a comprehensive analysis of the covert performance of the system with multiple detectors and jammers.…

Show abstract ⌄

[Preserving Covert Communication in Downlink NOMA Systems via Reradiating on Antennas](#) ↗
2023, IEEE Communications Letters

[AmBC Symbiotic Intelligent Transportation Systems Covertness Performance Analysis and Optimization](#) ↗
2023, TechRxiv

[State EstimationMovingWindow Gradient Iterative Algorithm for Bilinear Systems Using the ContinuousMixed p-norm Technique](#) ↗
2023, CMES - Computer Modeling in Engineering and Sciences

[Covert Communication in Downlink NOMA Systems With Channel Uncertainty](#) ↗
2022, IEEE Sensors Journal

[Filtering-based multi-innovation recursive identification methods for input nonlinear systems with piecewise-linear nonlinearity based on the optimization criterion](#) ↗
2022, Optimal Control Applications and Methods

>   View all citing articles on Scopus ↗

Hongxing Peng received his B. Sc. and M. Sc. degrees in industrial automation from and control theory and control engineering from Henan Polytechnic University, China in 1999 and 2003, respectively. He then received Ph. D. degree in detection technology and automatic equipment from Beijing Institute of Technology in 2009. He is currently an Associated Professor with the School of Physics and Electronic Information Engineering. His research interests include wireless communication and Internet-of-thing (IoT).

Wenjing He received the B.Sc. degree in electronic information engineering from the School of Physics and Electronic Information Engineering, Henan Polytechnic University, Jiaozuo, China, in 2020, where she is currently pursuing the M.Sc. degree in communication and information systems. Her current research interests include non-orthogonal multiple access (NOMA) and covert communication.

Yanliang Zhang received the B.Sc. Degree from Henan University, in 2001, and the Ph.D. degree from the School of Electronic Engineering, Xidian University, in 2011. He is currently an Associate Professor with the School of Physics and Electronic Information Engineering, Henan Polytechnic University, Jiaozuo, China. His research interests include machine vision and affective computing.

Xingwang Li (S'12-M'15-SM'20) received the M. Sc. and Ph. D. degrees from University of Electronic Science and Technology of China and Beijing University of Posts and Telecommunications in 2010 and 2015. From 2010 to 2012, he worked at Comba Telecom Ltd. In Guangzhou China, as an engineer. He spent one year from 2017 to 2018 as a visiting scholar at Queen's University Belfast, Belfast, UK. He is also a visiting scholar at State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications from 2016 to 2018. He is currently an Associated Professor with the School of Physics and Electronic Information Engineering, Henan Polytechnic University, Jiaozuo China. His research interests include MIMO communication, cooperative communication, hardware constrained communication, non-orthogonal multiple access, physical layer security, unmanned aerial vehicles, and the Internet of Things. He has served as many TPC members, such as the IEEE GLOBECOM, IEEE WCNC, IEEE VTC, IEEE ICCC and so on. He has also served as the Co-Chair for the IEEE/IET CSNDSP 2020 of the Green Communications and Networks Track. He also serves as an Editor on the Editorial Board for IEEE ACCESS, COMPUTER COMMUNICATIONS, PHYSICAL COMMUNICATION, KSII TRANSACTIONS ON INTERNET AND INFORMATION SYSTEMS and IET QUANTUM COMMUNICATION. He is also the Lead Guest Editor for the Special Issue on UAV-enabled B5G/6G networks: Emerging Trends and Challenges of PHYSICAL COMMUNICATION, Special Issue on Recent Advances in Physical Layer Technologies for the 5G-Enabled Internet of Things of WIRELESS COMMUNICATIONS AND MOBILE COMPUTING, and Special Issue on Recent Advances in Multiple Access for 5G-enabled IoT of SECURITY AND COMMUNICATION NETWORKS.

Yuan Ding (M'19) received his Bachelor's degree from Beihang University (BUAA), Beijing, China, in 2004, received his Master's degree from Tsinghua University, Beijing, China, in 2007, and received his Ph.D. degree from Queen's University of Belfast, Belfast, UK, in 2014, all in Electronic Engineering. He was a radio frequency (RF) Engineer in Motorola R&D Centre (Beijing, China) from 2007 to 2009, before joining Freescale Semiconductor Inc. (Beijing, China) as an RF Field Application Engineer, responsible for high power base-station amplifier design, from 2009 to 2011. He is now an Assistant Professor at the Institute of Sensors, Signals and Systems (ISSS) in Heriot-Watt University, Edinburgh, UK. His research interests are in antenna array, physical layer security, and 5G related areas. Dr. Ding was the recipient of the IET Best Student Paper Award at LAPC 2013 and the recipient of the Young Scientists Awards in General Assembly and Scientific Symposium (GASS), 2014 XXXIst URSI.

Varun G Menon is currently an Associate Professor in Department of Computer Science and Engineering, SCMS School of Engineering and Technology, India. He is a Senior Member of IEEE and a Distinguished Speaker of ACM Distinguished Speaker. Dr. Varun G Menon is currently a Guest Editor for IEEE Transactions on Industrial Informatics, IEEE Sensors Journal, IEEE Internet of Things Magazine and Journal of Supercomputing. He is an Associate Editor of IET Quantum Communications and also an Editorial Board Member of IEEE Future Directions: Technology Policy and Ethics. His research interests include Internet of Things, Fog Computing and Networking, Underwater Acoustic Sensor Networks, Cyber Psychology, Hijacked Journals, Ad-Hoc Networks, Wireless Sensor Networks.

Sandeep Verma received the Bachelor of Technology degree from the Department of Electronics and Communication Engineering (ECE), Ludhiana College of Engineering and Technology in 2010, the Master of Engineering degree in ECE from NITTTR Chandigarh in 2013, and the Ph.D. degree from the ECE Department, Dr. B. R. Ambedkar National Institute of Technology, Jalandhar, in 2020, where he is currently working as an Assistant Professor. His research interests include energy efficient wireless sensor networks, IoT-based architectures, and intelligent transportation systems. He has published more than 20 papers in reputed journals namely, IEEE TRANSACTION ON INDUSTRIAL INFORMATICS, IEEE INTERNET OF THINGS JOURNAL, Applied Soft Computing, Peer to Peer Networking and Applications, and many others. He has also reviewed more than 40 papers from the referred journals. He is a Professional Member of ACM.

View full text

Expert Systems | WILEY

Cognitive smart cities: Challenges and trending solutions

A smart city implies the realization of sustainable city growth enabled by technology-based intelligent solutions to give a good quality of life to its citizens. Information and communication technologies play a crucial role as the nerve centre of the smart city for collecting and analysing data from various sources, like mobile, social media, and sensors. Internet of things (IoT) and big data (BD) also play a critical role in the smart city infrastructures, changing the way we analyse patterns and trends in human behaviour. Smart cities generate a huge amount of data, and therefore need many flexible ways to implement data and processing gateways.

Recently, cognitive analytics have attracted the attention of researchers and practitioners worldwide as a technology-based smart solution. It is a novel approach to information discovery and decision making which uses multiple intelligent technologies such as statistical machine learning, deep learning, distributed artificial intelligence, natural language processing and visual pattern recognition to understand data and generate insights. A cognitive smart city refers to the convergence of emerging IoT and smart city technologies to realize cyber-physical social systems, their generated BD from sensing to communication and computing, and artificial intelligence techniques for all aspects of collaborative computing in sensors, actuators and human-machine interfaces. A cognitive city is one that learns and adapts its behaviour based on the past experiences and can sense, understand and respond to changes of a smart environment with many human and robotic elements. In cognitive cities, data flows not only from the citizens to city management systems (e.g., intelligent transportation systems and healthcare centres), but also from citizen to citizen. Citizens act as human sensors, and intelligence-enabled frameworks build a cyber-physical social system. Thus, consistent citizen engagement, ubiquitous data collection and sophisticated analytics are required to produce the best kind of cognitive city.

The implementation of cognitive smart city is highly context-dependent. Initiatives may range from incremental to disruptive and the deployment is shaped by many factors such as governance, economic, technology, social, environmental, legal and ethical issues. As smart cities projects become more pervasive across geography, technology and applications, it is imperative to identify key learnings to foster a deeper understanding of the technology evolution landscape and provide tangible benefits to smart city planners and key decision-makers. Viable intersection between technology solutions and digital urbanization design principles (people-centred and inclusive infrastructure, resilience and sustainability, interoperability and flexibility, managing risks and ensuring safety) need to be evaluated in order to provide balanced and replicable solutions.

In the research community, several works propose cognitive solutions that fit the needs of sustainable urban development. Academic literature, government consultation documents and policy papers articulate numerous challenges and research directions for incorporating cognition to realize new smart city services. Both qualitative and quantitative studies carefully consider BD analytics, semantic derivation and knowledge discovery, intelligent decision-making and on-demand service provision for a large number of smart city applications.

This special issue aims to stimulate discussion on the design, use and evaluation of self-correction and human cognition for continuous learning as key knowledge discovery drivers within socially connected urban ecosystems. The issue is focused on articles describing cognitive models for cyber–physical social systems with urban BD to leverage deeper insights from the vast amount of generated data with near real-time intelligence.

We received a very good response to our special issue call for articles. During the review process, each article was assigned to and reviewed by at least three experts in the field. After a rigorous multi-round review process, we were able to accept six excellent articles covering various topics related to cognitive smart cities. In the following, we will introduce those articles and highlight their main contributions.

The paper entitled, 'A taxonomy of energy optimization techniques for smart cities: architecture and future directions' discussed the need for devising more solutions for efficiently handling energy utilization associated challenges in smart cities. The paper presented a comprehensive survey on the energy optimization techniques in various systems, including the optimization techniques in block chain-based systems. Further, the paper presented a taxonomy that classifies energy optimization techniques and proposed an energy-efficient consensus mechanism, proof-of-high-performance optimization for high-performance computing-based ecosystems.

Authors in the paper, 'DFT: A deep feature-based semi-supervised collaborative training for vehicle recognition in smart cities' proposed a deep feature-based training (DFT) method for vehicle recognition in smart cities. DFT is also a semi-supervised collaborative training method on the basis of two base learners. DFT adjusts data pre-processing and training process, optimizes the constructing a disagreement encoding network, and expands on the recognition disagreement of pseudo-labelled samples-based training sets. Compared with the typical collaborative training methods, DFT greatly accelerates the model's training process by reducing the convergence time, and improves the efficiency of vehicle recognition, while remaining the recognition accuracy unchanged.

In the paper, 'Sensor data fusion for the industrial artificial intelligence of things' the authors discussed a new framework for addressing the different challenges of the artificial intelligence of things (AIoT) applications. The proposed framework is an intelligent combination of multi-agent systems, knowledge graphs and deep learning. Deep learning architectures are used to create models from different sensor-based data. Multi-agent systems can be used for simulating the collective behaviours of the smart sensors using IoT settings. The communication among different agents is realized by integrating knowledge graphs. Different optimizers based on constraint satisfaction as well as evolutionary computation are also investigated in the paper.

The paper entitled 'A formal method for privacy-preservation in cognitive smart cities' presented a discussion on a technique for privacy-preservation in smart cities based on pseudonymization, clustering, anonymization and differential privacy methods. The modified clustering algorithm selects the initial cluster-based on the concept of dissimilarity between the data sequences. The paper also assessed the functional correctness and preformation of the proposed model for privacy-preservation in smart cities.

In the paper, 'Soft computing for abuse detection using cyber-physical and social BD in cognitive smart cities' authors discuss the use of soft computing techniques for abuse detection in the complex cyber–physical–social BD systems in cognitive smart cities. The objective of the paper is to define and identify the diverse concept of abuse and systematize techniques for automatic abuse detection for cyber abuse detection on social media and real-time abuse detection using social IoT. The cyber abuse studies on social media platforms have further been categorized as cyber-hate and cyberbullying whereas the real-time abuse includes studies of cyber–physical social systems. As in a cognitive smart city, citizens expect more from their urban environments with minimal intervention, this study helps us to establish the need to capture situational context and awareness in real-time and foster the need to develop a proactive as well as reactive safety mechanism to mitigate the risks of online abuse.

The paper entitled 'A lightweight intelligent intrusion detection system for industrial IoT using deep learning algorithms' discusses the techniques for improving intelligent decision-making actions in the industrial IoT (IIoT) network in a sustainable city. Main cybersecurity attacks are predicted by applying a deep learning model in the paper. The various security and integrity features such as Denial of Service (DoS), malevolent operation, data type probing, spying, scanning, intrusion detection, brute force, web attacks and wrong setup are analysed and detected by a novel sparse evolutionary training mechanism-based prediction model.

We would like to express our sincere thanks to all the authors for submitting their papers and to the reviewers for their valuable comments and suggestions that significantly enhanced the quality of the articles. We are also grateful to Editor-in-Chief, Prof. Jon G. Hall and Special Issues & Reviews Editor Prof. Lucia Rapanotti for their great support throughout the whole review and publication process of this special issue, and, of course, all the editorial staff. We hope that this special issue will serve as a useful reference for researchers, scientists, engineers and academics in the field of cognitive smart cities' design and development.

Varun G. Menon[1]

Reza Khosravi[2]

Alireza Jolfaei[3]

Akshi Kumar[4]

Vinod P[5]

[1]*Department of Computer Science and Engineering, SCMS School of Engineering and Technology, Kochi, Kerala, India*

[2]*IEI Elites Division, Iran*

[3]*School of Computing, Macquaire University, Sydney, Australia*

[4]*Netaji Subhas University of Technology, Delhi, India*

[5]*University of Padova, Padua, Italy*

Correspondence

Varun G. Menon, Department of Computer Science and Engineering, SCMS School of Engineering and Technology, Kochi, Kerala, India.

Email: varunmenon@scmsgroup.org

Artificial Intelligence-Empowered Logistic Traffic Management System Using Empirical Intelligent XGBoost Technique in Vehicular Edge Networks

Monagi H. Alkinani, Abdulwahab Ali Almazroi, Mainak Adhikari, *Member, IEEE*, and Varun G. Menon, *Senior Member, IEEE*

Abstract— Recent advancements in computation and communication technologies and the increasing adoption of the Internet of Things (IoT) and Artificial Intelligence (AI) technologies have paved the way to tremendous developments in modern transportation systems. Driven by the massive number of connected vehicles and the stringent requirements of the public traffic management system, the transportation of data to and from the centralized cloud servers poses a great challenge. As a result, to meet the computational requirements and handle the massive amount of sensory data efficiently, the potential solution is to process/analyze the data at the edge of the network. Motivated by the challenges mentioned above, in this paper, we design a new empirically intelligent XGboost (EIXGB)-enabled logistic transportation system at the edge network for analyzing the data efficiently. Besides that, the proposed EIXGB technique intends to obtain real-time results based on the monitoring parameters of the public traffic management system with higher accuracy and minimum error. Extensive simulation results demonstrate the efficiency of the proposed EIXGB technique over the standard machine learning techniques using a set of parameters. The proposed technique achieves 87-97% accuracy over the different sets of features of a real-time dataset as per the simulation results.

Index Terms— Artificial intelligence, smart traffic management, intelligent transportation system, edge networks, gradient boost technique, accuracy.

I. INTRODUCTION

RECENT developments in technology along with rapid urbanization pose multiple challenges, and the efficient management of the increased traffic on roads is one among them [1], [2]. Various stakeholders ranging from governmental

Monagi H. Alkinani is with the Department of Computer Science and Artificial Intelligence, College of Computer Sciences and Engineering, University of Jeddah, Jeddah 21959, Saudi Arabia (e-mail: malkinani@uj.edu.sa).

Abdulwahab Ali Almazroi is with the Department of Information Technology, College of Computing and Information Technology, University of Jeddah, Jeddah 21959, Saudi Arabia (e-mail: aalmazroi@uj.edu.sa).

Mainak Adhikari is with the Department of Computer Science, Indian Institute of Information Technology, Lucknow 226002, India (e-mail: mainak.ism@gmail.com).

Varun G. Menon is with the SCMS School of Engineering and Technology, Karukutty 683576, India (e-mail: varunmenon@scmsgroup.org).

agencies to automotive manufacturers are seriously trying to overcome this challenge by deploying multiple Intelligent Transportation Systems (ITS). Different versions and varieties of Intelligent Transportation Systems (ITSs) have been deployed over these years, which has reduced traffic-related incidents [3]–[5]. Most of these systems work on collecting and processing a large amount of sensor data. With the recent expedition towards autonomous and smart vehicles, the number of internal and external sensors used with vehicles has increased tremendously.

An ITS is represented as a network of Road Side Equipment (RSE) and Internet of Vehicles, where the vehicles can communicate with each other in a multi-hop manner or access the connected services with or without the support of the infrastructure installed in the RSE [6]. Thus, efficient computation and communication cooperation on the road, including Vehicles-to-Vehicles or Vehicles-to-RSE, is a significant challenge in the transport system. A cooperative transport system aims to improve safety, efficiency, sustainability, and comfort beyond the scope of the stand-alone systems by taking advantage of the communication and cooperation between vehicles and RSE [7]. Further, autonomous systems such as self-driving cars and unmanned systems in logistic transportation activities are progressing quickly. The main challenging tasks in the transportation system are to ensure connectivity to the vehicles at every location due to the long span of the road network and to manage the flow of a large amount of traffic data through a network. Besides that, advanced artificial intelligence techniques have been used widely for the connected transportation system and autonomous vehicles to improve the efficiency and performance of the network with efficient decision-making [8], [9].

The modern ITS systems operate by creating a real-time connection between internal and external sensors, traffic speed sensors, traffic cameras, density sensors. Many ITS have deployed cloud servers to process the data from the sensors for real-time traffic management [10]. Cloud computing helps manage the variation in the speed and volume of data being collected and processed and provides offloading of processing and scaling-up as required. With the recent rise in the number of vehicles and sensors, the latency to cloud data centers has increased, and this is affecting the performance of applications deployed in ITS [11]. Also, in many situations, cloud data

centers find it tough to manage the velocity and variety of data with constraints in bandwidth. Edge computing/networks have been proposed recently as a potential solution to all these issues [12], [13]. As a result, to meet the computational requirements of the transportation system and handle the massive amount of sensory data efficiently, the potential solution in an intelligent transportation system is to transmit the data through advanced communication technology and process the data at the edge of the network with higher accuracy [14], [15].

A. Related Works

To meet the diverse computational requirements and to manage a large amount of sensed data, edge computing has been widely deployed in ITS [16], [17]. Edge computing/edge networks guarantee better performances with reduced delay to various ITS applications [18]. Recently, few edge-based solutions have been proposed to improve public traffic management efficiency. The authors in [5] present a solution for the optimization of the edge node deployment with the traffic data collection in ITS. The proposed work also focuses on keeping edge deployment's overall cost to a minimum. A traffic flow detection scheme for an intelligent transportation system based on deep learning on the edge node is proposed in [19]. The system mainly focuses on a real-time vehicle tracking counter with vehicle detection at the edge to manage the traffic flow.

A two-tier edge computing-based model for traffic monitoring and congestion detection is proposed in [20]. The systems try to provide more resources to the edge for video processing. An edge computing supported ITS system for traffic management in urban areas is proposed in [21]. The system employs a multi-agent reinforcement learning approach to help the edge devices conduct intelligent computation offloading individually. A joint computation offloading and caching solution for edge-supported ITS is proposed in [14] by moving the AI-inspired computing, caching, and communication resources to the proximity of smart vehicles. Authors in [22] give a comprehensive report of the application of edge computing technology in the collaborative optimization of ITS-based-on information and physical fusion. The paper discusses setting up observation points at various traffic junctions and uses long-term and short-term memory networks to collect data at each intersection. In [23], the authors have developed an edge-centric secured service provisioning mechanism using Block chain technology and a standard machine learning technique in an IoT-enabled maritime transportation system. The main objectives of this work are to increase network security while transferring the data to remote computing devices for further processing and analyze the data at the edge of the network with higher accuracy.

Authors in [24] present a discussion of edge computing strategies for Unmanned Aerial Vehicles (UAV). The main focus of the work was to increase the computation capacity of the network at the edge using a joint scheduling optimization model of communication and computation resources. A technique to use the vehicles that are parked for handling the offloaded tasks by edge servers is presented in [25]. The strategy try to make use of a large number of parked vehicles

to improve the efficiency in computations at the edge. Using a task scheduling algorithm some of the computations at the edge are being allocated to the parked vehicles. Authors in [26] discuss the challenges faced while enabling edge computing strategies in autonomous vehicular systems. The paper presents the numerous challenges with autonomous vehicles and edge computing and also discusses various possible solutions. A systematic survey on the recent developments in edge computing environments for intelligent Internet of Vehicles (IoV) is presented in [27]. The paper gives more focus on the major design issues with edge systems in IoV and also presents some use cases. Authors in [28] present a discussion on the various challenges posed in the incorporation of the edge computing strategies in space-air-ground integrated networks. The paper also discusses a possible edge computing framework that can be used in similar environments. Although many of the existing works discuss edge computing strategies in IoV's and vehicular networks, only limited discussions are provided on integrating edge computing along with AI into the public traffic management systems which can tremendously reduce the congestion of the network and the decision-making time. Further, analyzing the transportation data using an advanced AI technology with higher accuracy is a challenging task, and lightweight techniques should be preferred when designing the AI algorithms in vehicular edge networks.

B. Motivations

Driven by the massive number of connected vehicles and the stringent requirements of the data-intensive applications, it is not easy to process and analyze the transportation data at the centralized cloud servers. As a result, it is a critical challenge in the transportation system to process the sensory data at the network's edge. Besides that, another important task is to analyze the sensory data at resource-constraint edge devices using advanced machine learning techniques for making accurate decisions with higher prediction accuracy. Thus, from the analytical and theoretical point of view, the following key research questions need to be addressed while designing a logistic-driven edge-centric ITS for analyzing data on the set of distributed resource-constraint edge devices: *1. how to design an edge-centric intelligent framework for a transportation system for processing sensory data efficiently? 2. how can an advanced machine learning model be integrated with resource-constrained edge devices to analyze the sensory data with higher accuracy?*

The edge computing technology brings the cloud resources such as processing and storage resources at the edge of the network. Thus, integrating edge computing into the public traffic management system can reduce the congestion of the network and the decision-making time. However, due to the resource-constrained nature of the distributed edge devices, analyzing the transportation data using an advanced AI technology with higher accuracy is a challenging task. One of the potential solutions in this aspect is to integrate a lightweight machine learning technique with a set of distributed edge devices to develop an intellectual and logistic framework for

the transportation system that can analyze the data at the edge of the network with higher accuracy and lower precision.

C. Contributions

Motivating by the challenges mentioned above and meeting the requirements of the edge networks, in this paper, we incorporate a new empirically intelligent XGboost (EIXGB) technique that can analyze the sensory transportation data efficiently with higher accuracy while making the accurate decision. In the proposed intelligent edge framework, data collection and preprocessing have been performed on the distributed edge devices while analyzing the modified dataset on the centralized cloud servers using the new EIXGB technique with higher accuracy. Thus, the main novelty of the proposed method is to integrate edge networks in public traffic management systems for making correct decisions with minimum delay and introduce a new EIXGB technique for analyzing traffic management data with higher prediction accuracy. The main contributions of this work are summarized as follows.

- Design an intelligent framework of the EIXGB-enabled transportation system for increasing the compatibility of the logistic-driven public traffic management system. The proposed EIXGB technique assists in gathering data about current traffic conditions and processing them using the empirical XGBoost machine learning algorithm at the edge of the network. As a result, the proposed model can update itself based on current traffic conditions.

- The proposed EIXGB technique intends to obtain real-time results based on the monitoring parameters received from the local IoT devices of the transportation system. The empirically intelligent model contains the XGBoost model, which can automatically fine-tune its accuracy in predicting the transport model. This is feasible because the model training mode is activated automatically if the performance falls below a certain threshold, and it then empirically adjusts its parameters to increase accuracy.

- Extensive simulation results demonstrate the efficiency of the proposed EIXGB technique in the public traffic management system. The proposed technique is compared to existing standard machine learning algorithms such as random forest, decision tree, and K-nearest neighbor. The empirical evaluation over the various parameters is also provided, which aids this approach in predicting virtually precisely.

The remaining sections of the paper are organized as follows. Section II describes the system model followed by the proposed EIXGB technique for analyzing the network's transportation data at the edge. Section III demonstrates the efficiency of the proposed technique over the standard machine learning algorithms using various simulation parameters. Finally, Section IV concludes the work and light on the future direction of the proposed technique for the public traffic management systems.

II. SYSTEM MODEL AND PROPOSED TECHNIQUE

In this section, we elaborate on the proposed empirically intelligent XGboost (EIXGB)-enabled intelligent transporta-
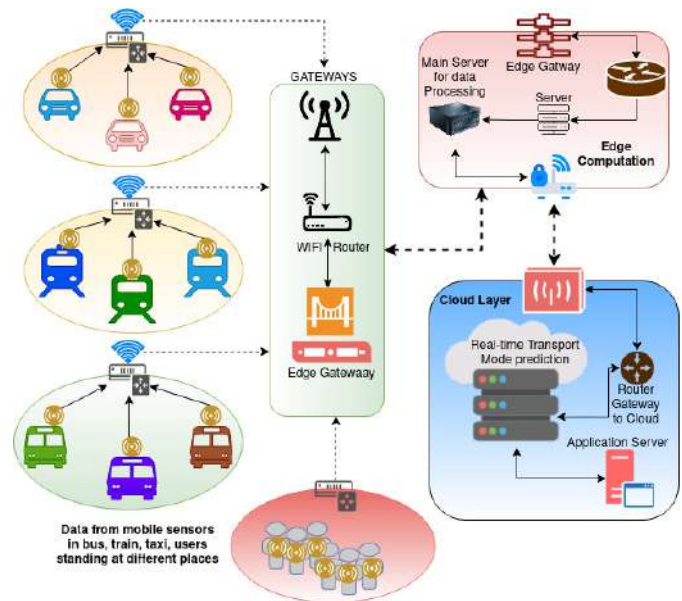
Fig. 1.   The proposed framework of EIXGB-enabled ITS.

tion system followed by the proposed EIXGB technique for the public traffic management system.

A. System Model

This section discusses the EIXGB model for detecting public transportation in an edge network. The overall system model is divided into four layers: the first layer collects data from sensors, the second layer performs feature extraction and selection, the third layer processes the data further for null values and normalization, and the processed data feeds into the EIXGB model. The proposed framework of the EIXGB-enabled intelligent transportation system is shown in Fig, 1.

One of the essential tasks in smart traffic management is traffic mobilization from a less required position to a more required position. In this technique, smart traffic management is accomplished by classifying and identifying public transportation using data from a sensor attached to the vehicle and the passenger's mobile phone. In this method, the sensor data from the passengers' mobile phones are utilized to categorize the public vehicle, and once classified, the information is sent to the distributed edge devices for data preprocessing. Then the information is updated on the centralized cloud server for further analysis. This information is updated with the web application to get it accessed by a passenger waiting at various locations to avail the public transportation. The suggested intelligent traffic management through the EIXGB technique is intended to use on the edge network.

The edge-centric smart traffic management model analyzes sensor data from IoT devices in real-time. The data is processed in the centralized cloud server using the proposed EIXGB technique following a round of processing at the edge server to determine the transport type and segment all comparable types of transport. As a result, the proposed framework is made up of several IoT sensor devices, i.e.,

$S_d = s_1, s_2, S_3, \ldots, s_d$. Here, an IoT sensor device represents the data from different sensors, where these sensors are from the mobile phones of passengers and sensors attached to that vehicle. Thus, $S_d \in (M_p \cup V_m)$, where M_p represents the number of mobile phone sensors and V_m represents the vehicle's sensors. These sensor devices in the smartphone or the vehicle are sending the data to the distributed edge devices for further preprocessing via a local gateway. The S_d number of IoT sensor devices passes the data through the G_n number of gateway, where, $G_n = G_1, G_2, G_3, \ldots, G_n$. These gateway devices upload the task to the edge server through the available wireless channel. The decision of uploading the sensor data, X_s, to the edge server depends on the following rule:

$$X_s = \begin{cases} 1 & length(S_d) \geq \theta_i \\ 0 & otherwise \end{cases} \quad (1)$$

where

$$\theta_i = \prod(M_1, M_2, M_3, \ldots), (V_1, V_2, V_3, \ldots) = \prod(M_p, V_m) \quad (2)$$

The workload, W_l of the edge server depends on its capability of virtual machine(VM), CP_{VM}, and size of the task available in a VM queue. Now as the main focus of this proposed work is to optimize the efficiency of the SMT through maximizing detection of the transport system from the sensor data and reducing the response time of SMT. The responses directly depend on the workload of the edge server. The task of the ith feature extraction, F_i and the detection of transport mode through EIEGB algorithm, AX, is denoted as T_f. This size of the task ST_f is the sum of all the tasks waiting in the queue and is calculated using Eqn.(3).

$$ST_f = \sum_{f=1}^{j}(F_i + Decision(AX(F_i)))_f = \sum_{f=1}^{j} ST_f \quad (3)$$

Hence, the Workload at t^{th} time, W_l is calculated as follows.

$$W_l(t_n) = \frac{ST_f^{t_n}}{CP_{vm}} \quad (4)$$

where,

$$CP_{vm} = \sum_{r=1}^{K} CP_{vm}^r \quad (5)$$

The relation between the workload and the execution of the decision-making is directly proportionate with each other. Faster training and quick decision-making reduce the workload of the system. So, achieving higher accuracy in the minimum time is expected and it is expressed using Eqn.(6).

$$Maximum_{accuracy}(AX(F_n)) \rightarrow Minimum(W_l) \quad (6)$$

B. Proposed EIXGB Technique for Public Traffic Management System

To achieve higher accuracy in the minimum time frame, in this paper, we develop a new EIXGB technique at the edge of the network for analyzing the real-time monitoring parameters of the public traffic management system. The proposed technique is divided into three phases: data collection and processing, feature selection, and EIXGB model. The phases of the proposed EIXGB technique are described in detail in the following sub-sections.

1) Data Collection and Processing: In this phase of smart traffic management, the vehicle and smartphone sensor data are used to precept the data from the gyroscope, magnetometer, accelerometer, audio, etc. The data precept in a specific time interval is then processed based on the fixed time window. Based on this window, features are extracted from each window. If the vehicle and smartphone used the same type of sensor, then based on the location of the data, the sum of those observed data is calculated, and the features are extracted from the received data set.

2) Feature Selection: The selection of the feature depends on the selection of the unique features. This process of feature selection incorporates Correlation-based selection. The correlation between two features X_1 and X_2 is calculated

$$CO(X_1, X_2) = \frac{\sum_{r=1}^{m}(X_1 - \mu_{X_1})(X_2 - \mu_{X_2})}{\sqrt{\sum_{r=1}^{m}(X_1 - \mu_{X_1})} \cdot \sqrt{\sum_{r=1}^{m}(X_2 - \mu_{X_2})}} \quad (7)$$

where, the mean values of X_1 and X_2 signal is performed as $\mu_{X_1} = \sum_{m=1}^{m} X_1$ and $\mu_{X_2} = \sum_{m=1}^{m} X_2$. The output value of $CO(X_1, X_2) \in (+1, -1)$ where the output value close to $+1$ indicates that X_1 and X_2 are similar and then one considered in the pattern and if it is close -1 proves the uniqueness in between X_1 and X_2. Each signal's characteristic is retrieved by examining a five-second window on it. The mean, minimum, maximum, and standard deviation values are derived from the signals of several sensors. The formula for calculating these is given below.

Mean: The means of the five seconds signal data is calculated as $Mean(SI)_i^t = (\frac{1}{n}\sum_{i=0}^{i} SI_i)$, where i^{th} signal value and t represent the window of the current iteration and N is the total number of signal values in the current window.

Minimum Value: The minimum value of each windows of signal is calculated as $Minimum(SI)_t^i = Minimum\{SI_1^t, SI_2^t, SI_3^t, \ldots, SI_i^t\}$.

Maximum Value: The maximum value of each symptom is defined as follows. $Maximum(SI)_t^i = Maximum\{SI_1^t, SI_2^t, SI_3^t, \ldots, SI_i^t\}$.

3) Empirically Intelligent XGboost Model (EIXGB) Model: The working of the EIXGB technique is supported through the exact gradient boosting algorithm. XGboost is one of the versions of this exact boosting algorithm, where continuous features are separated and this split is considered to be one of the problems in tree-based techniques. In a dataset D, there are i^{th} number of features, F_i and each consists of m no of attributes, is represented using Eqn.(8).

$$D = F_i = (A_i, B_i) \quad (8)$$

where, $A \in \mathbb{R}_m, B \in \mathbb{R}$.

In the EIXGB technique, multiple trees are created in the E space and the trees are ensembled where the features are added

to get the predicted output, which is expressed as follows.

$$\hat{B}_i = \sum_{k=1}^{K} G_k(A_i) \tag{9}$$

where G_k is the independent tree in the E space, with H number of leaf and each leaf is having weight W. Hence, $G_k = W_{E(A)}$, and here W_e represent the score of the leaf in an independent tree. The class decision for each tree is applied and it depends on the leaves. So, summing the score in the leaves is the only way to predict the output decision. Now, the loss in the XGboost tree is calculated as $R(\phi)$ which is equivalent to the sum of prediction $y(\hat{B}, B)$ and complexity of the tree δ, is represented as follows.

$$R(\phi) = \sum_i y(\hat{B}_i, B_i) + \sum_k \delta(G_k) \tag{10}$$

where $\delta(G_k)$ is the complexity of the kth independent tree which is with U no of the leaf. So, the complexity is because of the number of leaves in the tree and the weight assigned to it. This is represented using Eqn.(11).

$$\delta(G_k) = C_1 U + \frac{1}{2} C_2 ||W||^2 \tag{11}$$

In each iteration of the algorithm, the main focus is to minimize the loss function, which is represented as follows.

$$Minimizes(R(\phi))_t \tag{12}$$

In order to achieve the above mentioned target of Eq.(11), the tth iteration have $\hat{B}_i^t$ and it is required to add G_t for i^{th} instance, which is expressed in Eqn.(13).

$$R(\phi)^t = \sum_{i=1}^{n} R((\hat{B}_i^{t-1}, B_i) + G_t(A_i)) + \delta G_t \tag{13}$$

Now, performing the R-K methods second order approximation to get the loss function is as follows.

$$R(\phi)^t = \sum_{i=1}^{n} R((\hat{B}_i^{t-1}, B_i) + q_i G_t(A_i) + (\frac{1}{2} s_i G_t^2(A_i)) + \delta G_t \tag{14}$$

where q_i and h_i is the first and second-order gradient for loss function R^t. Eqn. (13) can be further explored and then for a fixed tree structure, leaf j the weight is W_j and is calculated as follows.

$$W_j = -\frac{\sum_{i \in I_j} q_i}{\sum_{i \in I_j} s_i + C_2} \tag{15}$$

Further, the loss function is represented as follows.

$$R(G)^t = -\frac{1}{2} \sum_{j=1}^{H} \frac{(\sum_{i \in I_j} q_i)^2}{\sum_{i \in I_j} s_i + C_2} + C_1 H \tag{16}$$

In the proposed method of the EIXGB algorithm, the loss function, $R(G_k)^t$ of G_k in t^{th} iteration is always checked by the critic, which stops the further exploration if the calculated accuracy, $Acc(G_k)^t$ for current loss function at that iteration

is less than the predefined threshold value, which is expressed using Eqn.(17).

$$R(G_k) \rightarrow Acc(G_k)^t = \begin{cases} 1 & Acc(G_k) \leq \xi \\ 0 & Acc(G_k) > \xi \end{cases} \tag{17}$$

where 1 signifies that the EIXGB process should continue the training process and 0 indicates that the EIXGB is ready for predicting the transport mode. In the case of the train status, the critic indicates for more rounds of the iteration where trees are further split into the number of trees, and loss function is evaluated further. This above mentioned process of EIXGB is discussed through the Algorithm 1.

Algorithm 1 Proposed EIXGB Strategy

INPUT: A_i: No of previously observed patterns of vehicle, B_i: Currently observed patterns from any vehicle.
OUTPUT: $Acc(G_k)^t$: Accuracy at t^{th} iteration, T_i: Transport mode
1: Initially fix $Acc(G_k)^t \rightarrow 0$, and iteration $t = m$
2: Now, $p_m \leftarrow t$
3: **if** $Acc(G_k)^m \leq \xi$ **then**
4:   **for** m:1 to p_m **do**
5:     Read the A_i patterns of the vehicle
6:     Calculate W_j of leaf j using Eqn.(15)
7:     calculate the loss function $R(G)^m$ using Eqn.(16)
8:   **end for**
9:   Now, $R(G_k) \rightarrow Acc(G_k)$
10: **end if**
11: **if** $Acc(G_k)^m > \xi$ **then**
12:   Do the classification using trained model G_k^m
13:   Forecast the $T_i : G_k^m \rightarrow T_i$
14: **end if**

In the edge server, the proposed Algorithm 1 is permanently active. The algorithm analyzed the data, which predicted the current logistic model in the intelligent transportation framework. However, if the forecast over the current timeline exceeds the threshold level, the algorithms preserve the data in their knowledge as an experience.

III. EXPERIMENTAL RESULT ANALYSIS

In this section, the experimental analysis of the proposed method is presented over the transport mode dataset[1] [29], [30]. EIXGB is mainly based on the XGBoost strategy, and there are several parameters in this strategy. In order to get the best performance of the proposed method, fine-tuning of these parameters is performed empirically. The readiness of the proposed method is decided by Eqn.17. Now, when the system is ready, the method's parameters are fixed, and accuracy in the detection of the transport mode is maximized. The methods are evaluated based on the solution's quality and stability.

The proposed method is developed on the Intel i5 10th generation system, with 8GB RAM and 6GB graphics memory, and it is also with 1TB SSD storage. The proposed method is

---

[1]https://raw.githubusercontent.com/fschwartzer/Udacity-Machine-Learning-Nanodegree/master/Capstone/dataset_5secondWindow%5B1%5D.csv

design and developed in a python programming language with NumPy, sklearn, scipy library and the results are plotted using the Matplotlib library of python. The method is compared with random forest, decision tree, and k-nearest neighbour algorithm. The dataset and feature selection process through Eqn.7 is presented in the dataset and feature selection section. The empirical evaluation and performance analysis strategy is explained in the following sub-section.

The transport mode dataset or TMD dataset is divided based on the number of features. In the experimental phase, the dataset is divided into three different datasets. The first dataset consists of 12 features and then the second one consists of 32 features, and the third dataset consists of 36 features. The dataset is divided into these three categories for analyzing the performance of the proposed method and other machine algorithms over different dimensions of the transport dataset. The performances of the proposed method and the standard machine algorithm are analyzed based on classification error, precision, recall, and accuracy. The performance of the proposed method is evaluated for solution quality, which is determined by calculating accuracy, which is formulated using Eqn.(18).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (18)$$

Other methods of performance measurement are precision and recall. The precision is the ratio between the true positive and the summation of false and true positive, which is defined as follows.

$$Precision = \frac{TP}{TP + FP} \quad (19)$$

The recall is the ratio of true positive over the sum of true positive and false negative, which is calculated using Eqn.(20).

$$Recall = \frac{TP}{TP + FN} \quad (20)$$

where TP = true positive, TN = True negative, FP = False Positive, and FN = False Negative. In the performance measurement, the high value of precision and recall reflects the highly accurate result.

The correlation value of the characteristics in the dataset is used to analyze them, and the correlation matrix is constructed based on Eqn. 7. Such matrix is shown in Fig. 2. The following rule is used during the correlation-based feature selection process.

If $CO(X_1, X_2) > 0.5$ *Then*
Remove X_2

Where X_1 and X_2 are the two different features. This rule states that the strongly correlated for the correlation matrix are eliminated. After eliminating the strongly linked features, the correlation matrix is re-plotted and shown in Fig. 3.

A. Empirical Evaluation of Parameters

The empirical evaluation of the parameter is carried out by determining the significant parameters of the proposed technique and then testing the algorithm with a range of alternative values for these parameters. The empirical test considers three parameters: η, γ, and maximum depth. These settings are
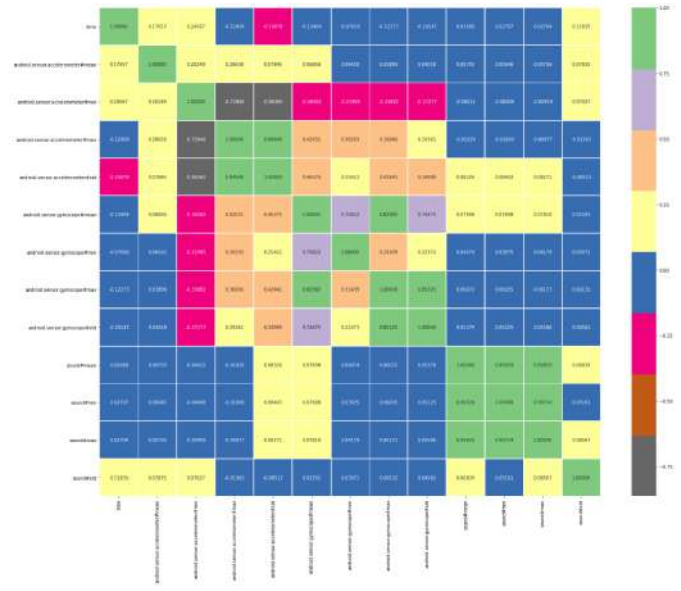
Fig. 2. The correlation map for dataset with 12 features.
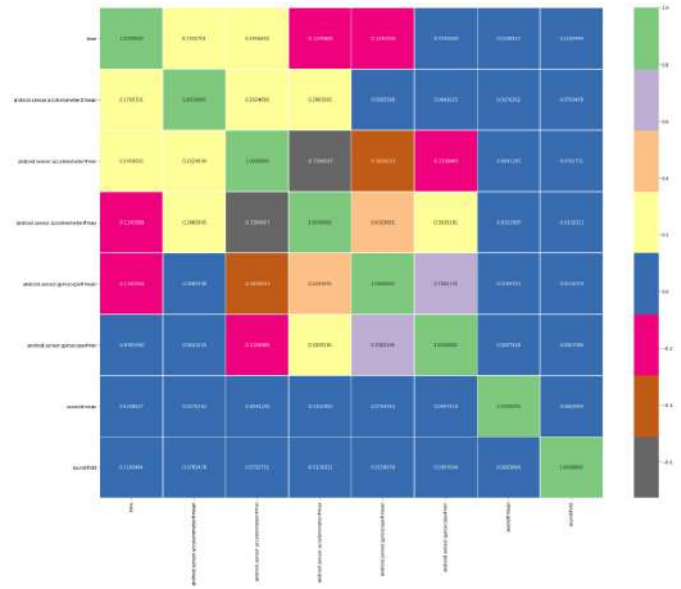
Fig. 3. The correlation map after removing the highly co-related features for dataset with 12 features.

evaluated across three distinct datasets, and the classification error is recorded for each iteration. In the empirical test, the parameter values for which the proposed technique provides the smallest classification error are fixed for that parameter.

Now, before settling on values for each parameter, a wide range of values is evaluated with the suggested approach, and the classification error is recorded each time. A restricted range of values for each parameter is chosen in order to demonstrate the empirical test of the proposed approach over different parameters. As a result, the eta parameter values are 0.1, 0.2, 0.25, 0.3, and 0.45. These values are 0.05, 0.15, 0.25, and 0.45 for the gamma parameter. The empirical performance of the suggested technique is evaluated across 5, 10, 15, 30, and 40 values of the maximum depth parameter. The empirical

TABLE I
OBSERVATION OF PROPOSED METHOD PERFORMANCE FOR DIFFERENT VALUES OF ETA PARAMETER

	0.1	0.2	0.25	0.3	0.45
12 Features	0.139	0.131	0.12	0.133	0.137
32 Features	0.132	0.123	0.11	0.143	0.148
36 Features	0.038	0.0423	0.036	0.038	0.048

TABLE II
OBSERVATION OF PROPOSED METHOD PERFORMANCE FOR DIFFERENT VALUES OF THE GAMMA PARAMETER

	0.05	0.15	0.25	0.45
12 Features	0.127	0.122	0.122	0.135
32 Features	0.133	0.144	0.13	0.144
36 Features	0.036	0.036	0.03	0.141

TABLE III
OBSERVATION OF PROPOSED METHOD PERFORMANCE FOR DIFFERENT VALUES OF THE MAXIMUM DEPTH PARAMETER

	5	10	15	30	40
12 Features	0.144	0.137	0.118	0.133	0.135
32 Features	0.344	0.341	0.243	0.346	0.383
36 Features	0.444	0.038	0.036	0.04	0.04

TABLE IV
FINAL RECOMMENDED VALUES OF DIFFERENT PARAMETERS OF THE PROPOSED METHOD

Parameters	Fixed values
η	0.25
γ	0.25
Maximum Depth	15

Fig. 4. Empirical adjustment of the η parameter of the proposed method for dataset with 12 feature.

Fig. 5. Empirical adjustment of the γ parameter of the proposed method for dataset with 12 feature.

test is conducted over 100 independent runs of the proposed method. Due to the limitation of the space, the testing figure over 12 and 36 features datasets are presented. The Fig. 4, Fig. 5, and Fig. 6 are represented by the test over 12 features dataset. Fig. 7, Fig. 8 and Fig. 9 is presented by the evaluation of the different parameter over 36 feature dataset.

It is discovered that when the η and gamma parameters are set to 0.25, the classification error is exceptionally low, Table I and Table II. Because these observations were made over three distinct dimension datasets, the 0.25 may be fixed and suggested as a result of the experiment. When the maximum depth parameter was set to 15, the proposed approach worked well in the experiment, Table III. The suggested approach achieves high accuracy over three different dimension datasets when the maximum depth is set to 15, and so this value is recommended from the experiment. From the empirical test, the final recommended values for each parameter in the proposed method are listed in Table IV.

B. Performance Analysis

The performance of the proposed method is measured in terms of solution quality and stability. The suggested EIXGB's solution quality is assessed by analyzing the method's precision, recall, and accuracy. The higher the value of these matrices can provide better answer for the provided dataset.
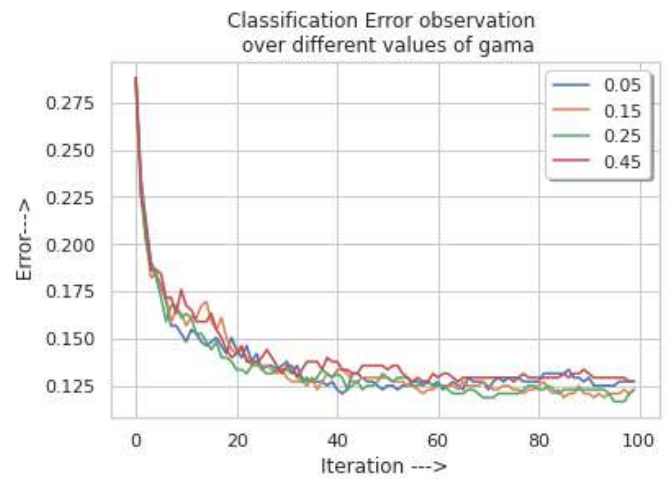
The proposed method is also tested with the same dataset but with varied dimensions, and the results are reported in each case. As a result, the proposed method is evaluated for solution stability, which is favorable if the proposed method performs well in each situation. The dataset is separated into three categories, each of which is constructed by altering the dataset's features. The first dataset contains 12 features, followed by another dataset containing 32 features, and finally, the dataset including 36 features. The proposed method is compared to the standard machine learning methods such as Decision tree [31], Random Forest [32], K-Nearest Neighbor [33], and Ensemble Learning [34] algorithms such as to show its outperforms. Over these datasets, standard algorithms were used to try and fix its parameter. The neighbor value 10 is used in the K-Nearest Neighbor algorithm, and it improves the performance. The depth of the tree is a Decision Tree parameter, which is set to 6 for this dataset. The Decision Tree's performance is assessed at various depths, with the final depth being set at 6. The depth of the Random Forest is also set
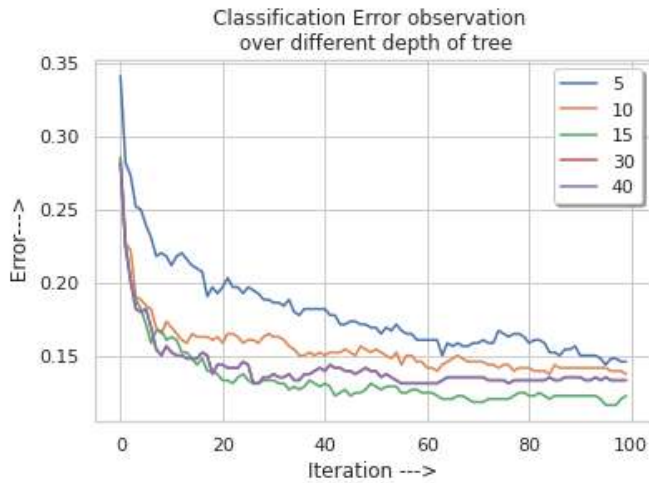
Fig. 6. Empirical adjustment over different maximum depth of tree for dataset with 12 feature.
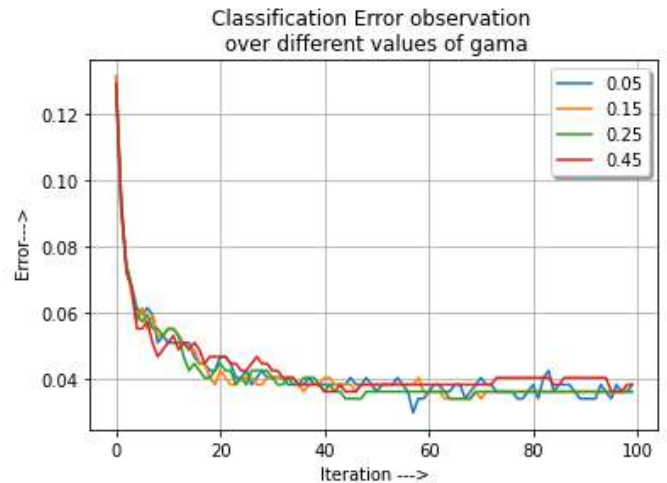
Fig. 8. Empirical adjustment of the γ parameter of the proposed method for dataset with 36 feature.
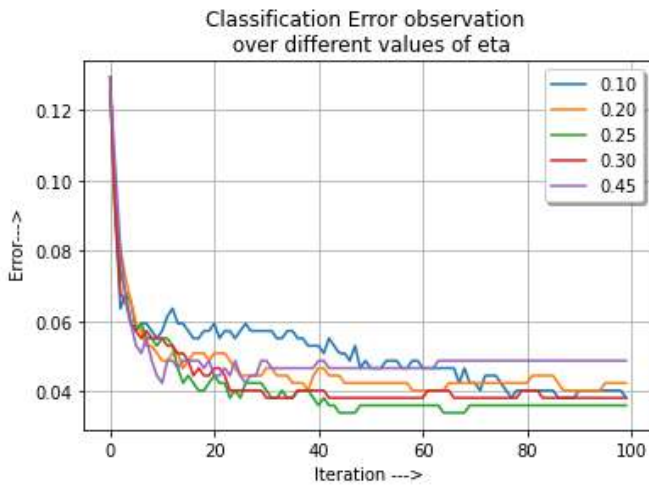
Fig. 7. Empirical adjustment of the η parameter of the proposed method for dataset with 36 features.
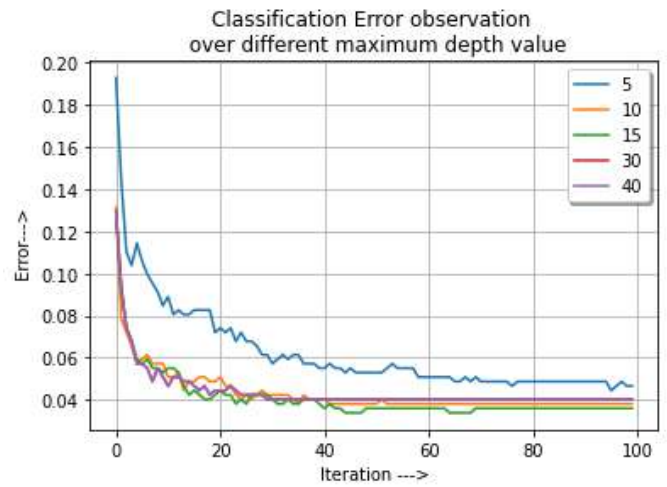
Fig. 9. Empirical adjustment over different maximum depth of tree for dataset with 36 feature.

to 5, and it is determined through empirical experimentation. The three datasets are divided into training and testing halves, with training accounting for 70% of the total and testing accounting for 30%.

The results in Table V show that the performance of the standard techniques including the proposed method is not optimal over the dataset with 12 features. The algorithm's performance suffers due to the fewer unique features to represent numerous classes of the benchmark dataset. The K-Nearest Neighbor algorithm's minimal performance results provide 46% prediction error. The proposed method has the best performance on this 12-feature dataset with a prediction error of 13%, which is 33% better than the K-NN. It's also been discovered that the proposed method outperforms the ensemble learning algorithm by 6%. Similarly, the proposed method outperforms Ensemble learning by 7% and 5% over 32 and 36 feature datasets, respectively. The performance of both algorithms is highly comparable. Table VIII and Fig. 10 shows

TABLE V

COMPARATIVE ANALYSIS OF THE PROPOSED METHOD WITH STANDARD ALGORITHM OVER 12 FEATURE DATASET

	Precision	Recall	Accuracy
Random Forest	0.71	0.71	0.7
Decision Tree	0.67	0.68	0.67
K-nearest Neighbour	0.54	0.54	0.54
Ensemble Learning	0.79	0.80	0.81
Proposed Method	0.88	0.87	0.87

a comparison of the random forest, decision tree, and k-nearest neighbor algorithms. The precision, recall, and accuracy values of several standard algorithms spanning different dimension dataset are represented in Tables V, Table VI, and Table VII.

The proposed algorithm yields a 13% inaccuracy across a dataset with 12 characteristics. It generates a 12% inaccuracy when evaluated over a 32-feature dataset; however, it provides better accuracy when tested over a 36-feature dataset. As a result, the suggested method's performance improves as more

TABLE VI
COMPARATIVE ANALYSIS OF THE PROPOSED METHOD WITH STANDARD
ALGORITHM OVER 32 FEATURE DATASET

	Precision	Recall	Accuracy
Random Forest	0.69	0.7	0.71
Decision Tree	0.65	0.66	0.67
K-nearest Neighbour	0.54	0.53	0.54
Ensemble Learning	0.81	0.81	0.81
Proposed Method	0.86	0.87	0.88

TABLE VII
COMPARATIVE ANALYSIS OF THE PROPOSED METHOD WITH STANDARD
ALGORITHM OVER 36 FEATURE DATASET

	Precision	Recall	Accuracy
Random Forest	0.82	0.82	0.82
Decision Tree	0.85	0.84	0.84
K-nearest Neighbour	0.8	0.8	0.8
Ensemble Learning	0.91	0.91	0.91
Proposed Method	0.97	0.96	0.96
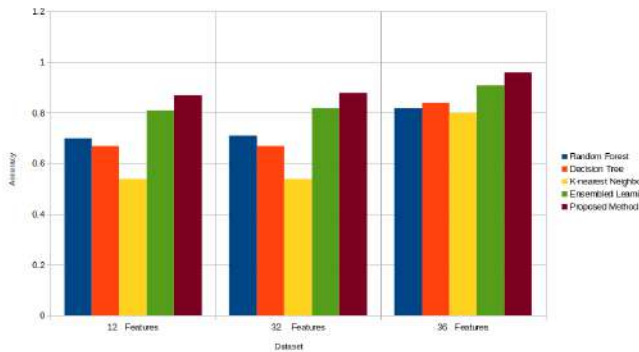
Fig. 10. Comparative analysis of the proposed method with other standard machine learning algorithm over different numbers of features.

TABLE VIII
COMPARATIVE ANALYSIS OF THE PROPOSED METHOD OVER DIFFERENT
MACHINE LEARNING ALGORITHMS

	12 Features	32 Features	36 Features
Random Forest	0.7	0.71	0.82
Decision Tree	0.67	0.67	0.84
K-nearest Neighbour	0.54	0.54	0.8
Ensemble Learning	0.81	0.82	0.91
Proposed Method	0.87	0.88	0.96

features are added, and in this example, it yields 0.04% error with 36 features. In the restricted simulation environment, this demonstrates the superiority of the proposed technique over the comparative machine learning methodology.

IV. CONCLUSION

In recent times, advanced artificial intelligence techniques have been employed at edge networks to make the public transportation system intelligent by supporting efficient decision-making and intelligent traffic control. Motivated by that, in this paper, we have designed a logistic-driven public traffic management system using the EIXGB technique at the edge networks. The main objective of this work is to analyze the real-time monitoring parameters in the resource constraint

edge devices with higher accuracy. The proposed EIXGB technique assists in gathering data about current traffic conditions and processing them using the empirical XGBoost machine learning technique. The empirically intelligent model contains the XGBoost model, which can automatically fine-tune its accuracy by predicting the transport model. To show the efficiency of the proposed technique, we have analyzed the proposed EIXGB model over a real-time public traffic management dataset. Extensive simulation results demonstrate the efficiency of the proposed technique over the standard machine learning algorithms such as random forest, decision tree, and k-nearest neighbor using various performance metrics. The simulation results prove that the proposed EIXGB technique achieves 87-97% accuracy over the real-time traffic management dataset. In the future, we will design an efficient computation and communication framework with collaborative edge and cloud servers for intelligent transportation systems at edge networks. Besides that, another future objective will be edge intelligence and computing with advanced communication technology for intelligent transportation systems with higher prediction accuracy, minimum latency, and energy consumption.

REFERENCES

[1] S. Guo *et al.*, "ROD-revenue: Seeking strategies analysis and revenue prediction in ride-on-demand service using multi-source urban data," *IEEE Trans. Mobile Comput.*, vol. 19, no. 9, pp. 2202–2220, Sep. 2020.

[2] C. Meng, X. Yi, L. Su, J. Gao, and Y. Zheng, "City-wide traffic volume inference with loop detector data and taxi trajectories," in *Proc. 25th ACM SIGSPATIAL Int. Conf. Adv. Geograph. Inf. Syst. (SIGSPATIAL)*. New York, NY, USA: Association for Computing Machinery, 2017, pp. 1–10, doi: 10.1145/3139958.3139984.

[3] Z. MacHardy, A. Khan, K. Obana, and S. Iwashina, "V2X access technologies: Regulation, research, and remaining challenges," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 1858–1877, 3rd Quart., 2018.

[4] M. Abbasi, H. Rezaei, V. G. Menon, L. Qi, and M. R. Khosravi, "Enhancing the performance of flow classification in SDN-based intelligent vehicular networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 7, pp. 4141–4150, Jul. 2021.

[5] C. Xiang *et al.*, "Edge computing-empowered large-scale traffic data recovery leveraging low-rank theory," *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 4, pp. 2205–2218, Oct. 2020.

[6] M. Adhikari, A. Munusamy, A. Hazra, V. G. Menon, V. Anavangot, and D. Puthal, "Security and privacy in edge-centric intelligent Internet of Vehicles: Issues and remedies," *IEEE Consum. Electron. Mag.*, early access, Sep. 30, 2021, doi: 10.1109/MCE.2021.3116415.

[7] B. R. Kiran *et al.*, "Deep reinforcement learning for autonomous driving: A survey," *IEEE Trans. Intell. Transp. Syst.*, early access, Feb. 9, 2021, doi: 10.1109/TITS.2021.3054625.

[8] F. Tang, B. Mao, N. Kato, and G. Gui, "Comprehensive survey on machine learning in vehicular network: Technology, applications and challenges," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 3, pp. 2027–2057, 3rd Quart., 2021.

[9] L. Zhu, F. R. Yu, Y. Wang, B. Ning, and T. Tang, "Big data analytics in intelligent transportation systems: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 1, pp. 383–398, Jan. 2019.

[10] K. Wang and Z. Shen, "Artificial societies and GPU-based cloud computing for intelligent transportation management," *IEEE Intell. Syst.*, vol. 26, no. 4, pp. 22–28, Jul. 2011.

[11] P. Arthurs, L. Gillam, P. Krause, N. Wang, K. Halder, and A. Mouzakitis, "A taxonomy and survey of edge cloud computing for intelligent transportation systems and connected vehicles," *IEEE Trans. Intell. Transp. Syst.*, early access, Jun. 7, 2021, doi: 10.1109/TITS.2021.3084396.

[12] A. Munusamy *et al.*, "Edge-centric secure service provisioning in IoT-enabled maritime transportation systems," *IEEE Trans. Intell. Transp. Syst.*, early access, Aug. 12, 2021, doi: 10.1109/TITS.2021.3102957.

[13] M. Adhikari, A. Munusamy, N. Kumar, and S. N. Srirama, "Cybertwin-driven resource provisioning for IoE applications at 6G-enabled edge networks," *IEEE Trans. Ind. Informat.*, early access, Jul. 14, 2021, doi: 10.1109/TII.2021.3096672.

[14] Z. Ning *et al.*, "Intelligent edge computing in Internet of Vehicles: A joint computation offloading and caching solution," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 4, pp. 2212–2225, Apr. 2021.

[15] F. Zhu, Y. Lv, Y. Chen, X. Wang, and F. Wang, "Parallel transportation systems: Toward IoT-enabled smart urban traffic control and management," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 10, pp. 4063–4071, Oct. 2020.

[16] X. Xu *et al.*, "Service offloading with deep Q-network for digital twinning-empowered Internet of Vehicles in edge computing," *IEEE Trans. Ind. Informat.*, vol. 18, no. 2, pp. 1414–1423, Feb. 2022.

[17] H. Fatemidokht, M. K. Rafsanjani, B. B. Gupta, and C.-H. Hsu, "Efficient and secure routing protocol based on artificial intelligence algorithms with UAV-assisted for vehicular ad hoc networks in intelligent transportation systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 7, pp. 4757–4769, Jul. 2021.

[18] Z. Zhou, A. Gaurav, B. B. Gupta, M. D. Lytras, and I. Razzak, "A fine-grained access control and security approach for intelligent vehicular transport in 6G communication system," *IEEE Trans. Intell. Transp. Syst.*, early access, Sep. 2, 2021, doi: 10.1109/TITS.2021.3106825.

[19] C. Chen, B. Liu, S. Wan, P. Qiao, and Q. Pei, "An edge traffic flow detection scheme based on deep learning in an intelligent transportation system," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 3, pp. 1840–1852, Mar. 2020.

[20] G. Liu *et al.*, "Smart traffic monitoring system using computer vision and edge computing," *IEEE Trans. Intell. Transp. Syst.*, early access, Sep. 22, 2021, doi: 10.1109/TITS.2021.3109481.

[21] S. Pan, P. Li, C. Yi, D. Zeng, Y.-C. Liang, and G. Hu, "Edge intelligence empowered urban traffic monitoring: A network tomography perspective," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 4, pp. 2198–2211, Apr. 2021.

[22] G. Yan and Q. Qin, "The application of edge computing technology in the collaborative optimization of intelligent transportation system based on information physical fusion," *IEEE Access*, vol. 8, pp. 153264–153272, 2020.

[23] A. Munusamy *et al.*, "Edge-centric secure service provisioning in IoT-enabled maritime transportation systems," *IEEE Trans. Intell. Transp. Syst.*, early access, Aug. 12, 2021, doi: 10.1109/TITS.2021.3102957.

[24] Y. Liu *et al.*, "Joint communication and computation resource scheduling of a UAV-assisted mobile edge computing system for platooning vehicles," *IEEE Trans. Intell. Transp. Syst.*, early access, Jun. 4, 2021, doi: 10.1109/TITS.2021.3082539.

[25] C. Ma, J. Zhu, M. Liu, H. Zhao, N. Liu, and X. Zou, "Parking edge computing: Parked-vehicle-assisted task offloading for urban VANETs," *IEEE Internet Things J.*, vol. 8, no. 11, pp. 9344–9358, Jun. 2021.

[26] S. Liu, L. Liu, J. Tang, B. Yu, and W. Shi, "Edge computing for autonomous driving: Opportunities and challenges," *Proc. IEEE*, vol. 107, no. 8, pp. 1697–1716, Jun. 2019.

[27] J. Zhang and K. B. Letaief, "Mobile edge intelligence and computing for the Internet of Vehicles," *Proc. IEEE*, vol. 108, no. 2, pp. 246–261, Feb. 2020.

[28] S. Yu, X. Gong, Q. Shi, X. Wang, and X. Chen, "EC-SAGINs: Edge computing-enhanced space-air-ground integrated networks for Internet of Vehicles," *IEEE Internet Things J.*, early access, Jan. 19, 2021, doi: 10.1109/JIOT.2021.3052542.

[29] Y. Vaizman, K. Ellis, and G. Lanckriet, "Recognizing detailed human context in the wild from smartphones and smartwatches," *IEEE Pervasive Comput.*, vol. 16, no. 4, pp. 62–74, Oct./Dec. 2017.

[30] Y. Vaizman, K. Ellis, G. Lanckriet, and N. Weibel, "Extrasensory app: Data collection in-the-wild with rich user interface to self-report behavior," in *Proc. CHI Conf. Hum. Factors Comput. Syst. (CHI)*. New York, NY, USA: Association for Computing Machinery, 2018, pp. 1–12, doi: 10.1145/3173574.3174128.

[31] T. Qiu, M. Zhang, X. Liu, J. Liu, C. Chen, and W. Zhao, "A directed edge weight prediction model using decision tree ensembles in industrial Internet of Things," *IEEE Trans. Ind. Informat.*, vol. 17, no. 3, pp. 2160–2168, Mar. 2021.

[32] P. Liu, Y. Zhang, H. Wu, and T. Fu, "Optimization of edge-PLC-based fault diagnosis with random forest in industrial Internet of Things," *IEEE Internet Things J.*, vol. 7, no. 10, pp. 9664–9674, Oct. 2020.

[33] D. Adidrana and N. Surantha, "Hydroponic nutrient control system based on Internet of Things and K-nearest neighbors," in *Proc. Int. Conf. Comput., Control, Informat. Appl. (ICINA)*, Oct. 2019, pp. 166–171.

[34] J. P. Barbin, S. Yousefi, and B. Masoumi, "Efficient service recommendation using ensemble learning in the Internet of Things (IoT)," *J. Ambient Intell. Humanized Comput.*, vol. 11, no. 3, pp. 1339–1350, Mar. 2020.

All   ⌄

🔍

ADVANCED SEARCH

ADFL: A Poisoning Attack Defense Framework for Horizontal Federated Learning

Publisher: IEEE   Cite This   📄 **PDF**

Jingjing Guo   ; Haiyang Li   ; Feiran Huang   ; Zhiquan Liu   ; Yanguo Peng   ; Xinghua Li   ; Jianfeng Ma   ; Varun G. Menon…   **All Authors** •••

Abstract

Document Sections

I. Introduction

II. Preliminaries

III. Proposed Method

IV. Experiment and
Discussion

V. Related Works

Show Full Outline ⌄

Authors

Figures

References

Citations

Keywords

Metrics

More Like This

Footnotes

Abstract: Recently, federated learning has received widespread attention, which will promote the implementation of artificial intelligence technology in various fields. Privacy-pre... **View more**

▶ **Metadata**

Abstract:

Recently, federated learning has received widespread attention, which will promote the implementation of artificial intelligence technology in various fields. Privacy-preserving technologies are applied to users' local models to protect users' privacy. Such operations make the server not see the true model parameters of each user, which opens wider door for a malicious user to upload malicious parameters and make the training result converge to an ineffective model. To solve this problem, in this article, we propose a poisoning attack defense framework for horizontal federated learning systems called ADFL. Specifically, we design a proof generation method for users to generate proofs to verify whether it is malicious or not. An aggregation rule is also proposed to make sure the global model has a high accuracy. Several verification experiments were conducted and the results show that our method can detect malicious user effectively and ensure the global model has a high accuracy.

Jingjing Guo
Xidian University, Xian, China

PDF

Jingjing Guo received the B.S., M.Sc., and Ph.D. degrees in computer science and technology from the School of Computer Science and Technology, Xidian University, Xian, China, in 2009, 2012, and 2015, respectively.

She is currently an Associate Professor with the School of Cyber Engineering, Xidian University. Her research interests include trust management, wireless networks security, and artificial intelligence security.

Jingjing Guo received the B.S., M.Sc., and Ph.D. degrees in computer science and technology from the School of Computer Science and Technology, Xidian University, Xian, China, in 2009, 2012, and 2015, respectively.

She is currently an Associate Professor with the School of Cyber Engineering, Xidian University. Her research interests include trust management, wireless networks security, and artificial intelligence security. **View more**

Haiyang Li
Xidian University, Xian, China

Haiyang Li received the B.S. degree in cyber engineering in 2020 from the School of Cyber Engineering, Xidian University, Xi'an, China, where he is currently working toward the master's degree.

His research interests include security of wireless network and artificial intelligence system.

Haiyang Li received the B.S. degree in cyber engineering in 2020 from the School of Cyber Engineering, Xidian University, Xi'an, China, where he is currently working toward the master's degree.

His research interests include security of wireless network and artificial intelligence system. **View more**

Feiran Huang
Jinan University, Guangzhou, China

Feiran Huang (Member, IEEE) received the B.S. degree in physics from the School of Physics and Electronics, Central South University, Changsha, China, in 2011, and the Ph.D. degree in computer science and technology from the School of Computer Science and Engineering, Beihang University, Beijing, China, in 2019.

He is currently an Associate Professor with the College of Cyber Security, Jinan University, Guangzhou, China. H... Show More

Feiran Huang (Member, IEEE) received the B.S. degree in physics from the School of Physics and Electronics, Central South University, Changsha, China, in 2011, and the Ph.D. degree in computer science and technology from the School of Computer Science and Engineering, Beihang University, Beijing, China, in 2019.

He is currently an Associate Professor with the College of Cyber Security, Jinan University, Guangzhou, China. H... **View more**

Zhiquan Liu
Jinan University, Guangzhou, China

Zhiquan Liu received the B.S. degree in mathematics from the School of Science, Xidian University, Xi'an, China, in 2012, and the Ph.D. degree in computer science and technology from the School of Computer Science and Technology, Xidian University, in 2017.

He is currently an Associate Professor with the College of Cyber Security, Jinan University, Guangzhou, China. His research interests include trust management and priva... Show More

Zhiquan Liu received the B.S. degree in mathematics from the School of Science, Xidian University, Xi'an, China, in 2012, and the Ph.D. degree in computer science and technology from the School of Computer Science and Technology, Xidian University, in 2017.

He is currently an Associate Professor with the College of Cyber Security, Jinan University, Guangzhou, China. His research interests include trust management and priva... **View more**

Yanguo Peng
Xidian University, Xian, China

Yanguo Peng (Member, IEEE) received the B.Sc. degree in network engineering from North University of China, Taiyuan, China, in 2009, the M.S. degree in computer software and theory from Guizhou University, Guiyang, China, in 2012, and the Ph.D. degree in computer systems organization from Xidian University, Xi'an, China, in 2016.

He is currently an Associate Professor with the School of Computer Science and Technology, Xid... Show More

Yanguo Peng (Member, IEEE) received the B.Sc. degree in network engineering from North University of China, Taiyuan, China, in 2009, the M.S. degree in computer software and theory from Guizhou University, Guiyang, China, in 2012, and the Ph.D. degree in computer systems organization from Xidian University, Xi'an, China, in 2016.

Xinghua Li
Xidian University, Xian, China

Xinghua Li received the M.E. and Ph.D. degrees in computer science from Xidian University, Xi'an, China, in 2004 and 2007, respectively.

He is currently a Full Professor and Ph.D. Supervisor with Xidian University. His research interests include wireless networks security, privacy protection, cloud computing, software defined network, and security protocol formal methodology.

Xinghua Li received the M.E. and Ph.D. degrees in computer science from Xidian University, Xi'an, China, in 2004 and 2007, respectively.

He is currently a Full Professor and Ph.D. Supervisor with Xidian University. His research interests include wireless networks security, privacy protection, cloud computing, software defined network, and security protocol formal methodology. **View more**

Jianfeng Ma
Xidian University, Xian, China

Jianfeng Ma (Member, IEEE) received the M.E. and Ph.D. degrees in computer software and communications engineering from Xidian University, Xi'an, China, in 1988 and 1995, respectively.

He is currently a Full Professor with Xidian University. His research interests include information security, coding theory, and cryptography.

Jianfeng Ma (Member, IEEE) received the M.E. and Ph.D. degrees in computer software and communications engineering from Xidian University, Xi'an, China, in 1988 and 1995, respectively.

He is currently a Full Professor with Xidian University. His research interests include information security, coding theory, and cryptography. **View more**

Varun G. Menon
SCMS School of Engineering and Technology, Ernakulam, India

Varun G. Menon (Senior Member, IEEE) received the Ph.D. degree in computer science and engineering from Sathyabama University, Chennai, India, in 2017.

He is currently an Associate Professor with the Department of Computer Science and Engineering, SCMS School of Engineering and Technology, Kerala, India. He has published more than 50 research papers in peer reviewed and highly indexed International Journals and Conferences... Show More

Varun G. Menon (Senior Member, IEEE) received the Ph.D. degree in computer science and engineering from Sathyabama University, Chennai, India, in 2017.

He is currently an Associate Professor with the Department of Computer Science and Engineering, SCMS School of Engineering and Technology, Kerala, India. He has published more than 50 research papers in peer reviewed and highly indexed International Journals and Conferences... **View more**

Konstantin Kostromitin Igorevich
South Ural State University National Research University, Celabinsk, Russian Federation

Konstantin Kostromitin Igorevich received the B.S., M.S., and Ph.D. degrees in physics from Chelyabinsk State University, Chelyabinsk, Russia, in 2008, 2010, and 2013, respectively.

He is currently a Researcher with the Department of Information Security, South Ural State University, Chelyabinsk. His research interests include physics and Internet of Things security.

Konstantin Kostromitin Igorevich received the B.S., M.S., and Ph.D. degrees in physics from Chelyabinsk State University, Chelyabinsk, Russia, in 2008, 2010, and 2013, respectively.

He is currently a Researcher with the Department of Information Security, South Ural State University, Chelyabinsk. His research interests include physics and Internet of Things security. **View more**

Contents

SECTION I.
Introduction

It is anticipated that Big Data-driven artificial intelligence (AI) will soon be applied to many aspects of our daily life, including medical care, agriculture, transportation systems, etc. [1]– [4]. When integrating AI into a variety of Internet of Things applications, distributed machine learning is preferred for many data-processing tasks. Federated learning is a recent advance in distributed machine learning by distributing the training work to distributed users [5]– [7]. In this learning paradigm, each user trains a local model based on its local data. Then, users' model updates (uploaded parameters) of each iteration will be uploaded to an aggregation server to update a global model according to a certain aggregation rule. During the whole learning process, none of the users' raw data and training process is exposed to others, including the aggregation server, which makes the procedure of global model aggregating and updating more vulnerable to poisoning attacks.

Currently, the most widely used aggregation rule (e.g., AvgFed) takes the weighted average of each user's local model parameters as the global model [8]. In this case, the global model can be arbitrarily manipulated even if just one user's device is compromised. A lot of attacks have been proved to be effective on these models [9]– [12]. In this work, we focus on the users performing a model poisoning attack, in which the adversary (Byzantine) users can send arbitrary model updates (gradients) to the aggregation server during the training process. The aim of them is to ensure the training algorithm converges to ineffective models [12].

In order to resist the attack of malicious users, several aggregation rules were proposed recently to robust the global model [12]– [15]. There are generally two kinds of defence strategies based on several observation results from the training process. For the first defence strategy, the similarity of each user's uploaded parameters should be measured first using a proper method (such as the Euclidean distance, Jaccard Index, and so on). Users with sufficient similar uploaded parameters will be considered to be put in the same group. With the assumption that the majority of the users are benign users, the local model submitted by users at the largest group will be used to update the global model, so that to avoid the negative impact caused by the local model submitted by Byzantine users. Here, benign users are considered to be users that upload real local model parameters obtained by training their local uncontaminated data. The second strategy is that the aggregation server will test the received parameters by its own testing set. If the test result shows the accuracy of the tested model is lower than a predefined threshold, these parameters will not be used to update the global model, and users who send these parameters will also be regarded as malicious users.

We can see that the prerequisite of using the two strategies is to make sure the aggregation server is able to observe all users' uploaded parameters at each iteration. It has already been proved that it is possible to learn privacy information from users' uploaded parameters (gradients) [16]. In order to preserve users' privacy, several schemes have been proposed to prevent the aggregation server from observing the true uploaded parameters [17], [18]. These schemes make the identification of the malicious uploaded parameters impossible to implement by the two strategies mentioned above.

Aiming at the situation that available works cannot identify malicious uploaded parameters in the privacy-preserving federated learning system, in this article, we propose a novel poisoning attack defense scheme for the privacy-preserving federated learning system. In particular, secure multiparty computation (MPC)-based privacy-preserving federated learning scheme proposed by Keith Bonawitz *et al.* [17] is the federated learning system we focus on, in which the uploaded parameters are masked in a particular way and the actual uploaded parameters will not be revealed to the aggregation server. With the proposed method, the aggregation server is able to detect whether a parameter vector uploaded by a user is malicious. In the model aggregation process, malicious uploaded parameters will not be used to update the global model, so as to guarantee the reliability of the global model.

In nutshell, the contributions of this article are threefold and are given as follows.

1) We propose a malicious user detection method for privacy-preserving federated learning system. With the proposed method, aggregation server can identify the malicious parameters uploaded to it even if the server can only observe users' masked parameters.

2) We propose an aggregation rule for the aggregation server to eliminate the influence of the malicious users on the global model.

3) We implement the prototype of the proposed method and conduct several experiments. The results show that our method can detect malicious users effectively and ensure the global model has a high accuracy.

The rest of this article is organized as follows. Section II shows the preliminaries of this article. Section III introduces the proposed method in detail. In Section IV, we explain the experiments we performed to verify the efficiency of our approach, followed by the related works of this article in Section V. Finally, Section VI concludes this article.

SECTION II.
Preliminaries

In this section, we discuss the preliminaries and assumptions needed for our scheme.

A. Horizontal Federated Learning and Problem Formalization

The horizontal federated learning system consists of K users and an aggregation server (AS). Each user owns local data $\mathcal{D}_i$ and keeps $\mathcal{D}_i$ private, i.e., $\mathcal{D}_i$ is not shared with other users and server AS. We have $|\mathcal{D}_i| = l_i$ and $\sum_i l_i = l$. The server is attempting to train a machine learning model with global parameter vector $\mathbf{w}_G \in \mathbb{R}^n$, where n is the dimensionality of the parameter space.

The learning process is shown in Fig. 1. At each iteration t, a random subset of m users is chosen for synchronous aggregation. Each user u_i ($i \in [1, m]$) minimizes the empirical loss over its local data $\mathcal{D}_i$ by running a machine learning algorithm for E epochs with a batch size of B. At the end of this iteration, each user obtains a local model $\mathbf{w}_i^{t+1}$ and computes a parameter vector $\boldsymbol{\delta}_i^{t+1}$ (it can be the local model, the gradients obtained during the local training, and so on), which will be sent back to the server. To obtain the updated global model for the next iteration, a proper aggregation rule should be used to get $\mathbf{w}_G^{t+1}$. The mostly used aggregation rule now is the averaging-based rule proposed in [8]: $\mathbf{w}_G^{t+1} = \mathbf{w}_G^t + \sum_{i \in [1,m]} \alpha_i \delta_i^{t+1}$, where $\alpha_i = l_i / l$ and $\sum \alpha_i = 1$.
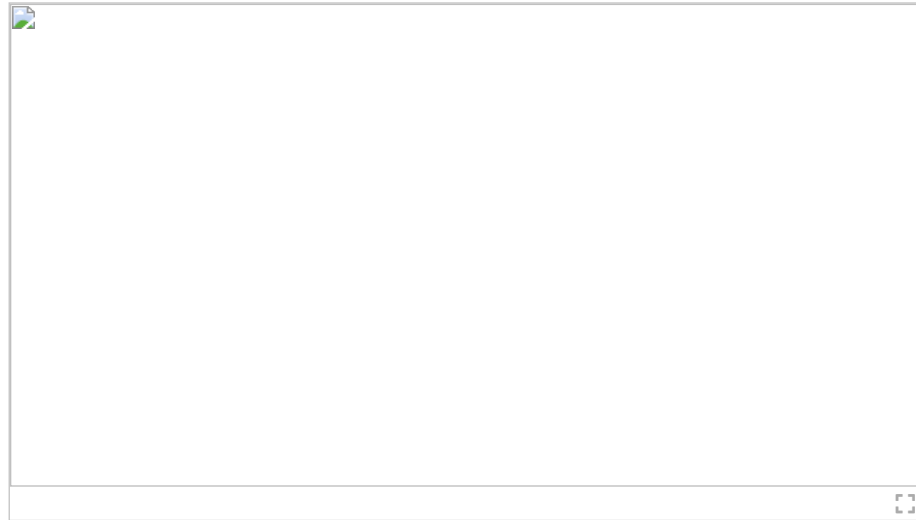
Fig. 1.
Learning process of horizontal federated learning system.

Most aggregation rules assume that all δ_i^t are the real training result of user u_i using $\mathcal{D}_i$, while in order to obtain the global model or simply to attack the system, adversaries having no local data may generate the uploaded parameter vector arbitrarily. In this case, the server can measure the distance among all users to detect the abnormal uploaded parameter vectors. While, in order to preserve the privacy of each user, the uploaded parameter vector of each user is usually protected encrypted by a proper privacy-preserving technology, such as the secure multiparty computing, which will be introduced in the following section. So, the problem our work focuses on is to detect the encrypted abnormal uploaded parameter vectors, so that to eliminate the negative impact of them on the global model, i.e., $\mathrm{DE}(E(\{\delta^h\}), E(\{\delta^a\})) = \{u^a\}$, where DE is the proposed detection method, $\{\delta^h\}$ and $\{\delta^a\}$ mean the set of uploaded parameter vectors of benign users and malicious users, respectively, $\{u^a\}$ is the set of malicious users, and $E()$ is the privacy protection technology.

B. Cryptographic Primitives and MPC Operation

In the following, we introduce some cryptographic primitives and core operations related to the privacy-preserving federated learning based on MPC.

1) Key Generation

Key generation consists of two algorithms (**KA.param, KA.gen**).

- **KA.param(k)$\rightarrow pp$**: Algorithm **KA.param** produces some public parameters using a secret parameter k. Specifically, the parameters $pp = (\mathbb{G},' g, q, H)$, where $\mathbb{G}'$ is a group of prime order q, along with a generator g, and H is a hash function.

- **KA.gen**$(pp) \rightarrow (s_u^{SK}, s_u^{PK})$: Algorithm **KA.gen** allows any party u to generate a private–public key pair. u should first choose a random $x \leftarrow \mathbb{Z}_q$ as its private key s_u^{SK}, and then calculate g^x as the public key s_u^{PK}.

2) Key Agreement

Key agreement is completed by an algorithm **KA.agree**, with which two parties are able to negotiate their shared secret key.

- **KA.agree**$(s_u^{SK}, s_v^{PK}) \rightarrow s_{u,v}$: User u and v can obtain their shared secret key by calculating $H((g^{(s_v^{SK})})^{(s_u^{(SK)})}) = H((s_v^{PK})^{(s_u^{SK})})$, which can be implemented by the Diffie–Hellman key agreement protocol. It is obvious that **KA.agree** (s_u^{SK}, s_v^{PK})=**KA.agree** (s_v^{SK}, s_u^{PK}).

3. Mask Code Generation

Before user u's local model x_u uploading to the server, it should be masked in a particular way. For a given local model, its parameters can be arranged in a certain order to form the parameter vector of the model. So, here, we use the parameter vector to describe u's local model x_u. Assume that all users of the federated learning system are in order, and each pair of users (u and v) agree on some random vector $r_{u,v}$. If u uploads $x_u + r_{u,v}$ and v uploads $x_v - r_{u,v}$ to the server, then the mask code $r_{u,v}$ will be canceled when their masked models are added, so that the server can also calculate the global model with the true local models of u and v not being revealed. So, the mask code generation can be described as (1)– (3).

$$p_{u,v} = \begin{cases} \Delta_{u,v} \times \mathrm{PRG}(s_{u,v}), & u \neq v \\ 0, & u = v \end{cases} \tag{1}$$

$$\Delta_{u,v} = \begin{cases} 1, & u > v \\ -1, & u < v \end{cases} \tag{2}$$

$$p_u = \mathrm{PRG}(b_u). \tag{3}$$

View Source ⊘

In (1), $p_{u,v}$ is a vector whose elements are the output of a secure pseudorandom generator, PRG, with the shared secret key $s_{u,v}$ as seed. From (2), we can see that $p_{u,v} + p_{v,u} = 0$. For p_u in (3), it is the individual mask code for user u, which is also the output of the PRG with a random seed b_u. In addition, the length of $p_{u,v}$ and p_u should be the same as that of the user's local model (parameter vector).

The masked model of user u (y_u) is shown in (4), where x_u is u's true local model, and p_u and $p_{u,v}$ are the mask code of u described above.

$$y_u = x_u + p_u + \sum_{v \neq u} p_{u,v}. \tag{4}$$

View Source ⊘

4. Secure Model Aggregation

When the aggregation server received masked models from users, it will conduct the aggregation rule to get the global model of this iteration by (5), where V is the set of users who drop the system before uploading their local model, and U is the set of users uploading the masked model to the server.

$$z = \sum_{u \in U} y_u - \sum_{u \in U} p_u + \sum_{v \in V, u \in U} p_{v,u}. \tag{5}$$

View Source ⓘ

For p_u and $p_{v,u}$ ($u \in U$, $v \in V$), the server will construct b_u and s_u^{SK} to generate them using the reconstruction algorithm **SS.recon** of Shamir's t-out-of-n secret sharing scheme [19]. For details about the secret sharing process, refer to [17].

C. System Architecture

As shown in Fig. 2, our system model consists of three kind of entities, namely $Trust\quad Authority$ (TA), $User$, and $Aggregation\quad Server$ (AS), respectively.

TA: TA is mainly in charge of the initialization of the system, including the public parameters generation and public and private key assignment.

$User$: A user is a device that trains a local model using the data owned by itself, and then sends the masked parameter vector to AS during each iteration. The local model can be any kind of machine learning model, such as convolutional neural network, logistic regression, and so on. For a given user, its uploaded parameters can be arranged in a certain order to form the uploaded parameter vector of the user, so that we can use the parameter vector to describe its uploaded parameters. In the rest of this article, the local model and the uploaded parameter vector mean the same data. Users in the system can be divided into two categories based on their behaviors. A user that uploads the real parameter vector by training its own data to AS is called benign user ($u1$, $u2$, ..., ui in Fig. 2). If a user has no local data and sends a made-up parameter vector arbitrarily to AS which aims to break the accuracy of the global model, it will be called malicious user (user m in Fig. 2). In our scheme, in order to prevent the parameter vector sent by malicious user from influencing the accuracy of the global model, each user should generate a $Proof$ corresponding to AS's requirement.

AS: The aggregation server aggregates the masked parameter vectors uploaded by users using a certain rule and sends the results (global model) to each user. In addition, it also takes charge of the detection of malicious parameter vector sent by malicious user.
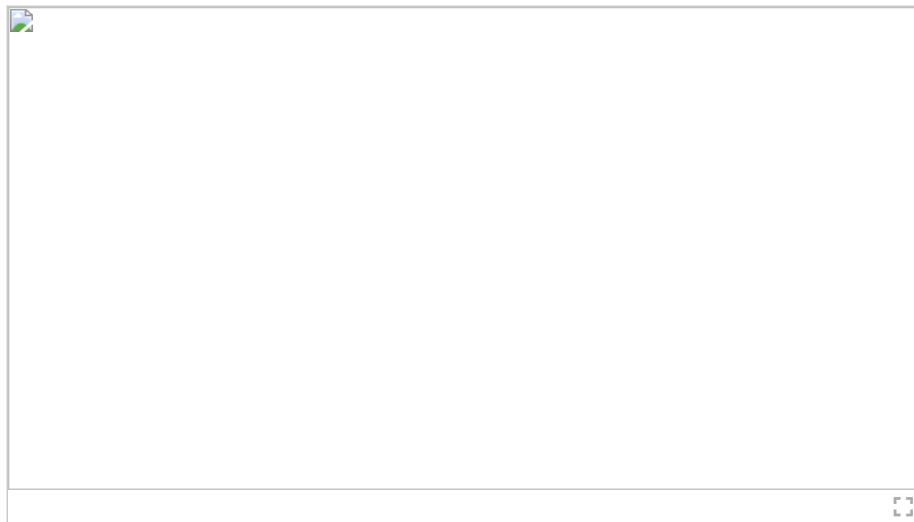
Fig. 2.
System architecture of this work.

D. Adversarial model

For TA and AS, we assume they are trustworthy and will not collude with any entity. In addition, we assume all users will not collude with each other and there are at least two benign users in the system. The local data of all users in the system, except malicious users, are independent and

identically distributed (IID). The malicious users in our work have no local training data and aim to attack the system deliberately, so it will generate the uploaded parameter vector arbitrarily (sometimes maybe well-designed) and add the mask code on it. Then, the masked false parameter vector will be sent to AS. For the problem we focus on, the situation with some users dropout is the same as that with no user dropout. So, we do not consider the dropout problem of users during the training process for simplicity.

SECTION III.
Proposed Method

In this section, we will introduce the proposed method in detail. The method is run between a TA, an AS, and a set of users $\{u_1, u_2, \ldots, u_n\}$. We assume the identity of all users is ordered through the subscript. Fig. 3 shows the process of the proposed method, which consists of four steps. In brief, the main task of each step is shown as follows.

Step 1 Key generation. In this step, all the parameters and keys which will be used in the following steps are generated.

Step 2 Masked model generation and uploading. In this step, each user will train the local model based on its local data and generate its mask code and uploaded parameter vector. If the user is malicious, it will fabricate its uploaded parameter vector. Then, the fabricated uploaded parameter vector will be masked with the mask code and uploaded to AS.

Step 3 Malicious user detection. A *garble_agg_model* will be generated by AS. Each user will generate proof based on this *garble_agg_model*, which will be sent to AS to verify whether the user is malicious.

Step 4 Global model updating. In the fourth step, AS will aggregate the global model using the received benign models in previous step. Finally, users will update their local model based on the received global model. The system will back to step 2 to begin the next iteration until the model accuracy reaching the predefined threshold or the iteration number exceeding the predefined number.
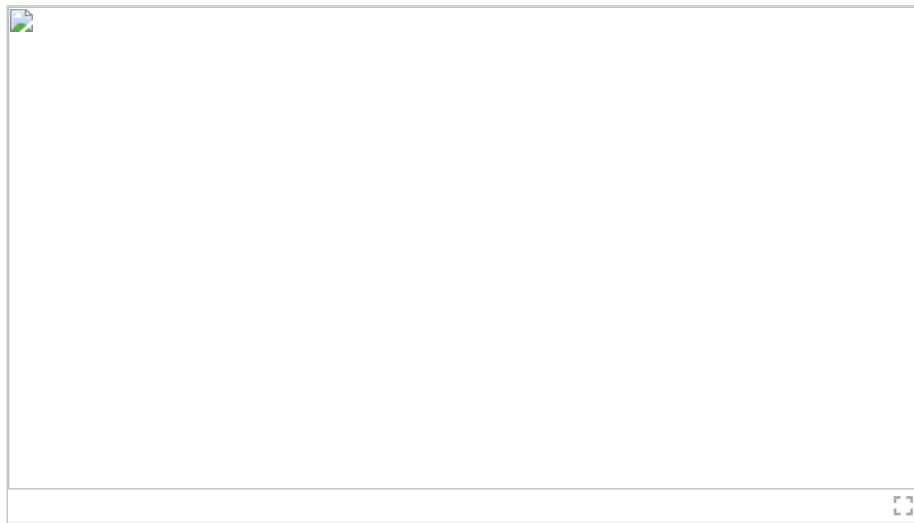
Fig. 3.
Process of the proposed method.

In the following sections, we will introduce each step in detail. For ease of discussion, we list the symbols used in the following sections and their meanings in Table I.

TABLE I Summary of Symbols and Notations

Notation	Meaning
n	number of users
$\{u_i \mid i \in [1,n]\}$	set of users
d_i	the amount of local data of u_i
$s_{u_i}^{sk}$	user u_i's private key
$s_{u_i}^{pk}$	user u_i's public key
$s_{i,j}$	shared secret key between u_i and u_j
$m_{i,j}, mm_{i,j}$	two mask codes of u_i
LM_{u_i}	local model of u_i
$MLM_{i,j}$	masked local model of u_i using $m_{i,j}$
MLM_i	masked local model of u_i using $mm_{i,j}$
ULM_i	the masked model of u_i that uploaded to AS
MMA	masked local model array
agg_model	aggregation model list generated by AS
L	length of agg_model
$garble\_agg\_model$	the random combination of agg_model and $random\_list$
$garble\_model\_length$	length of $garble\_agg\_model$
$o-posiiton$	a set records the position of all elements of agg_model in $garble\_agg\_model$
$s-position_i$	the proof generated by u_i
$inter-position$	intersection of $o-posiiton$ and $s-position_i$
th	a predefined threshold to verify malicious user
m	the percentage of malicious user in the system

A. Key Generation

In this phase, all users should first register their identity at TA. First, user u_i ($i \in [1,n]$) sends $[d_i, MAC_{u_i}, nonce]$ to TA, where the three parameters are the amount of local data, MAC address of u_i and a random number selected by u_i. Then, TA will generate a unique account (identity) id_{u_i} for u_i and send it to u_i and AS. Afterwards, TA uses function **KA.***param* to generate the parameters $pp = [\mathbb{G}, p, g, H]$ of the system and sends them to all users. After that, u_i will generate its private/public key $s_{u_i}^{SK}$ and $s_{u_i}^{PK}$ by function **KA.***gen*(pp) and send its public key to AS. AS will forward the received public keys to all other users. Using the public keys of other users and u_i's private key $s_{u_i}^{SK}$, u_i is able to generate shared secret key with all other users by using **KA.** $agree(u_i, u_j) (i \neq j)$.

B. Masked Model Generation and Uploading

For user u_i, it will first train its local model LM_{u_i} using its local data. Then, it will enter the mask code generation phase. For each user u_j ($j \neq i$), u_i will generate two mask codes ($m_{i,j}$ and $mm_{i,j}$) using (6)– (8), where $s_{i,j}$ is the shared secret key of u_i and u_j. Both the mask codes have the same length as that of the local model's parameter vector.

$$m_{i,j} = \begin{cases} \Delta_{i,j} \times PRG(s_{i,j}), & i \neq j \\ 0, & i = j \end{cases} \tag{6}$$

$$mm_{i,j} = \begin{cases} \Delta_{i,j} \times PRG(s_{i,j} + 1), & i \neq j \\ 0, & i = j \end{cases} \tag{7}$$

$$\Delta_{i,j} = \begin{cases} 1, & i > j \\ -1, & i < j \end{cases}. \tag{8}$$

View Source

Then, u_i will calculate a set of masked local models by (9) and 10. These masked local models will be combined into a masked local model list ULM_i by (11). Then, ULM_i will be uploaded to AS.

$$MLM_{i,j} = LM_{u_i} - m_{i,j}, j \in [1,n] \text{ and } i \neq j \tag{9}$$

$$MLM_i = LM_{u_i} - \sum_{k \in [1,n], k \neq i} mm_{i,k} \tag{10}$$

$$ULM_i = [MLM_{i,1} || \cdots || MLM_{i,i-1} || MLM_{i,i+1} || \cdots || MLM_{i,n} || MLM_i]. \tag{11}$$

View Source

C. Malicious User Detection

In this section, we will introduce the detail of the third step of the proposed method. From Fig. 3, we can see there are mainly four phases in this step, namely model aggregation, garble aggregation model list generation, model verification, and malicious user verification.

Model Aggregation. First, AS combines all users' masked local model list ULM_i to get a masked local model array MMA with n rows and n columns shown in (12). With MMA, AS will generate an aggregation model list agg_model by masked local model aggregation algorithm shown in Algorithm 1. From Algorithm 1, we know the first C_n^2 elements in agg_model are the average value of local models (LM_{u_i}) of any two users in the system. The last element of agg_model is the average value of all users' local models.

$$MMA = \begin{pmatrix} MLM_{1,2} & MLM_{1,3} & \cdots & MLM_{1,n} & MLM_1 \\ MLM_{2,1} & MLM_{2,3} & \cdots & MLM_{2,n} & MLM_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ MLM_{n,1} & MLM_{n,2} & \cdots & MLM_{n,n-1} & MLM_n \end{pmatrix}. \tag{12}$$

View Source ⓘ

Algorithm 1: Masked Local Model Aggregation Algorithm.

Input: MMA;

Output: agg_model // aggregation model list;

1:   create integer variables $k = 0$, i, j;

2:   **for** $i \leftarrow 1 \sim n-1$ **do**

3:   **for** $j \leftarrow i+1 \sim n$ **do**

4:   $agg\_model[k] = \frac{MLM_{i,j} + MLM_{j,i}}{2}$;

5:   k++;

6:   **end for**

7:   **end for**

8:   $agg\_model[k] = \frac{\sum_{t=1}^{n} MLM_t}{n}$;

9:   **return** agg_model;

Garble Aggregation Model List Generation. After obtaining the aggregation model list agg_model, AS will generate a garble aggregation model list $random\_list$ randomly, the length of which is a multiple of that of agg_model. Then, AS combines $random\_list$ and agg_model disorderly to form a new list $garble\_agg\_model$ and stores the position of all elements of agg_model in $garble\_agg\_model$ in a set $o-position$. Then, $garble\_agg\_model$ should be sent to all users.

Model Verification. When a user u_i receives $garble\_agg\_model$ from AS, it will run the model verification process shown in Fig. 3. In this process, u_i first tests the accuracy of each element of $garble\_agg\_model$ using its local data. Then, it selects $C_n^2 + 1$ elements from $garble\_agg\_model$ with the highest accuracy, and records the position of these elements in a set $s-position_i$ in order of their accuracy from the largest to the smallest. Finally, u_i sends $s-position_i$ to AS.

Malicious User Verification. For user u_i, AS computes the intersection of $o-position$ and $s-position_i$ and stores the result in set $inter-position_i$. If there exists empty set in the $inter-position$ of all users, AS will consider the system is suffering serious attacks and the learning process will be terminated. If all users' $inter-position$ sets are not empty, then AS will verify whether user u_i is malicious according to the length of $inter-position_i$. If

$|inter - position_i| > th \times |agg\_model|$, u_i will be seen as a benign user and its local model will also be seen as normal model, and vice versa. Here, notion $|\cdot|$ means the cardinal number of a set, and th is a predefined threshold which is a proportional value of the length of agg_model.

D. Model Updating

This step includes two phases, namely global model updating and local model updating.

Global Model Updating. AS will aggregate the normal local models to update the global model. First, all benign users make up a benign user set $bs = \{u_i \mid u_i \quad is \quad benign \quad user\}$. Then, AS will update the global model GM^k of the current iteration (assume it is the kth iteration) using (13), where $lb = |bs|$.

$$GM^k = \frac{\sum_{u_i, u_j \in bs, i \neq j} \left(\frac{d_i + d_j}{2 \times \sum_{u \in bs} d_u} \times MLM_{i,j} \right)}{C_{lb}^2}. \tag{13}$$
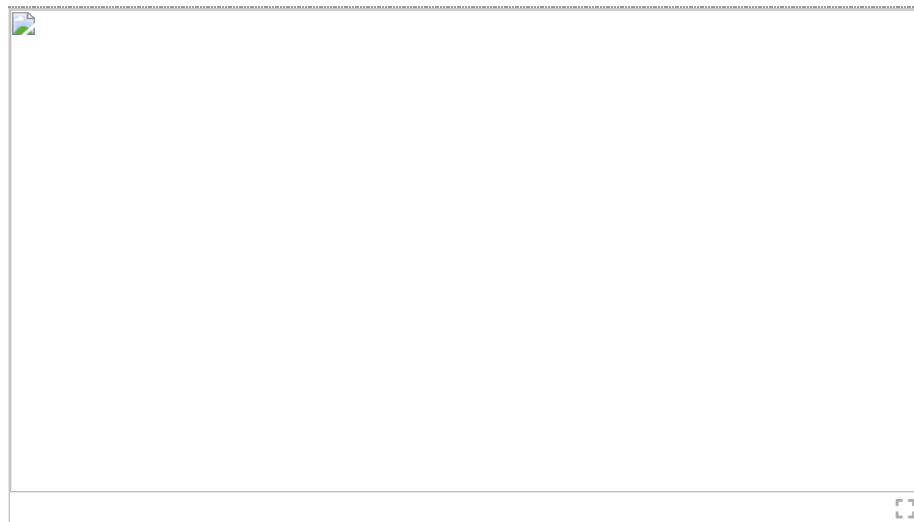
View Source ⓘ

Local Model Updating. When user u_i receives the global model GM^k from AS, it will compare the accuracy of GM^k and its local model LM_{u_i}. Then, the model with higher accuracy will be the local model in the next iteration.

SECTION IV.
Experiment and Discussion

We implement the proposed method to evaluate the performance of our ADFL. The experiment platform and environment are shown in Table II. We deploy the multiuser distributed learning system in a stand-alone environment. Each user runs the same training algorithm to generate its local model. All the training data are selected from MNIST database [1] which has a training set of 60 000 examples, and a test set of 10 000 examples. In the following, we will first verify the functionality of our scheme, and then measure its resource cost with different system parameters. Finally, we compare AFDL with two state-of-the-art works to illustrate the advantages of ADFL.

TABLE II Experiment Environment and Parameters

A. Functionality

In order to verify the performance of the proposed scheme, we test the classification accuracy of the proposed method by conducting logistic regression network and neural network (NN) on MNIST database at user side.

Fig. 4 shows the result when the local model is a logical regression model. We can see when there are 20% malicious users in the system, the growing trend in accuracy of the global model is almost the same as that when $m = 0$ (no malicious user), which means the proposed method is able to detect the malicious users and eliminate their negative impact in the aggregation process. In terms

of the exact accuracy value, before malicious users are detected, the accuracy of the global model will be influenced negatively, so the exact value of the classification accuracy is a little bit smaller when $m = 0.2$. In addition, after the malicious users are detected, because there are more benign users when $n = 20$ than when $n = 10$, the accuracy is higher in the situation of $u = 20$ from 14th round.
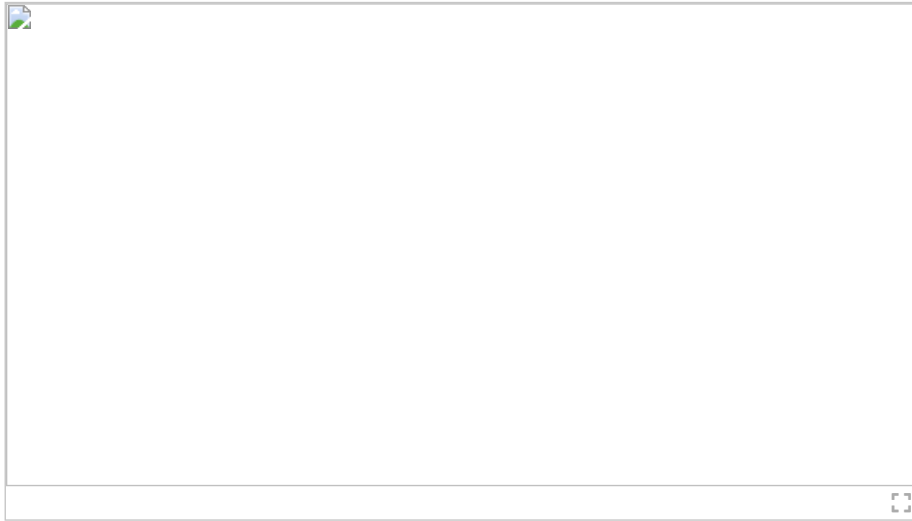
Fig. 4.
$th = 0.5$, $garble\_model\_length = 3\,L$ classification accuracy on MNIST database with three layers logical regression as local model.

Fig. 5 shows the result when the local model is a four-layers NN. When $m = 0.2$, we can see that the classification accuracy of the global model is almost the same as that when $m = 0$, which means the proposed method can also detect the malicious user effectively. Moreover, from Figs. 4 and 5, we can see the structure of the local model will not influence the performance of our scheme.
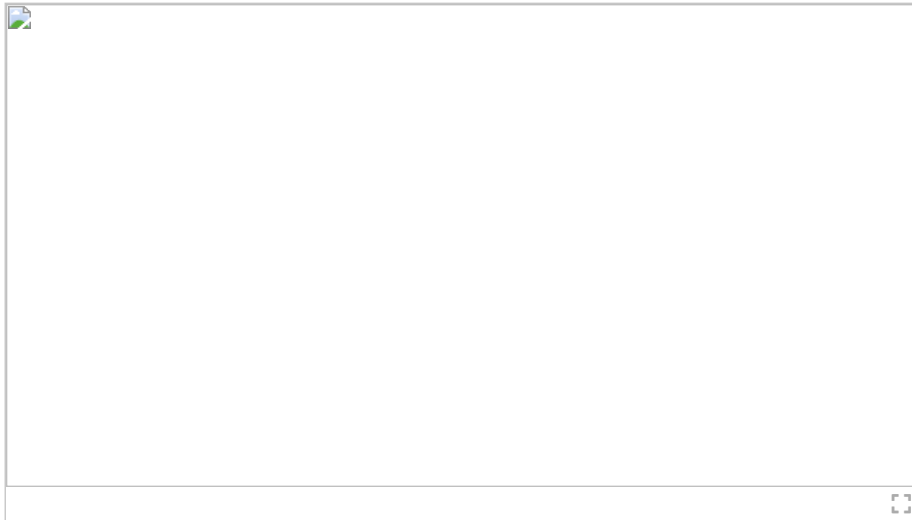
Fig. 5.
$th = 0.5$, $garble\_model\_length = 3\,L$ classification accuracy on MNIST database with four layers neural network as local model.

B. Performance Analysis

Definitely, the performance of the proposed method is closely related to three factors, i.e., the proportion of the malicious users (m), the threshold th, and $garble\_model\_length$. To analyze the influence of these factors on the performance of the proposed method, we analyze the classification accuracy of the global model using AFDL under different value of m, th, and $garble\_model\_length$, respectively. In this section, we use the logistic regression model as the local model of each user.

Fig. 6 shows the classification accuracy with different proportion of malicious users. We assign the value of the proportion of malicious users as 0.1, 0.2, 0.3, 0.4, 0.5, and 0.6. The result shows that

when $m = 0.6$, the accuracy of the global model still reaches around 0.8 from 12th round, which is not significantly different from that when $m=0.1$. In addition, we can see from Fig. 6 that no matter what the value of m, the accuracy of the global model is basically between 0.7 and 0.8, which means our scheme can effectively detect the malicious users to eliminate the negative influence of them on the accuracy of the global model.
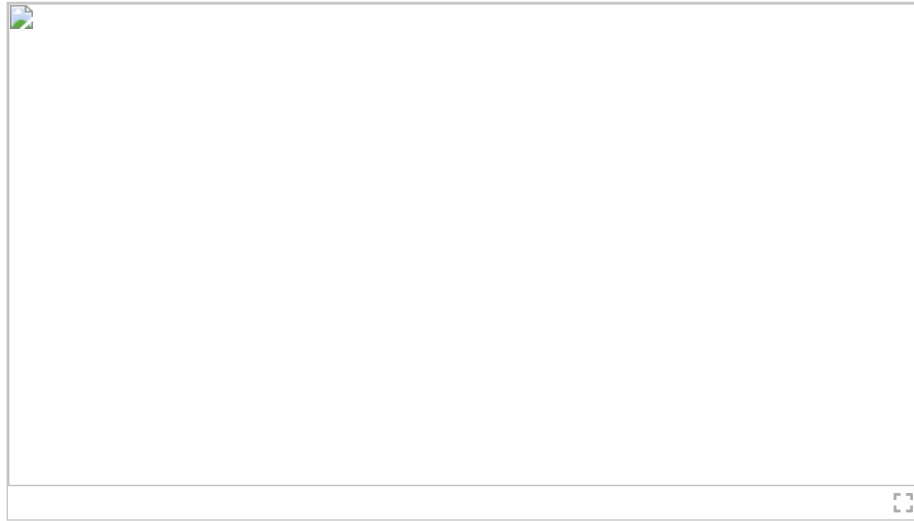
Fig. 6.
$n = 10$, $th = 0.5$, $garble\_model\_length = 3\,L$, influence of m on the classification accuracy.

Regarding the influence of th, the value of th is not the higher the better, which is shown in Fig. 7. That is because setting th too high will cause benign users' model to be detected as malicious due to data differences and training errors. Fig. 8 shows the classification accuracy with different value of $garble\_model\_length$. We can see the value of $garble\_model\_length$ is also not the higher the better. That is because too long the $garble\_model\_length$ will also increase the possibility of benign user being detected as a malicious user.
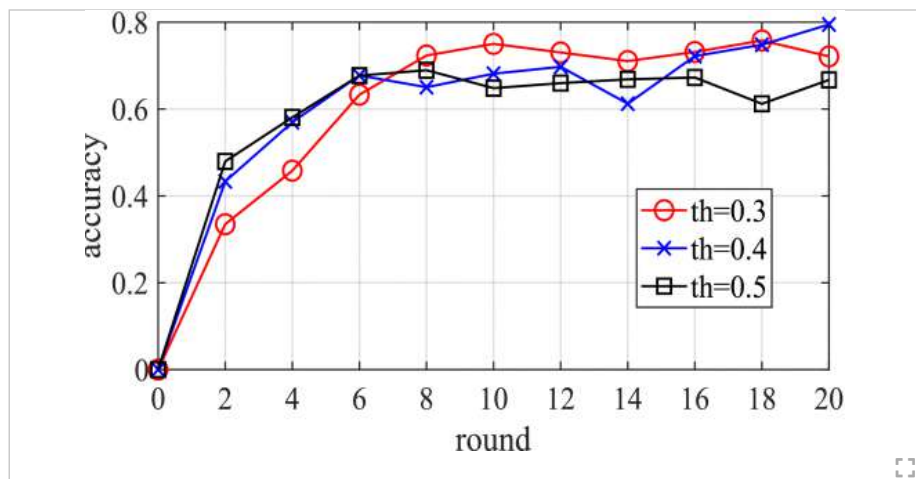
Fig. 7.
$n = 10$, $m = 0.4$, $garble\_model\_length = 3\,L$, influence of th on the classification accuracy.
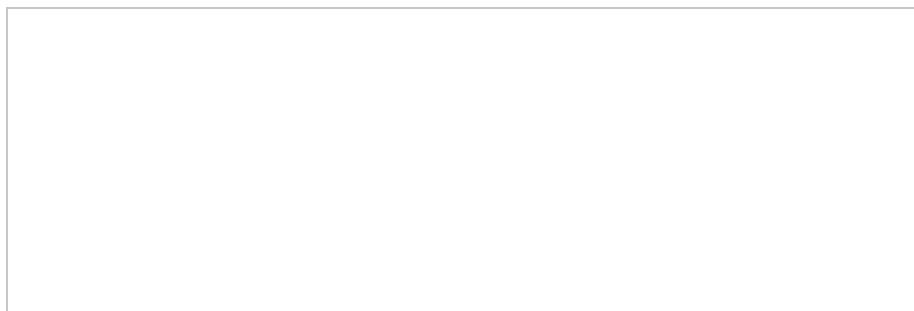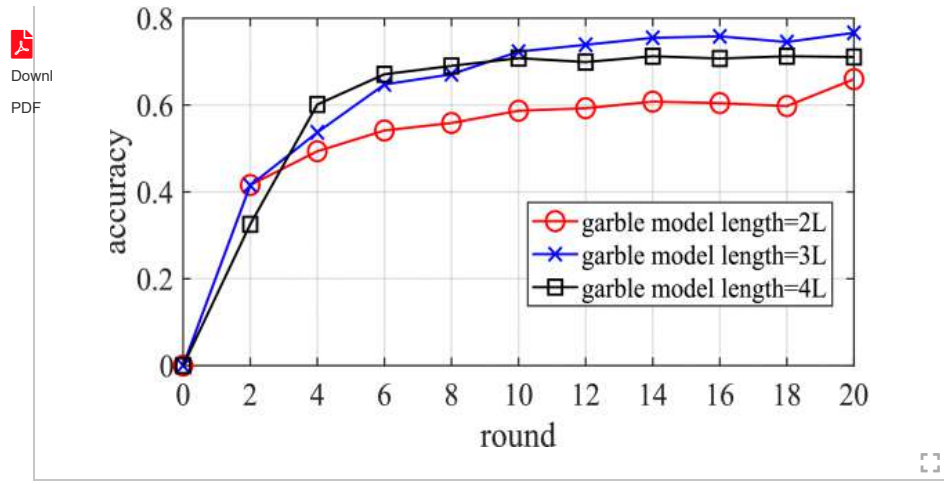
Fig. 8.

$n = 10$, $th = 0.5$, $m = 0.4$, influence of $garble\_model\_length$ on the classification accuracy.

C. Resource Consumption Analysis

In this section, we will analyze the resource consumption of the proposed method. First, we analyze the computational and communication complexity of the proposed method at both AS side and user side. Then, we will give the experimental result in terms of the computational and communication cost with the logistic regression model as the local model of each user.

Regarding the computational complexity, Table III shows the computational complexity in each step of the proposed method at both server side and user side. We can see that the overall computational complexity at both server side and user side is the same, that is, $O(n^2)$. Here, n means the number of users in the system.

TABLE III Computational Complexity on Service Side and User Side

From Fig. 3, we can see that during Step 1, the communication costs at service side and user side are $O(n^2)$ and $O(1)$, respectively. In Step 2, each user sends its masked local model list ULM_i to AS, so the communication costs at the server side and user side are 0 and $O(n)$, respectively. In Step 3, AS sends $garble\_agg\_model$ to each user. Thus, the communication cost at the server side is $O(garble\_model\_length) = O(L) = O(n^2)$. In addition, each user sends its $s - position$ to AS. For the user side, the communication cost is $O(|s - position|) = O(C_n^2 + 1) = O(n^2)$. In Step 4, the communication cost is unrelated to n, which should be $O(1)$ at both the service side and user side. Overall, the communication complexities at the service side and user side are $O(n^2)$ and $O(n^2)$, respectively.

In the following, we analyze the experimental result in terms of the resource consumption. Fig. 9 compares the computational cost of the whole proposed method and Step 3 of the method with different $garble\_model\_length$ values. From Fig. 9, we can see that given the number of users, the computational cost of the whole method and that devoted to Step 3 (malicious user detection)

is proportional to the value of $garble\_model\_length$. That is because users have to test more
models accuracy to get $s - position$ with the increasing of $garble\_model\_length$.
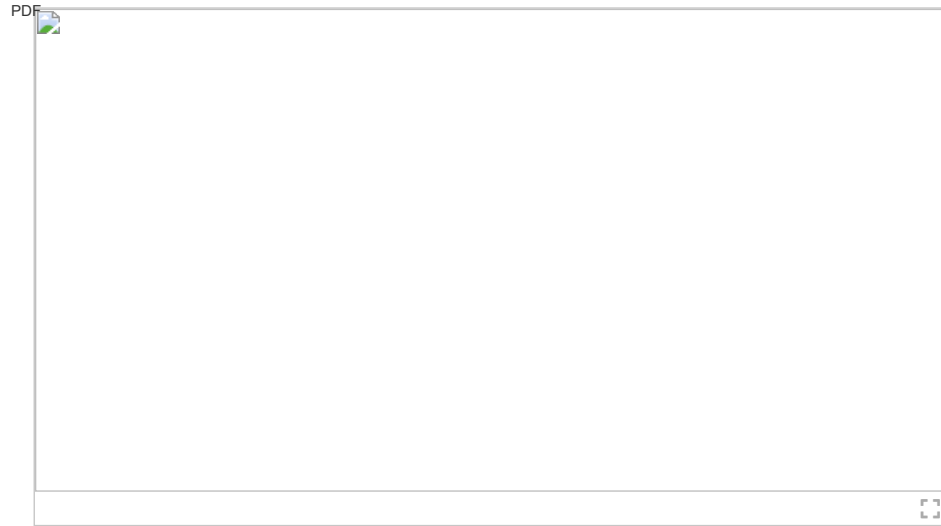
Fig. 9.
$n = 10$, $th = 0.5$, $m = 0.4$, computational cost of user with different
$garble\_model\_length.$

The proportion of the computational cost during Step 3 in the computational cost of the whole
proposed method with different number of users is shown in Fig. 10. Obviously, with the change of
the number of users, the proportion of the computational cost at both server side and user side
during Step 3 (malicious user detection) in the computational cost of the whole proposed method
remains constant. At the server side, the computational cost of Step 3 accounts for a larger
proportion of the overall cost than that at the user side.
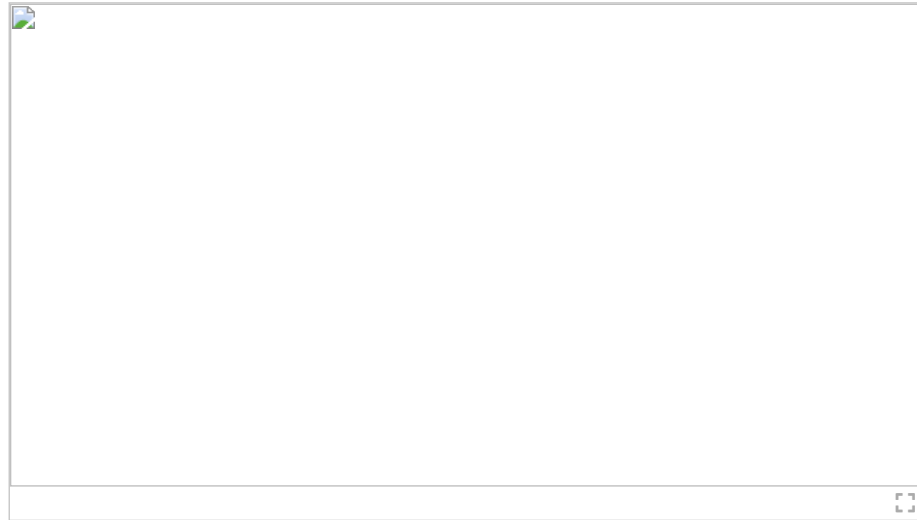
Fig. 10.
$th = 0.5$, $m = 0.4$, $garble\_model\_length = 3\,L$, proportion of computational cost
during Step 3 in that of the whole method with different number of users.

Figs. 11 and 12 show the communication cost on the user side and server side when the system has
different numbers of users and different $garble\_model\_length$ values. Here, we use the amount of
data transmitted to measure the communication cost, rather than the space occupied by these data.
That is because the space occupied by the data may be different in different environments.
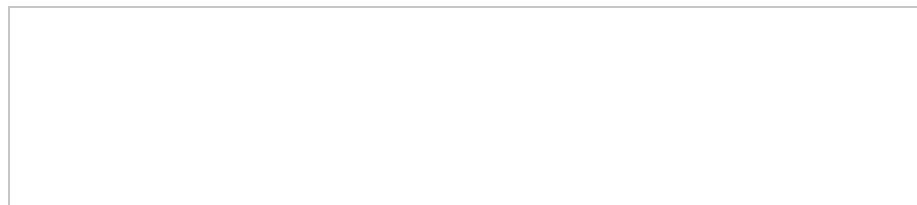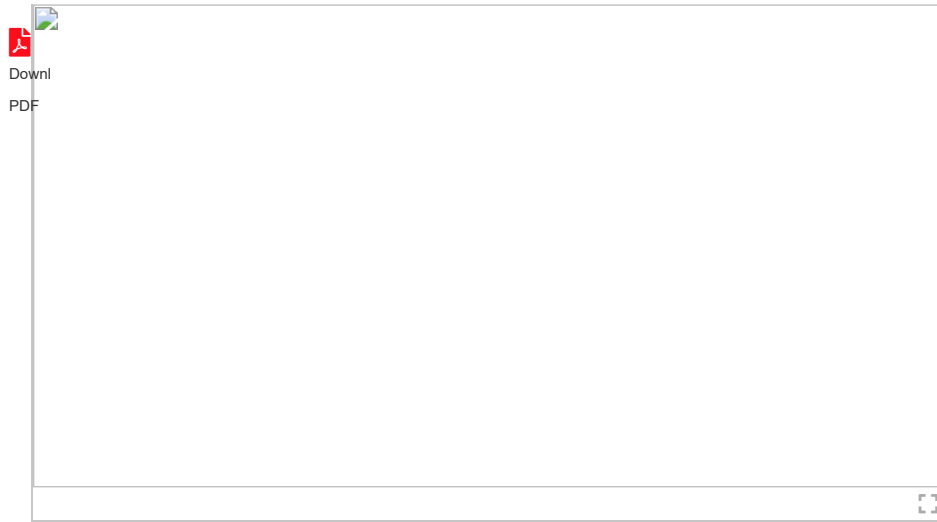
Fig. 11.
$th = 0.5$, $m = 0.4$, communication cost at user side with different number of users and $garble\_model\_length$ values.
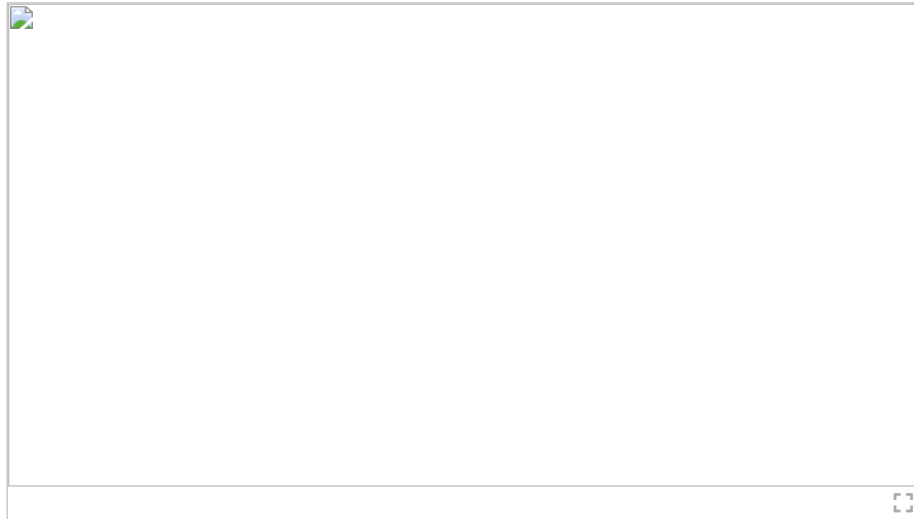
Fig. 12.
$th = 0.5$, $m = 0.4$, communication cost at server side with different number of users and $garble\_model\_length$ values.

In Fig. 11, it is obvious that during the model verification phase of Step 3, with the increase of the number of users, the communication overhead will increase in the order of quadratic. In addition, given the number of users, the cost of the model verification phase is linear with the value of $garble\_model\_length$. For the whole proposed scheme, the overall cost is directly proportional to the number of users, and is basically independent of the value of $garble\_model\_length$.

Fig. 12 gives the communication cost at the server side. We can see that on the server side, the main communication cost comes from the process of garble aggregation model list generation. Both the overall cost and the cost of the phase garble aggregation model list generation have a quadratic relationship with the number of users in the system. In addition, given the number of users, they are linear with the value of $garble\_model\_length$.

D. Comparison With Available Works

In this section, we compare the performance of our ADFL with two available works, namely AvgFed algorithm [8] and Li's autoencoder-based scheme [20]. From Fig. 13, we can see, the compared two schemes cannot resist the poisoning attack from malicious users who have no local dataset. Because AvgFed algorithm did not consider the malicious behaviors of users and all uploaded models, it will contribute to the global model. For Li's scheme, although autoencoder is a helpful technique in the abnormal detection area, the mask code added in the local model will destroy the property of the local model (both benign and malicious). This will make the autoencoder not distinguish malicious model from benign model.
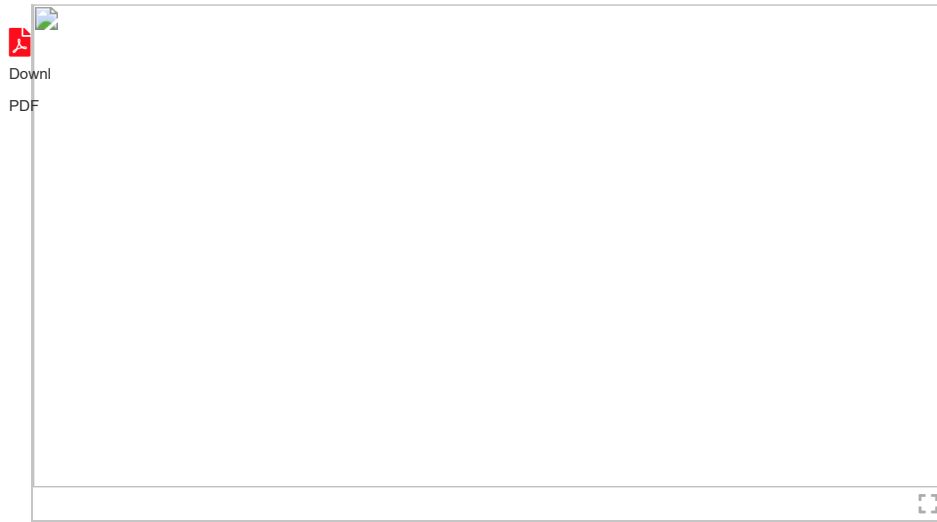
Fig. 13.

$n = 10$, $th = 0.5$, $m = 0.4$, performance comparison with available works.

SECTION V.
Related Works

In this section, we introduce the latest related works of federated learning in terms of privacy-preserving and malicious user detection.

A. Privacy-Preserving Federated Learning

In federated learning, users and aggregation server collaboratively train a unified machine learning model. In each iteration, each user uploads its local model parameters (gradients) to the server, which aggregates all received models and returns the aggregation results to each user. Although each update is ephemeral and contains no information about the user's private dataset, it is still possible to learn individual privacy by inspecting the user's most recent upload. To keep server from accessing any individual user's privacy, several privacy-preserving federated learning frameworks have been proposed, using techniques mainly including homomorphic encryption (HE) [18], differential privacy [21], [22], MPC (Secure Multi-Party Computation) [17]. Phong *et al.* [18] proposed a privacy preserving deep learning algorithm based on additively HE. Wei *et al.* [21] proposed a federated learning framework to prevent information leakage based on differential privacy, in which artificial noise is added to parameters at the user's side before uploading to the server. Bonawitz *et al.* [17] proposed an MPC-based privacy-preserving federated learning. In their work, the true local model of each user is masked by a mask code derived from the shared key of each two users, so that the aggregation server can only know the weighted averages of the uploaded local models, without knowing any individual user's local model.

Either techniques mentioned above makes the server unable to observe the users' true local model (parameters), while this also brings more opportunities to adversaries to hide their malicious models. So, in this article, we focus on the malicious local model detection problem in MPC-based federated learning system.

B. Abnormal Detection for Federated Learning

Because federated learning has the characteristics of multiuse distributed cooperation and privacy protection, it may suffer from various threats [10]. In the server side, aggregation server may send false global model to each user. In the user side, users who have no training data may upload arbitrarily generated local model to the server. In addition, adversaries may poison the local data of users to make them get a low accuracy local model. In order to find the abnormal behaviors of all parties (including aggregation server and users) in the federated learning process, researchers have proposed some solutions, which are summarized in Table IV.

TABLE IV Survey of Available Methods for Abnormal Behaviors Detection in Federated Learning

available works	server side	user side	support security aggregation	attacks defended	core idea
[2]	✓	✗	✓	global model tampering attack	homomorphic encryption
[24]	✓	✗	✓	global model tampering attack	linearly homomorphic hash
[25]	✓	✗	✓	global model tampering attack	Lagrange interpolation
[26]	✗	✓	✗	Sybil-based poisoning attacks	Sybil users have the same upload parameter vectors
[9]	✗	✓	✗	local model poisoning	find the maximum clique of upload parameter vectors
[27]	✗	✓	✓	Byzantine attack	distance based outlier detection
[21]	✗	✓	✓	Byzantine attack	autoencoder based detection
[28]	✗	✓	✓	Byzantine attack	truth discovery technique

Researchers have proposed several methods to verify the integrity of the global model [2], [23], [24]. Xu *et al.* [2] proposed a scheme named VerifyNet to verify the correctness of the aggregated results returned from the server. Fu *et al.* [24] used Lagrange interpolation to elaborately set interpolation points for verifying the correctness of the aggregated gradients.

For the verification of the users' reliability, several schemes have also been proposed. Fung *et al.* [25] proposed a scheme to defend against Sybil-based poisoning attacks for the federated learning setting based on interclient contribution similarity. Cao *et al.* [9] proposed a scheme to eliminate poisoned local models from malicious participants during training. They identify benign local models by solving a maximum clique problem, and suspected (poisoned) local models will be ignored during global model updating.

The abovementioned schemes require that aggregation server is able to observe all users' local models at each iteration. However, this is not possible in the privacy-preserving federated learning schemes mentioned in the previous section. So, several researchers proposed malicious model detection schemes for privacy-preserving federated learning. So *et al.* [26] proposed a malicious model detection scheme which utilized the distance-based outlier detection approach to guarantee the robustness against Byzantine users. Li *et al.* [20] proposed a method to detect anomalous clients at the server side, in which the autoencoder is used to detect the local model uploaded by users. Xu *et al.* [27] designed a solution to reduce the negative impact of irregular users in federated learning, which is based on truth discovery technique. In this scheme, server should first initialize the value of each parameter of the global model and send it to all users, which needs the experience of the server or a public auxiliary database to estimate the global model.

SECTION VI.
Conclusion

In this article, we proposed a solution to detect malicious model uploaded to aggregation server in MPC-based privacy-preserving federated learning system. During the learning process, the aggregation server can only access the masked local models sent by users. In order to detect the malicious local model, aggregation server needs to send a challenge to all users in each iteration, and then each user should generate a response to this challenge, which will be used by aggregation server to verify whether a user is malicious. A global model aggregation method is also proposed in which malicious local models will not be used in the global model aggregation process. To verify the performance of our method, several experiments were conducted and the results showed that our method can detect malicious local models effectively and ensure the global model has a high accuracy. In addition, we analyzed the resource consumption of the proposed method.

In the future, we will continue to find solution on malicious user detection for federated learning systems using other privacy-preserving techniques. In addition, we will also focus on the abnormal detection of local models in Non-IID horizontal federated learning system.

Authors ^

Jingjing Guo
Xidian University, Xian, China

Jingjing Guo received the B.S., M.Sc., and Ph.D. degrees in computer science and technology from the School of Computer Science and Technology, Xidian University, Xian, China, in 2009, 2012, and 2015, respectively.

She is currently an Associate Professor with the School of Cyber Engineering, Xidian University. Her research interests include trust management, wireless networks security, and artificial intelligence security.

Jingjing Guo received the B.S., M.Sc., and Ph.D. degrees in computer science and technology from the School of Computer Science and Technology, Xidian University, Xian, China, in 2009, 2012, and 2015, respectively.

She is currently an Associate Professor with the School of Cyber Engineering, Xidian University. Her research interests include trust management, wireless networks security, and artificial intelligence security. **View more**

Haiyang Li
Xidian University, Xian, China

Haiyang Li received the B.S. degree in cyber engineering in 2020 from the School of Cyber Engineering, Xidian University, Xi'an, China, where he is currently working toward the master's degree.

His research interests include security of wireless network and artificial intelligence system.

Haiyang Li received the B.S. degree in cyber engineering in 2020 from the School of Cyber Engineering, Xidian University, Xi'an, China, where he is currently working toward the master's degree.

His research interests include security of wireless network and artificial intelligence system. **View more**

Feiran Huang
Jinan University, Guangzhou, China

Feiran Huang (Member, IEEE) received the B.S. degree in physics from the School of Physics and Electronics, Central South University, Changsha, China, in 2011, and the Ph.D. degree in computer science and technology from the School of Computer Science and Engineering, Beihang University, Beijing, China, in 2019.

He is currently an Associate Professor with the College of Cyber Security, Jinan University, Guangzhou, China. His research interests include multimodal data analysis, social media analysis, and network security.

Feiran Huang (Member, IEEE) received the B.S. degree in physics from the School of Physics and Electronics, Central South University, Changsha, China, in 2011, and the Ph.D. degree in computer science and technology from the School of Computer Science and Engineering, Beihang University, Beijing, China, in 2019.

He is currently an Associate Professor with the College of Cyber Security, Jinan University, Guangzhou, China. His research interests include multimodal data analysis, social media analysis, and network security. **View more**

Zhiquan Liu
Jinan University, Guangzhou, China

Zhiquan Liu received the B.S. degree in mathematics from the School of Science, Xidian University, Xi'an, China, in 2012, and the Ph.D. degree in computer science and technology from the School of Computer Science and Technology, Xidian University, in 2017.

He is currently an Associate Professor with the College of Cyber Security, Jinan University, Guangzhou, China. His research interests include trust management and privacy preservation in vehicular networks and UAV networks.

Zhiquan Liu received the B.S. degree in mathematics from the School of Science, Xidian University, Xi'an, China, in 2012, and the Ph.D. degree in computer science and technology from the School of Computer Science and Technology, Xidian University, in 2017.

He is currently an Associate Professor with the College of Cyber Security, Jinan University, Guangzhou, China. His research interests include trust management and privacy preservation in vehicular networks and UAV networks. **View more**

Yanguo Peng
Xidian University, Xian, China

Yanguo Peng (Member, IEEE) received the B.Sc. degree in network engineering from North University of China, Taiyuan, China, in 2009, the M.S. degree in computer software and theory from Guizhou University, Guiyang, China, in 2012, and the Ph.D. degree in computer systems organization from Xidian University, Xi'an, China, in 2016.

He is currently an Associate Professor with the School of Computer Science and Technology, Xidian University. His research interests include cloud security, data privacy protection, and blockchain.

Yanguo Peng (Member, IEEE) received the B.Sc. degree in network engineering from North University of China, Taiyuan, China, in 2009, the M.S. degree in computer software and theory from Guizhou University, Guiyang, China, in 2012, and the Ph.D. degree in computer systems organization from Xidian University, Xi'an, China, in 2016.

He is currently an Associate Professor with the School of Computer Science and Technology, Xidian University. His research interests include cloud security, data privacy protection, and blockchain. **View more**

Xinghua Li
Xidian University, Xian, China

Xinghua Li received the M.E. and Ph.D. degrees in computer science from Xidian University, Xi'an, China, in 2004 and 2007, respectively.

He is currently a Full Professor and Ph.D. Supervisor with Xidian University. His research interests include wireless networks security, privacy protection, cloud computing, software defined network, and security protocol formal methodology.

Xinghua Li received the M.E. and Ph.D. degrees in computer science from Xidian University, Xi'an, China, in 2004 and 2007, respectively.

He is currently a Full Professor and Ph.D. Supervisor with Xidian University. His research interests include wireless networks security, privacy protection, cloud computing, software defined network, and security protocol formal methodology. **View more**

Jianfeng Ma
Xidian University, Xian, China

Jianfeng Ma (Member, IEEE) received the M.E. and Ph.D. degrees in computer software and communications engineering from Xidian University, Xi'an, China, in 1988 and 1995, respectively.

He is currently a Full Professor with Xidian University. His research interests include information security, coding theory, and cryptography.

Jianfeng Ma (Member, IEEE) received the M.E. and Ph.D. degrees in computer software and communications engineering from Xidian University, Xi'an, China, in 1988 and 1995, respectively.

He is currently a Full Professor with Xidian University. His research interests include information security, coding theory, and cryptography. **View more**

Varun G. Menon
SCMS School of Engineering and Technology, Ernakulam, India

Varun G. Menon (Senior Member, IEEE) received the Ph.D. degree in computer science and engineering from Sathyabama University, Chennai, India, in 2017.

He is currently an Associate Professor with the Department of Computer Science and Engineering, SCMS School of Engineering and Technology, Kerala, India. He has published more than 50 research papers in peer reviewed and highly indexed International Journals and Conferences. His research interests include Internet of Things, fog computing and networking, underwater acoustic sensor networks, scientometrics, educational psychology, ad-hoc networks, wireless communication, opportunistic routing, and wireless sensor networks.

Varun G. Menon (Senior Member, IEEE) received the Ph.D. degree in computer science and engineering from Sathyabama University, Chennai, India, in 2017.

He is currently an Associate Professor with the Department of Computer Science and Engineering, SCMS School of Engineering and Technology, Kerala, India. He has published more than 50 research papers in peer reviewed and highly indexed International Journals and Conferences. His research interests include Internet of Things, fog computing and networking, underwater acoustic sensor networks, scientometrics, educational psychology, ad-hoc networks, wireless communication, opportunistic routing, and wireless sensor networks. **View more**

Konstantin Kostromitin Igorevich
South Ural State University National Research University, Celabinsk, Russian Federation

Konstantin Kostromitin Igorevich received the B.S., M.S., and Ph.D. degrees in physics from Chelyabinsk State University, Chelyabinsk, Russia, in 2008, 2010, and 2013, respectively.

He is currently a Researcher with the Department of Information Security, South Ural State University, Chelyabinsk. His research interests include physics and Internet of Things security.

Konstantin Kostromitin Igorevich received the B.S., M.S., and Ph.D. degrees in physics from Chelyabinsk State University, Chelyabinsk, Russia, in 2008, 2010, and 2013, respectively.

He is currently a Researcher with the Department of Information Security, South Ural State University, Chelyabinsk. His research interests include physics and Internet of Things security. **View more**

Figures ⌄

References ⌄

Citations ⌄

Keywords ⌄

Metrics ⌄

Footnotes ⌄

More Like This

Big Data Privacy and Challenges for Machine Learning

2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)

Published: 2020

Manufacturing Data Privacy Protection System for Secure Predictive Maintenance

2022 5th International Conference on Data Science and Information Technology (DSIT)

Published: 2022

Show More

Downl
PDF

IEEE Personal Account   **Purchase Details**   **Profile Information**   **Need Help?**   **Follow**

CHANGE   PAYMENT OPTIONS   COMMUNICATIONS   US & CANADA: +1 800
USERNAME/PASSWORD   PREFERENCES   678 4333
                          VIEW PURCHASED
                          DOCUMENTS   PROFESSION AND   WORLDWIDE: +1 732
                                                   EDUCATION   981 0060

                                                   TECHNICAL INTERESTS   CONTACT & SUPPORT

About IEEE *Xplore* | Contact Us | Help | Accessibility | Terms of Use | Nondiscrimination Policy | IEEE Ethics Reporting ↗ | Sitemap |
IEEE Privacy Policy

IEEE Account

» Change Username/Password
» Update Address

Purchase Details

» Payment Options
» Order History
» View Purchased Documents

Profile Information

» Communications Preferences
» Profession and Education
» Technical Interests

Need Help?

» **US & Canada:** +1 800 678 4333
» **Worldwide:** +1 732 981 0060
» Contact & Support

About IEEE *Xplore* | Contact Us | Help | Accessibility | Terms of Use | Nondiscrimination Policy | Sitemap | Privacy & Opting Out of Cookies

A N2CNN-Based Anomaly Detection Method for Cardiovascular Data in Cyber-Physical System

Raju Pal, Ashish Kumar Tripathi, *Member, IEEE*, Avinash Chandra Pandey, *Member, IEEE*, Mohammad Ayoub Khan, *Senior Member, IEEE*, Varun G. Menon, and Himanshu Mittal

Abstract—Internet-of-medical-things is the new means of monitoring patient health remotely. However, the real-time detection of anomalies in the patient data is a challenging task, especially on ECG-data. To ease the same, a novel method, NSGA-II based convolution neural network, is presented in this paper for efficient anomaly detection. In the proposed method, non-dominated sorting genetic algorithm-II is emto obtain optimal hyper-parameters of CNN by evaluating three objective functions namely, accuracy, precision, and recall. Further, the performance validation of the proposed method is conducted on two public datasets and compared against seven state-of-the-art methods. Experimental results affirm that the proposed method outperforms the considered methods with an accuracy of 94.83% and 94.96% on MIT-BIH arrhythmia dataset and INCART dataset, respectively. Therefore, it can be claimed that the proposed method is an efficient alternative for anomaly detection.

Index Terms—Cyber-physical system, remote health monitoring, convolutional neural network, elitist non-dominated sorting genetic algorithm, anomaly detection.

I. INTRODUCTION

CYBER-physical systems are connecting the world intelligently by interconnecting devices with people through Internet of Things (IoT). This has facilitated different domains of human lifestyle such as entertainment, shopping, and health [1]. With the ability to capture real-time information by sensors, there has been a boom in the IoT-based applications. One of the prominent applications is healthcare. In healthcare, patients and physicians are closely associated through IoT, which advantages in better health monitoring. Moreover, this collaboration of healthcare with IoT has initiated the era of remote health monitoring. In this, patients are geared with a number of health monitoring sensors to capture data like, pulse rate, calories burned, etc. While working with real-time data, this helps in intelligent health management by classifying alarming conditions in a pervasive manner. Fig. 1 demonstrates a remote health monitoring in cyber-physical system from capturing the sensory data to analyzing it for anomaly detection. For efficient analysis, this paper presents a new anomaly detection method for effective health monitoring in the cyber-physical system.

In general, health monitoring can be viewed as a classification process of labeling health signals to different health conditions. One of the most common devices in practice for health monitoring is the electrocardiogram (ECG). ECG captures cardiovascular movements in the form of electrical signals by attaching sensors to the human chest. Generally, ECG signals resemble a pattern of heart movement, which is analyzed to determine the heart condition. However, this analysis requires expertise and is a challenging task for a naive person. In literature, deep-learning models have been quite efficient, especially convolutional neural networks (CNN). However, CNN demands an optimal value for large number of hyper-parameters such as, number of layers, number of nodes, filter size, etc. Moreover, incorrect initialization of CNN hyper-parameters have resulted in local optimal solutions and determination of optimal values for them is a non-trivial and time-consuming task. Recent works have demonstrated that meta-heuristic algorithms are efficient approach to handle the same.

Meta-heuristic based methods are the mathematical models of the nature behaviour for attaining optimization [2]. To attain optimal solution, these algorithms employ intensification (exploitation) and diversification (exploitation). In exploration, existing solutions are updated with a large step-size to examine the search region, while exploitation performs the local search by taking small step-size. This searching behavior helps in finding an optimal solution. Generally, meta-heuristic algorithms consider single-objective function to attain the solution. However, single objective functions considers single perspective of the data which is not efficient for CNN hyper-parameters optimization [3]. On the contrary, multi-objective meta-heuristic algorithms consider multiple criteria such as accuracy, recall, precision, etc. and have exhibited better performance in optimization of CNN hyper-parameters. The considered objectives are generally highly correlated where multi-objective algorithms performs better than single-objective algorithms. Certainly, elitist non-dominated sorting
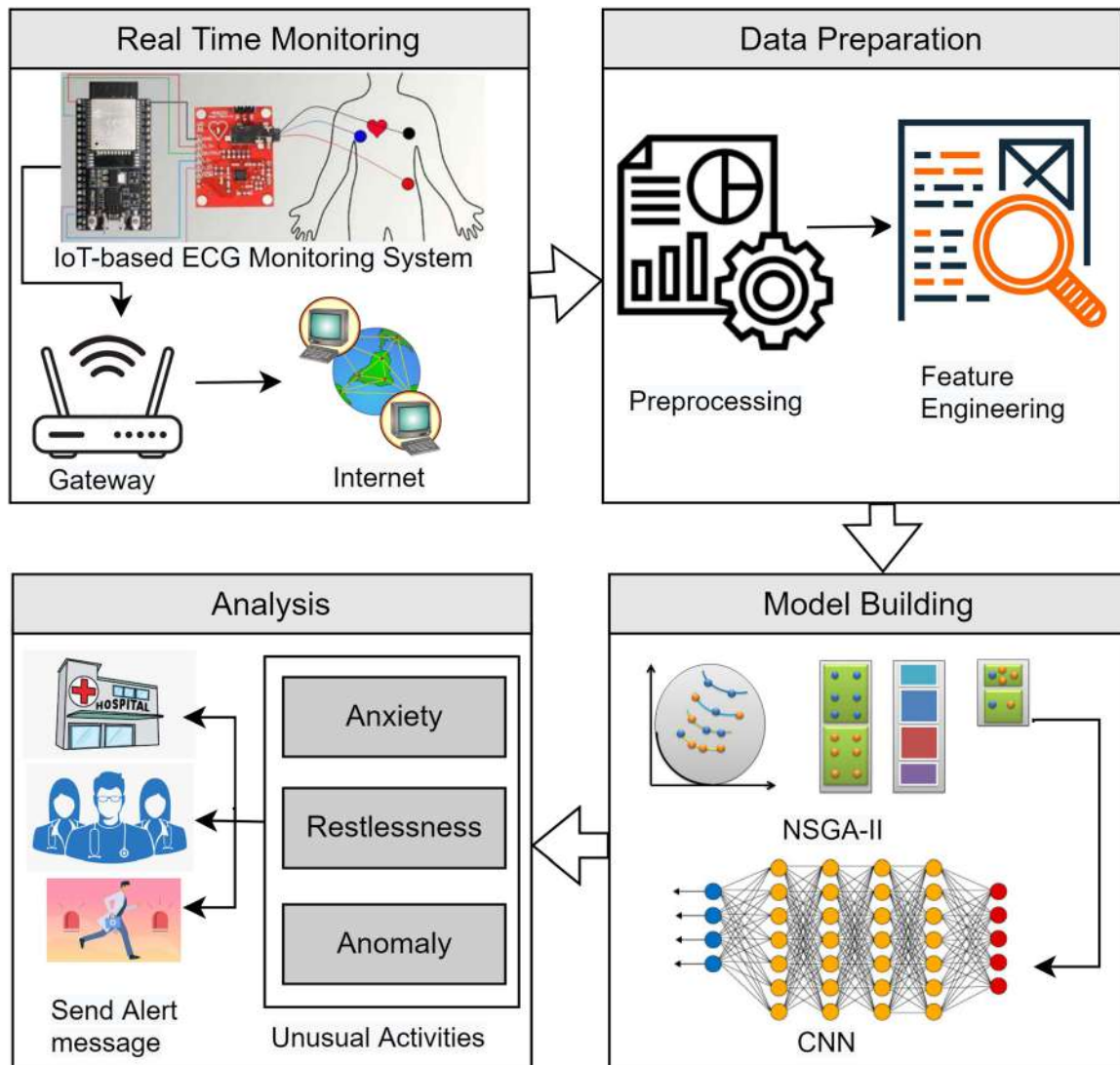
Fig. 1.   Framework of the remote health monitoring in cyber-physical system.

genetic algorithm-II (NSGA-II) is one of the successful multi-objective meta-heuristic algorithms. It uses the concept of non-domination sorting for organizing the solutions into different fronts according to multiple objectives. The perseverance of elite solutions and maintaining the diversity in the next generation solutions using crowding distance operators are the main advantages of NSGA-II. Therefore, this paper presents a new automated method to generate an optimized CNN model by leveraging the advantages of NSGA-II.

Thus, the overall contribution of this paper is two folds; (i) Multi-objective NSGA-II is leveraged to optimize the hyper-parameters of CNN model and (ii) The new optimized CNN architecture is employed for automated categorization of sensor generated health signals. An extensive experimental analysis of the proposed method, NSGA-II based CNN, is performed on two public datasets, namely MIT-BIH arrhythmia dataset and INCART dataset. The experimental results are compared against seven state-of-the-art classification based methods.

The organization of the paper is as follows; Section II briefs the related work on the ECG-based methods and optimization of

CNN. Section III details the proposed method for the classification of ECG-data. Section IV discusses the considered datasets, while evaluation metrics are presented in Section V. Section VI performs the ablation study on the proposed model and Section VII analysis the experimental performance of the proposed method. Lastly, the conclusion is drawn in Section VIII.

II. RELATED WORK

In literature, several ECG-based classification methods are presented, which can be mainly classified into machine learning, deep learning, and metaheuristic-based methods [4]. Support vector machine (SVM), linear regression, and linear discriminant are the popular machine learning algorithms used for cardiac arrhythmic classification [5], [6]. Park *et al.* [7] employed SVM to classify the heartbeats into normal and abnormal categories. Garcia *et al.* [8] detected the abnormal heart movements using a support vector machine and temporal vector cardiogram. Lannoy *et al.* [9] used weighted SVM to classify the MIT-BIH imbalance database into respective

categories. Huang *et al.* [10] employed a hierarchical SVM with a maximum voting approach and found considerable improvements in the accuracy of the classification tasks.

Literature witnessed that SVM-based approaches showed better results on balanced datasets. However, they show poor performance for datasets having imbalanced classes [11]. To eliminate the class imbalance problem, a two-stage approach has been proposed. The proposed method first employs the synthetic minority over-sampling technique (SMOTE) [12] to balance the skewed datasets and then uses CNN for heartbeat classification [13]. Besides, many approaches based on deep learning have been presented to classify heartbeats into normal and abnormal categories [14] [15]. Hammad *et al.* [16] developed a rule-based ECG classification model on ECG-extracted features through Pan–Tompkins algorithm. In [17], Acharya *et al.* presented a CNN-based classification method to analyze ECG records. Similarly, Amrani *et al.* [18] employed a CNN model with Q-Gaussian multi-class SVM for heart monitoring. Further, Sannino *et al.* [14] applied a deep neural network after performing pre-processing on the ECG data. Mousavi *et al.* [19] combined the recurrent neural network with CNN to develop the ECG classification method. Moreover, Kachuee *et al.* [20] employed the concept of transfer learning by using a trained CNN on the ECG dataset. In [21], Wang *et al.* employed a dual fully connected neural network for the detection of abnormalities in the heart. Ebrahimi *et al.* [22] presented a concise review of CNN for ECG-based signal classification.

Deep-learning models, particularly CNN, have proven to classify heartbeats effectively. However, it requires an optimal value for many hyper-parameters, such as the number of layers, nodes, and filter size, which is a tedious task to determine. Therefore, several meta-heuristic algorithms have been extensively leveraged for the optimization of CNN to solve numerous real-time problems. Goel *et al.* [23] presented an automatic CNN-based COVID-19 diagnosis system in which hyper-parameter values are optimized through a grey-wolf optimizer. Zhang *et al.* [24] fine-tuned the convolutional neural network with a whale optimization algorithm to diagnose skin cancer. Wang *et al.* [25] employed particle swarm optimization for the hyper-parameter's optimization of CNN. Likewise, Navaneeth and Suchetha [26] optimized the architecture of 1D-CNN with a support vector machine by employing particle swarm optimization and classified chronic kidney diseases. However, multi-objective meta-heuristic algorithms have gained increasing attention for CNN optimization due to their ability to achieve optimal solutions with conflicting objectives. Azad *et al.* [27] utilized a multi-objective genetic algorithm on a neural network to optimize power generation and water desalination in a solar chimney system.

Xie *et al.* [28] optimized the neural network using multi-objective swarm approach based on decomposition for finding optimal feed rate of the spindle in computerized numerical control milling. Bakar *et al.* [29] presented NSGA-II based CNN to design the low Reynolds number airfoil. Recently, Goel *et al.* [30] performed COVID-19 diagnosis by presenting a two-layer deep neural network which is optimized through multi-objective grasshopper optimization algorithm. Moreover, Kanwal *et al.* [31] evolved the convolutional auto-encoder to obtain optimal architecture by using multi-objective particle swarm optimization. Singh *et al.* [32] optimized the hyper-parameters of CNN with multi-objective DE (differential evolution) for medical image classification. A comprehensive survey on existing multi-objective meta-heuristic algorithms for hyper-parameter optimization of different deep neural networks is studied in [33], [34].

III. PROPOSED METHOD

In the proposed approach, we are using convolution neural network (CNN) optimized with enhanced NSGA-II with three objective functions, namely accuracy, specificity, and sensitivity. Moreover, three health monitoring datasets are leveraged to validate the efficacy of the proposed N2CNN model. A simplified architecture of the proposed approach is depicted in Fig. 2. Further, each component of the N2CNN is described in the following sections.

A. Convolution Neural Network

CNN is one of the popular deep learning model, equipped with a number of distinct layers. In this paper, a fully connected nonlinear CNN variation with nine layers including input, multiple convolutions, softmax, batch normalization, and a class output has been used. The first input layer of the CNN takes the image data. This means the size of the input layer has to be the same as that of the input data. The convolution function of the convolution layer comprises of sliding a kernel (or filter) over the input to capture different kinds of local information [35]. The values of the filters are chosen by learners during the training stage [36]. The size of the produced feature map depends on different factors such as the number of filters, size of the filter, stride of the filter process and the padding (i.e., adding additional zero value to fit the filter on bordering elements). The feature map produced by the entire convolution layer can be said to be a map of $n + 1$ dimension (where n denotes the input dimension) [37]. The output of the entire convolution layer can be given by

$$C_i = \epsilon(k_i \circledast f + b_i) \tag{1}$$

here, for every i (kernel number in the convolution layer), C is the feature map produced, $\circledast$ denotes the convolution operation, k represents the kernel (or filter), f denotes the input data, whereas bias is represented by b. The non linear layer also makes use of the RelU (rectified linear unit) function to induce non linearity into the network. This is accomplished by replacing all negative values with zeros, which increases the classification accuracy [38]. The function is defined as

$$f(x) = \begin{cases} 0, & x < 0 \\ x, & x \geq 0 \end{cases} \tag{2}$$

here, x represents the input given to the function.

A commonly used method for reducing the dimensionality of the input data is achieved using a pooling layer. In this paper, the max pooling operation has been employed since it
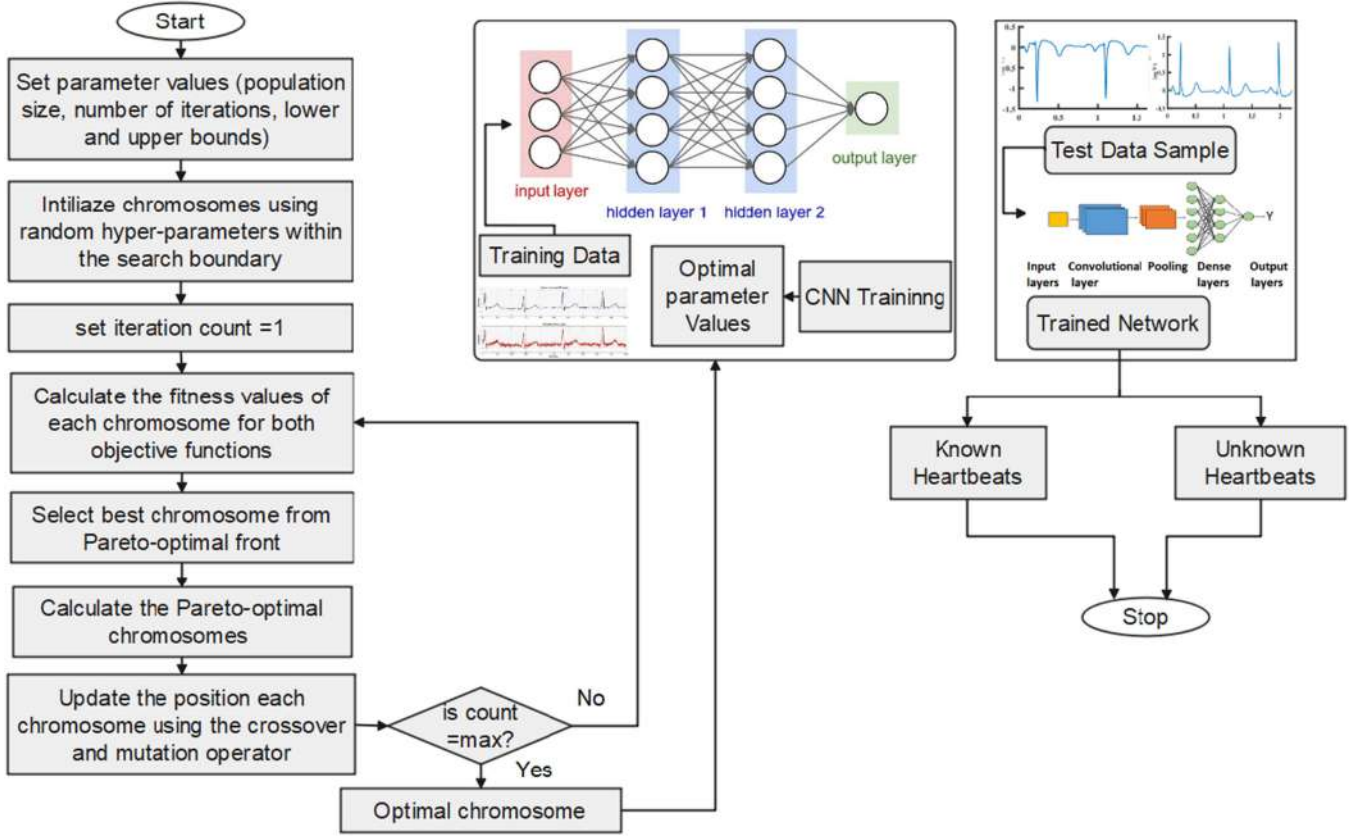
Fig. 2.   Flow diagram of N2CNN-based Anomaly Detection Method.

is known to give better results. However, an average operation can also be used as an alternative [39], [40].

$$f(x) = \max(x) \tag{3}$$

here, x defines the input given to this layer. Further, to reduce the sensitivity of network initialization and enhance the training speed, a batch normalization layer has been added. Then the fully connected layer is meant to help the model learn about the nonlinear parameters. This layer consists of neurons that are connected to every other neuron. The function for each neuron can be given as follows:

$$f(x) = \epsilon \left(\sum_{i}^{n} x_i \times w_i \right) + b \tag{4}$$

Here, ϵ denotes the nonlinear function, and i is the input number, input is defined by x, whereas weight and bias are represented by w and b respectively. The fully connected layer is followed by a normalized exponential function (also known as the SoftMax layer). This layer is used for categorical classification. It creates a probabilistic distribution over n different possible outcomes. The function can be given by Eq. (5).

$$f_{softmax}(x) = \frac{e^x}{\sum_{i}^{n} e^{x_i}} \tag{5}$$

Finally, the desired class for the given input is generated by the class out layer.

B. Nsga-Ii

Elitist non-dominated sorting genetic algorithm (NSGA-II) [41] is an enhanced and fast version of NSGA algorithm. It is one of the popular multi-objective optimization algorithm. Mathematically, a multi-objective optimization problem are defined as follows:

$$Minimize/Maximize \quad fun_m(X), \tag{6}$$

where $m = 1, 2, \ldots, M$. Whereas:

$$e_j(X) \geq 0, \tag{7}$$
$$ie_k(X) = 0, \tag{8}$$
$$X = (x_1, x_2, \ldots, x_i, \ldots, x_n)^T \tag{9}$$

where, $j = 1, 2, \ldots, J$ denotes inequality constraints, $k = 1, 2, \ldots, and K$ denotes equality constraints, $X^{lower} \leq X_i \leq X^{Upper}$ (search bounds),and $i = 1, 2, \ldots, n$. X^{lower} and X^{upper} are the minimum and maximum dimension values respectively. NSGA-II is applied to the problems formulated as defined above and find the Pareto-optimal solutions which are considered as the best solutions considering all the objective functions. The key features of NSGA-II are as follows:

- Elite preserving property: The best solution of current population is carry forwarded to the next iteration. This may sometimes helps the algorithm to escape from local optima.

- Maintaining the diversity among the Pareto-optimal fronts by the use of crowding distance.
- More focus on the non-dominated solutions in each generation.

The step by step process of NSGA-II is defined as given below:

1) Let P_t and Q_t denotes the child and offspring populations (having N solutions each) and their combination is called R_t which can be defined as $R_t = P_t \cup Q_t$ (having 2 N solutions).
2) Apply the non-domination sorting [41] on R_t and ensure the elite solutions.
3) The best solutions are arranged into different fronts as per their ranks. The best solutions are resides in Front F_1.
4) Now, to formulate new population P_{t+1}, all solutions resides in F_1 are selected. If $F_1 < N$, then the solutions from F_2 fronts will be chosen and so on.
5) From selecting the best solution from the last front, crowding distance sorting is applied within that front and less crowded solution will be picked to maintain the diversity in the population.
6) Once P_{t+1} is formulated, all the GA operators can be applied on it.

The above set of steps are applied for the predefined number of iterations or till the sufficient error precision is not achieved. Finally, the resultant solutions are called the Pareto-optimal solutions. Out of these solutions, one solution is selected as per application need.

C. NSGA-II for CNN Hyper-Parameter Optimization

The performance of CNN can be improved by optimizing its hyper-parameters. In this paper, two hyper-parameters of CNN have been optimized namely, number of layers and filter size. To perform optimization, Elitist Non-dominated sorting genetic algorithm (NSGA-II) [41], a popular multi-objective optimization, has been employed in the proposed method. This benefits over single objective optimization as it is suitable for optimization with multiple objectives [3]. We have considered three objectives for the optimization process, namely accuracy(acc), recall (rec), and precision (prc) which are defined by Eq. (10), (11), and (12) respectively.

$$Accuracy(Acc) = \frac{\sum_{i=1}^{k} \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i}}{k} \tag{10}$$

$$Recall(rec) = \frac{\sum_{i=1}^{k} TP_i}{\sum_{i=1}^{k}(TP_i + FN_i)} \tag{11}$$

$$Precision(prc) = \frac{\sum_{i=1}^{k} TP_i}{\sum_{i=1}^{k}(TP_i + FP_i)} \tag{12}$$

Where, i represent any class and there are total k classes, TP_i represents the true positive samples or number of hits for any class i, TN_i, FP_i, and FN_i denote the true negative, false positive, false negative samples for respective class i.

Let us assume there is a population of N solutions. Each solution consists of x hyper-parameters generated randomly between upper bound (x_i^U) and lower bound (x_i^L). Let there are k objective functions (in this paper, $k = 3$), then vector Z of objective functions can be represented as Eq. (13).

$$Z(x) = (Z_1(x), Z_2(x), \ldots, Z_k(x)) \tag{13}$$

subject to j constraint functions which are denoted by vector f in Eq. (14).

$$f(x) = (f_1(x), f_2(x), \ldots, f_j(x)) \tag{14}$$

Fig. 2 presents the overall workflow of the proposed method. The inputs to each layer are represented by in, and outputs are represented by out. Each generation, having a population P_t is composed of chromosomes that are generated using mutation and crossover of hyper-parameters. Once training and testing is completed for the current population, the next population $(P_t + 1)$ is created using Pareto fronts and crowding distance in objective space with three objectives Acc, rec, and prc. The following steps can be used to summarize the implemented technique:

Step 1: A population P_0 containing N generated randomly chromosomes.

Step 2: Each chromosome represents the hyper-parameters of CNN networks. This CNN has a fixed input and output layer by default regardless of the structure decided by the genetic algorithm (GA). The GA has to select the number of layers between fixed layers, kernel size, layer type, pooling size, stride, and count of neurons of a fully connected layer.

Step 3: After the hyper-parameters have been generated, the weights of the network are trained by ADAM [42] using two-fold approach and applied on the considered dataset. To use as objective parameters, the values of Acc, rec, and pre are averaged.

Step 4: The parent population is sorted by the objective parameters as a non-domination sorting, where $P = P_0$.

Step 5: Further, N new offspring population are created using crossover and mutation [43].

Step 6: Combined population, R_t (equals 2 N) consists of Q_t and P_t.

Step 7: The combined population is further sorted by fast non-dominated sorting, and solutions are organized into different fronts (Fr_i).

Step 8: If $\mid Fr_i \mid \leq N$, then include all solutions of ith front to the next generation else go to step 10.

Step 9: $i = i + 1$ and go to step 8.

Step 10: Define the crowding distance as illustrated by Deb *et al.* [41].

Step 11: A crowded comparison operator is used in case of population size being greater than N. This operator is leveraged in decreasing order to develop the population size of size N.

Step 12: Repeat the process from step 5 until termination conditions are fulfilled (i.e., 50 generations).

TABLE I
DISTRIBUTION OF HEARTBEATS AMONG FIVE CLASSES

	MIT-BIH Dataset	INCART dataset
Class	Total samples	Total samples
N	90,589	1,53,618
S	2,779	1,959
V	7,236	20,041
F	803	219
Q	8,039	7
Number of heartbeats	1,01,407	1,75,844

TABLE II
DISTRIBUTION OF HEARTBEATS AMONG FIVE CLASSES

	MIT-BIH Dataset	INCART dataset
Class	Total samples	Total samples
Known heartbeats	101407	175837
Unknown heartbeats	8039	7
Number of heartbeats	109446	175844

TABLE III
ABLATION STUDY OF N2CNN ON MIT-BIH DATASET

CNN	Accuracy	Precision	Recall	F1-Score
With NSGA-II	94.70	79.72	85.42	82.26
W/o NSGA-II	89.02	65.24	77.58	75.36

TABLE IV
ABLATION STUDY OF ADAM ALGORITHM IN N2CNN ON MIT-BIH DATASET

N2CNN	Accuracy	Precision	Recall	F1-Score
With ADAM	94.70	79.72	85.42	82.26
With GD	91.86	76.40	81.9	77.92

IV. CONSIDERED DATASET

1) **MIT-BIH arrhythmia database:** The MIT-BIH database [44] is a freely available for public access, which has been widely used in comparable studies. Two ECG recording channels recorded at 360 Hz by 47 human male and female volunteers ranging in age from 23 to 89 years are included in this collection. Each set has 22 recordings which is separated into training (DS1) and testing (DS2) parts. Two cardiologists classified each heartbeat into one of 19 categories.

2) **INCART dataset:** INCART is a popular dataset that is also available for free on the Physionet website [45]. There are 75 annotated recordings in this collection, which have been derived from 32 Holter recordings. Each record lasts 30 minutes and features 12 standard leads. Every lead is sampled at a rate of 257 Hz, with gains ranging from 250 to 1,100 ADC units per millivolt. There are over 175,000 beat annotations in the reference annotation files. Each heartbeat is divided into 19 categories. The class annotations of the dataset have been generated by an automated algorithm.

Initially, ECG heartbeats were classified into 19 groups. However, AAMI reclassified these categories into five primary groups. As per the AAMI standard, ECG heartbeats are divided in 5 classes namely, unknown (Q), atrial premature and supraventricular ectopic (S), fusion of ventricular (F), normal (N), premature ventricular contraction and ventricular ectopic (V). Furthermore, Table I also indicates the count of heartbeats in each of the five classes (N, S, V, F, and Q) for the considered datasets. Out of five heartbeat categories, N, S, V, and F belong to the known heartbeat, while Q belongs to the unknown heartbeat. Therefore, to vindicate the efficacy of the proposed N2CNN based method, heartbeats are grouped into two categories, namely known beats and unknown beats as shown in Table II. The known type comprises heartbeats of N, S, V, and F classes, and the unknown type contains the heartbeat of the Q class. It is clearly visible in the table that the class distribution of the datasets are imbalanced. For a fair analysis, modified synthetic minority oversampling technique (MSMOTE) has been used to enhance the frequency of the minority class. To test the efficacy of the proposed approach, 70% of randomly chosen data samples has been employed for training and the rest data is utilized for testing.

V. EVALUATION METRICS

The most popular parameter for evaluating the performance and efficacy of the ECG classification approach is accuracy. However, the heartbeat categories are very unbalanced; accuracy alone cannot be enough to investigate the efficacy of the proposed N2CNN. Therefore, to assess the efficacy of the proposed approach, four fundamental metrics, namely F1-score, precision, and recall, are also utilized along with accuracy. Further, to make a uniform environment for comparing results, the AAMI standard recommends that the value of F1-score, precision, and recall should be high in successful ECG heartbeat classifications.

VI. ABLATION STUDY OF THE PROPOSED METHOD

In this paper, ablation study is conducted to analyze the effect of various components on the efficiency of the proposed method. This process is analogy to the brain ablation procedure [46] and has been considered a viable approach to determine the robustness of the deep neural network [47]. To perform this study, MIT-BIH dataset has been considered according to the settings mentioned in Section IV.

Table III depicts the quantitative verification of the optimization of the proposed model with NSGA-II in terms of the considered evaluation metrics. In the table, the proposed model with NSGA-II has performed much better than without it which clearly validates the performance advantages of optimizing the CNN with NSGA-II. Further, the functional verification of the ADAM algorithm in the proposed model (N2CNN) is explored in Table IV. For comparison, gradient descent (GD) algorithm has been considered. It is clearly visible that the use of ADAM algorithm has resulted in efficient performance, especially in terms of accuracy.

VII. EXPERIMENTAL ANALYSIS

A. Confusion Matrix

In this work, four standard multi-objective optimization methods other than proposed NSGA-II based CNN (N2CNN)

TABLE V
CONFUSION MATRICES RETURNED BY ALL COSIDERED METHODS FOR MIT-BIH DATASET AND INCART DATASET

Methods		MIT-BIH dataset		INCART dataset	
	Classes	KH	UH	KH	UH
N2CNN	KH	97654	2046	166572	2
	UH	3753	5993	9265	5
MOGA-CNN	KH	95321	3103	159224	4
	UH	6086	4936	16613	3
MOPSO-CNN	KH	94897	3257	158132	3
	UH	6510	4782	17705	4
MOGWO-CNN	KH	94925	3188	158673	3
	UH	6482	4851	17164	4
MOWOA-CNN	KH	94879	3136	158714	4
	UH	6528	4903	17123	3

TABLE VI
PERFORMANCE ANALYSIS WITH EXISTING MULTI-OBJECTIVE OPTIMIZATION METHODS ON MIT-BIH DATASET

Method	Accuracy	Precision	Recall	F1-measure
MOGA-CNN	91.60	70.82	77.70	73.60
MOPSO-CNN	91.08	69.52	76.53	72.29
MOGWO-CNN	91.16	69.78	76.98	72.62
MOWOA-CNN	91.17	69.85	77.28	72.76
N2CNN	**94.70**	79.72	85.42	82.26

TABLE VII
PERFORMANCE ANALYSIS WITH EXISTING MULTI-OBJECTIVE OPTIMIZATION METHODS ON INCART DATASET

Method	Accuracy	Precision	Recall	F1-measure
MOGA-CNN	90.55	50.01	66.70	47.54
MOPSO-CNN	89.93	50.01	73.54	47.37
MOGWO-CNN	90.24	50.01	73.69	47.46
MOWOA-CNN	90.26	50.01	66.56	47.46
N2CNN	**94.73**	50.03	83.08	48.70

Fig. 3. Radar Chart for MIT-BIH dataset.

Fig. 4. Radar Chart for INCART dataset.

method, are used in the anomaly detection. These are Multi-objective Genetic Algorithm (MOGA) [48], Multi-objective particle swarm optimization (MOPSO) [49], Multi-objective Grey-wolf Optimization (MOGWO) [50], and Multi-objective whale optimization algorithm (MOWOA) [51]. The confusion matrices are generated to analyze the classification performance of proposed N2CNN methods against other considered methods and depicted in Table V. For MIT-BIH dataset, out of 109446 heartbeat samples, 97654 are correctly identified as known heartbeat (KH). Further, 3753 samples are false positives which are considered as the anomalies. Moreover, for INCART datset, 166572 samples out of 175844 samples are correctly identified as the known heartbeats by the proposed N2CNN method. The values of false positives, true positives, true negatives, and false negatives for all the considered methods are shown in Table V. The false positive cases are considered as detected anomalies. It can be envisioned from the table that the N2CNN based approach shows better results for all the considered performance parameters.

B. Performance Evaluation

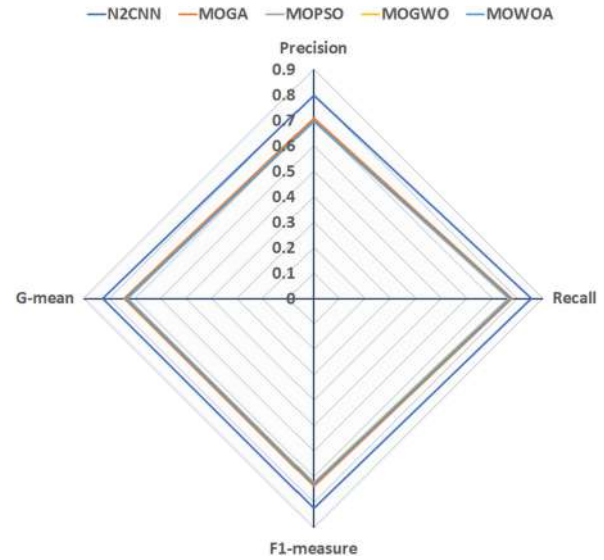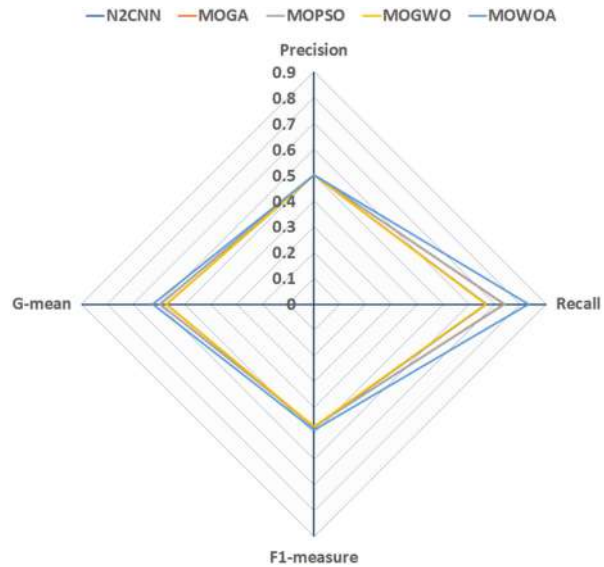The deep learning networks are trained over ECG signals. From the overall dataset, 70% signals are leveraged for training and rest are used for validation. The performance is measured in terms of and F1-measure, precision and recall and the values for these parameters are provide in Table VI and VII for the considered multi-objective optimization methods for MIT-BIH and INCAR datasets respectively. The values of these parameters indicates that the NSGA-II optimized CNN provides the best accuracy on both of the considered datasets. Moreover, Figs. 3 and 4 show the radar charts to validate the efficacy of N2CNN against other methods for BIT-MIH and INCART datasets respectively. The proposed N2CNN method covers the larger area in the radar chart for both the datasets which clearly depicts its efficacy over other considered methods. Therefore, the proposed network can be reliably used to detect the anomaly in real-time applications.

TABLE VIII
COMPARISON OF THE PROPOSED APPROACH WITH STATE-OF-THE-ART TECHNIQUES OVER MIT-BIH DATASET

Authors	Methods	Accuracy (%)	F1 score (%)	Precision (%)	Recall (%)
Martis et al. [52], 2014	DWT and SVM	94.52	——	——	——
Li et al. [53], 2016	WPE and random forest	94.61	——	——	——
Garcia et al. [8], 2017	SVM and swarm optimization (PSO)	92.40	——	——	81.10
M. Kachuee et al. [20], 2018	Deep residual CNN	93.40	——	——	——
Ramirez et al. [54], 2019	Hybrid neural network	93.80	——	——	——
Marinho et al. [55], 2019	NB, SVM, MLP, and OPF	94.30	——	——	——
Bidias et al. [56], 2020	Conditional entropy of ordinal patterns	93.62	——	——	——
Proposed approach, 2021	N2CNN	94.70	94.77	94.70	94.70

TABLE IX
COMPARISON OF THE PROPOSED APPROACH WITH STATE-OF-THE-ART TECHNIQUES OVER INCART DATASET

Authors	Methods	Accuracy (%)	F1 score (%)	Precision (%)	Recall (%)
Chazul et al. [57], 2004	Heartbeat Interval Features	84.70	77.05	77.05	77.05
Zhang et al [58], 2014	Disease-specific feature and SVM	86.24	79.36	79.36	79.36
Huang et al. [10], 2014	Random projections and RR intervals	87.43	81.14	81.14	81.14
Chu et al. [59], 2019	CNN and LSTM	88.56	81.17	82.80	82.84
Proposed Method, 2021	N2CNN	94.73	94.74	94.73	94.73
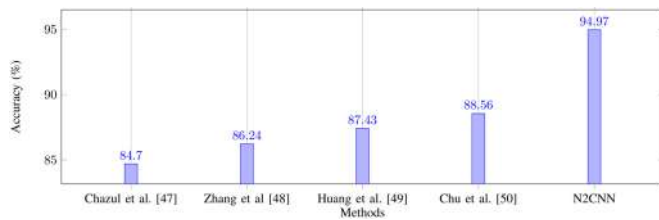
Fig. 5. Comparison of accuracy of the proposed approach and other state-of-the-arts over INCART dataset.

C. Comparative Analysis With State-of-The-Art Methods

The proposed N2CNN has also been validated against four other state-of-the-art approaches as shown in Table VIII and IX. In Table VIII, results of the N2CNN and other methods over the MIT-BIH dataset are tabulated, while in Table IX, performance metrics of the proposed approach and state-of-the-art methods over INCART dataset has been presented.

From the results of Table VIII and IX, It can be observed that the N2CNN outperforms other state-of-the-art approaches in terms of accuracy, precision, and F1 score. The F1-score is a useful performance metric, especially when dealing with unbalanced data. It can be seen from the above analysis that the proposed N2CNN surpassed the existing state-of-the-art approaches for the considered metrics, including the F1 score, which confirm the efficacy of the proposed N2CNN based approach. Furthermore, for better visualization, the accuracy of the proposed and seven other considered methods for both datasets is also depicted in Fig. 5 and 6, respectively. It is visible from the figure that the N2CNN has clearly surpassed the other methods on both datasets. Thus, it can be established from the experimental results that the proposed N2CNN based approach is superior to seven other state-of-the-art methods in terms of accuracy.

VIII. CONCLUSION

This paper introduces a new NSGA-II based convolution neural network (N2CNN) for anomaly monitoring in remote
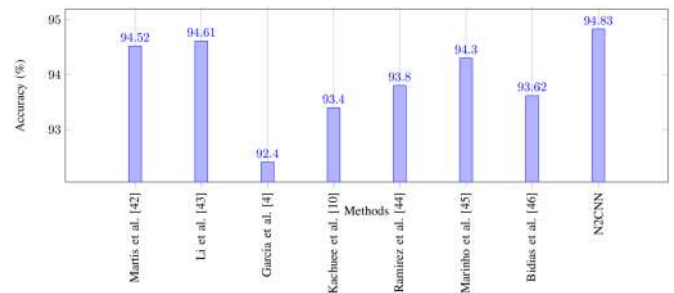

Fig. 6. Comparison of accuracy of the proposed approach and other state-of-the-arts over MIT-BIH dataset.

patient. The proposed method utilizes the strength of multi-objective NSGA-II for hyperparameters optimization of CNN layers. The classification accuracy of the proposed method is tested on two public datasets, namely, INCART, and MIT-BIH. The experimental results witnessed that the proposed N2CNN surpassed the seven other state-of-the-art methods on both datasets. The proposed method has achieved 94.70% accuracy in the MIT-BIH dataset, whereas 94.73% accuracy in INCART dataset, whereas, runner method has given 91.60% and 90.55% accuracy on both the datasets respectively. Thus, it can be elucidated that the proposed N2CNN model can be efficiently utilized to monitor patients' health conditions in a real-time environment. The employment of NSGA-II on CNN has optimized the network in terms of number of layers and filter size. This has resulted in better performance of proposed N2CNN. Moreover, the proposed N2CNN method can detect the any tempering or anomaly in the cardiovascular data of the patients. Thus, the security of the remote patient data can be efficiently handled used the proposed N2CNN based method. In the future, the proposed method can be employed on massive real-time datasets using Big-data methods. Moreover, new wearable sensors may be used for creating data to ensure the health of patients.

REFERENCES

[1] F. J. Dian, R. Vahidnia, and A. Rahmati, "Wearables and the Internet of Things (IoT), applications, opportunities, and challenges: A survey," *IEEE Access*, vol. 8, pp. 69 200–69 211, 2020.

[2] H. Mittal and M. Saraswat, "An optimum multi-level image thresholding segmentation using non-local means 2D histogram and exponential kbest gravitational search algorithm," *Eng. Appl. Artif. Intell.*, vol. 71, pp. 226–235, 2018.

[3] S. S. Mostafa, N. Horta, A. G. Ravelo-Garcia, and F. Morgado-Dias, "Analog active filter design using a multi objective genetic algorithm," *AEU- Int. J. Electron. Commun.*, vol. 93, pp. 83–94, 2018.

[4] E. J. D. S. Luz, W. R. Schwartz, G. Cámara-Chávez, and D. Menotti, "ECG-based heartbeat classification for arrhythmia detection: A survey," *Comput. Methods Programs Biomed.*, vol. 127, pp. 144–164, 2016.

[5] C. M. Bishop and N. M. Nasrabadi, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.

[6] D. Bhamare and P. Suryawanshi, "Review on reliable pattern recognition with machine learning techniques," *Fuzzy Inf. Eng.*, vol. 10, no. 3, pp. 362–377, 2018.

[7] K. Park *et al.*, "Hierarchical support vector machine based heartbeat classification using higher order statistics and hermite basis function," in *Proc. Comput. Cardiol.*, 2008, pp. 229–232.

[8] G. Garcia, G. Moreira, D. Menotti, and E. Luz, "Inter-patient ECG heartbeat classification with temporal VCG optimized by PSO," *Sci. Rep.*, vol. 7, no. 1, pp. 1–11, 2017.

[9] G. D. Lannoy, D. François, J. Delbeke, and M. Verleysen, "Weighted SVMs and feature relevance assessment in supervised heart beat classification," in *Proc. Int. Joint Conf. Biomed. Eng. Syst. Technol.*, 2010, pp. 212–223.

[10] H. Huang, J. Liu, Q. Zhu, R. Wang, and G. Hu, "A new hierarchical method for inter-patient heartbeat classification using random projections and RR intervals," *Biomed. Eng. Online*, vol. 13, no. 1, pp. 1–26, 2014.

[11] N. Japkowicz and S. Stephen, "The class imbalance problem: A systematic study," *Intell. Data Anal.*, vol. 6, no. 5, pp. 429–449, 2002.

[12] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, 2002.

[13] H. Zhang, L. Huang, C. Q. Wu, and Z. Li, "An effective convolutional neural network based on SMOTE and Gaussian mixture model for intrusion detection in imbalanced dataset," *Comput. Netw.*, vol. 177, 2020, Art. no. 107315.

[14] G. Sannino and G. De Pietro, "A deep learning approach for ECG-based heartbeat classification for arrhythmia detection," *Future Gener. Comput. Syst.*, vol. 86, pp. 446–455, 2018.

[15] S. K. Pandey and R. R. Janghel, "Classification of electrocardiogram signal using an ensemble of deep learning models," *Data Technol. Appl.*, vol. 55 no. 3, pp. 446–460, 2021.

[16] M. Hammad, A. Maher, K. Wang, F. Jiang, and M. Amrani, "Detection of abnormal heart conditions based on characteristics of ECG signals," *Measurement*, vol. 125, pp. 634–644, 2018.

[17] U. R. Acharya *et al.*, "A deep convolutional neural network model to classify heartbeats," *Comput. Biol. Med.*, vol. 89, pp. 389–396, 2017.

[18] M. Amrani, M. Hammad, F. Jiang, K. Wang, and A. Amrani, "Very deep feature extraction and fusion for arrhythmias detection," *Neural Comput. Appl.*, vol. 30, no. 7, pp. 2047–2057, 2018.

[19] S. Mousavi and F. Afghah, "Inter-and intra-patient ECG heartbeat classification for arrhythmia detection: A sequence to sequence deep learning approach," in *Proc. ICASSP IEEE Int. Conf. Acoust., Speech Signal Process.*, 2019, pp. 1308–1312.

[20] M. Kachuee, S. Fazeli, and M. Sarrafzadeh, "Ecg heartbeat classification: A deep transferable representation," in *Proc. IEEE Int. Conf. Healthcare Informat.*, 2018, pp. 443–444.

[21] H. Wang *et al.*, "A high-precision arrhythmia classification method based on dual fully connected neural network," *Biomed. Signal Process. Control*, vol. 58, 2020, Art. no. 101874.

[22] Z. Ebrahimi, M. Loni, M. Daneshtalab, and A. Gharehbaghi, "A review on deep learning methods for ecg arrhythmia classification," *Expert Syst. With Appl.*, vol. 7, 2020, Art. no. 100033.

[23] T. Goel, R. Murugan, S. Mirjalili, and D. K. Chakrabartty, "Optconet: An optimized convolutional neural network for an automatic diagnosis of COVID-19," *Appl. Intell.*, vol. 51, no. 3, pp. 1351–1366, 2021.

[24] N. Zhang, Y.-X. Cai, Y. -Y. Wang, Y.-T. Tian, X.-L. Wang, and B. Badami, "Skin cancer diagnosis based on optimized convolutional neural network," *Artif. Intell. Med.*, vol. 102, 2020, Art. no. 101756.

[25] Y. Wang, H. Zhang, and G. Zhang, "cPSO-CNN: An efficient PSO-based algorithm for fine-tuning hyper-parameters of convolutional neural networks," *Swarm Evol. Comput.*, vol. 49, pp. 114–123, 2019.

[26] B. Navaneeth and M. Suchetha, "PSO optimized 1-D CNN-SVM architecture for real-time detection and classification applications," *Comput. Biol. Med.*, vol. 108, pp. 85–92, 2019.

[27] A. Azad, E. Aghaei, A. Jalali, and P. Ahmadi, "Multi-objective optimization of a solar chimney for power generation and water desalination using neural network," *Energy Convers. Manage.*, vol. 238, 2021, Art. no. 114152.

[28] J. Xie *et al.*, "Multi-objective feed rate optimization of three-axis rough milling based on artificial neural network," *Int. J. Adv. Manuf. Technol.*, vol. 114, no. 5, pp. 1323–1339, 2021.

[29] A. Bakar, K. Li, H. Liu, Z. Xu, M. Alessandrini, and D. Wen, "Multi-objective optimization of low reynolds number airfoil using convolutional neural network and non-dominated sorting genetic algorithm," *Aerospace*, vol. 9, no. 1, 2022, Art. no. 35.

[30] T. Goel, R. Murugan, S. Mirjalili, and D. K. Chakrabartty, "Multi-COVID-Net: Multi-objective optimized network for COVID-19 diagnosis from chest x-ray images," *Appl. Soft Comput.*, vol. 115, 2022, Art. no. 108250.

[31] S. Kanwal, I. Younas, and M. Bashir, "Evolving convolutional autoencoders using multi-objective particle swarm optimization," *Comput. Elect. Eng.*, vol. 91, 2021, Art. no. 107108.

[32] D. Singh, V. Kumar, Vaishali and M. Kaur, "Classification of COVID-19 patients from chest CT images using multi-objective differential evolution–based convolutional neural networks," *Eur. J. Clin. Microbiol. Infect. Dis.*, vol. 39, no. 7, pp. 1379–1389, 2020.

[33] E. -G. Talbi, "Automated design of deep neural networks: A survey and unified taxonomy," *ACM Comput. Surv.*, vol. 54, no. 2, pp. 1–37, 2021.

[34] M. Feurer and F. Hutter, "Hyperparameter optimization," in *Automated Machine Learning*. Cham, Switzerland: Springer, 2019, pp. 3–33.

[35] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[36] J. Ren and L. Xu, "On vectorization of deep convolutional neural networks for vision tasks," *Proc. AAAI Conf. Artif. Intell.*, vol. 29, no. 1, pp. 1840–1846, 2015.

[37] D. Stutz, "Understanding convolutional neural networks," Seminar Rep., Vis. Comput. Inst., RWTH Aachen Univ., pp. 1–23, 2014.

[38] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. Int. Conf. Mach. Learn.*, 2010, pp. 807–814.

[39] D. Scherer, A. Müller, and S. Behnke, "Evaluation of pooling operations in convolutional architectures for object recognition," in *Proc. Int. Conf. Artif. Neural Netw.*, 2010, pp. 92–101.

[40] H. Mittal and M. Saraswat, "A new fuzzy cluster validity index for hyperellipsoid or hyperspherical shape close clusters with distant centroids," *IEEE Trans. Fuzzy Syst.*, vol. 29, no. 11, pp. 3249–3258, Nov. 2021.

[41] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Trans. Evol. Comput.*, vol. 6, no. 2, pp. 182–197, Apr. 2002.

[42] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Representations*, 2015, pp. 1–15.

[43] K. Deb, "Multi-objective optimisation using evolutionary algorithms: An introduction," in *Multi-Objective Evolutionary Optimisation for Product Design and Manufacturing*. London, U.K.: Springer, 2011, pp. 3–34.

[44] G. B. Moody and R. G. Mark, "The MIT-BIH arrhythmia database on CD-ROM and software for use with it," in *Proc. IEEE Comput. Cardiol.*, 1990, pp. 185–188.

[45] A. L. Goldberger *et al.*, "Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[46] J. Wolf, "Speech recognition: Invited papers presented at the 1974 IEEE symposium," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 25, no. 2, pp. 207–207, Apr. 1977.

[47] R. Meyes, M. Lu, C. W. de Puiseau, and T. Meisen, "Ablation studies in artificial neural networks," 2019, *arXiv:1901.08644*.

[48] R. Pal, A. Malik, S. Yadav, and R. Karnwal, "Optimal cluster head election in industrial WSNs using the multi-objective genetic algorithm," in *Applications of Advanced Optimization Techniques in Industrial Engineering*. Boca Raton, FL, USA: CRC, 2022, pp. 169–180.

[49] C. C. Coello and M. S. Lechuga, "Mopso: A. proposal for multiple objective particle swarm optimization," in *Proc. IEEE Congr. Evol. Comput. (Cat. No 02TH8600)*, 2002, vol. 2, pp. 1051–1056.

[50] S. Mirjalili, S. Saremi, S. M. Mirjalili, and L. D. S. Coelho, "Multi-objective grey wolf optimizer: A novel algorithm for multi-criterion optimization," *Expert Syst. With Appl.*, vol. 47, pp. 106–119, 2016.

[51] J. Wang, P. Du, T. Niu, and W. Yang, "A novel hybrid system based on a new proposed algorithm–multi-objective whale optimization algorithm for wind speed forecasting," *Appl. Energy*, vol. 208, pp. 344–360, 2017.

[52] R. J. Martis, U. R. Acharya, C. M. Lim, K. Mandana, A. K. Ray, and C. Chakraborty, "Application of higher order cumulant features for cardiac health diagnosis using ECG signals," *Int. J. neural Syst.*, vol. 23, no. 04, 2013, Art. no. 1350014.

[53] T. Li and M. Zhou, "ECG classification using wavelet packet entropy and random forests," *Entropy*, vol. 18, no. 8, pp. 1–16, 2016.

[54] E. Ramirez, P. Melin, and G. Prado-Arechiga, "Hybrid model based on neural networks, type-1 and type-2 fuzzy systems for 2-lead cardiac arrhythmia classification," *Expert Syst. With Appl.*, vol. 126, pp. 295–307, 2019.

[55] L. B. Marinho, N. de MM Nascimento, J. W. M. Souza, M. V. Gurgel, P. P. Reboucas Filho, and V. H. C. de Albuquerque, "A novel electrocardiogram feature extraction approach for cardiac arrhythmia classification," *Future Gener. Comput. Syst.*, vol. 97, pp. 564–577, 2019.

[56] J. B. B.à Mougoufan, J. A. E. Fouda, M. Tchuente, and W. Koepf, "Adaptive ECG beat classification by ordinal pattern based entropies," *Commun. Nonlinear Sci. Numer. Simul.*, vol. 84, 2020, Art. no. 105156.

[57] P. De Chazal, M. O'Dwyer, and R. B. Reilly, "Automatic classification of heartbeats using ECG morphology and heartbeat interval features," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 7, pp. 1196–1206, Jul. 2004.

[58] Z. Zhang, J. Dong, X. Luo, K.-S. Choi, and X. Wu, "Heartbeat classification using disease-specific feature selection," *Comput. Biol. Med.*, vol. 46, pp. 79–89, 2014.

[59] J. Chu, H. Wang, and W. Lu, "A novel two-lead arrhythmia classification system based on CNN and LSTM," *J. Mechanics Med. Biol.*, vol. 19, no. 3, 2019, Art. no. 1950004.

Avinash Chandra Pandey (Member, IEEE) is currently an assistant professor with the Department of Computer Science & Engineering of Indian Institute of Information Technology Design and Manufacturing, Jabalpur, India. He has more than eight years of teaching and research experience. He has guided many M.Tech. dissertations and B.Tech. projects. He has authored or coauthored more than 20 journal and conference papers in the area of text mining, data mining, and soft computing. His research interests include data analytics, text mining, and soft computing.

Mohammad Ayoub Khan (Senior Member, IEEE) received the Master of Technology degree in computer science and engineering from Guru Gobind Singh Indraprastha, New Delhi, India, and the Ph.D. degree in computer engineering from Jamia Millia Islamia, New Delhi, India. He is currrentlly an associate professor with the University of Bisha, Saudi Arabia, with interests in the Internet of Things, RFID, wireless sensor networks, adhoc networks, smart cities, industrial IoT, signal processing, NFC, routing in network-on-chip, real time and embedded systems. He has more than 14 years of experience in his research areas. He has authored or coauthored many research papers and books in reputed journals and international IEEE conferences. He contributes to the research community by undertaking various volunteer activities in the capacity of Editor for many journals and as a Conference Chair.

Raju Pal received the Ph.D. degree in computer science & engineering from the (Jaypee Institute of Information Technology), JIIT, Noida, India. He is currently an assistant professor with JIIT. He has more than eight years of teaching and research experience. His research interests inlcude medical image analysis, meta-heuristic algorithms, and wireless sensor networks. He has also been the co-Principal Investigator of the research project on histopathological image analysis funded by the SERB, Department of Science and Technology, India. He was the Program Committee Member of various international conferences and reviewer of many international journals of good repute.

Varun G. Menon received the M.Tech. degree in computer and communication (with University first rank), the Diploma in training and development, the M.Sc. degree in applied psychology, the MBA degree in human resource management, and the Ph.D. degree in computer science and engineering. He is currently a Professor and the Head of the Department of Computer Science Engineering, SCMS School of Engineering and Technology, Palissery, India. He is currently a Distinguished Speaker of the Association of Computing Machinery (ACM). His research interests include sensor technologies, Internet of Things, Green IoT, wireless communication, fog computing and networking.

Ashish Kumar Tripathi (Member, IEEE) received the M.Tech. and Ph.D. degrees from the Department of Computer Science and Engineering, Delhi Technological University, Delhi, India, in 2013, and 2019 respectively. He is currently an assistant professor with the Department of Computer Science & Engineering, Malaviya National Institute of Technology, Jaipur, India. His research interests include Big Data analytics, social media analytics, Internet of Things, vedio and Image data processing.

Himanshu Mittal received the Ph.D. degree in the field of computer vision under the supervision of Dr. Mukesh Saraswat. He is currently an assistant professor with the Computer Science Department of Jaypee Institute of Information Technology (JIIT), Noida, India. His research interests include image analysis, machine learning, and evolutionary algorithms. He has an excellent academic record and also research background with the papers in reputed journals including IEEE Transa'ctions. He has also been the co-Principal investigator of the research project funded by the Science and Engineering Research Board, Department of Science and Technology, India.

SPRINGER LINK

Menu   Search   Cart

A novel fuzzy clustering-based method for human activity recognition in cloud-based industrial IoT environment

Himanshu Mittal, Ashish Kumar Tripathi, Avinash Chandra Pandey, P. Venu, Varun G. Menon & Raju Pal ✉

Wireless Networks (2022)

Abstract

With the advancement of technology such as video monitoring, Internet-of-things, cloud, and machine learning, Industry 4.0 is working continuously to ensure the security of workers. The workers are equipped with sensors to analyze their activities. In general, the recognition of human activities in cloud-based industrial scenario is leveraged to monitor the safety of the workers. This paper introduced a new optimal clustering method for the activity recognition of workers in industry using cloud based IoT environment. The proposed method uses the temporal and spatial features of human workers in industry. The proposed method

is tested on publicly available dataset of different activities maintained into three groups, namely movement, gestures, and object handling, in the context of the medium and small industrial environment. The experimental findings validate that the proposed method achieves 80.2%, 81.05% and 80.19% of average accuracy for movement, gesture, and object handling activities, which clearly outperformed the fuzzy c-means, particle-swarm optimization, and HMM-based activity recognition methods.

This is a preview of subscription content, <u>access via your institution</u>.

Access options

Buy article PDF

39,95 €

Price includes VAT (India)

Instant access to the full article PDF.

<u>Rent this article via DeepDyve.</u>

<u>Learn more about Institutional subscriptions</u>

Data availability

The datasets analysed during the current study are available from the second author on reasonable request.

References

1. Abbasi, M., Tahouri, R., & Rafiee, M. (2019). Enhancing the performance of the aggregated bit vector algorithm in network packet classification using gpu. *PeerJ Computer Science, 5*, e185.

2. Abbasi, M., Najafi, A., Rafiee, M., et al. (2020). Efficient flow processing in 5g-envisioned sdn-based internet of vehicles using gpus. *IEEE Transactions on Intelligent Transportation Systems, 22*(8), 5283–5292.

3. Ahmed, I., Zhang, Y., & Jeon, G., et al. A blockchain-and artificial intelligence-enabled smart iot framework for sustainable city. *International Journal of Intelligent Systems*.

4. Chen, J., Sun, Y., & Sun, S. (2021). Improving human activity recognition performance by data fusion and feature engineering. *Sensors, 21*(3), 692.

5. Chen, K., Zhang, D., Yao, L., et al. (2021). Deep learning for sensor-based human activity recognition: Overview, challenges, and opportunities. *ACM Computing Surveys (CSUR), 54*(4), 1–40.

6. Chowdhury, A. K., Tjondronegoro, D., Chandran, V., et al. (2017). Physical activity recognition using posterior-adapted class-based fusion of multi-accelerometers data. *IEEE Journal of Biomedical and Health Informatics, 99*, 1–1.

7. Dallel, M., Havard, V., Baudry, D., et al. (2020). Inhard-industrial human action recognition dataset in the context of industrial collaborative robotics. In *2020 IEEE International Conference on Human-Machine Systems (ICHMS), IEEE* (pp. 1–6).

8. Dang, L. M., Min, K., Wang, H., et al. (2020). Sensor-based and vision-based human activity recognition: A comprehensive survey. *Pattern Recognition, 108*(107), 561.

9. Eltaeib, T., & Mahmood, A. (2018). Differential evolution: A survey and analysis. *Applied Sciences, 8*(10), 1945.

10. Hu, J., Pan, Y., Li, T., et al. (2020). Tw-co-mfc: Two-level weighted collaborative fuzzy clustering based on maximum entropy for multi-view data. *Tsinghua Science and Technology, 26*(2), 185–198.

11. Khosravi, M. R., & Samadi, S. (2019). Reliable data aggregation in internet of visar vehicles using chained dual-phase adaptive interpolation and data embedding. *IEEE Internet of Things Journal, 7*(4), 2603–2610.

12. Khosravi, M. R., & Samadi, S. (2021). Bl-alm: A blind scalable edge-guided reconstruction filter for smart environmental monitoring through green iomt-uav networks. *IEEE Transactions on Green Communications and Networking, 5*(2), 727–736.

13. Kilany, M., Hassanien, AE., & Badr, A. (2015). Accelerometer-based human activity classification using water wave optimization approach. In *2015 11th International Computer Engineering Conference (ICENCO), IEEE* (pp. 175–180).

14. Liao, X., Zheng, D., & Cao, X. (2021). Coronavirus pandemic analysis through tripartite graph clustering in online social

networks. *Big Data Mining and Analytics,* *4*(4), 242–251.

15. Liu, Y., Pei, A., Wang, F., et al. (2021). An attention-based category-aware gru model for the next poi recommendation. *International Journal of Intelligent Systems, 36*(7), 3174–3189.

16. Liu, Y., Li, D., Wan, S., et al. (2022). A long short-term memory-based model for greenhouse climate prediction. *International Journal of Intelligent Systems, 37*(1), 135–151.

17. Maitre, J., Bouchard, K., & Gaboury, S. (2021). Alternative deep learning architectures for feature-level fusion in human activity recognition. *Mobile Networks and Applications* 1–11.

18. Mittal, H., & Saraswat, M. (2019). An automatic nuclei segmentation method using intelligent gravitational search algorithm based superpixel clustering. *Swarm and Evolutionary Computation, 45*, 15–32.

19. Mittal, H., & Saraswat, M. (2020). A new fuzzy cluster validity index for hyper-ellipsoid or hyper-spherical shape close clusters with

distant centroids. *IEEE Transactions on Fuzzy Systems.*

20. Mittal, H., Pandey, A. C., Pal, R., et al. (2021). A new clustering method for the diagnosis of covid19 using medical images. *Applied Intelligence, 51*(5), 2988–3011.

21. Nandy, S., Adhikari, M., Khan, M. A., et al. (2021). An intrusion detection mechanism for secured iomt framework based on swarm-neural network. *IEEE Journal of Biomedical and Health Informatics.*

22. Nunes, U. M., Faria, D. R., & Peixoto, P. (2017). A human activity recognition framework using max-min features and key poses with differential evolution random forests classifier. *Pattern Recognition Letters, 99,* 21–31.

23. Nweke, H. F., Teh, Y. W., Al-Garadi, M. A., et al. (2018). Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges. *Expert Systems with Applications, 105,* 233–261.

24. Pal, R., Mittal, H., & Saraswat, M. (2019).
    Optimal fuzzy clustering by improved
    biogeography-based optimization for
    leukocytes segmentation. In *2019 Fifth
    International Conference on Image
    Information Processing (ICIIP), IEEE* (pp.
    74–79).

25. Pal, R., Saraswat, M., & Mittal, H. (2021).
    Improved bag-of-features using grey relational
    analysis for classification of histology images.
    Complex & Intelligent Systems, 7(3), 1429–
    1443.

26. Pandey, AC., Tripathi, AK., Pal, R., et al.
    (2019). Spiral salp swarm optimization
    algorithm. In *2019 4th International
    Conference on Information Systems and
    Computer Networks (ISCON), IEEE* (pp. 722–
    727).

27. Pant, M., Zaheer, H., Garcia-Hernandez, L., et
    al. (2020). Differential evolution: A review of
    more than two decades of research.
    *Engineering Applications of Artificial
    Intelligence, 90*(103), 479.

28. Price, K., Storn, R. M., & Lampinen, J. A.
    (2006). *Differential evolution: A practical
    approach to global optimization*. Springer.

29. Raju, P., Subash, Y., Rishabh, K., et al. (2020). Eewc: Energy-efficient weighted clustering method based on genetic algorithm for hwsns. *Complex & Intelligent Systems, 6*(2), 391–400.

30. Roitberg, A., Perzylo, A., Somani, N., et al. (2014). Human activity recognition in the context of industrial human-robot interaction. *Signal and Information Processing Association Annual Summit and Conference (APSIPA)* (pp. 1–10). IEEE: Asia-Pacific.

31. Ronao, C. A., & Cho, S. B. (2016). Human activity recognition with smartphone sensors using deep learning neural networks. *Expert Systems with Applications, 59,* 235–244.

32. Saraswat, M., Arya, K., & Sharma, H. (2013). Leukocyte segmentation in tissue images using differential evolution algorithm. *Swarm and Evolutionary Computation, 11*, 46–54.

33. Singh, G., & Singh, A. (2020). A hybrid algorithm using particle swarm optimization for solving transportation problem. *Neural Computing and Applications, 32*(15), 11699–11716.

34. Storn, R., & Price, K. (1997). Differential evolution-a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization, 11*(4), 341–359.

35. Tripathi, A. K., Sharma, K., & Bala, M. (2018). Dynamic frequency based parallel k-bat algorithm for massive data clustering (dfbpkba). *International Journal of System Assurance Engineering and Management, 9*(4), 866–874.

36. Tripathi, A. K., Sharma, K., & Bala, M. (2018). A novel clustering method using enhanced grey wolf optimizer and mapreduce. *Big Data Research, 14,* 93–100.

37. Tripathi, A. K., Sharma, K., & Bala, M. (2019). Parallel hybrid bbo search method for twitter sentiment analysis of large scale datasets using mapreduce. *International Journal of Information Security and Privacy (IJISP), 13*(3), 106–122.

38. Tripathi, A. K., Sharma, K., Bala, M., et al. (2020). A parallel military-dog-based algorithm for clustering big data in cognitive industrial internet of things. *IEEE Transactions on Industrial Informatics, 17*(3), 2134–2142.

39. Tripathi, A. K., Mittal, H., Saxena, P., et al. (2021). A new recommendation system using map-reduce-based tournament empowered whale optimization algorithm. *Complex & Intelligent Systems, 7*(1), 297–309.

40. Tu, P., Li, J., Wang, H., et al. (2021). Non-linear chaotic features-based human activity recognition. *Electronics, 10*(2), 111.

41. Weiss, G. M., & Lockhart, J. (2012). The impact of personalization on smartphone-based activity recognition. In *Workshops at the Twenty-Sixth AAAI Conference on Artificial Intelligence.*

42. Xue, Z., & Wang, H. (2021). Effective density-based clustering algorithms for incomplete data. *Big Data Mining and Analytics, 4*(3), 183–194.

43. Zappi, P., Lombriser, C., Stiefmeier, T., et al. (2008). Activity recognition from on-body sensors: accuracy-power trade-off by dynamic sensor selection. In *European Conference on Wireless Sensor Networks* (pp. 17–33). Springer.

44. Zdravevski, E., Lameski, P., Trajkovik, V., et al. (2017). Improving activity recognition accuracy in ambient-assisted living systems by automated feature engineering. *IEEE Access, 5,* 5262–5280.

45. Zhang, H., Babar, M., Tariq, M. U., et al. (2020). Safecity: Toward safe and secured data management design for iot-enabled smart city planning. *IEEE Access, 8,* 145256–145267.

46. Zheng, X., Wang, M., & Ordieres-Meré, J. (2018). Comparison of data preprocessing approaches for applying deep learning to human activity recognition in the context of industry 4.0. *Sensors, 18*(7), 2146.

Author information

Authors and Affiliations

Jaypee Institute of Information Technology, Noida, India

Himanshu Mittal & Raju Pal

Malaviya National Institute of Technology, Jaipur, India

Ashish Kumar Tripathi

Indian Institute of Information Technology, Design and Manufacturing, Jabalpur, India

Avinash Chandra Pandey

College of Engineering, University of Bahrain, Isa-Town, Bahrain

P. Venu

SCMS School of Engineering and Technology, Ernakulam, India

Varun G. Menon

Corresponding author

Correspondence to Raju Pal.

Additional information

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Rights and permissions

Reprints and Permissions

About this article

Cite this article

Mittal, H., Tripathi, A.K., Pandey, A.C. *et al.* A novel fuzzy clustering-based method for human activity recognition in

cloud-based industrial IoT environment. *Wireless Netw*

(2022). https://doi.org/10.1007/s11276-022-03011-y

AN OVERVIEW OF SOCIO - ECONOMIC CONDITIONS OF KUDIYA'S COMMUNITY AT KAMMADI COLONY OF KASARAGOD DISTRICT, KERALA

Akhil Baby

Assistant Professor, Department of Basic Science and Humanities, SCMS School of Engineering and Technology, Karukutty, Kerala, India

Abstract : Tribals, the original inhabitants of India have contributed much to the nation's culture, history and heritage. Under the constitutional provisions of Directive Principles, the States' major concern for tribes has been their welfare and development. It is ironical that despite a large number of well-meaning constitutional provisions aimed at protecting and safeguarding the welfare and interest of the tribal communities, the process of marginalization of the tribal's has gone on unabated. In this context, the paper has made an overview on the socio-economic conditions of tribals belonging to the Kudiya's (Malakkudiyas), which has the lowest standard of living, even though being the bulk of the tribal populace in Kerala.

IndexTerms - Socio economic conditions, Kudiya (Malakkudiyas) Original inhabitants

I. INTRODUCTION

The constitution India has provided many privileges to the scheduled tribes considering their complex problems in terms of geographical isolation, socio economic backwardness, distinctive culture, poor infrastructure facilities, language, religion and exploitation by various groups. Tribal population is aboriginal inhabitants of India who have been living a life based on natural environment and have cultural patterns congenial to their physical and social environment. The term "scheduled tribe" is primarily administrative and constitutional concept.

The Tribal groups in India have distinct cultural patterns. They form the indigenous group are also called 'upajati' in some regions of India. A large chunk of Indians belong to Tribal group (Tribal community) – Adivasis, as they are called, they occupy distinct regions, especially in the hilly and forested areas. Scheduled Tribes in India form the largest proportion of the total population in Lakshadweep and Mizoram followed by Nagaland and Meghalaya. Madhya Pradesh has the largest number of Scheduled Tribes followed by Bihar. Bastar district of Madhya Pradesh consists of largest number of Scheduled Tribes. There are no Scheduled Tribes in Punjab, Delhi, Chandigarh, Pondicherry, Haryana and Jammu and Kashmir. Tribal population in India constitutes 8.2% of Indian population according to census of India, 2001. India has the largest total tribal population compared to any other country in the world.

According to the Constitution (Scheduled Castes) Orders (Amendment) Act, 1990, Scheduled Castes can only belong to Hindu or Sikh or Buddhist religions. There is no religion bar in case of Scheduled Tribes. The Sachar Committee report of 2006 revealed that scheduled castes and tribes of India are not limited to the religion of Hinduism. The 61st round survey of the NSSO found that 90% of the Buddhists, one-third of the Sikhs, and one-third of the Christians in India belonged to the notified scheduled castes or tribes of the Constitution.

Tribal groups (communities) in Kerala

Tribes in Kerala (Adivasis of Kerala) are the indigenous population found in the southern Indian state of Kerala. Most of the tribal people of Kerala live in the forests and mountains of Western Ghats, bordering Karnataka and Tamil Nadu. According to 2001 census of India, the Scheduled Tribe population in Kerala is 3, 64,189. Wayanad has the highest number of tribals (1, 36,062). Idukki (50973) and Palakkad (39665) districts are the next two that make the loin portion of the native tribal people groups in the state. The Paniya (Paniyar) are the largest of the 35 major tribes. A total of 72.8% of tribal population is concentrated in six districts: Wayanad, Idukki, Palakkad, Kasaragod, Trivandrum and Kannur.

The list of Scheduled Trines in Kerala was revised on the basis of the Scheduled Castes and Scheduled Tribes Orders (Amendment) Act, 1976. This revised list was adopted in 1981, 1991 and 2001 census. In Kerala there are still 37 Scheduled Tribes out of 48 tribal communities.

Scheduled Tribe People-in Kasaragod

Kasaragod is one among the 14 districts in the state of Kerala and is considered as one of the most backward region. Tribes are also an integral part of Kasaragod; the district consisted of almost nine tribal communities. The socio-economic conditions of the tribes are very pathetic. The central as well as the state governments are implementing various policies for the upliftment of the tribes. But the condition of the tribes still remains as backward. Major proportion of the tribes inhabit in the backward or isolated areas of the districts.

Following are the scheduled tribes in the Kasargod district :

- Koragas
- Marattis
- Kudiya/Malakkudiyas
- Mavilan
- Malavettuvan
- Malayaraya

According to the census of 2011, the total population of scheduled tribes in Kasargod district is 48857. Of them 23950 are male and 24907 are female people.

The Koragas are one of the most primitive tribes, living in Kasaragod. The traditional occupation of these people is basket and scope making. They are financially, educationally and culturally backward. But the Maratis are educationally and culturally forward people than the other tribes. They were once transferred to the OEC category in 2003, but later in 2013 again they came under the scheduled Tribes category through strong protest from the part of Maratis. In relation to other tribes of Kasaragod a major part of Marati tribes are having well-accompanished jobs, even Government jobs. The Tribal Development Department is implementing many schemes for the development of these tribes in the district. Like the Scheduled Caste people, Scheduled Tribe people also have their own customs and rituals.

Functions of Tribal Development Office, Kasaragod

The Tribal Development Departments are formed by the government in order to protect various interests of scheduled tribes. Following are the various functions performed by the Tribal Development Office, Kasargod;

- ➢ Implementation of Economic Development Schemes and Welfare Schemes of Scheduled Tribes.
- ➢ Distribution of Educational concession to pre-matric and post-matric Scheduled Tribe students.
- ➢ Conduct medical awareness camps in the settlements.
- ➢ Conduct youth festivals and cultural programmes.

The study is about socio economic conditions of "Kudiya" tribe settled at Kammadi colony of Panathadi Panchayath. Kammadi colony is a one of the sole kudiya colony in Kerala. The Kudiya community also known as "Malakkudy" tribe. In Kerala according to census of 2011, their population can be estimated to be 785. They are live in Kasargode district and main occupation is agriculture. The rest of the tribe is found in Koorg (Kodak) the neighboring district of Karnataka. The Kudiya's mothertounge is 'Tulu' although children are learn both Kannada and Malayalam. The major economic resources for Kudiyas are forest and land. Hunting and gathering from the forest used to be their main stay along with labour work. After Kerala land reforms act passed many of them got their own land which they cultivated. The Kudiya women are active in agricultural activity and animal husbandary. They go for work at construction sites. The Kudiya profess the

Hindu faith and worship village regional deities besides under Parthean now. Festivals are of religious significance and the main one is Sivarathri.

Among the tribes of Kasaragod, we are exclusively focusing on Kudiya/Malakudiya tribes. They are also found in the South Karnataka. The Kudiya/Malakudiya tribes are the early migrants from Karnataka. Those who live in lowland are called Kudiya and highlanders are the Melakudi. Traditionally they were hunters and gatherers and in the course of time, they turned into marginal farmers. They subsist on rearing pigs, poultry, cattle and wok as agricultural labourers. They also make baskets. Those living in remote and isolated areas do not have adequate infrastructural facilities for their total development.

STATEMENT OF THE PROBLEM

The tribals, the original inhabitants of India have contributed much to the nation's culture, history and heritage. The principles entailed development along the lines of their own genius, respect of tribals rights in land and forest, training and building up a team of their own people to do the work of administration and development, to their social and cultural institutions. It is ironical that despite a large number of well-meaning constitutional provisions aimed at protecting and safeguarding the welfare and interest of the tribal communities, the process of marginalization of the tribals has gone on unabated. In this context the researcher has made an attempt to examine the socio-economic conditions of tribals belonging to the Kudiya's community with special reference to Kammadi colony of Kasaragod district.

SIGNIFICANCE OF THE STUDY

In India, as per census 2O11 population of SC and ST together forming quarter of the total population. Kerala is a homeland of number of tribal communities. Thirty six communities listed in the scheduled tribes list of the state. Of them "Kudiya/Malakudiya" community is seen only in Kasaragod district. Kudiya community is distributed in 7 gramapanchayats. They lives in remote and isolated areas and they don't have adequate infrastructural facilities. Kasaragod is a socially and economically backward district in Kerala. The place has been trying to escape from the evils of backwardness for many years. Even though the central and state governments have been providing various services for the upliftment of tribal people their problems are not fully solved. So it is important to analyses the socio-economic life of the various tribal communities in general and Kudiya/ Malakudiya in particular. The review of literature revealed a research gap, that is, no one studied Kudiyas of Kammady colony at Panathady. Since study on socio-economic conditions of Kudiya community is limited, it is necessary to identify their social welfare and economic conditions. This study has very much importance because the Kudiya community of Kammady colony did not get benefits from various schemes and projects proposed by the Government. So the study will help to understand the challenges and prospects of this tribe. It is a humple attempt at a minimum level to fill the research gap.

OBJECTIVES OF THE STUDY

Following are the objectives of this study:

1. To understand socio-economic conditions of "Kudiyas" at Kammady colony of Panathady Grama panchayath.
2. To understand challenges and prospects faced by this tribe.
3. To suggest measures to improve the socio-economic conditions of "Kudiyas".

METHODOLOGY OF STUDY

To carry out the research, a separate interview schedule carried out to collect primary data from the government officials. Questionnaire was also used to collect details from the beneficiaries of the targeted Municipality. The beneficiaries include the entire population of the beneficiaries of Kammadi colony of Kasaragod district, Kerala. Hence the sample size includes 41 respondents. Secondary sources such as journals, government reports, newspapers, documents from the government offices etc. were used. Primary data are collected from 41 families of Kammady colony which is large enough to give us enough information. Data are collected through well designed questionnaire which contains 70 questions which covers all areas of our study and which are capable of giving a clear picture about socio-economic status of Kudiyas settled at Kammady colony.

REVIEW OF LITERATURE

D.N.Majumdar, T.N.Madan (1970): It is about charecteristics of tribal economy. The socio economic structure of tribal communities marked different from that of the non- tribal or advanced groups of people. They have a very simple technology which fits well with their ecological surroundings and conservation outlook. Their economy said to be subsistence type. (George Dalton). There is absence of technological aids, based on barter and exchange, absence of profit motive in economic dealings, co-operative and collective endeavors, low or no progress of innovation, monopoly is absent, existence of regular market, absence of specialization, the nation of property is closely related to display and expenditure of wealth rather than to its accumulation.

L.P. Vidyarthi with Binay Kumar Rai (1977) : According to L P Vidyarthi and Binay Kumar Rai tribes can be described in different ways. The tribes in India are classified into various types on the basis of geography/location, languages, race, on the basis of integration as well as economic and cultural classifications.

V.Nandakumar (2004): V.Nandakumar tells about tribes in Kerala in the following manner. The majority of tribes are living in rural areas and within the forest lands of the Western Ghats. Most of the tribal people are engaged in agriculture and allied sectors. The economy of the tribes can be treated as 'agrarian', since the tribes depend on primary occupation like cultivation, live stock, hunting, fishing etc.

R. Mercy "Constraints in the education of the tribal pupils of Kerala" -Thesis. Department of Education, University of Calicut (2005)

G.S. Deoganger (1979): The most important tribal communities living in Kasaragod region are the Koragas and Maratis. They are mainly engaged in basketry. A tribal is a born borrower. His ignorance is exploited by many who come into contact with him, 'Hardworking, god fearing, honest and humble theirs is a society ridden with superstition and ignorance which makes them vulnerable and extremely backward.

P.R.G.Mathur (1977): The tribes are known in different names as 'Adivasies', 'Kattujathi' or 'Kattukurman', 'Vanavasi', 'Girijans', 'Forest dwellers' etc. The tribal communities living in different regions in the state may be classified into seven major territorial groups taking into account their historical, ethnic and socio-cultural relations.

DATA ANALYSIS

The Kudiya tribes are the largest Scheduled Tribes of Kerala. Even though Kudiya form the leading tribal community in Kasaragod, they are the most backward in every respect. They are the downtrodden group holding very little land and having little amenities of life. The researcher has taken Kudiya tribe as the main focus of the study. Kudiya/Malakudiya is one among the major tribes of Kasaragod district. They are inhabited in the colonies of Kammady puthikai and Paivalike Panchayath and also a few numbers of families are settled in Nileshwaram. These colonies are under the supervision of Panathady, Enmakaje as well as Nileshwaram tribal extension office. Panathady which consisted the Kammady colony, having 41 families residing there and have a population of about 167. In order to analyse socio-economic conditions of Kudiya/Malakudiya community we visited Kammady colony of Panathady panchayath and data collected from 41 households from the Kammady are taken into consideration. The data collected through interview using well-structured questionnaire.

1. Gender wise Analysis

Details of the sex wise classification of primary data are shown in the following table and graph.

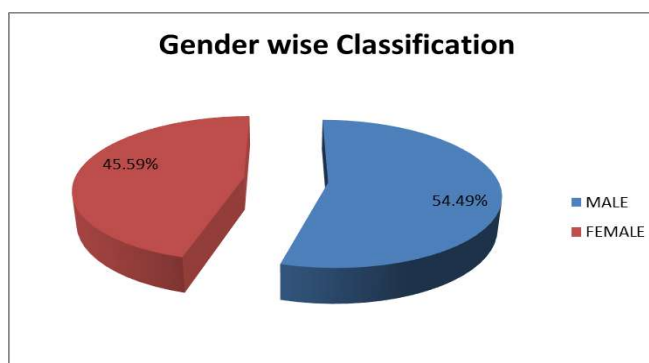| Sl. no	Sex	No .of persons	Percentage
1	Male	91	54.49%
2	Female	76	45.59%
3	Total	167	100%

Table 1   **Figure 1**

The selected 41 families of Kammadi colony contains 167 tribal populations; 91 male population (54.49%) as well as 76 female population (about 45.59%). The samples are dominated by male population rather than females.

2. Marital Status of Respondents

Marital status is another aspect studied and details are presented in the following table and graph.

Sl. no	Marital status	No. of persons	Percentage
1	Married	93	55.78%
2	Un married	74	44.31%
	Total	167	100%

Table 2
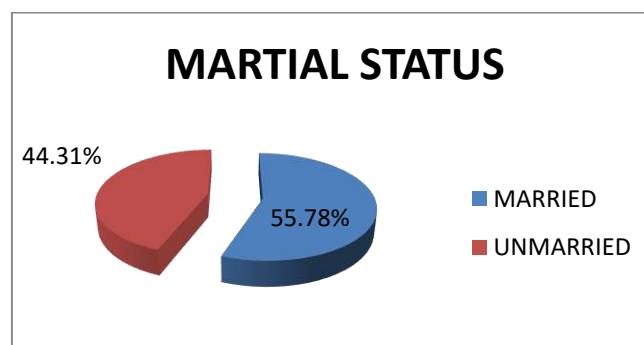
Figure 2

Out of 167 families of Kudiya community in Kammadi colony 93 (55.78%) are married people and 74 (44.31%) are unmarried. As the table and diagram depicts it is clear that most of the people are married. The remaining people are still youth and children.

3. Age wise distribution

Age wise distribution of data are shown in the following table and graph from this analysis   we may come to know how much is the active population and how much is the dependent one.

Sl. no	Age group	No. of house hold	Percentage
1	Below 5	8	4.79%
2	5 – 15	22	13.17%
3	15- 25	29	17.37%
4	25-35	27	16.17%
5	35-45	28	16.77%
6	45- 55	20	11.98%
7	55- 65	20	11.98%
8	Above 65	13	7.78%
	Total	167	100%

Table 3
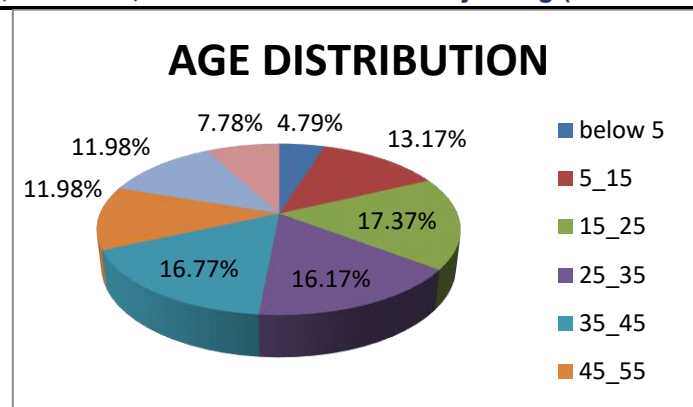
Figure 3

The given table conveys information about the age status of Malakudiya tribes. Majority of the tribal population are belonging to the age group 25-55(44.92%), the population belongs to these age group is considered as the working population. At the same time there is high dependency ratio. Among the dependency ratio, 17.96% of population come under the age category of 1-15 and 11.98% as well as 7.78% are also considered as dependents belongs to the age groups of 56-65 and age above 65 respectively.

4. Category wise details

Survey tried to collect category wise details of inhabitants in the Kammadi colony. the details are as follows:

Sl. No.	Category	No. of families	Percentage
1	SC	1	2.44%
2	ST	39	95.12%
3	OBC	1	2.44%
4	General	0	0%
	Total	41	100%

Table 4
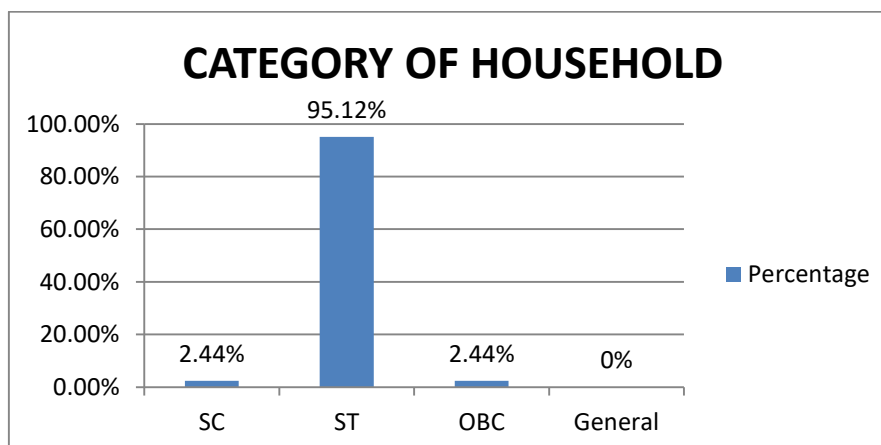
From the table 5.4 depicts category of households, 95.12% of households are belongs to ST (Scheduled Tribes) and 2.44% are belongs to SC(Scheduled Casts) and the remaining 2.44% are belongs to OBC(Other Backward Class). That is, out of 41 families of Kammadi, 39 families are ST, and there exist one SC family and one OBC family.

5. Economic classification of Families

In this area we discuss, economic classification of various households : survey revealed that 96.5% of the inhabitants belonged to the AAY category, showing the weak economic condition of this tribe .

Sl. No	Economic category	No. of families	Percentage
1	APL	0	0%
2	BPL	1	2.44%
3	AAY	40	97.565
	Total	41	100%

Table 5

ECONOMIC CATEGORY OF HOUSEHOLDS
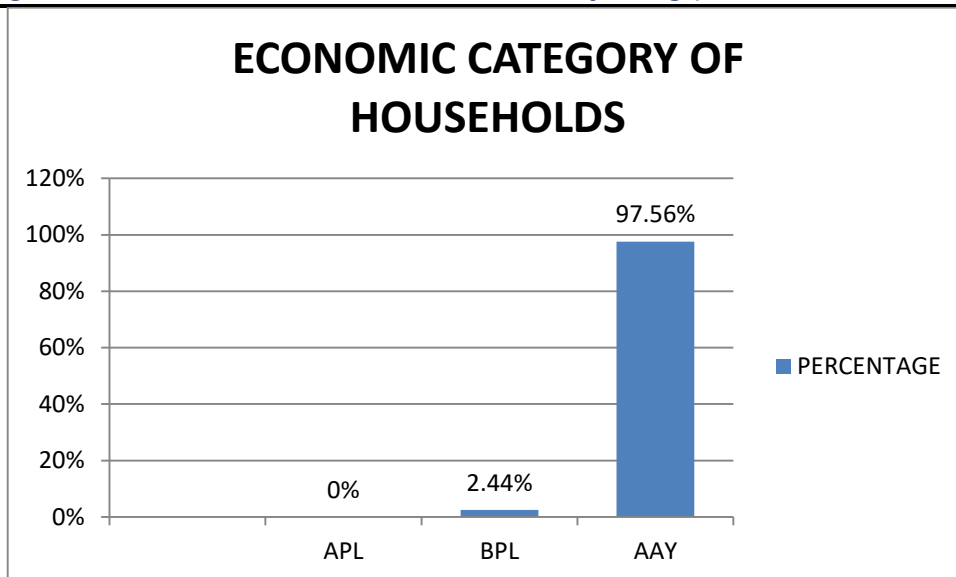
Figure 5

The table 5.6 and figure 5.6 shows the distribution of households on the basis of their economic category. As depicted above most of the families of Kammady are in the category of AAY (Antyodaya Anna Yojana) which is a Government of India sponsored scheme to provide highly subsidised food to millions of the poorest family. 2.44% are belongs to the category of BPL(Below Poverty Line). So it is clear that most of the families are very poor. The whole Kudiya/ Malakudiya tribal families are depending on Public Distribution System (PDS) for purchasing the basic amenities of life at a subsidized rate, which were allotted to them by the government. The PDS of the colony is working in an efficient manner.

6. Educational status

According to the survey, more than 42% are found to be the illiterates. Only 4.4% studied above higher secondary level. This reveals the poor educational attainments of this tribe.

Sl. no	Literacy level	No. of persons	Percentage
1	Illiterate	68	42.77%
2	Below 8th	57	35.85%
3	SSLC, Plus two	27	16.98%
4	Higher education	7	4.40%
	Total	159	100%

Table 6

Education level
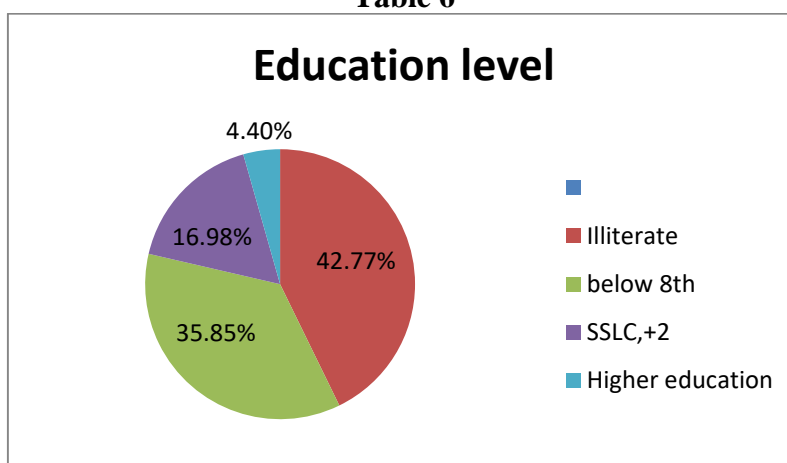
Figure 6

From the table and diagram we can understand that there exist greater illiteracy in Kammadi colony (42.77%) Among the total illiterates majority are old aged persons, especially women. The present generation have primary education (35.85%). 16.98% have the educational qualification of SSLC/Plus Two. At the same time only 4.40% of the tribals have acquired higher education by overcoming all the difficulties.

7. Occupational status of families

Primary survey enquired about the nature of occupation of people at the Kammadi colony and results are presented in the table and graph:

Sl. no	Occupation	No. of persons	Percentage
1	Agriculture	2	1.19%
2	Labourer	73	43.71%
3	Private sector	3	1.79%
4	Govt. sector	1	0.59%
5	NRK (Non Residence of Kerala)	2	1.19%
6	Students	30	17.96%
7	Unemployed	18	10.77%
	Total	167	100%

Table 7
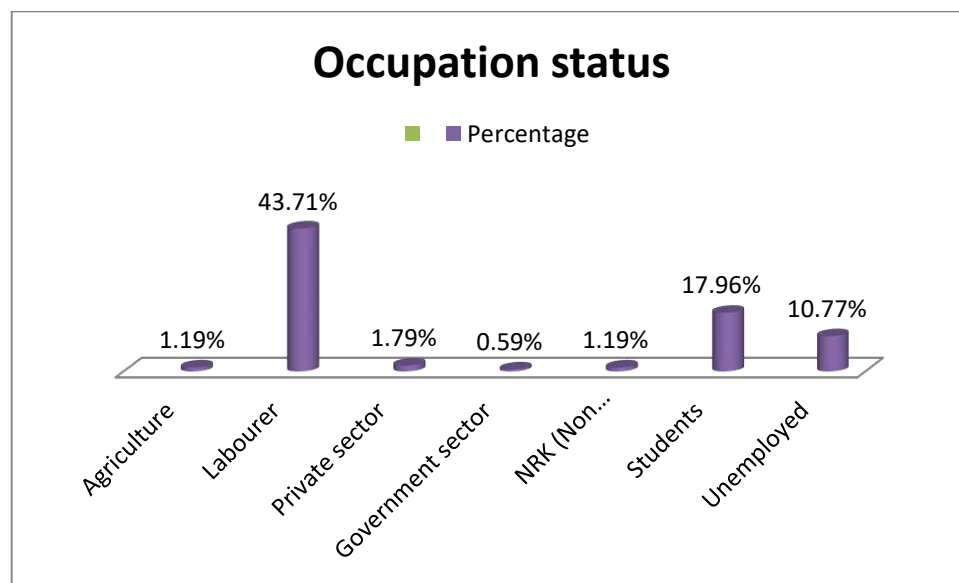
Figure 7

The above illustration has shared the occupation details of the Kudiya / Malakudiya tribes. Majority of the tribal population are labourers(43.71%), only about 0.59% of them are having government jobs. Apart from this the tribes are depending on allied activities including animal husbandry, poultry farming. 1.79% of them working in private sector and 1.19% of the tribal population are Non Residence of Kerala. Only 1.19% people are depending agriculture for their living. 17.96% of the total population are students. About 10.77% of the population is remained as unemployed, there is no regularity in job availability.

8. Details of Monthly Income

How much income is generated in a month is another aspect studied. the results are as follows: 51% are earning between Rs.2000 to 4000par month. Only 2.4% earns above Rs. 10,000/-. This shows poor economic condition of the vast majority of tribals in this colony.

Sl. No	Monthly income	No. of families	Percentage
1	Below 2000	9	21.90%
2	2000 – 4000	21	51.20%
3	4000 – 6000	6	14.60%
4	6000 – 8000	3	7.30%
5	8000 – 10000	1	2.40%
6	Above 10000	1	2.40%
	Total	41	100%

Table 8
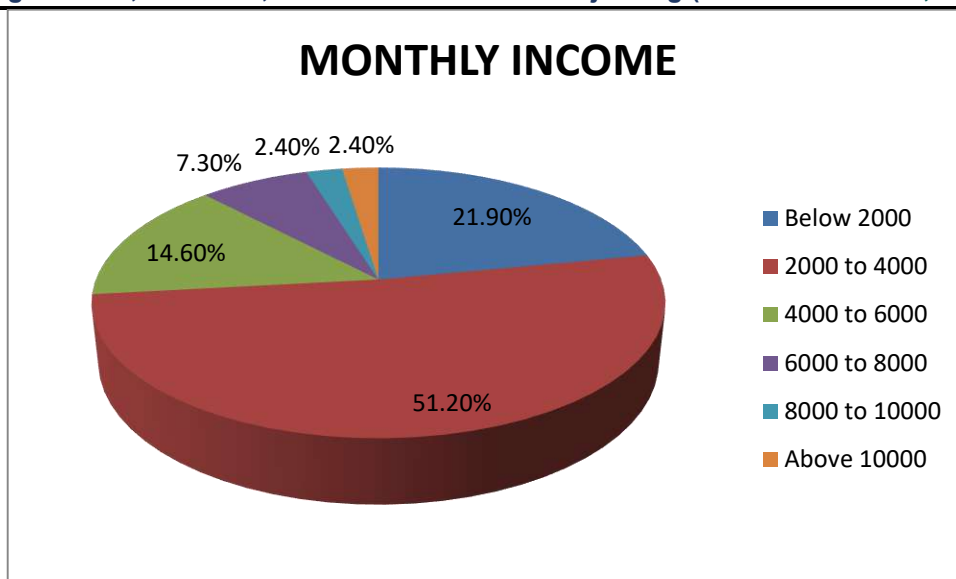
Figure 8

From the above illustration it is clear that majority of the Kudiya/Malakudiya tribes are belonging to the monthly income group ranging between 2000-4000, about 21 families (51.2%). The average family income of Kudiya/Malakudiya tribes is about 3149 and the percapita income of the community is 773. As we know in current scenario this income is insufficient for a family to survive. The Kudiya/Malakudiya tribes of Kammady colony are mainly depending on labour sector for their livelihood. A major proportion of the people are depending on Rubber tapping, which is seasonal oriented. They have a wage rate of 250-300 per day, but there may be inconsistencies in wage rates. During Monsoon and other rainy seasons it seems very difficult for the people of Kammady. There is only one person having government and no one having other white collar jobs.

9. Average monthly savings

Capability to save is an important indicator of economic position of a household. The survey results in this respect are discussed below. The survey reveled that majority of households (63.4%) have very meagre savings of less than Rs. 500 per month.

Sl. No	Monthly savings	No. of families	Percentage
1	Below 500	26	63.40%
2	500 - 1000	2	4.87%
3	1000 - 1500	5	12.10%
4	1500 – 2000	3	7.31%
5	2000 – 2500	5	12.10%
6	Above 2000	0	0%
	Total	41	100%

Table 9
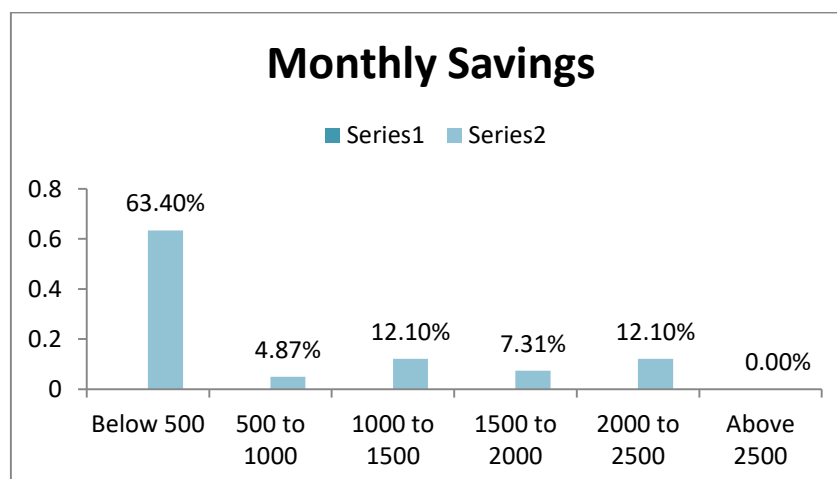
Figure 9

Table shows the monthly savings of Kudiya community. The table and figure depicts that the saving habit of the tribe is very less. About 63.40% of the tribe having amount of less than 500 as savings or no saving. 4.87% have small savings which ranges between 500-1000. 12.10% of the total population have 1000-1500 as monthly savings. 7.31% saves an amount ranging between 1500-2000 and 12.10% have savings ranging between 2000-2500. No one saves more than 2500.

10. Access to Electricity

Sl. no	Electricity	No. of families	Percentage
1	Yes	39	95.13%
2	No	2	4.87%
	Total	41	100%

Table 10
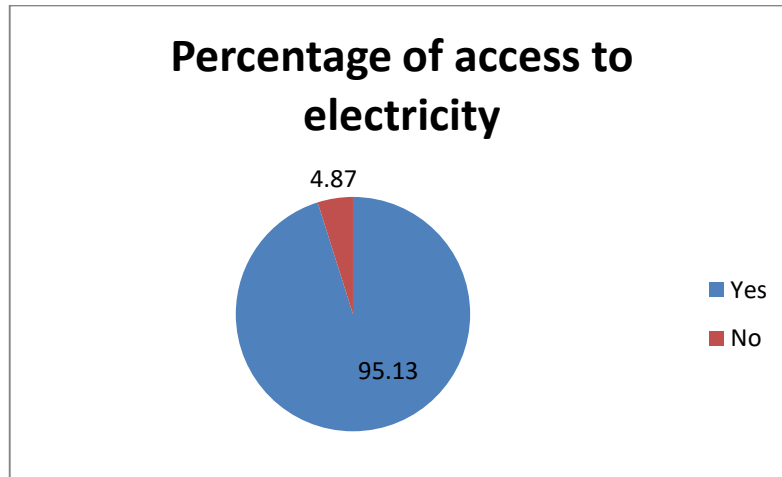
Figure 10

The table and figure shows the household's access to electricity. Majority of the families of Kudiya/Malakudiya tribes are having electricity (95.13%) remaining only 4.87% are experiencing the absences of such facilities. The government has taken the initiatives for electrifying each house in the colony.

11. Access to cooking gas

Another indicator of welfare is access to cooking gas . According to the survey 75% of the families have access to cooking gas.

Sl. no	Gas	No. of Families	Percentage
1	Yes	31	75.7%
2	No	10	24.39%
	Total	41	100%

Table 11
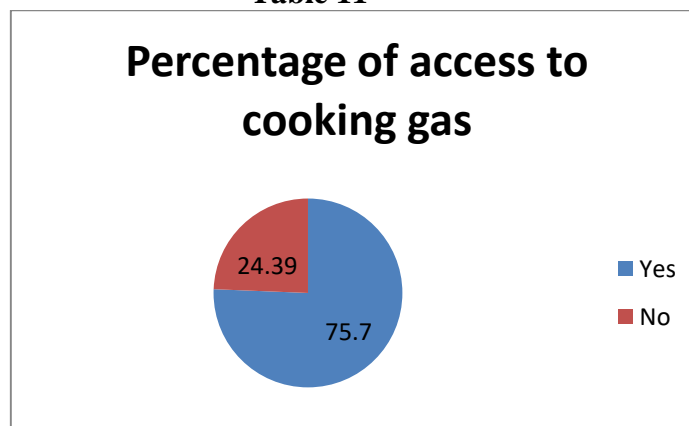
Figure 11

The above representation discusses the family's details on consumption of cooking gas. About 75.7% of families are having gas connection and they are also using wood (100%). Now there are improvements even in the life style of the tribes. Today the Malakudiya/ Kudiya tribes are not isolated even though they are settled in such a remote or isolated area they are having relationship with the outside world. The tribal behavior has almost disappeared. Majority of the families are having electricity and gas connection.

12.Toilet facilities

From the hygienic point of view, access to toilet is important. survey results showed that only 73% have proper toilet facility. In the days of "Swatch bharath" here more than 27% families have no toilet facility.

Sl. No	Toilets	No. of families	Percentage
1	Yes	30	73.18%
2	No	11	28.82%
	Total	41	100%

Table 12
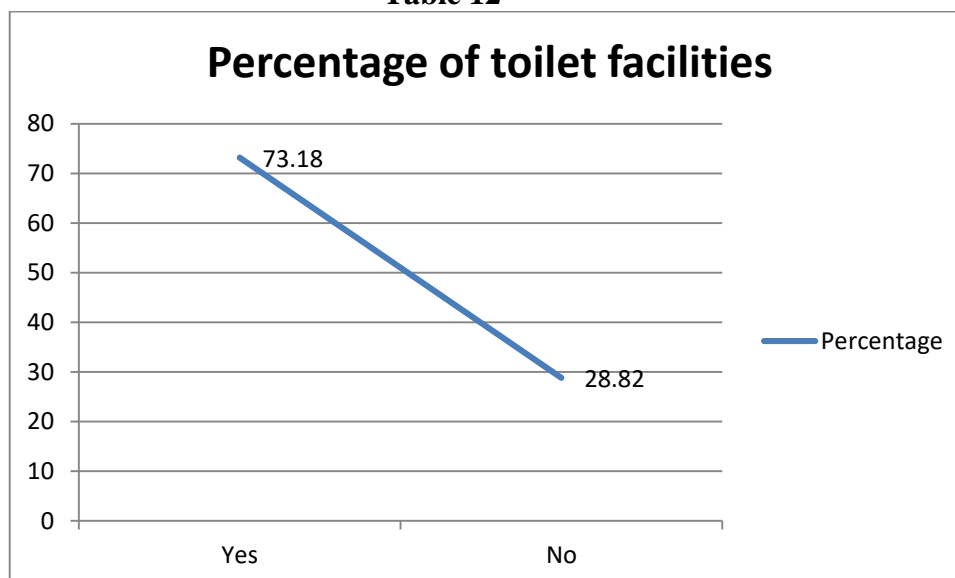
Figure 12

13. MGNREGP

Mahatma Gandhi National Rural Employment Guarantee Programmes (MGNREGP) aims to enhance livelihood security in rural areas by providing 100 days of wage employment in a financial year. It has a large impact in rural areas of Kerala. Details on MGNREGP in Kammadi are shown below:

Sl. No	MGNREP	No. of persons	Percentage
1	Yes	9	5.3%
2	No	158	94.7%
	Total	**167**	**100%**

Table 13
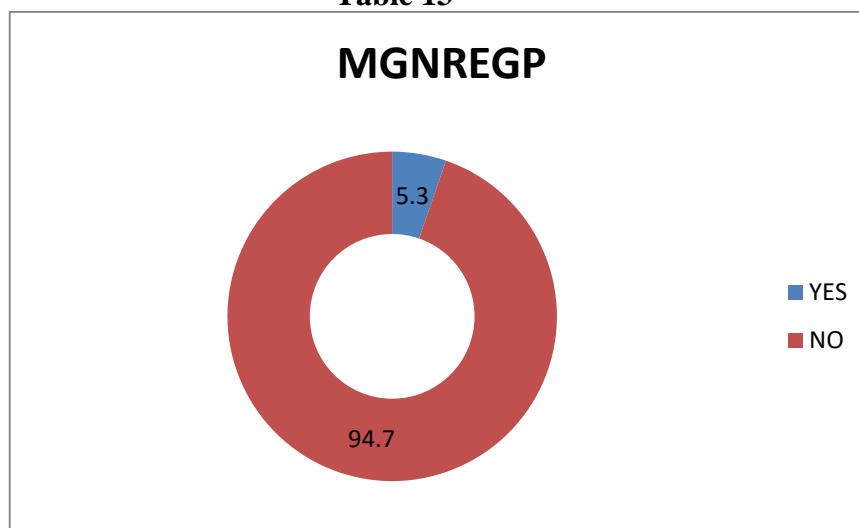
Figure 13

The above illustration clearly indicates that, only a few numbers of families are depending on Mahatma Gandhi National Rural Employment Guarantee Programmes (MGNREGP) as a source of livelihood. Major proportion of the tribal families are depending on agriculture and other allied activities. They are the agricultural labourers who depend on Rubber tapping.

FINIDINGS

- Number of Males are relatively more in this tribe. In the sample of 41 families the total population is 167, of them, 91 (54.49%) male population and 76 (45.51%) female population.
- In the total population ,93 (55.78%) are married people and 74 (44.31%) are unmarried.
- Majority of the tribal population are belonging to the age group 25-55 (44.92%), they are the working population. 17.96% come under the age group of 1-15 and 19.76% are above the age 55, which means there exist high dependency ratio.
- In kammady, 95.12% of households are belonging to the ST(Scheduled Tribe) category , 2.44% belongs to SC(Scheduled Cast) , and 2.44% belongs to OBC(Other Backward Cast)'
- Most of the families are in the category of households having 4 to 6 members(44.46%). 29.27% having 2 to 4 members, 21.95% having family members more than 6. Only 7.2% of total households having 1-2 members.
- Majority of the families are belonging to AAY category(97.66%).
- Education is not benefitted majority of this tribe. Still 57% are not literate. 35.85% have basic education and 16.98% completed secondary level of education and 4.40% are having higher education. About 42.77% are illiterate and 57.23% are literate.
- Only 0.59% of total population is having government jobs. Majority of the population (43.71%) are labourers and 1.19% are depending only agriculture sector for living. 1.79% are working in private sector and 1.19% are Non Residents of Kerala. 10.77% of the total population remains as unemployed.
- Major proportion of the tribal families are having monthly income ranging between Rs.2000-4000 (51.20%).
- The average monthly income is about Rs 3149 and Per Capita Income is Rs 773. The monthly income is insufficient for the families to acquire basic necessities of life.
- 60.90% of the total population is having monthly expenditure ranging between Rs.2000-4000. 21.90% have expenditure slab of 4000-6000. 4.80% incurs expenditure more than 8000.
- Saving habit of this tribe is very low. About 63.40% of the tribe having amount less than Rs.500 as saving or no saving. 4.87% have small savings. 12.10% have amount ranging between 1000-1500 as savings. No one saves more than Rs 2500 and 12.10% have monthly savings ranging between 2000-2500.
- Majority of the families of Kudiya tribe have electricity(95.13%). Remaining 4.87% are experiencing the absence of such facilities.
- About 75.7% of families are having cooking gas connection.
- Major proportion of Kudiya/Malakudiya tribal families are having proper toilet facilities (73.81%), they are using toilet facilities properly. The remaining 28.82% of families are not having adequate toilet facilities. They require proper toilet facilities.
- Majority of the Kudiya families are having AADHAR Card(96.41%). 76. 94% have Voter ID Card and all families have Ration card. But more than half of them are not having Health Insurance Card(53.56%).
- About 31.83% have owned vehicles and 68.17% are not owning any type of vehicles.
- About 64.67% have bank accounts.
- Major proportion of the households are having Kudumbasree memberships (64.7%).
- Only about 5.3% of the total population are depending on MGNREGP (Mahathma Gandhi National Rural Employment Generation Programme).

SUGGESTIONS

- The Kammady colony needs proper road facilities. Nearly 5k.m proper road is absent . More funds should be allocated urgenty to build a proper road
- Their higher education is adversely affected by the lack of proper road facilities and the isolation of the colony.
- Proper job training should be given to the educated persons.
- Majority of them are not getting welfare schemes such as old age pension, widow pension and so on. So proper monitoring systems should be ensured.
- There is no regularity in medical camp and hence participation by the tribes.
- There is no nearby Public Health Centers (PHC). Major proportions of the tribal families are depending upon private hospitals far away from their colony.
- They have no awareness about Health Insurance card and utility of that. So the availability of such cards should be given into priority.
- The lack of proper latrine facilities is a major obstacle to development in this area.This can be rectified by the proper implementation of the "Swatch Bharath Mission".

CONCLUSION

Malakudiya/Kudiya tribe is one of the tribes living in Kasaragod district. When compared to the previous years, their living standard shows substantial improvements. The government is taking the initiatives for the development of Malakudiya tribes. But in Kammady colony government programmes and policies are not successfully implemented. In the opinion of the tribes of Kammady, they are better known as Kudiyas from traditional period onwards, but now officially they are recognized as Malakudiyas. Malakudiya/Kudiya tribes are good at their heart; they are loveable people having attachment with their family especially with the elders of the community. The tribes who settled in Kammady colony are mainly depending on agriculture and other allied activities for their livelihoods. The Kudiya/Malakudiya situated at the colony of Kammady have better standard of living in comparison to their primitive years. At the same time they have to face certain challenges in the form of inadequate road facilities, low income and uncertainty in working days and wage rates. They are loveable people. The present generation of Malakudiya tribes had achieved better educational qualification.

REFERENCE

1. Balakrishnana, E P "Economics Of Tribals And Their Transformation, A Study In Kerala", M G University Online Research Library, Issue Date October 29, 2010.
2. Banerjee, Rahul "Actualizing Adivasi Self- Rule- The Only Panacea", Yojana, Vol. No. 58, January 2014, p. 22- 45.
3. Battacharee, Suparna "Emergence Of Marketting Of Handloom And Handicraft Product Of Monappa Tribe In Tawang Of Arunachal Pradesh", Southern Economist, Vol. No. 52, June 15, 2013, p. 26-30.
4. Bhagat, R B "Conditions Of SC/ ST Households- A Study Of Unequal Improvements", Economic And Political Weekly, Vol. No. 48, October 12, 2013, p. 36-38.
5. Bharathi, Mudunuri "Tribal Women's Perspective On The Land Acquisition Bill", Economic And Political Weekly, Vol.No 40, May 19, 2012, p. 34-36.
6. Chakraparty, Gurupadai "Scheduled Castes And Scheduled Tribes in Rural India- Their Income, Education And Health Status", Economic And Political Weekly, Vol. No. 26, June 2010, p. 43-47.
7. Geetha, T. K "Political Socialization Of Tribes", MG UniversityResearch Library, Issue Year 1998.

EPRA International Journal of

ECONOMICS, BUSINESS

AND

MANAGEMENT STUDIES

Save the Earth

Research

Awareness

Protection

Environmental

EPRA

Indexed By:

Published By :EPRA Publishing

CC License

AN OVERVIEW OF SOCIO - ECONOMIC CONDITION OF PANIYA COMMUNITY IN ARALAM FARM OF KANNUR DISTRICT, KERALA

Akhil Baby

Assistant Professor, Department of Basic Science and Humanities,
SCMS School of Engineering and Technology, Karukutty, Kerala

ABSTRACT

Tribals, the original inhabitants of India have contributed much to the nation's culture, history and heritage. Under the constitutional provisions of Directive Principles, the States' major concern for tribes has been their welfare and development. It is ironical that despite a large number of well-meaning constitutional provisions aimed at protecting and safeguarding the welfare and interest of the tribal communities, the process of marginalization of the tribal's has gone unabated. In this context, the paper has made an overview on the socio-economic conditions of tribals belonging to the Paniyas (Paniyas), which has the lowest standard of living, even though being the bulk of the tribal populace in Kerala.

KEYWORDS: - *Socio economic conditions, Paniya, Kudumb, Original inhabitants*

INTRODUCTION

Tribal population is identified as the original inhabitants of our country. Broadly speaking the tribes may be geographically located in four groups:

- North Eastern India - Assam, Manipur, Nagaland
- Eastern India - West Bengal, Bihar, Orissa
- Central and Western India - Madhya Pradesh, Rajasthan
- Southern India - Tamil Nadu, Kerala, Andhra Pradesh

The socio-economic condition of the tribal's – a primary survey including 270 paniya tribal's households from pulpally and mullankolly villages of Sulthan Bethery. The mainly five major tribal communities have their origin in wayanad and Kannur. The paniya tribe was largest among them. They are the largest ST of Kerala. The majority of the paniya tribal population (71.95%) is in Wayanad.

PANIYAS

The etymological meaning of the term "paniya" indicates that they earn their livelihood from labour as the term "pani" in Malayalam means "labour" or "worker".

PHYSICAL APPERANCE

The colour of the paniyas varies from dark to dark brown. They speak a debased form of Malayalam mixed with tamil words. Thurston observes that they are dark skinned tribe, short in stature, with broad noses and wavy hair.

SUBDIVISIONS

The Paniyas have two sub divisions according to the dwelling pattern. Paniyas of the plain land have contact with only their own tribal people living in their forests. The latter group is mainly found in Nilambur forests of Malappuram district.

DRESS

Normally the paniya settlements (paddies) may be a cluster of few huts (Pire or chala).The dress of paniya is simple and different from other tribal groups. The males wear a cloth which is wrapped around the waist and small mundu

on the shoulders covering the body. The paniya women (panichi) wear a long cloth and a smaller one above the bosom through the ampits.

FOOD
Traditionally paniya's were food gathers enjoying the freedom and self sufficiency of nomadic life in the inferior forests. They use edible roots, leaves etc . The major food items include tapioca, vegetables, animal flesh, fish etc . Besides drinking tea or coffee, they are addicted to drinking toddy and some varieties of alcoholic beverages.

RELIGION
Kuttichathan, Kuliyan, Kattubhagavathi or kali , mariyamma and ayyappan are the chief god of paniyas. Paniyas give offering s to their gods during rituals and festive occasions. They observe a number of festivals in different seasons. Some of these festivals are exclusively their own and some others spread regionally throughout the state.

OCCUPATION
Paniyas are primarily agricultural laboures but very few families possess land holding. A paniya male get Rs 85 per day as wage where as paniya female get Rs 50 as wage in the 1970-80. Traditionally they were engaged in bounded labour (Kundal pani). Now a days the economy of paniyas is well –knit, with that of the non –tribal settlers to control the occupational pattern and economic conditions.

FUNERAL RITES
The place of burial of Paniyas is usually near their paddy. Mourning is observed by the members of the family for seven days. The mourner in the house is supposed to perform the rituals on the seventh day.

MARRIAGE CUSTOMS
The most common form of marriage among the Paniya is marriage by negotiation. Marriage by negotiation is always accompanied by the payment of bride price by the bride groom. It is a custom among the Paniya community that the husband has to make an annual subscription known as Thalapattam to his wife parents on the annual festival day.

SOCIAL CUSTOMS
Ear- boring ceremony is usually performed when the child attains the age of three. A mixture of ginger oil and turmeric paste is used for heeling the wound.

INDEBTNESS
Often debts are incurred for meeting the expenses of marriage ceremonies, death rites etc. In other words, Money borrowing is the major factor behind land alienation.

STATEMENT OF THE PROBLEM
The tribals, the original inhabitants of India have contributed much to the nation's culture, history and heritage. The principles entailed development, along the lines of their own genius respect of tribals rights in land and forest, training and building up a team of their own people to do the work of administration and development, to their social and cultural institutions. It is ironical that despite a large number of well-meaning constitutional provisions aimed at protecting and safeguarding the welfare and interest of the tribal communities, the process of marginalization of the tribals has gone unabated. In this context the researcher has made an attempt to examine the socio-economic conditions of tribals belonging to the Paniya community with special reference to Aralam farm of Kannur district.

OBJECTIVES
- To analyze the socio-economic conditions of Paniya community.
- To identify the role of the government in the Paniya tribal development.
- To give suggestion to improve efficacy of government programmes.

METHODOLOGY
This study is based on the both primary and secondary data. Primary data collected from Aaralam farm through personal interviews and surveys. Source of secondary data include books, journals, newspaper etc.

REVIEW OF LITERATURE

P S Lenin, Research scholar, Bharathiar University, "Empowerment of tribal children of attapady, Kerala"
This is a study of the need of addressing the educational constraints and empowers the tribal children of Attapady, Kerala. This research is aim to find out the various educational availability of Tribal children, to analyze the educational problems, to find out need of counseling and guidance, to assess the method of conducting programmes for tribal children and thus empower the children belongs to scheduled tribes.

M S Sreerekha, "Challenges before Kerala's land less: the story of Aralam farm", Whether from a class perspective or from a community identity perspective, it is undeniably the biggest failure that decades after the land reforms, A good majority of the dalits and Adivasis in Kerala remain fully landless. The author is grateful to M G Geethandhan for inspiration and help. The Kerala government bought Aralam farm from the central government as a part of a post –muthanga agreement in 2004.

Kempson et al (2000) identified the range of physical and geographical barriers to financial inclusion factors that can contribute to financial exclusion for different products and individuals under certain circumstances. There are a number of "dimensions " or "forms" of financial exclusion that have been identified.

Leeladhar V (2005) Opined in his lecture that a holistic approach on the part of the banks in creating awareness about financial products, education, and advice on money management, debt counseling, savings and credit will be addressing financial exclusion.

Aditya P (2007) pointed out that HDFC bank has a deeply felt commitment to financial inclusion and has been the pioneer in using the business correspondent model to achieve this purpose.

ANALYSIS OF PRIMARY DATA
A sample survey has been conducted and data were collected through questionnaire method from 40 households on the basis of the socio – economic condition of Paniya community in Aralam farm. The collected data were analyzed in this chapter by using various tables and diagrams.

1.   INCOME WISE CLASSIFICATION
The respondents were also classified on the basis of their income; the following table shows the income wise classification.

Table – 1
Income Wise Classification

Sl. No	Income	No of Respondents	Percentage
1	0 - 10000	40	100%
2	10000 – 50000	0	0
3	50000 – 100000	0	0

Primary Data
In the above table, Among the 40 Respondents 100% has below 10000 incomes.

Diagram -1
Income Wise Classification
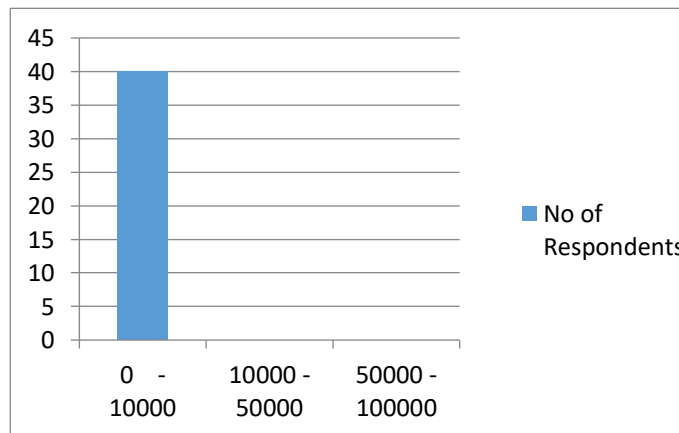
2. CASTE WISE CLASSIFICATION

The respondents were also classified on the basis of their caste, in the following table.

Table – 2
Caste Wise Classification

Sl .No	Castes	No of Respondents	Percentage
1	SC	8	20%
2	ST	32	80%
3	OEC	0	0
4	Others	0	0

Primary Data

The table 2 shows that 32 percentages of the sample respondents are under the Scheduled Tribe and 20 percentages under Scheduled Cast.

Diagram – 2
Caste wise classification
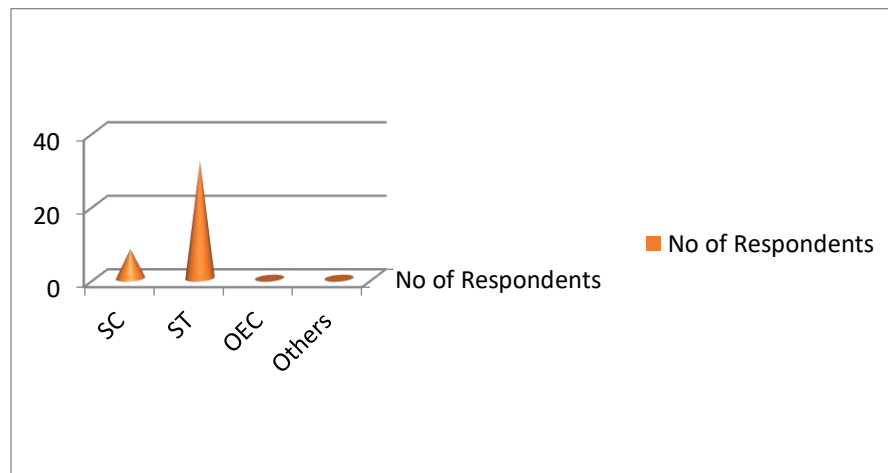
3. CLASSIFICATION BASED ON THE OTHER SOURCES OF INCOME

Classification based on the other sources of income is shown in the given table.

Table - 3
Classification based on the other sources of income

Sl .NO	Particulars	No of Respondents	Percentage
1	Land	40	100%
2	Vehicles	0	0
3	Consumer durables	0	0
4	Others	0	0

Primary Data

From the 40 respondents, 100% are using Land for various income purposes.

Diagram -3
Classification based on the other sources of income
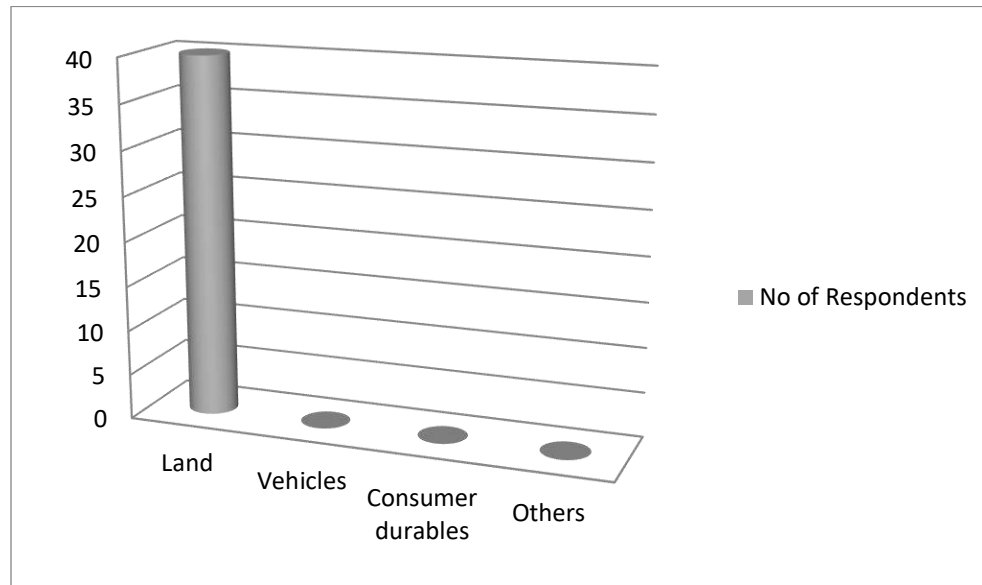
4.   SATISFACTION OF EXISTING WAGE RATE

The wage rate of the households is very poor. Under the satisfaction of existing wage rate are shown in the given table.

Table – 4
Satisfaction of Existing wage rate

Sl . No	Particulars	No of Respondents	Percentage
1	Yes	10	25%
2	No	30	75%

Primary Data

From the above table 25% of households satisfied existing wage rate and 75% are not satisfied.

Diagram – 4
Satisfaction of Existing wage rate
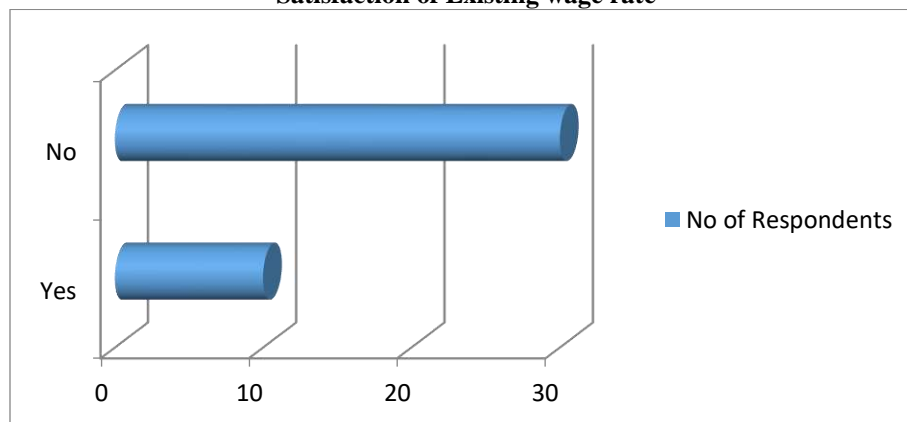
5.   CATEGORY CLASSIFICATION
The 40 respondents are classified on the basis of category shows following table.

Table – 5
Category classification

Sl .No	Category	No of Respondents	Percentage
1	BPL	40	100%
2	APL	0	0

Primary Data

Here 100% of households are under the BPL category.

Diagram – 5
Category classification
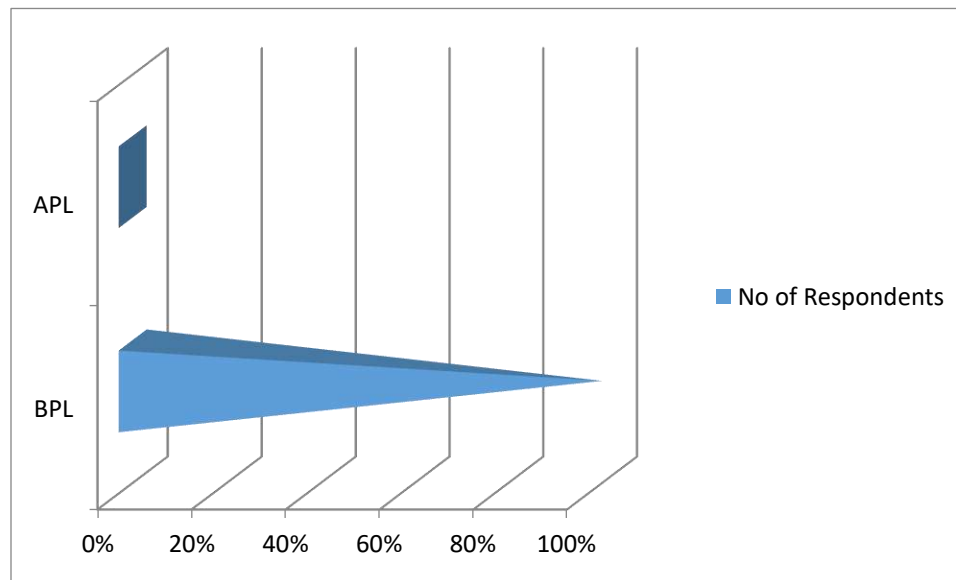
6.   CLASSIFICATION OF BANK ACCOUNT
Classification based on the bank account is given in the table

Table – 6
Classification of Bank Account

Sl . No	Particulars	No of Respondents	Percentage
1	Yes	40	100%
2	No	0	0

Primary Data

In the above table 100% of households having bank account.

Diagram – 6
Classification of Bank Account
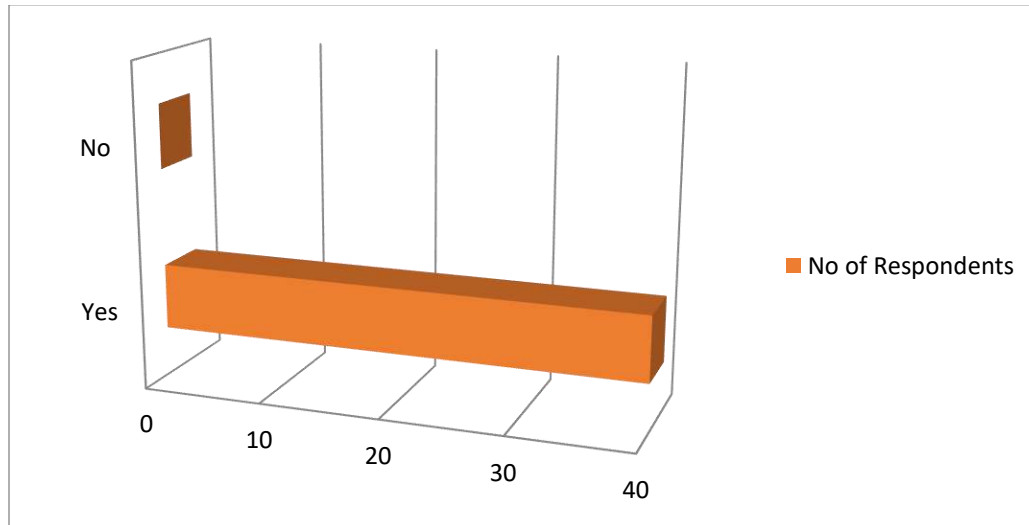
7.　CLASSIFICATION OF NATURE OF HOUSE
The nature of house is shown in the below table.

Table No – 7
Classification of Nature of House

Sl. No	Nature	No of Respondents	Percentage
1	Roof tiles	11	28%
2	Sheet	13	32%
3	Concrete	16	40%

Primary Data

Among the 40 respondents 28% having nature of house is roof tiles, 32% having Sheet and 40% having Concrete.

Diagram – 7
Classification of Nature of House
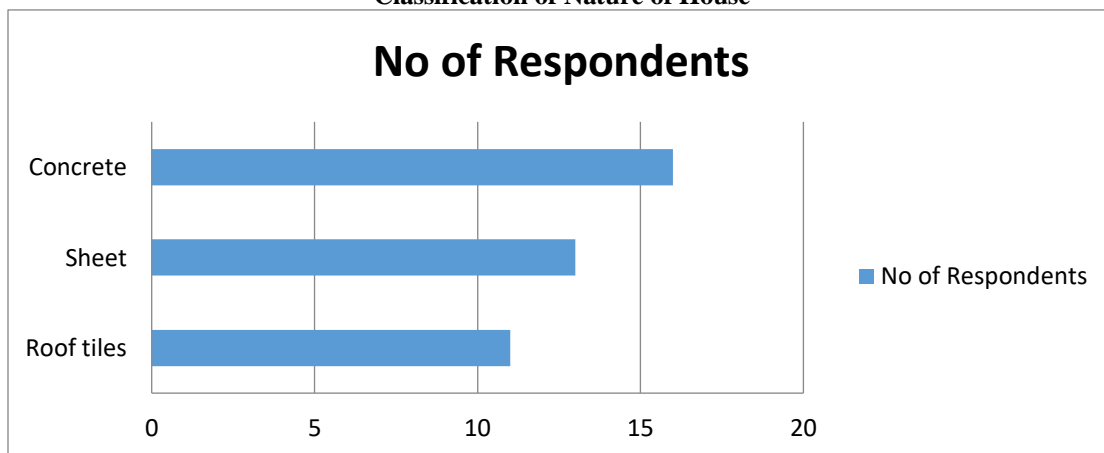
8. OPINION ABOUT SANITATION FACILITY

The sanitation facility is very poor. Under the opinion about sanitation facility are shown in the given table.

Table – 8
Opinion About Sanitation Facility

Sl. No	Particulars	No of Respondents	Percentage
1	Satisfied	30	75%
2	Unsatisfied	10	25%
3	No Opinion	0	0

Primary Data

Among the 40 respondents 75% having satisfied the sanitation facility and 25% having unsatisfied the sanitation facility.

Diagram – 8
Opinion about Sanitation Facility
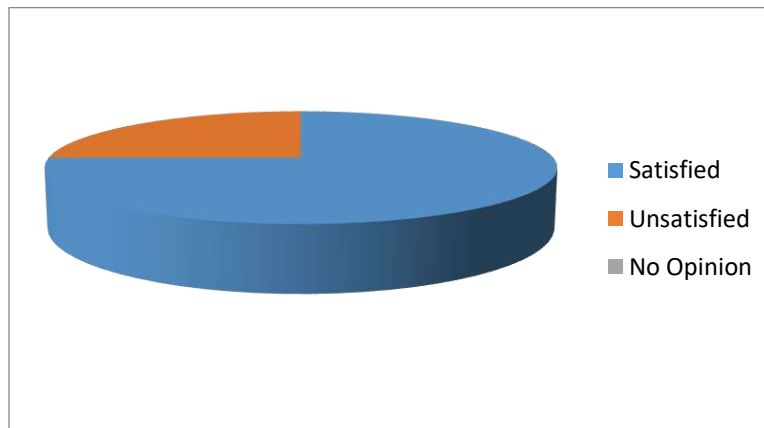
9. OPINION ABOUT GOOD WATER FACILITIES

The water facilities available to the respondents are satisfied on the basis of those the opinion of respondents about water facilities is shown in the table below.

Table No – 9
Opinion About Good Water Facilities

Sl. No	Particulars	No of Respondents	Percentage
1	Satisfied	12	30%
2	Unsatisfied	25	62%
3	No Opinion	3	8%

Primary Data

From the above table 30% have satisfied, 62% have unsatisfied and 8% have no opinion.

Diagram – 9
Opinion About Good Water Facilities
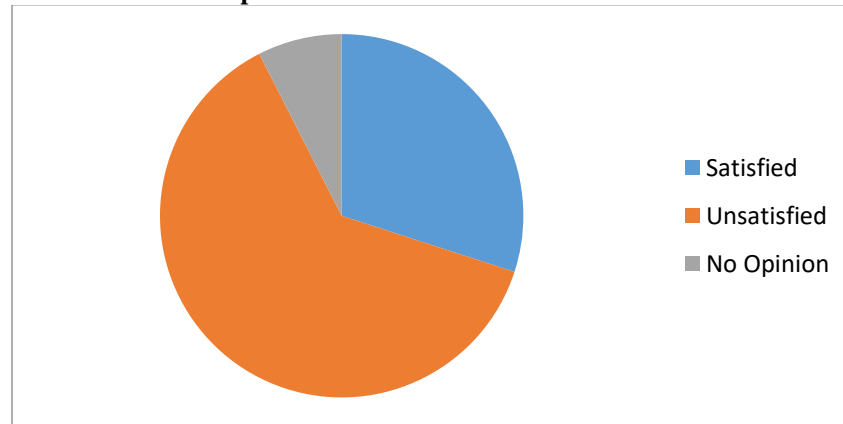
FINDINGS AND SUGGESTIONS
FINDINGS
The research findings of the study are based on the Socio – Economic Condition of Paniya Community in Aralam Farm.

- The Majority of the people in study are BPL families.
- The major source of income in the study area is Land.
- Most of the households are Agricultural labours concentrating on their farms.
- Total annual income of the 100% of households is below 10000.
- In the case of sanitation facilities 75% has satisfied and 25% has unsatisfied.
- The Opinion about good water facilities is 30% has satisfied 62% has unsatisfied and 8% has No opinion.

SUGGESTIONS
Based on the findings of the study the following suggestions can improve the Socio- Economic condition of Paniya community in Aralam Farm.

- ❖ Authorities may take necessary steps to include financial power (JAN DHAN YOJANA, MGNREGP etc..) in households.
- ❖ Aralam farm to provide more employment programmes and to improve the wage rate.
- ❖ Central government and local government should provide welfare schemes related to work.
- ❖ To promote bonus and allowances to the tribal people.
- ❖ To implement health facilities in the study area.

CONCLUSION
The study attempts to analyze the Socio –Economic Condition of Paniya community in Aralam Farm. The study was based on a sample survey using sample size of 40 households. After analyzing data, the researcher can state that the socio – economic condition of Paniya community in Aralam farm is not satisfactory. On the basis of this study, the researcher wants to suggest that urgent attention should be given in the growth by local governments. Which is essential for a good social and economic life in the area.

BIBLIOGRAPHY
1. *P. Panda (2011) "Paniya voices: A Participatory poverty and health assessment among a marginalized south Indian population".*
2. *Dr. M. Sukumaran (2014)" Education versus Deprivation: A case of Tribal's in Kerala".*
3. *M.S Sreerekha (2010) "Challenges before Kerala's Landless: The story of Aralam Farm".*
4. *P.S Lenin (2015) "Empowerment of Tribal children of Attapady, Kerala "*
5. *Dr. Sibi Zacharias (2003) "The micro level impact of tribal development programmes among the Kadar tribe of Kerala."*
6. *Mrs. Remya Krishnan "A study of financial inclusion and financial literacy among the tribal people in Wayanad district of Kerala."*

PACS numbers: 61.46.-w, 62.20.Qp, 81.05.Ni, 81.40.Pq, 81.65.Kn, 82.45.Bb

Investigation of Mechanical, Wear, and Corrosion Properties of Al−BN−SiC−RHA Hybrid Composites Synthesized Through Powder Metallurgy Process

G. R. Raghav, D. Muthu Krishnan[*], K. J. Nagarajan[*],
Vidya Chandran, R. Suraj, R. Sujith, S. Dhanesh, and M. S. Anoop

SCMS School of Engineering and Technology,
Vidya Nagar, Palissery, Karukutty, Ernakulam, 683582 Kerala, India
[*]*K.L.N. College of Engineering, Department of Mechanical Engineering,*
630612 Pottapalayam, Sivagangai District, Tamil Nadu, India

Present study explores the effect of Rice Husk Ash (RHA) reinforcements in mechanically alloyed Al−BN−SiC hybrid composites through experimental characterization. The physical, tribological and corrosion resistance properties of Al−BN−SiC−RHA hybrid composites are studied. In the outset the composite pellets are synthesized using powder metallurgy process. The homogenously amalgamated composite powders are then squeezed together to form multiple green composite pellets each measuring 8 mm in diameter using a hydraulic press at a pressure of 500 MPa. The Al−BN−SiC−RHA hybrid composites are characterized using FE-SEM, EDS, and XRD analysis for facilitating the studies on the microstructural as well as the elemental properties. Microhardness and compressive strength of the Al−5BN−5SiC−5RHA hybrid composites is found to improve by 6 and 13.6% respectively compared to those of Al−5BN−5SiC hybrid composites. The RHA reinforced hybrid composites have also experienced reduction in density and an enhancement in the wear resistance of the composite materials while compared to those for their non-RHA counterparts. The electrochemical corrosion analysis of the Al−5BN−5SiC−5RHA hybrid composites also confirms that the Al−5BN−5SiC−5RHA hybrid composites have better corrosion resistance. It is also evident that the corrosion resistance increases with increase in RHA rein-

Corresponding author: G. R. Raghav
E-mail: raghavmechklnce@gmail.com

Removal of organic dye from aqueous solution by using MnO$_2$ doped Fly ash as a photocatalyst and its comparison studies

Lakshmi Priya*

^{a} Department of Civil Engineering, SCMS School of engineering and technology ,Karukutty

Abstract: A commercial organic dye Drimaren Black PS-SG was obtained from a dying industry in Tamil Nadu, India. MnO$_2$-Fly ash was used as a photocatalyst for dye removal. Characterization studies like SEM and FTIR were performed on the catalyst. Influence of various parameters like variation of initial dye concentration, pH, H$_2$O$_2$ concentration and Fe (II) dosage were studied after the addition of MnO$_2$-Fly ash catalyst. These results were compared with H$_2$O$_2$+UV-C and Fenton+UV-C processes of dye degradation. The optimum pH value using MnO$_2$-Fly ash, H$_2$O$_2$+UV-C and Fenton+UV-C degradation was around 3.5. The rate of degradation elevated with the concentration of H$_2$O$_2$ and dropped after 10mM for H$_2$O$_2$+UV-C and Fenton+UV-C processes. The rate of degradation elevated with Fe (II) dosage and then dropped after 30mg/l for Fenton+UV-C process. The optimum initial dye concentration for all the processes were 10mg/l. The optimum catalyst dosage was found to be 1.5g/l. The dye degradation efficiencies were found to be 71%, 33% and 63% for MnO$_2$-Fly ash, H$_2$O$_2$+UV-C and Fenton+UV-C processes respectively.

Keywords: Dye, Photocatalyst; H$_2$O$_2$+UV-C, Fenton+UV-C

1. Introduction

Wastewater from industries is emancipated into streams in a large quantity. They are not given ample treatment before disposal [1]. This can craft a grave threat to the environment including the underwater species. Sunlight cannot infiltrate into the water because of the turbidity formed in water. Conventional methods of waste water treatment include physical and chemical methods. However these methods have a constraint that they do not translate contaminant into an inert substance, they merely alter the phase only and as a result further treatment becomes mandatory. Biological treatment however posses a disadvantage that it behaves differently with assorted pollutants. The experiment was conducted on an organic reactive dye Drimaren Black PS-SG, which is a real textile dye used in a textile dying unit located at Erode, India. Industries, which discharge a huge amount of chemicals into streams, make it necessary to perform studies on efficient treatment methods. Advanced Oxidation Process (AOP) is performed on the dye solution. AOP is a type of oxidation process in which hydroxyl radical is the oxidizing species. The oxidation potential of hydroxyl radical is 2.8 V. The theory behind AOP is transition state theory. Photocatalysis using MnO$_2$-Fly ash as a photocatalyst, H$_2$O$_2$+UV-C and Fenton+UV-C processes were carried out. The existence of manganese ions with variable oxidation state enables Manganese Dioxide (MnO$_2$) to portray significant role in reduction-oxidation reactions [6]. MnO$_2$ has got wide applications in

Micro-plastics in the Vicinity of an Urban Solid Waste Management Facility in India: Assessment and Policy Implications

Anupama S[1] · Nisha Luckins[1] · Ratish Menon[1] · Sruthy Robert[1] · Thushara Sudheish Kumbalaparambi[1]

Abstract

The study focuses on analyzing the presence and quantity of microplastics near an unscientific solid waste dumping site located at Kochi city in India and thereby assessing the contribution of massive open dumping towards microplastic pollution in the river. Two sets of sampling with three sediment samples from nearby Kadambrayar river and five topsoil samples from various parts of the Brahmapuram waste dumping yard were carried out during January – February 2019. The samples were analyzed as per the US National Oceanic and Atmospheric Administration (US NOAA) protocol. ATR-FTIR and DSC analysis based characterization indicated that polyethylene dominated the micro plastics followed by polypropylene and polystyrene. The size distribution of particles showed that comparatively larger particles of size range between 2.36 – 4.75 mm were present in the topsoil than that in river sediments which had more number of particles in size range below 2.36 mm. The study confirmed the presence of an average of 100 microplastic pieces per 100 gram of sediments of Kadambrayar river bordering the waste dumping yard at Brahmapuram. The topsoil of Brahmapuram waste dumping yard was found to contain 178 pieces of microplastics per 100 gm of soil. Microplastics are found to be high in the river stretch near the open dumping site which has a significant role in the pollution, causing a major threat to the entire ecosystem.

Introduction

Plastic pollution in the environment is becoming a challenging issue in the environment due to its longevity causing long distance transportation and accumulation due to its biodegradable nature (Avio et al. 2016; Depledge et al. 2013; Eriksen et al. 2014). Microplastics are small plastic pieces of less than 5 mm in size, which can be either primary or secondary in origin (Free et al. 2014; Sruthy and Ramaswamy 2016; Avio and Cardelli 2017). The primary sources include plastic microbeads in personal care products and synthetic fibers from the textile industry (Free et al. 2014). The secondary sources include the degradation of synthetic polymers like high-density polyethylene, low-density polyethylene, polystyrene, polypropylene, PET, etc. by physical, chemical or biological ways (Barnes et al. 2009; Wright et al. 2013; Halle et al. 2017; Fisher & Paglialonga 2016; Qiu & Tan 2016; GESAMP 2017). Studies about microplastics in the freshwater environment are limited in Kerala, located at the southernmost part of India. The Kochi city was selected as the broad study area as it is the second most urbanized city on the west coast of India (Naidu et al. 2018) and also due to its high density of population, large riverine discharge, and industrial and marine discharges. According to the 'Water and Air Quality Directory 2010' published by the SPCB Kerala, it can be concluded that the river Kadambrayar is polluted.

The key objective of this study was to identify the presence of microplastics in the topsoil and in the sediment of river Kadambrayar and characterize them to understand the nature of microplastics pollution in the vicinity of the solid waste management facility at Brahmapuram village on the outskirts of Kochi city in Kerala. The River Kadambrayar flows through an urbanizing zone of the city and it is a major

✉ Nisha Luckins
  nisha@scmsgroup.org

[1] SCMS School of Engineering & Technology, Karukutty, Ernakulam, Kerala, India

RESEARCH ARTICLE

Assessment of urban air quality from Twitter communication using self-attention network and a multilayer classification model

Thushara Sudheish Kumbalaparambi[1] · **Ratish Menon**[1] · **Vishnu P Radhakrishnan**[2] · **Vinod P Nair**[3]

Abstract

Social media platforms are one of the prominent new-age methods used by public for spreading awareness or drawing attention on an issue or concern. This study demonstrates how the twitter responses of public can be used for qualitative monitoring of air pollution in an urban area. Tweets discussing about air quality in Delhi, India, were extracted during 2019–2020 using a machine learning technique based on self-attention network. These tweets were cleaned, sorted, and classified into 3-class quality viz. poor air quality, good air quality, and noise or neutral tweets. The present study used a multilayer classification model with first layer as an embedding layer and second layer as bi-directional long-short term memory (BiLSTM) layer. A method was then devised for estimating $PM_{2.5}$ concentration from the tweets using 'spaCy' similarity analysis of classified tweets and data extracted from Continuous Ambient Air Quality Monitoring Stations (CAAQMS) in Delhi for the study period. The accuracy of this estimation was found to be high (80–99%) for extreme air quality conditions (extremely good or severe) and lower during moderate variations in air quality. Application of this methodology depended on perceivable changes in air quality, twitter engagement, and environmental consciousness among public.

Keywords BiLSTM · Delhi · Air pollution · PM2.5 · spaCy · Deep learning

Introduction

Increasing air pollution has become a major concern for the environmental quality of life in the urban areas. Recent studies visualize the grim reality of air pollution and its health risks across the cities of the world (Leffel et al. 2022; Castells-Quintana et al. 2021; WHO 2020; Pant et al. 2019). For strategizing any control or eradication program, first step would be to monitor air quality at possibly high temporal and spatial resolutions to identify the source, location, quality and quantity of pollutants (Gholami et al. 2021; Shanavas et al. 2020; Xu et al. 2018). The existing regulatory

✉ Ratish Menon
   ratishmenon@scmsgroup.org

1 Department of Civil Engineering, SCMS School of Engineering & Technology, Kochi, Kerala, India

2 Department of Computer Science & Engineering, SCMS School of Engineering & Technology, Kochi, Kerala, India

3 Department of Computer Applications, Cochin University of Science & Technology, Kochi, Kerala, India

monitoring networks, especially in developing countries such as India, however, make sparse spatial and temporal measurements (CPCB 2020). Any additional information in real time, even if qualitative, would be an advantage for air quality management initiatives in countries like India which has 6 out of the 10 most polluted cities in the world (IQAir 2019).

With increased access to internet and emergence of social media platforms, people are now expressing their views and relation to the events around them like never before. Social media have grown tremendously in number and popularity ever since its emergence and now include multitude of platforms such as Twitter, Facebook, LinkedIn, YouTube channels, blogs, chat rooms, and discussion forums. India with its population of ≃1.3 billion people has over 560 million internet users and is the second largest online market in the world. The internet accessibility and use in the country largely varied based on factors like gender and socio-economic divide (Statista 2020b). Analyzing public behaviour on such platforms are not new to researchers and have been part of routine business analytics (Hassani and Mosconi 2022; Fan and Gordon 2014). Many studies have also explored their potential for monitoring environmental

ORIGINAL PAPER

Fatigue Response of Industrial Waste Mixes for Use as Cemented Base Materials in Flexible Pavement

Rahul R. Pai[1] · S. Patel[2] · J. T. Shahu[3]

Abstract

The fatigue response of cemented base layers is very essential in assessing the required maintenance frequency and the life cycle of bound layers. The flexural properties of industrial waste mixes, namely, steel slag–fly ash–lime (SFL), copper slag–class F fly ash–lime (CFL), and copper slag–class C fly ash (CCF) mixes are highlighted in this study. Unconfined compressive strength (UCS) of 0, 7, 14, 28, 90, 180, and 270 days cured waste mix specimens was determined to study the influence of the curing period on strength enhancement. The shear strength parameters, flexural strength, flexural modulus, and fatigue life were determined for 28 days cured waste mix specimens. The laboratory flexural and fatigue tests were carried out at stress levels of 40, 60, and 80%. Power function gave good relationship for flexural strength with UCS ($R^2 = 0.91$) and cohesion ($R^2 = 0.88$). Furthermore, a reasonably good power relationship was established between flexural modulus and flexural strength ($R^2 = 0.80$). Nondestructive structural evaluation using falling weight deflectometer (FWD) was performed on seven different flexible pavement test sections built with the aforementioned cemented waste mixes in base layers for impulse loads of 40, 55, and 70 kN. Stress-based fatigue performance models based on logarithmic as well as power function were used to predict the fatigue life of these cemented layers in the field based on the results obtained from initial FWD test. The fatigue life of the waste mixes was predicted in the range of 536,657 (70 kN) cycles to 218,623,854 (40 kN) cycles with an anticipated modulus reduction of 50% after fatigue failure.

Keywords Steel slag · Copper slag · Fly ash · Flexural modulus · Fatigue life · Falling weight deflectometer · Field study

Introduction

Natural aggregates are indispensable for the construction, maintenance, and widening of highway projects. The construction of flexible pavement consumes around 15,000 tons of natural aggregates per km length of the road, leading to the inadequacy of the most essential construction component. The extraction of natural aggregates from quarries is a major concern as far as ecological balance is concerned. On the other hand, in a developing country like India, rapid industrialization is leading to the generation of a large quantity of slags from copper/steel industries, and fly ash from coal/lignite-based thermal power stations. Huge open dumps of these industrial wastes not only make the valuable land futile, but also are a threat to the ecology and public health. For every tonne of steel and copper, 0.2–0.3 tons of steel slag and 1.8–2.2 tons of copper slag are generated. At present in India, around 28 million tons of steel slag and 6 million tons of copper slag are generated by the steel and copper industries every year [1]. Moreover, from the 226 million tons of fly ash produced by the 197 thermal power stations in the year 2020, only 19% was being utilized for the construction of roads and embankments [2]. A befitting solution to the aforementioned issues is the mass application of industrial wastes for embankments and road applications.

In the past, extensive laboratory investigations were conducted by several researchers to validate the feasibility of

✉ Rahul R. Pai
  rahul22by7@gmail.com

  S. Patel
  spatel@amd.svnit.ac.in; satya24may@gmail.com

  J. T. Shahu
  shahu@civil.iitd.ac.in; jtshahu@gmail.com

1 Civil Engineering Department, SCMS School of Engineering and Technology, Kochi, Kerala, India

2 Department of Civil Engineering, S. V. National Institute of Technology, Surat, Gujarat 395007, India

3 Department of Civil Engineering, Indian Institute of Technology, Delhi 110 016, India

STABILISATION OF MARINE CLAYEY SOIL USING WATER HYACINTH FIBER

Dr. Santhosh G

Associate Professor, Department of Civil Engineering, SCMS School of Engineering and Technology, Kochi, Kerala, India.

Dr. Sanju Sreedharan

Associate Professor, Department of Civil Engineering, SCMS School of Engineering and Technology, Kochi, Kerala, India.

Akshaya Ramesh, Amal Benny, Amal Hemaraj, Jiya Joy

B. Tech Civil Engineering, SCMS School of Engineering and Technology, Kochi, Kerala, India

ABSTRACT-The main objective of this study is to investigate the use of water hyacinth as a natural fiberin geotechnical applications and to evaluate the effects fibers on strength of clay by carrying out tests on sample. Soil is generally a mixture of four basic types: gravel, sand, clay and silt. Soil typically has low tensile and shear strength and it specific characteristics depends on whether it is currently dry or wet.Replacement of soil with higher quality materials is not always the best alternative because of cost and environmental considerations. Soil stabilisation is a method to improve the properties of weak soils and meet the design requirements.When stabilising soil, there are a number of factors to consider, including the chemical and mineralogical composition, the water content, physical properties of the soil to be stabilised and the intended use of the stabilised soil.

Keywords: Hyacinth fibre, clayey soil,

I. INTRODUCTION

Foundation is an important aspect for any land-based structure and has to be strong to withstand the entire structure; hence the soil around it plays a very critical role. So, to work with soils, we need to have proper knowledge about their properties and factors which affect their behaviour. The process of soil stabilization helps to achieve the required properties in a weak soil needed for the construction work.

Most of the Cochin area, is undergoing rapid industrialization, consists of extremely soft marine clay calling for expensive deep foundations. This paper presents a study on the physical properties and engineering characteristics of Cochin marine clays and its improvement.

Soil stabilization is the process of changing the properties of soil means of mechanical or chemical in order to produce an improved soil material which meets all the desired engineering properties. Soils are generally stabilized to increase their strength and durability and hence prevent erosion. The main aim is the creation of a soil material or system that will hold under the design use conditions and for the designed life of the engineering project.

The main benefits of soil stabilization are increasing strength, reduced susceptibility to swelling and shrinkage, improved durability to weather and traffic, reduces environmental impact, improved handling and compaction characteristics, reduction in plasticity,suitable for stockpiling and subsequent reuse.

Analytical Study on Strength Recovery in Vertically Curved Deficient Tubular Frame Reinforced with CFRP

[1] Tennu Syriac, [2] Shevin N. V

[1]Assistant Professor, [2] P. G Scholar
[1]Civil Engineering Department,
[1] SCMS School of Engineering & Technology, Karukutty, Ernakulam, India

Abstract: Tubular steel structures are having various application in construction. In majority of cases they are used as vertically curved members for bridges, trusses etc. Vertically curved structures are chosen because of their structural performance as well as architectural importance. Deficiencies may appear on structures due to manufacturing defects, deterioration, etc. which results in serious reduction in strength and durability of structures. The most adopted technique for retrofit is by use of Carbon Fiber Reinforced Polymer. This study is focused on the effectiveness of Carbon Fiber Reinforced Polymer (CFRP) in recovering the reduction in strength of a vertically curved steel tube due to deficiency. The deficiency is in the form of a small slit of width 5mm at the bottom of tube at center. The increase in load carrying capacity of frame and the change in behaviour of crack at various configuration of CFRP were analysed using finite element method. Unidirectional CFRP is used for strengthening steel tube.

Index Terms - Ultimate load capacity, Fiber reinforced polymer, Stress intensity factor, Finite element, steel tube.

I. INTRODUCTION

Steel structure is a metal structure which is made of structural steel components connect with each other to carry loads and provide full rigidity. Because of the high strength grade of steel, this structure is reliable and requires fewer raw materials than other types of structure like concrete structure and timber structure. Although steel structures have been in existence since mid-19th century, to date there has been little guidance covering the design of curved steel elements. When forces and moments are applied to a curved steel member, number of effects not present with straight members need to be considered. Different effects must be considered if the member is curved. During the last two decades of 20th century, demand for curved steel members in building structures increased considerably. The main advantage of using curved members is the undoubted aesthetic appeal of curved steel section and tubes. Modern building construction frequently uses curved steelwork for aesthetic appeal, structural efficiency or both.

When a structure becomes deteriorated and/or unable to withstand the applied loads, there are two possible solutions: replace the structure, or repair the current one. However, full structure replacement has become an unfavorable choice due to the tight budgets and low resources of most local, state, and federal agencies. Also, full structure replacement imposes other disadvantages due to the disruption of construction, such as: the need for detours and traffic problems. From this it can be concluded that when the opportunity of a cost-effective and easy to implement repair is available, structural engineers, owners, and operators will opt to repair or upgrade the structure rather than replace it.

A wide variety of techniques are available for strengthening of structures. Among the methods the most promising ones are FRP (Fiber Reinforced Polymer) based retrofitting schemes. The FRP's used vary in shape, size, configuration and components present in them. These are effectively implemented for flexural and shear strengthening of components. The main objectives of this work were to study the behavior and effectiveness CFRP strengthening in recovering the ultimate load capacity on a deficient arch frame and to study the behavior of deficiency (crack) on CFRP strengthened deficient arch frame on various configuration of CFRP.

II. LITERATURE REVIEW

Amir Hamzeh Keykha (2019), conducted a study on effect of CFRP strengthening on the structural behavior of the vertical curved SHS steel beams under concentrated load or uniformly distributed load using finite element investigation. To analyze and study the effects of CFRP strengthening on the behavior of the vertical curved steel beams, three-dimensional modeling and nonlinear static analysis were used. The results indicated that using CFRP to strengthen the vertical curved steel beams in some specific locations has a significant effect on increasing the ultimate capacity of these beams. Nonlinear static analysis was carried out to achieve the specimen its failure mode.

Taylor & Francis
Taylor & Francis Group

Check for updates

Impacts of Flood on Water Quality of Periyar River and Remediation Using Natural Fibers

Praseeja Ayyappan Vasantha, Lakshmi Sindhu Anil, Nasna Mohamedali, Radhu Dileep, and Vishnuja Sajive Nair

Department of Civil Engineering, SCMS School of Engineering and Technology, Ernakulam, India

ABSTRACT

Kerala witnessed a major flood in August 2018, which brought severe effects on the environment and ecosystem, especially affecting the quality of rivers. This study aims in analyzing the impact of the flood on water quality of the Periyar River. The samplings for the study is taken from ten points along the banks of the river Periyar and were tested for its physical and chemical properties. The present water quality of the river was compared with that of the previous year and also with water quality at the time of flood to analyze the impact of flood. The tested samples were found to be polluted, and this study suggests the application of natural fibers for treating the turbid water for safe drinking and household purposes. The fibers are added to the polluted water in proportion of 1 mg/L with a settling time of 30 minutes. In this study, the remediation using natural fibers such as rice husk ash and coir was done, and the results indicate that the application of these fibers improves the water quality parameters such as pH and turbidity.

KEYWORDS

Coir fibers; natural fibers; rice husk ash; water pollution; water quality

关键词

椰子纤维; 天然纤维; 稻壳灰; 水污染; 水质

摘要

喀拉拉邦在2018年8月发生了一场大洪水, 给环境和生态系统带来了严重影响,尤其是影响了河流的质量. 本研究旨在分析洪水对Periyar河水质的影响. 这项研究的样本取自Periyar河沿岸的十个点,并对其物理和化学性质进行了测试. 与前一年以及洪水时的水质进行了比较,以分析洪水的影响. 检测样本被发现受到污染,本研究建议使用天然纤维处理混浊水,用于安全饮用和家庭用途. 将纤维按1mg/L的比例添加到污水中,沉淀时间为30分钟. 在本研究中,使用稻壳灰和椰壳等天然纤维进行修复,结果表明,这些纤维的应用改善了水质参数,如pH值和浊度.

Introduction

Kerala, the southernmost state of India which lies on the banks of the Arabian Sea and on the tropic region, experiences a humid tropical wet climate due to the presence of a large number of rainforests. The rainfall in the state is influenced by the South-west and North-east monsoons, and about 90% of the rainfall occurs during the monsoon months. Heavy rain and strong wind are always common every year during the monsoon season in Kerala. The high-intensity storms that occur during the monsoon months lead to heavy discharge in the rivers. Rainfall intensity is found to be increasing in plain regions (Jain et al. 2017), which can affect Kerala even though it has an asymmetric topography. In addition, the continuous and heavy precipitation in the terrains ultimately flows to the rivers, leading to flooding condition. Kerala experienced heavy rainfall from 1 June 2018 to 19 August 2018, which was 53% surplus of normal rainfall during monsoon season (Mishra et al. 2018), which resulted in severe flooding in the 14 districts of the state. The heavy floods led to the death of people, displacement of large number of people, demolition of buildings, landslides, falling of trees, blocking

Fluoride Removal From Aqueous Solutions Using Different Parts Of Phyllanthus Emblica

Ms. Roshni K R

Assistant Professor
Civil Engineering Department
SCMS School of Engineering and Technology, Karukutty, Kerala, India

Abstract : Fluoride is a prominent contaminant in many states of India. It creates some health hazards, so it should be removed from the drinking water sources. In this paper the suitability of Phyllanthus emblica -an ecofriendly adsorbent to remove fluoride from water is analysed. Different parts of Phyllanthus emblica tree like bark, seed, stem and leaf were pretreated to act as an adsorbent, then the treatment efficiency was analysed under varying conditions of adsorbent dosage, pH and contact time. Maximum fluoride removal efficiency of 79.433% was shown by Phyllanthus emblica seeds at an adsorbent dosage of 2 g, initial fluoride concentration of 3 mg/L, contact time of 30 minutes and a pH of 5. The isotherms analysis was also conducted.

IndexTerms - **Fluoride, Phyllanthus emblica, isotherms**

I. INTRODUCTION

The ionic form of fluorine is widely seen in environment. It is an electronegative and highly reactive element in halogen family. It can combine with other elements via ionic or covalent bonds. Most of the monovalent fluoride compounds like NaF, KF etc. are water soluble and some divalent compounds like CaF_2 and PbF_2 is insoluble in water. Fluoride is present in soil strata so it will penetrate to ground water as fluoride ion. Industrial effluents, sewage, coal combustion, precipitation, volcanic activity etc. contribute fluoride in surface and subsurface water bodies.

Presence of fluoride in the surface water, sub surface water and ground water is a serious issue in areas which lie in the geographic belt where abundant fluoride mineral deposits are present. According to IS 10500-2012, the acceptable limit of fluoride in drinking water is 1 mg/L and permissible limit in the absence of an alternate source is 1.5 mg/L. Excessive consumption of fluoride contaminated water may result in harmful effect on body tissues and other body parts especially teeth, bones and soft tissues. Higher concentration of fluoride may cause teeth mottling and some severe health hazards like skeletal damage and neurological problems (Sachan et al. 2014). Dental fluorosis includes discoloration in teeth, pitting and mottling of dental enamel. A small concentration of 0.5 mg/L will adversely affect invertebrates and fishes. Therefore, control of fluoride is necessary for providing safe drinking water for the community. The government of India had initiated National Sub-Mission to provide safe drinking water for people. One of its major area is to control TDS, iron, arsenic, nitrates and fluoride from water. Many states including West Bengal, Rajasthan, Telangana, Jharkhand, Bihar, Uttar Pradesh etc. were affected by fluoride contamination. Government of India also initiated some programs namely National Rural Drinking Water Program, Integrated Fluorosis Mitigation in Madhya Pradesh, Mitigation of Fluorosis in Nalgonda District Villages, Fluorosis Mitigation in Nuapada District in Orissa, Sachetana Plus Mission in Karnataka etc. to control fluoride contamination (Swati Dubey, et al., 2018).

Cost effective and eco-friendly techniques has to developed to remove fluoride from water. There are different commonly used treatment methods that can be adopted for fluoride removal, but in that adsorption technique has been applied widely and offers a highly efficient, simple and low-cost process for fluoride removal from aqueous solutions (Junyong He, et al., 2020). Carbon based adsorbents, metals, biomaterials, metal organic frame work, minerals, etc. can be used as adsorbents for removing fluoride. Phyllanthus emblica, commonly called as Amla is a biomaterial adsorbent. In this study, different forms of Phyllanthus emblica i.e. the leaves, seeds, barks and stem are used as adsorbent and its suitability to remove fluoride is analysed.

II MATERIALS AND METHODS
2.1 Adsorbent

Leaves, stem, bark and seeds of Phyllanthus emblica was collected and washed in tap water to remove dust or sand particle before drying and were sun dried for 3 days. They were powdered and the fraction passing through 300-micron sieve and retained on 75-micron sieve was taken. Powdered leaves, bark, seed and stem was treated in 1N HCl solution and kept in glass beakers for 24 hours. The solution was filtered using filter paper and washed again with distilled water. Then dried in an oven at 1000C for 3 hours to remove excess moisture. Each item was stored separately in air tight containers.
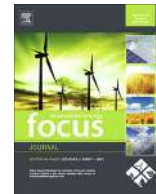
Sustainable power conversion topology based STATCOM for reactive power compensation

Smrithi K. [a,c,*], Jayanand B. [b]

[a] Department of Electrical & Electronics Engineering, Government Engineering College, Thrissur, Kerala, India
[b] Department of Electrical & Electronics Engineering, SCMS School of Engineering and Technology, Ernakulam, Kerala, India
[c] Affiliated to APJ Abdul Kalam Technological University, Kerala, India

ARTICLE INFO

ABSTRACT

The stability of AC voltage is one of the vital drawbacks affecting the power transfer capability. FACTS devices open a new control technology for mitigating this problem to a great extent. But the size, DC link utilization, and stability of the controlling device is still a major concern. The main thrust is to improve the quality of reactive current injected for voltage enhancement in the power system. This paper proposes a new inverter based STATCOM with fewer components and high DC link utilization. This will reduce voltage stresses across the devices. The required DC link voltage is being curtailed by 1.57 times. When compared to multilevel level inverters, less number of switches and one high-frequency switching is required per phase. The simultaneous switching of devices in the same leg eliminates the shoot-through problem. The proposed method improves the quality of injected current for reactive power compensation and load voltage has attained THD (total harmonic distortion) 2%. MATLAB- Simulink platform is used for simulation and verified with experimental results.

Introduction

Power quality is an extensive problem, and it takes years to build up variations through the progression. Quality of power is getting deteriorated with the extended use of power electronic equipment. Power quality complications include current harmonics, voltage harmonics, current imbalance, reactive power compensation, voltage flicker, electromagnetic interference, voltage sag, voltage swells, etc [1–2].

Notable methodical research is going on in this area to reduce these power quality complications. Reactive power causes exorbitant voltage drops in the system and reduces the quality of power. Reactive power compensation can attenuate voltage drop and enrich the power quality, and the system effectiveness. Conventionally passive filters were used for voltage profile improvement. But it is not advisable to use these in variable load applications. With the advancement in power electronic circuits, active power filters are extensively used for mitigating power quality hurdles [3,4].

The generation of well-regulated reactive power using different power converters without the usage of capacitors or inductors was made by Gyugyi in 1976 [1–2 4–5]. In 1988 Hingorani introduced the concept of FACTS technology, and it ameliorate the quality of power in the transmission and distribution system. The reliability as well as steady-state and transient stability of the transmission systems also can be enhanced. FACTS controllers categorize as converter based and impedance-based FACTS controllers [6–8].

STATCOM, a converter-based FACT controller, is represented as a shunt-connected synchronous voltage source. The power transfer capacity within a particular transmission system can be improved by introducing a STATCOM. The reactive output power (capacitive or inductive) of the compensator is amended for regulating the voltage at intended locations of the transmission network. This helps to sustain the expected power flow on contingencies and system instabilities. STATCOM can distribute active power to the system, by drawing regulated real power from an energy source such as a super-conducting magnetic storage, ample capacitor, a battery

Interfaced circuit using a non- destructive method for moisture measurement

Divya Nath K[a],*, Prabhu Ramanathan[b]

[a] SCMS School of Engineering & Technology, Karukutty, Ernakulam, Kerala 683582, India
[b] Universitetslektor, Department of Engineering Science, University West, Nohabgatan 18A, Building 73, Trollhättan, Sweden

ARTICLE INFO

ABSTRACT

Analysing the moisture in stored products like harvested cereal grains and their products, peas, beans, oil-seeds, copra, cocoa beans, spices etc. is very much important to avoid the fungi growth. Moisture can be present in grain in more than one state, i.e. as bound, adsorbed or absorbed water. A designed, integrated circuit was interfaced with personal computer to measure the capacitance which in turn help to calculate the moisture content of rice. The interfaced circuit was tested by measuring the capacitance of different ceramic capacitor. This technique is fast, reliable, accurate and gives hundred set of readings in few seconds. Moisture contents are measured in percentage. The error correction was done with the help of mat - lab programming.

1. Introduction

Agriculture is a main source of income in a developing country like India. The moisture content of grains depends upon the humidity and the temperature of the atmosphere. As the humidity and temperature of the atmosphere varies, it is very much necessary to measure the moisture content of grains frequently. Moisture content is a factor critical to the shelf life. The moisture content in food grains like paddy, rice, etc. should be known by the farmers for its trading, processing and storage else fungi affect the grains and become useless [1,2]. At the time of harvesting, 20% - 40% will be the approximate amount of water content in food items and that has to be reduced to 10% −13% for trading, processing, and storing [3].

The direct or indirect method can be used to measure the moisture content. In direct technique moisture content is calculated using the standard equation (1) or (2).

Dry basis % MC

$$= \frac{(\text{Wet weight of sample } - \text{Dry weight of Sample})}{\text{Dry weight of sample}} \quad (1)$$

Wet basis % MC

$$=, \frac{(\text{Wet weight of sample } - \text{Dry weight of Sample})}{\text{Wet weight of sample}} \quad (2)$$

The difference in the two weights (loss on drying) is then compared with either the original weight (wet-base test) or final weight (dry-base test) and the moisture content are calculated. The direct method does not require calibration, help to measure high levels of moisture and are cheaper than more sophisticated instruments. The disadvantage is that all volatiles are evaporated causing improper moisture content calculations. In this the operator error contributes inaccurate readings. Since in this method heating of grains are done, to conduct this method significant amount of space are required.

In the indirect technique moisture content is calculated by measuring any of the physical property having relation with moisture. For example consider the resistance of the food grains as a physical property which has the relation with moisture, ie resistance will be more when moisture content in grains is less. Using this relation moisture content in food items are indirectly calculated. Moisture content can also be measured either by destructive method or non-destructive method. In non - destructive method, samples are directly used where as in destructive method the food items are cleaned and shells are removed. The time and labour cost are less for non – destructive method of moisture measurement.

The capacitive type moisture meters were used to measure moisture content of food items in the early 20th century. The electrical grain moisture meters were developed considering the dielectric property of the grains. Barley and rice were used as samples [4–6,]. When current is passed through the samples like soil and wheat whose MC have to be measured, samples got cracked and were wasted in electric resistance method. This technique is

* Corresponding author.
E-mail address: divyanathk@gmail.com (D.N. K).

Analyzing the cause of human electrical accidents using Swiss Cheese model

Beena Puthillath

*Department of Electrical Engineering, School of Engineering,
Cochin University of Science and Technology, Kochi, India*

Bhasi Marath

*School of Management Studies, Cochin University of Science and Technology,
Kochi, India, and*

Babu Chembakthuparambil Ayappan

*Department of Electrical Engineering, School of Engineering,
Cochin University of Science and Technology, Kochi, India*

Abstract

Purpose – This study aims to explore the factors influencing electrical accidents. Here, the authors aim to understand and model the causes of electrical accidents at multiple levels.

Design/methodology/approach – In the study, the authors have tried to put causes of accidents in the electricity distribution segment, in the framework of the Swiss Cheese model. Delphi kind of expert survey was conducted to find the Cheese Slice (level) and the causes (holes) for electrical accidents. Inputs from a hundred experts having more than five years of experience in electrical utility companies have been used to find Cheese Slice and holes, to explain the occurrence of an electrical accident.

Findings – Effective training for safe work practices, safe knowledge and closer supervision would go a long way to plug the holes in the Cheese Slice in human factors. The difference in perception of managers, supervisors and workers on the importance of various causes of electrical accidents are also presented and discussed.

Research limitations/implications – This research is based on expert opinion and survey where respondent perception is reported. Actual accident data has not been used here.

Practical implications – The holes or causes of accidents at different levels (Cheese Slice) have been identified for plugging or removal for better safety.

Social implications – Electrical energy is widely used, and therefore, electrical safety is a social concern and also improving it is a social need.

ENERGY GENERATION FROM DAIRY EFFLUENT USING MFC TECHNOLOGY AND PIEZO ELECTRIC TRANSDUCER

Beena Puthillath

Department of EEE, SCMS School of Engineering & Technology, Ernakulam

ABSTRACT: Today the world is facing so many challenges regarding the energy crisis. As the population increases the demand for energy also get elevated. There is increase in energy demand for portable electronic devices and harnessing of alternative renewable source of energy is of interest as it is eco-friendly. This made us think about harnessing energy from different sources like from a microbial fuel cell utilizing dairy waste and vibrational energy from footsteps. For this the use of Piezo-electric crystals are being used for absorbing the pressure energy from the footsteps and converting the corresponding mechanical energy to electrical energy. Another method adopted for energy generation is the use of Microbial Fuel Cell (MFC). This uses the bacterial action to derive bio-electricity from waste. The MFC can generate electricity as well as it contributes to the effective waste management on a large scale. The electricity produced is of low power.

KEYWORDS: Microbial Fuel Cell, Dairy byproduct, Whey, Renewable energy, Piezo-electric sensor

Introduction

The world we are seeing today is facing a lot of undeniable setbacks. For thousands of years we have been relying on fossil fuel for power generation to meet the demand. As the fossil fuels acts as the primary means of energy generation it also causes many complications as well. Climate change is one of the serious issues caused by the over dependence on fossil fuel. The abundant usage causes hike in the pollutants concentration in the atmosphere which will leads to massive increase in temperature on the earth's surface. This is a huge threat to the environment. This is what led us to make us look into the greatly and rapidly developing fields of renewable energy.

The use of Piezo electric modules for the application on footpath can contribute to the vast generation of electricity by installation. Basically, we don't require any expensive machinery and it widely uses the pressure generated from the footsteps of the pressure from the moving vehicles as in advanced applications.

The Piezo crystals has the ability to generate voltage by absorbing the pressure energy and converting to electrical energy. It is called as Piezo-electric phenomenon. The main advantage is that the areas which are needed to generate energy is not isolated or divided from the surroundings, it uses the active surroundings to generate electricity, saving the cost for land and any other premises.

The Prototype can be installed in any place especially crowded region, from road pavements, schools and other educational institutions, offices etc. The practical implementation is such a prototype will inculcate feelings of using cleaner energy, conserving the planet and efficiently utilizing all resources god has gifted us with.

The microbial fuel cell can be used for waste treatment as well as for biological power generation. Microbial Fuel Cell (MFC) uses bacterial metabolism to producing electrical current from organic substrates. Here the byproduct obtained from Paneer making process in a dairy is the fuel to the Microbial Fuel Cell and the analysis of the MFC with different varieties of bacteria has been carried out.

We are using the action of Klebsiella Pneumonae bacteria for the whole process as it yields comparatively high amount of output from other micro-organisms. The Whey paneer used for power generation is treated with different microorganisms in a two compartment cell without any mediators in the microbial fuel cell.

The microorganisms utilized carbonate especially the lactose present in the Whey for generation of bioelectricity. The microbial fuel cell has no intermediary process and hence efficient bioelectricity is produced by this process.

Ref No : IJCRT/Vol 9 / Issue 4/ 332

To,
Beena Puthillath

Subject: Publication of paper at International Journal of Creative Research Thoughts.

Dear Author,

With Greetings we are informing you that your paper has been successfully published in the International Journal of Creative Research Thoughts - IJCRT (ISSN: 2320-2882). Thank you very much for your patience and cooperation during the submission of paper to final publication Process. It gives me immense pleasure to send the certificate of publication in our Journal. Following are the details regarding the published paper.

Thank you very much for publishing your article in IJCRT.

Editor In Chief
International Journal of Creative Research Thoughts - IJCRT
(ISSN: 2320-2882)

REVIEW OFF THERMOELECTRIC POWER GENERATION

DEEPA.S

Associate Professor, Electrical and Electronics Engineering

SCMS SCHOOL OF ENGINEERING AND TECHNOLOGY, KARUKUTTY, ERNAKULAM, INDIA

Abstract: Electrical energy has a vital role in the development of civilization as it is the most refined form of energy. Load demand of the power system varies with the electricity consumption. As the power generation has to meet the varying load demand for an efficient system, various technologies like thermoelectric power generation can be used. Thermal energy can be recovered as electricity by using thermoelectric generator module. It can be related to energy conservation and reduces the maximum demand of energy with power energy management. This paper contains the basic concepts of thermo electric power generation and its applications are reviewed and discussed here.

Key words – TEG, thermoelectric generator, Seebeck effect, thermoelectric materials

1. INTRODUCTION:

For the development of any country, electrical energy has a vital role. Some of the advantages of electrical energy are bulk power generation, economic transmission over long distance and can be efficiently distributed for domestic and commercial loads. Conventionally, electrical energy is obtained by the conversion from burning fuels or by the fusion of nuclear material or by using hydro energy Hydro energy is a replenishable source but is also limited in terms of power.

When exponentially rising population and increasing per capital energy consumption is considered, the non- replenishable fuel sources are not likely to last for a long time. Hence a co-ordinated plan is required so that energy supplied satisfies load requirement with low economical cost. In thermo electric generators, heat energy is directly converted to electrical energy. Some of the advantages of thermoelectric materials are

1) As there is no moving part, the working is smooth and silent.
2) Less maintenance is required
3) They are compact in size and pollution free.
4) It can withstand elevated temperatures for hours of steady state operations.
5) In rural areas where electricity is a problem, this technology is suited for small scale and remote applications.

As TEG has low conversion efficiency of 5% to 8 %, its usage is restricted to specialized fields where reliability is a major concern. Drawbacks are expensive, large output resistance and large variation in temperature .In this paper, fundamental concepts of thermo electric power are discussed and a brief overview of several applications of thermoelectric, which include thermoelectric wooden stove, power generation from waste energy etc.

2. METHODOLOGY

a. Seebeck effect

Heat is applied to two different conductor materials with a junction of difference in temperature which will generate electric current. The voltage magnitude depends on temperature difference and type of the material used. Figure (1) shows the voltage generation using Seebeck effect.

Generation of Electricity from Waste Heat

Deepa S[1], Mohammad Rizan[2], Amaldev M Lal[3], Tritin Thomas[4], Fathima Hussain[5]

Associate Professor, Department of Electronics and Electrical Engineering[1]

Students, Department of Electronics and Electrical Engineering[2,3,4,5]

SCMS School of Engineering and Technology (of KTU), Ernakulam, Kerala, India

Abstract: *Thermo electric generation converts heat energy into electrical energy . Power generated from TEG depends on the temperature difference between hot and cold surface . To improve the efficiency of TEG, MPPT algorithm with boost converter is used . Maximum power is obtained in the system when the output resistance of the system matches with the input resistance of TEG. By modelling the power variations generated from TEG system in series and parallel were minimized . The proposed system consists of TEG with boost converter having P& O MPPT . This paper presents simulation model of TEG module using MATLAB and is successful in generating a stable output.*

Keywords*:* TEG, MPPT, MATLAB

I. INTRODUCTION

Heat energy from waste heat source is directly converted into electrical energy in a thermoelectric generator .Seebeck affect is the basic principle behind the thermo electric generation [1]. By using series and parallel combination of the TEGs , power generation can be increased.[2].

The principal advantages of TEG are that they do not have any moving part and they do not produce greenhouse gases. Low efficiency is the major drawback of TEG.[3].Waste heat energy from different sources can be tapped and converted into useful energy. By using TEG system, energy efficiency of automobiles can be improved.[4]

This proposed work uses converters that perform both maximum power point tracking (MPPT) and power regulation.. Impedance matching is obtained by making internal resistance of TEG equal to the load resistance [6] .MPPT algorithms are used to maximize the power generation in the systems. The algorithms used for MPPT process in these converters is Perturb and observation (P&O) algorithm [8-9].

From the data sheet provided by the manufacturer maximum power, current, and voltage of TEG for impedance matching are known. They ensure hot and cold surface temperatures that can be reached by TEGs. Seebeck coefficient indicated in the data sheet is one of the most important criteria. Simulation of TEG is done using temperature difference and Seebeck coefficient which in turn give the power value taken from the TEG.

II. TEG

A TE is formed by combining p and n type semiconductors. The TEs are connected in series with each other to increase the voltage. TEGs are made up by connecting the TEs in parallel with the help of ceramic plates in order to raise thermal conductivity. When a temperature difference between the ceramic plates is generated, heat transfer from the hot surface to the cold surface becomes possible. The heat transfer leads to electron flow from the n-type to the p-type semiconductors[1]. Thus, a voltage generated is as follows:

$$VOC = \alpha.\Delta T \quad (1)$$

where VOC is the open circuit voltage (V) of the TEG, α is the Seebeck coefficient (V/K), and ΔT is the temperature difference (K). The temperature difference is between the hot, Th , and the cold surfaces, Tc , ($\Delta T = Th - Tc$)

III. TEG MODELLING WITH MATLAB

3.1 TEG Array

TEG properties: Specifications of TEG are given by the manufacturer.

Hot surface temperature, T_h	250∘C

Automated Robotic Electric Vehicle Charging Machine with Digital Payment

[1]Dr. Divya Nath K, [2]Akshay Dhanesh,

[1]Associate Professor, [2] Student, Department of Electrical Engineering SCMS School of Engineering and Technology, Karukutty, India.

A P J Abdul Kalam Kerala Technological University

ABSTRACT

Day by day the demand for hybrid and electric vehicles are increasing. It became necessary to develop an innovative and human friendly charging infrastructure for steady charging. Today charging of electric vehicle is done by. In future there will be autonomous car, the charging of the car should also be automated. This paper deals with robotic electric vehicle charging machine with digital payment. It explains the pros of automated charging system and its necessity in the coming future. The next part explains the systems which are in presently available in market. The last part deals with the proposed system and it future scope and modifications.

Keywords: electric, vehicles, charging, Automatic, station

I.   INTRODUCTION

Even though the number of electric vehicles is very less compared to IC engine vehicles, the charging process is time consuming and costly as man power is required. Lot of money and effort is invested to develop an autonomous and customer friendly charging system. There are lot of advantages for contactless method of charging. Marketability of automated robotic charging system is being looked into, by the manufacturers. There are many issues like complex vehicle adaption, energy losses, environmental impacts on flora and fauna, electromagnetic radiations need to be solved. Driving long distances with short intervals of charging is necessary to make travelling in electric vehicle more enchanting and user-friendly. In inductive system, high charging ability and low power transmission performance are not feasible. Fast charging or quick charging or DC quick charging is done with conductive process. Next option is battery switch stations, which replace the discharged battery with charged one. Due to the difficulty in integration of grid to vehicle and also of high cost, battery switch option is not implemented.

The Figure 1 highlights number of kilometers / hour loading capacity that the electric vehicle can cover based on the energy consumption and different loading methods [1, 2, 3]. The losses due to discharging and charging are not considered in the comparison. When comparing low and fast charging technologies, the charging time can be reduced significantly for fast charging system for the load up to 170 KW.
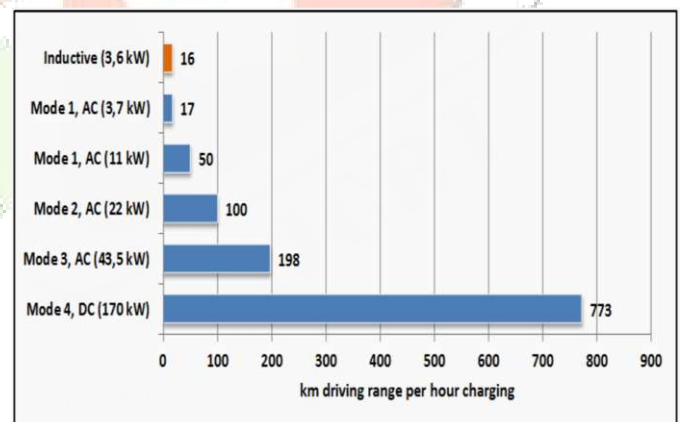

Figure 1: Charging Capacity by Charging Method

The people face difficulty for charging their electric vehicle when the cable becomes unhandy, and hard. This happens when the diameter of cable size increases. The increase in cable occur when powerful direct current is used for charging [4]. The automated conductive charging process is attractive and customer friendly because of this possibility to load vehicles. New opportunities are offered Autonomous vehicles along with automatic charging systems provides new opportunities [5].

Overtaking Assistance Using Wireless Camera and Sensors (Zenarch Tech)

Priya Venugopal[1] and Deepa S[2]

Department of Electrical Engineering

SCMS School of Engineering and Technology, Karukutty, India

Kerala Technological University, Thiruvananthapuram, Kerala

priyavenugopal@scmsgroup.org[1] and deepas@scmshroup.org[2]

Abstract: *The major area of concern in the advancement of intelligent vehicles is to improve the safety of driving system. This paper comes up with a new way of using visual system for driving assistance by replacing convex mirror with wireless camera and sensors. The new system uses vehicle-to-vehicle OAS which contains both transmitter and receiver systems. It will assist the driver to avoid accidents during overtaking. This is a driver support system which helps you maintain the correct speed and safe distance from other vehicles. .It helps in reduction of accidents and provides safe journey on all weather conditions.*

Keywords: Advanced Driver Assistance System, vehicle-to-Vehicle Communication, Zenarch Tech

I. INTRODUCTION

Due to exponential growth in the number of vehicles, road safety is a compelling issue nowadays. Advanced Driver Assistance System (ADAS) concept in the practical realization of transportation system. It helps to improve the awareness of driver about surroundings and decrease accidents, the proposed system uses low cost cameras as sensors which decreases market price of driving assistance system.

More than 200,000 people lost their life in road accident every year in India. When compared with developed countries, 400 mortalities occur in a day. The main reason for these traffic accidents are careless driving at turnings and dangerous over takings at curves. Government has set up convex mirrors which are fitted at curve points and T sections. But these mirrors are not so effective, mainly due to weather conditions, moist etc. Its effectiveness also depends on a person's distant viewing ability for all these problems, our team found out a solution that causes a drastic change in reducing traffic accidents and to enable safe journey on every weather condition

A road crash occur when a vehicle hit with another vehicle , human, animals or static hindrance such as seedling , post or structure . Traffic crash results in mortality, property impair. It affects economy of both society and the person involved .According to 2013 global survey, there were 125 million traffic mortalities with grievous injuries and breathing with conflicting health impacts. Traffic injuries are evaluated to be the ninth noted threat of life of all age group, worldwide, and are to be forecasted to be seventh leading as it is evident in the statistics majority of the demise from road accidents occur while overtaking. Therefore, by this project we aim to ensure safety during overtaking and thereby reducing the accidents caused due to careless driving. Overtaking assistance using V2V TECHNIQUE is a wirelessly transmitting visual technology that sends video information to the display connected in the vehicle wirelessly. The video can be recorded like a black box method used in many applications so that it can be used for future purposes. An adaptive system can be developed so that people can overtake safely without any trouble on turnings and hair pins. The system makes use of Raspberry pi module to ensure safe ride.

II. MATERIALS REQUIRED

- Raspberry pi 3b+
- 180-degree pi camera

SINUSOIDAL COMMUTATION OF THREE PHASE BRUSHLESS DC MOTOR USING LOW -COST LINEAR HALL EFFECT SENSORS

VARUN JOSE , Asst.Professor

Department of Electrical & Electronics Engineering

S.C.M.S School of Engineering & Technology, Kerala, India

Abstract: Torque ripples in BLDC motor can be eliminated to a great extend by implementing sinusoidal commutation.The paper presents a Low-cost Sinusoidal Commutation drive for BLDC motor actuator for 3rd stage of the PSLV. The method implemented as part of paper uses Linear Hall Sensors to sense the magnetic field intensity and thereby use this information to generate the gating signals .The technique of implementation proposed in [1] is used to drive the BLDC motor in this paper. This method has the advantage of being simple and implementation flexibility . Also the use of Linear Hall Sensors in driving sinusoidally commutated BLDC motors is proposed in [2] . This work uses the information from the Linear Hall Sensors to get the information about the rotor position in conjugation with proposed technique in [2] to drive the sinusoidally commutated BLDC motor actuator system . The simulation of mentioned work is realized in MATLAB SIMULINK .The implementation of proposed method is done using analog control circuits.

Key words : BLDC,SIMULINK , PWM, EPROM .

I. INTRODUCTION

       In Sinusoidal commutation BLDC motor controllers drive the 3 phase motor winding's with 3 phase currents that vary sinusoidally & smoothly as the rotor rotates eliminating torque ripples, commutation spikes present if trapezoidal commutation is employed. Exact information about rotor position is essential to maintain accurate sinusoidal modulation of stator currents as the rotor rotates. The ordinary hall effect sensors provides just a rough measurement of rotor position and are ineffective. High resolution sensors like absolute encoder & resolver can be used instead for extracting exact position information of the rotor. Three phase currents through the winding's can be made sinusoidal by means of PWM inverter . Since above mentioned sensors are really expensive in order to minimize the cost a sinusoidal drive system for BLDC motor has been proposed in this paper which utilizes Low Cost Linear Hall Effect sensor IC as rotor position sensor.

II. BLOCKDIAGRAM

       The overall block diagram of proposed sinusoidal commutation technique is shown in **Fig 2.1**. A set of 3 linear hall effect sensors(kept 120 degrees apart) are used for position detection of rotor.The system consists of linear hall effect sensor IC's ,triangular wave generator ,pwm pulse generation circuit, pwm inverter, comparators for region selection as well as pwm pulse generation, driver circuit , potentiometric arrangement , PI compensator , amplifier + notch filter and current limiter

Role of traversing speed and axial load on the properties of friction stir welded dissimilar AA6101-T6 AND AA1350 aluminium alloys

Ashok Kumar R, R. Pandiyarajan, Raghav G. R, K. J. Nagarajan, Sathish Rengarajan, Satheesh Pandian D & K. Sudhakar

View supplementary material

Published online: 10 Nov 2022.

Submit your article to this journal

View related articles

View Crossmark data

Taylor & Francis
Taylor & Francis Group

Check for updates

Role of traversing speed and axial load on the properties of friction stir welded dissimilar AA6101-T6 AND AA1350 aluminium alloys

Ashok Kumar R[a], R. Pandiyarajan [b], Raghav G. R [c], K. J. Nagarajan [d], Sathish Rengarajan[e], Satheesh Pandian D[a] and K. Sudhakar [f]

[a]Department of Mechanical Engineering, K.L.N. College of Engineering, Sivagangai, India; [b]Department of Mechatronics Engineering, Agni College of Technology, Thalambur, India; [c]Department of Mechanical Engineering, SCMS School of Engineering and Technology, Cochin, India; [d]Department of Mechatronics Engineering, Thiagarajar College of Engineering, Madurai, India; [e]Department of Mechanical Engineering, St. Joseph's College of Engineering, Chennai, India; [f]Faculty of Mechanical and Automotive Engineering Technology, Universiti Malaysia Pahang, Malaysia

ABSTRACT

Traditionally, AA6101-T6 and AA1350 aluminum alloys find prominent place in many industrial applications like electrical bus bar utilities for the sake of better conductivity. Friction Stir Welding is a non-traditional solid-state welding technique that was invented and developed by The Welding Institute. In this analysis, AA6101-T6 and AA1350 alloys have been put on experimentation with friction stir welding by varying axial load (5 kN to 9 kN) and traversing speed (60 mm/min to 120 mm/min). In order to examine the weld quality, properties such as tensile, wear and microstructural analyses have been carried out. From the results, it was observed that the weldment made with 7 kN axial load and 90 mm/min traversing speed exhibited better joint properties. Similarly, the weldment made with 5 kN axial load and 120 mm/min traversing speed exhibited poor joint properties. This has been attained due to the variations in heat input, metals softening and grain coarsening at weld nugget. Taguchi L_9 orthogonal array with ANOVA and Signal to Noise ratio have been adopted to understand the contribution of axial load and traversing speed on the properties of weldments. Results reveal that traversing speed is statistically significant on the properties of weldment, when compared to axial load.

1. Introduction

Friction stir welding (FSW) is a thermo-mechanical method that applies a special and hard-wearing tool to weld the metals (Paidar et al. 2020). It does not require any gas shields or filler materials. In FSW, the rotating tool probe penetrates into the metals; the tool shoulder meets the top surfaces of the metals and traverses in the weld line to generate the frictional heat. **This generated heat softens the material while the rotation of tool probe moves the softened material on both sides resulting with the coalescences of the plates** (Ashok Kumar and Thansekhar 2014). As the aluminum alloys have better strength-to-weight ratio, they are widely applied in many industrial productions especially transport and aerospace industries. **Since solidification under fusion welding is prone to create hot cracking, residual stress, porosity, and embrittlement it is not recommended for welding aluminum alloys**. FSW is highly pertinent to weld aluminum and its alloys, because it eliminates liquefaction **allied issues**. Heat origination during joining determines the properties of joints **and it** highly relies on the welding parameters (Ashok Kumar and Thansekhar 2019).

Higher traversing or welding speed minimizes the heat generation that results in lack of stirring and mixing of materials **leading to** weld flaw at weld nugget (Elangovan and Balasubramanian 2008). Akbari et al. (2016) reported

that decreasing the traversing velocity leads to higher heat input resulting in the formation of the flash defects. Shanmuga Sundaram and Murugan (2010) have found that an upsurge in traversing velocity increases the tensile elongation, whereas the increase in axial load decreases the tensile elongation. Kumar and Kailas (2008) have reported that an increase in axial load improves the coalescence of transferred material in weld nugget and reduces the defect formation during welding. Elangovan, Balasubramanian, and Valliappan (2008) have determined that the flash formation increases with the gradual increment of downward axial force during welding. They have also found that **lesser** as well as higher axial loads **resulted with** reduced tensile strengths.

Various research works have been conducted to join the aluminum alloys by varying traversing speed and axial load. But still there is a lack of knowledge about the **influence** of traversing speed and axial load on the properties of weldments. Hence, this investigation has been exclusively conducted to analyze the effects of traversing speed and axial load on AA6101-T6 and AA1350 joints. Joint properties have been measured by carrying out tensile and wear tests and microstructural analyses. **S/N ratio and ANOVA have been employed** to examine the importance of each parameter on responses.

2. Experimental procedure

In this investigation, AA1350 and AA6101-T6 plates of size 100 mm x 50 mm x 6 mm were employed. **These** plates were **positioned** at retreating and advancing sides respectively (Ashok Kumar and Thansekhar 2019). The High Carbon High Chromium steel tool was applied for joining. The tool geometry used for this study **consists of** 7.5 mm shoulder radius (R), 5.6 mm pin height (h), 2.5 mm pin radius (r) and hexagon shaped pin. Joints were made by varying the axial load or downward thrust force (5 kN to 9 kN) and traversing speed or traversing velocity (60 mm/min to 120 mm/min) and keeping the remaining parameters constant.

To compute the tensile strength of weldment, tensile samples were cut normal to weld direction as per the ASTM E8M-04 standard. Computer controlled 5 Ton Universal Testing Machine was employed to conduct the tensile tests. Wilson Wolpert micro Vickers Hardness apparatus was utilized for measuring the hardness values across the various regions of joints. Abrasion resistance (Rotary drum) tester was applied for carrying out the wear tests **by cutting** 16 mm diameter circular pieces were cut at weld nugget. De-winter inverted trinocular metallurgical microscope was used for microstructural analyses and specimens were prepared by polishing and etching using Keller's reagent.

3. Taguchi technique

Taguchi is a powerful technique **used to** improve the performance of a process with lesser number of experiments and **to find** the optimal values of the responses. **This technique has** a specially designed table called an orthogonal array which minimizes the number of experiments (Table 1). The response values based on the orthogonal array are converted into *S/N* ratios to measure the performance characteristics.

In practice, there are three different functional categories under S/N ratio namely, larger the best, smaller the best, and nominal the best. The sole objectives of this work are to increase the tensile strength and to decrease the weight loss through optimal conditions in both cases. In accordance with these objectives, this study has adopted larger the best and smaller the best functions respectively. As a result, with the derived increase in tensile strength, the *S/N* ratio for *i*th performance characteristic for the *n* observations y_i in each trial can be arrived by the following equation:

Table 2. Response table for *S/N* ratios for Tensile strength.

Level	Traversing speed (mm/min)	Axial load (kN)
1	36.21	35.11
2	38.36	37.54
3	34.36	36.28
Delta	4.01	2.43
Rank	1	2

Table 3. Response table for *S/N* Ratios for Weight loss.

Level	Traversing speed (mm/min)	Axial load (kN)
1	−18.15	−18.61
2	−17.14	−17.67
3	−19.46	−18.46
Delta	2.32	0.94
Rank	1	2

$$\frac{S}{N} = -10 log_{10} \frac{1}{n} \sum_{i=1}^{n} \frac{1}{y_i^2}. \tag{1}$$

Similarly, with the derived decrease in weight loss, the *S/N* ratio for *i*th performance characteristic for the *n* observations y_i in each trial can be arrived by the following equation:

$$\frac{S}{N} = -10 log_{10} \frac{1}{n} \sum_{i=1}^{n} y_i^2. \tag{2}$$

HT The experimental results and respective *S/N* ratio for each response have been **illustrated** in Table 1. Thus, the best level of controlled variables for enhancing tensile strength and minimizing wear are 90 mm/min traversing velocity and 7 kN downward thrust force. **As per the response tables for *S/N* ratios shown in** Table 2 **and** Table 3**, traversing speed is the more dominant welding parameter for tensile strength and weight loss**.

The ANOVA has been mainly adopted to validate the significance of designed models and measure the percentage contribution of controlled variables to objective functions. This analysis has been employed with 95% confidence level. Tables 4 and Table 5 reveal that R^2 values lie between 0.9 and 1, **thereby justifying** the designed models **as** significant and adequate. **Thus, when compared with axial load**, it has been found that traversing speed has much influence upon tensile and wear properties, **as illustrated in** Tables 4 **and** Table 5.

Table 1. Taguchi L$_9$ Orthogonal array.

			Tensile Strength		Weight loss	
Sample	Traversing speed (mm/min)	Axial load (kN)	Result (kN/m²)	S/N ratio (dB)	Result (%)	S/N ratio (dB)
1	60	5	68.3 × 10³	36.6884	7.76	−17.7972
2	60	7	73.6 × 10³	37.3376	7.54	−17.5474
3	60	9	53.8 × 10³	34.6156	9.02	−19.1041
4	90	5	69.7 × 10³	36.8647	7.82	−17.8641
5	90	7	92.7 × 10³	39.3416	6.81	−16.6629
6	90	9	87.9 × 10³	38.8798	6.99	−16.8895
7	120	5	38.8 × 10³	31.7766	10.20	−20.1720
8	120	7	62.7 × 10³	35.9454	8.72	−18.8103
9	120	9	58.5 × 10³	35.3431	9.33	−19.3976

Table 4. ANOVA for Tensile strength.

Factors	DF	SS	MS	F ratio	P value	PCR [%]
Traversing speed	2	1378.9	689.4	4.98	0.05	62.42
Axial load	2	455.8	227.9	0.78	0.5	20.63
Error	4	374.4	93.6			16.95
Total	8	2209.1				

$R^2 = 0.99$

Table 5. ANOVA for Weight loss.

| Factors	DF	SS	MS	F ratio	P value	PCR [%]
Traversing speed	2	7.410	3.7051	7.50	0.023	71.44
Axial load	2	1.410	0.7050	1.4938	0.645	13.59
Error	4	1.553	0.3883			14.97
Total	8	10.373				

$R^2 = 0.98$

DF: Degree of freedom, SS: Sum of squares, MS: Mean of squares, PCR: Percentage of contribution.

4. Results and discussions

4.1. Consequences of traversing speed and axial load on tensile properties

During the joining process, downward thrust force determines the frictional heat generation and traversing velocity **dictates the duration of the exposure of the tool** to the workpieces. **From** Table 1**, it can be observed** that the joint fabricated with a traversing velocity of 120 mm/min and a downward thrust force of 5 kN has poor tensile properties (Sample 7). **This is due to the 5 kN downward thrust force generating inferior frictional heat, and 120 mm/min traversing speed resulting a lower amount of heat effected by shorter exposure time of the tool with workpieces**. Insufficient heat input produces poor plastic deformation, which reduces the breaking and transportation of materials from retreating side to advancing side and **vice versa** during stirring process. Consequently, improper mixing and lack of consolidation occur. This creates tunnel defect and poor diffusion at weld nugget (Figure 1). Hence, poor tensile strength has been observed.

On the other hand, **sample 3 is associated** with **higher frictional heat (9 kN downward thrust force) and lower cooling rate or more interaction time of tool with workpieces (60 mm/min traversing velocity). This combination of parameters** can significantly increase the heat input during welding. Excessive heat input produces an extensive degree of softening (Akbari et al. 2016) that results in **flow of** excessive materials during stirring process and incomplete coalescence. This creates volume defect and weaker bonding at the weld interface (Figure 2) which reduces the tensile properties of the joint. At the same time, 7 kN axial load and 90 mm/min
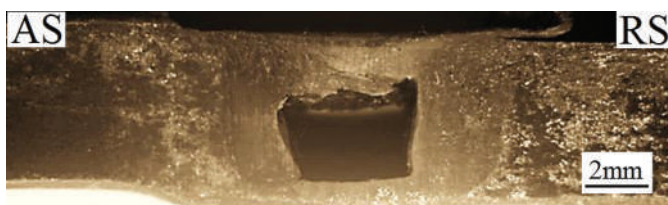
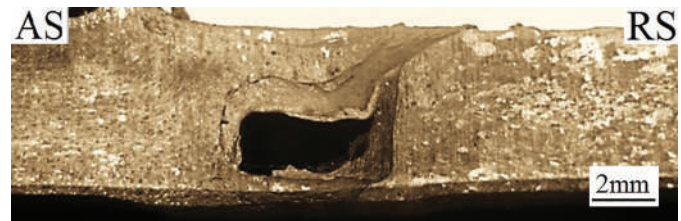Figure 1. Macrostructure of sample 7.

Figure 2. Macrostructure of sample 3.

Figure 3. Macrostructure of sample 5.

traversing speed (Sample 5) result in sufficient frictional heat and adequate contact time of tool with workpieces by which better plastic deformation **is produced**. This improves the breaking and transportation of materials on both sides during stirring. Also, better mixing of materials and complete coalescence at the interface zone occur (Figure 3). Hence, better tensile strength has been observed. **Under the research observations, other samples end up with insufficient or excessive heat input produced by the various axial loads and traversing speeds. Such varied parameters tend to cause improper mixing and poor consolidation or excessive flow of materials, and incomplete coalescence at the stirred area. Consequently, these samples produced lower tensile strength. However, as mentioned previously, sample 5 (7 kN axial load and 90 mm/min traversing speed) could yield the desired result**.

4.2. Consequences of traversing speed and axial load on wear properties

Hardness and percentage of weight loss of stirred zone of entire samples are shown in Figure 4 and Figure 5. From the figures, it **can be** noticed that there are more variations in
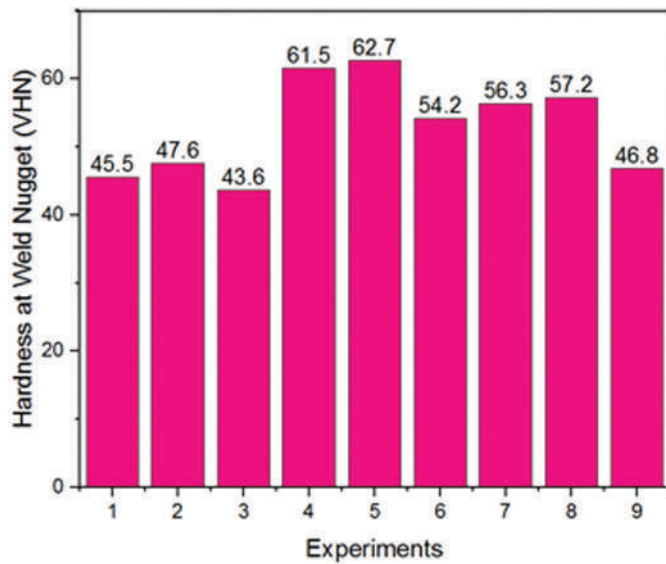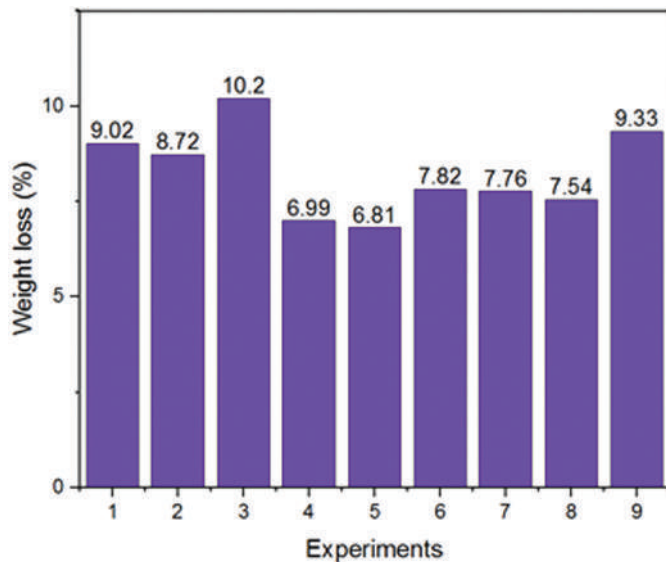
Figure 4. Hardness details of all samples.

Figure 5. Weight loss details of all samples.

drum wear testing. The greater precipitate formation is exposed to the rotational drum, which reduces friction **leading to the attainment of** minimal weight reduction (Figure 5).

At the same time, the precipitates are coarsened and dissolved at a lower heat input produced by the **combined effect** of 120 mm/min traversing velocity and 5 kN downward thrust force, which accelerates the rate of grain growth. As a result, an increase in the size of grains is observed at weld nugget (Figure 6), which reduces the hardness (Figure 4). Therefore, the least amount of restriction **enables** dislocation movement, which softens the metal and increases friction at the weld nugget. Consequently, a higher wear rate is observed (Figure 5). **On the other hand**, 9 kN axial load and 60 mm/min traversing speed make materials get extremely softened due to the high heat input and give additional grain growth after deformation due to the prolonged cooling time (Figure 7). Lower hardness is consequently noted (Figure 4). Hardness and wear resistance are always inversely correlated (Ashok Kumar and Thansekhar 2019). **Due**

Figure 6. Microstructural analysis of stirred zone of sample 7.

hardness and wear rates among all samples. Mg_2Si precipitates occur in the stirred region during the joining of AA6101-T6 and AA1350 alloys (Ashok Kumar and Thansekhar 2019). **Further, in response to heat input, the precipitates occurred thus are coarsened, dissolved, and reprecipitated. The experimental observation shows that** 7 kN axial load and 90 mm/min traversing speed **are capable of generating** sufficient heat to dissolve and cause the reappearance of the precipitates. More reappearance of precipitates reduces grain growth rate (Wu and Zhang 2017). As a consequence, finer grains are found in stirred area (Figure 8) which improves hardness (Figure 4). **Also**, these reappearing precipitates serve as impediments to the mobility of dislocation. **Normally**, the mobility of dislocation determines the plastic deformation of metals. Thus, metal becomes harder **and the harder** surfaces **possess** lower friction during wear testing. **Further**, the reappearing precipitates in weld nugget act as load supporting elements during rotary
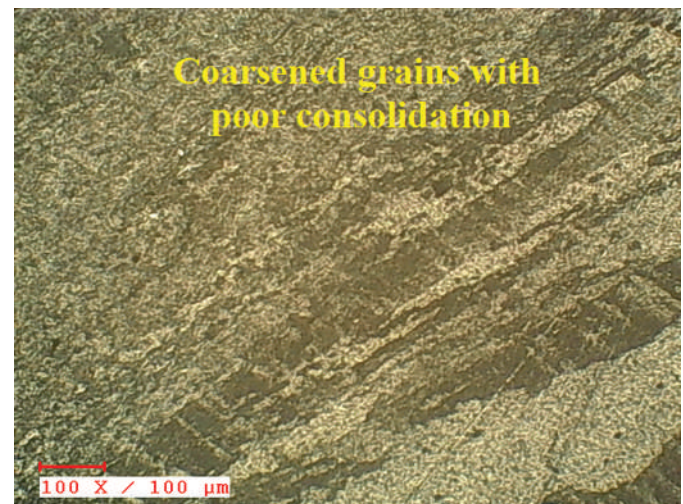

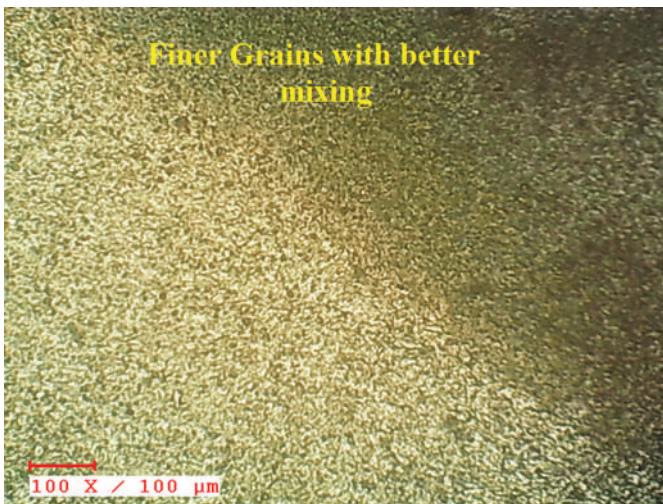Figure 7. Microstructural analysis of stirred zone of sample 3.

Figure 8. Microstructural analysis of stirred zone of sample 5.

to the decreased hardness, less wear resistance is provided during rotary drum wear testing. Thus, greater weight loss is obtained (Figure 5). When compared to the combination of 7 kN axial load and 90 mm/min traversing speed (Sample5), other combinational parameters create unsatisfactory hardness as well as wear qualities. **These drawbacks happen in other samples** due to the coarsening and dissolving of precipitates or the excessive softening of materials (Figure 4 and Figure 5).

5. Conclusions

- A mathematical model for welding of AA6101-T6 and AA1350 alloys has been developed by incorporating welding process parameters using Taguchi orthogonal array, *S/N* ratio and ANOVA.
- From the statistical analysis, when compared to axial load, it has been found that the traversing speed has higher impact on tensile and wear properties.
- Sound weld joints can be achieved with the combination of moderate traversing speed (90 mm/min) and moderate axial load (7 kN). This is associated with sufficient frictional heat, adequate exposure time of tool with workpieces and dynamic recrystallization.
- Similarly, the combinational parameters of 5 kN axial load and 120 mm/min traversing speed have exhibited poor weld properties. This is related to the lower heat input, shorter interaction time of tool with workpieces and coarsening of precipitates.
- Due to the poor plastic deformation or excessive softening of materials, other combinational parameters have produced poor mechanical properties.

ORCID

R. Pandiyarajan http://orcid.org/0000-0002-4026-1163
Raghav G. R http://orcid.org/0000-0001-6028-3979
K. J. Nagarajan http://orcid.org/0000-0002-1153-4214
K. Sudhakar http://orcid.org/0000-0002-4867-2362

Nomenclature

ANOVA	Analysis of variance
FSW	Friction stir welding
h	Pin height
r	Pin radius
R	Shoulder radius
R^2	Correlation coefficient
S/N ratio	Signal to noise ratio

References

Akbari, R., S. Mirdamadi, A. Khodabandeh, and M. Paidar. 2016. "A Study on Mechanical and Microstructural Properties of Dissimilar FSWed Joints of AA5251–AA5083 Plate." *International Journal of Materials Research* 107 (8): 752–761. doi:10.3139/146.111400.

Ashok Kumar, R., and M. R. Thansekhar. 2014. "Effects of Tool Pin Profile and Tool Shoulder Diameter on the Tensile Behaviour of Friction Stir Welded Joints of Aluminium Alloys." *Advance Material Research* 984-985: 586–591.

Ashok Kumar, R., and M. R. Thansekhar. 2019. "Wear Behaviour of Friction Stir Welded Dissimilar Aluminium Alloys." *Metallophysics and Advanced Technologies* 41 (2): 203–211. doi:10.15407/mfint.41.02.0203.

Elangovan, K., and V. Balasubramanian. 2008. "Influences of Tool Pin Profile and Welding Speed on the Formation of Friction Stir Processing Zone in AA2219 Aluminum Alloy." *Journal of Materials Processing Technology* 200: 163–175. doi:10.1016/j.jmatprotec.2007.09.019.

Elangovan, K., V. Balasubramanian, and M. Valliappan. 2008. "Influences of Tool Pin Profile and Axial Force on the Formation of Friction Stir Processing Zone in AA6061 Aluminium Alloy." *International Journal of Advanced Manufacturing Technology* 38: 285–295. doi:10.1007/s00170-007-1100-2.

Kumar, K., and V. Kailas. 2008. "On the Role of Axial Load and the Effect of Interface Position on the Tensile Strength of a Friction Stir Welded Aluminium Alloy." *Materials & Design* 29: 791–797. doi:10.1016/j.matdes.2007.01.012.

Paidar, M., R. V. Vignesh, A. Moharrami, O. O. Ojo, A. Jafari, and S. Sadreddini. 2020. "Development and Characterization of Dissimilar Joint between AA2024-T3 and AA6061-T6 by Modified Friction Stir Clinching Process." *Vacuum* 176: 1–6. doi:10.1016/j.vacuum.2020.109298.

Shanmuga Sundaram, N., and N. Murugan. 2010. "Tensile Behavior of Dissimilar Friction Stir Welded Joints of Aluminium Alloys." *Materials & Design* 31: 4184–4193. doi:10.1016/j.matdes.2010.04.035.

Wu, Q., and Z. Zhang. 2017. "Precipitation-Induced Grain Growth Simulation of Friction-Stir-Welded AA6082-T6." *The Journal of Materials Engineering and Performance* 26: 2179–2189. doi:10.1007/s11665-017-2639-1.

On the design of a compact emergency mechanical ventilator with negative expiratory exit pressure for COVID-19 patients

Sheeja Janardhanan, Vidya Chandran & Rajesh Rajan

Published online: 27 Jul 2022.

Submit your article to this journal ⇗

View related articles ⇗

View Crossmark data ⇗

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

Check for updates

On the design of a compact emergency mechanical ventilator with negative expiratory exit pressure for COVID-19 patients

Sheeja Janardhanan[a], Vidya Chandran[b] and Rajesh Rajan[c]

[a]School of Naval Architecture and Ocean Engineering, Indian Maritime University, Visakhapatnam, India; [b]Department of Mechanical Engineering, SCMS School of Engineering and Technology, Karukutty, Ernakulam, India; [c]Department of Cardiology, Sabah Al Ahmed Cardiac Centre, Kuwait City, Kuwait

ABSTRACT

The present work deals with the design of a cylinder-piston arrangement to deliver the required tidal volume (TV) of air to the patient through the respiratory tract especially in the setting of severe acute respiratory syndrome corona virus 2 (SARS CoV-2) or corona virus disease (COVID-19). The design ensures that only the desired volume of air is delivered in each breath and a negative pressure is retained at the delivery point in a separate cylinder. The frequency of piston motion is the same as that of the average human respiratory rate (RR). The effect of negative pressure on time of evacuation under the present condition has been verified. The present design provides a compact ventilator unit with a surface area of 0.8×0.4 m^2 with a minimal power requirement of 116.48 W. An RR of 16 is obtained with a volume flow rate in lit/s by using a twin cylinder arrangement with bore diameter 0.1 m and length 0.4 m. The ratio of inspiration time to expiration time is designed to be 1:2 by controlling the stroke frequency as 16 and piston speed 0.32 m/s. The present design provides promising quantitative information on the design of an automated continuous mechanical ventilator (CMV), which is different from bag mask valve (BMV) operated ventilators, and on preventing and minimising barotrauma.

1. Introduction

We are all stranded in our homes due to the severe acute respiratory syndrome corona virus 2 (SARS CoV-2) corona virus disease (COVID-19) pandemic, as a measure for containing the virus. Due to lack of proper immunisation against this novel virus, the entire human race is at risk. Moreover, with the number of infected people doubling every few days, our medical systems and devices are being overburdened. As a result, infected people are not receiving adequate care and treatment, resulting in more and more fatalities. Lack of sufficient ventilators is a serious concern. Only those patients under 60 years of age were considered for treatment with ventilators while the rest were left without [1]. Such heart-rending situations give us much motivation to produce and provide the most needed equipment for such an hour of crisis. A typical emergency mechanical ventilator comprises an artificial manual breathing unit (AMBU) bag used to pump air into the nose of the patient through a mask [2]. In cases where the patient is incapable of initiating a breath, supplying air through an endotracheal tube (ETT) becomes necessary [3]; this is an invasive procedure. In the case of an AMBU bag, it is normally compressed manually to discharge air into the patient's lungs. During a pandemic of an infectious disease, such as COVID-19, human operation of the ventilator would put the health personnel at risk of contracting the virus [4]. The only solution is thus to produce compact and power-driven emergency mechanical ventilators operating on the principles of a continuous mechanical ventilator (CMV).

This paper presents the complete methodology of operation of the proposed device with evidence of its operation in realistic pressure ranges. On a broader perspective, the paper provides input on the design of an emergency ventilator that offers not only a safe and convenient mode of operation but also ensures there is no side-effect to the patients, which is otherwise a significant drawback of mechanical emergency ventilators [5].

CONTACT Vidya Chandran ✉ vidyachandran@scmsgroup.org ▣ Department of Mechanical Engineering, SCMS School of Engineering and Technology, Vidya Nagar, Karukutty, Ernakulam, Kerala 683576, India

2. Methods

For an efficient design of ventilators, an understanding of the pressures on the inside and outside of human lungs is important. Figure 1 shows the fluid pressures under which a normal human lung operates.

Pressure inside the lungs is expressed in gauge values. Pressure in the thorax, P_T is always negative. Pressure in the lung, P_L is positive when inhaling (inspiration), and reduces to a value slightly above atmospheric pressure, P_{atm} when exhaling (expiration) [6]. The process is clearly explained in the literature [7]. After expiration, there will be a positive pressure corresponding to the volume of air remaining in the lungs called the intrinsic or auto positive end expiratory pressure (PEEP). In the case of patients with airflow obstruction following a medical condition such as acute respiratory distress syndrome (ARDS) or pneumonia, auto PEEP will eventually be a higher value [8], leading to further distress known as barotrauma.

Total pressure in the thorax and lungs, P_{LT}, is given by the following equation:

$$P_{LT} = P_L + P_T \tag{1}$$

The pressure variation inside the lungs during inspiration and expiration and over a respiratory cycle is shown in Figures 2 and 3, respectively.

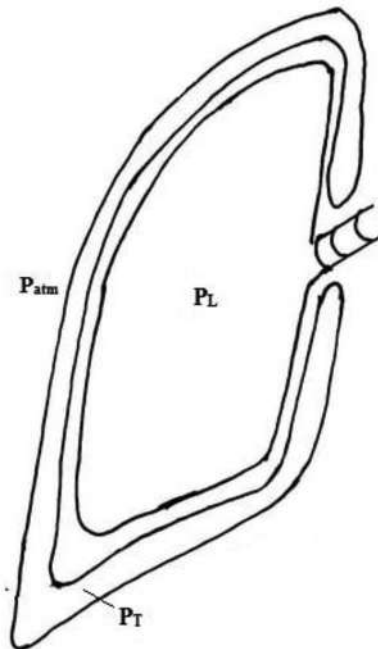Here, the following interpretations are made: lungs and the thorax are coupled together; intrapleural pressure, P_{Plueu}, is the same as the thorax pressure given by the following equation:

$$P_{Plueu} = P_T \tag{2}$$

Intrapulmonary pressure, P_{Pul}, refers to the lung pressure and is given by the following equation:

$$P_{Pul} = P_L \tag{3}$$

A comprehensive new design of a motor-driven piston compresses air inside a cylinder while executing horizontal simple harmonic motions with a frequency comparable to the respiratory rate (RR) [9]. The cylinder compresses the air to be delivered through the ETT or the mask and sucks back the expired air creating a negative end expiratory pressure which aids in easy expiration. It has been found by clinical experts that COVID-19 results in ARDS and pneumonia [10]. Also, the exhaled breath is proven to be a cause of airborne diseases due to the presence of droplets [8]. Hence, in this work a separate cylinder and piston arrangement is proposed for discharging exhaled air for COVID-19 patients which is likely to contain minute droplets of the infected fluid from lungs. The system ensures that the minimum required tidal volume (TV) of air is discharged into the lungs of the patient with respiratory distress which starts as dyspnoea.

The paper presents the mechanical construction and working of a compact cylinder-piston compression
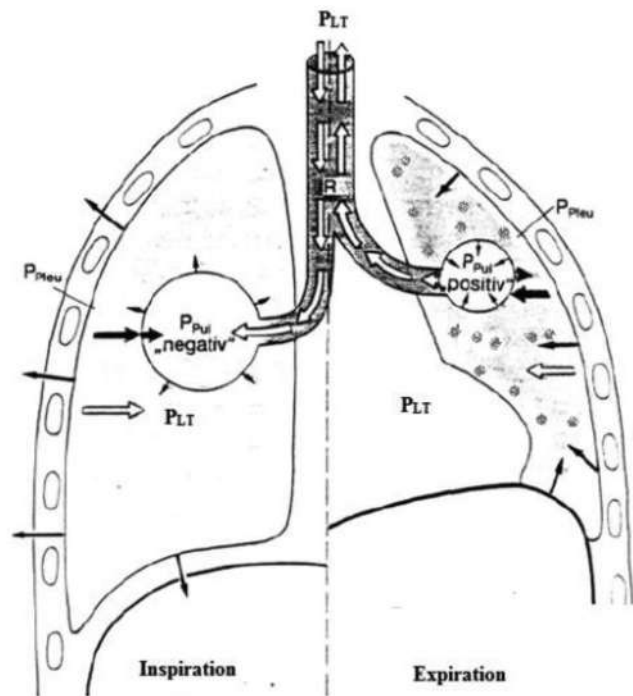

Figure 1. Pressure inside a lung.

Figure 2. Pressure variation inside a lung during inspiration and expiration [6].
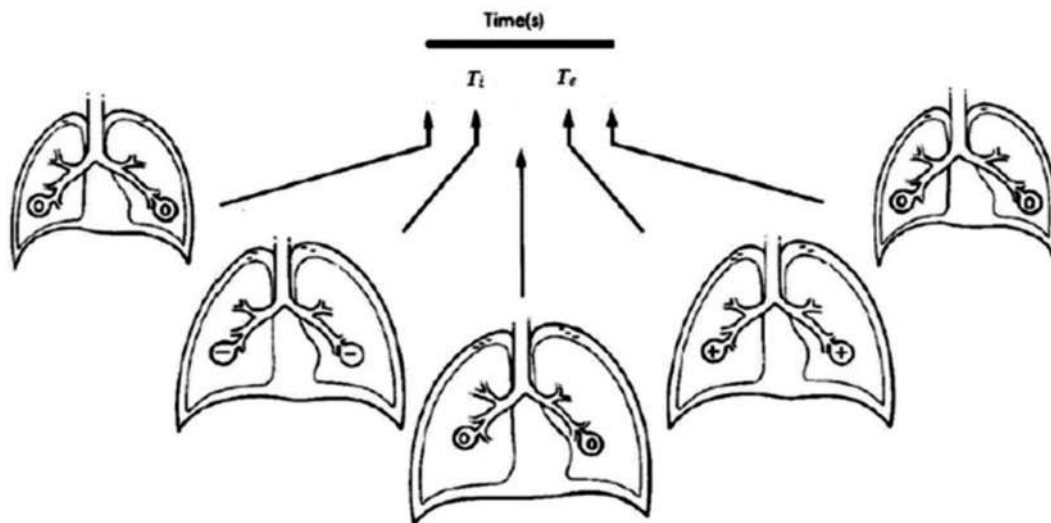
Figure 3. Pressure changes in the lungs during normal breathing during a respiratory cycle [9].

system for delivering the required TV of air to the distressed lungs, meeting the physiological requirements of the respiration process. As expiration needs more time than inspiration, care should be exercised not to overburden the lungs with the next volume of fresh air before most of the exhaled air is drained from the lungs. When the lungs expand to hold the excess volume of air, the inner tissues of the lungs, the alveoli, collapse, resulting in another medical condition of the lungs called barotrauma. The emergency mechanical ventilators built to be easily portable do not usually possess apparatus to measure and control flow. The focus of every design so far has been on the pumping mechanism. The major drawback of such devices is that the inspiratory cycle commences well before the end of the expiratory cycle. Thus, it is no surprise that barotrauma is a side-effect of such emergency ventilators.

The present paper addresses this issue by proposing a compact mechanical device which not only pumps air to the lungs during the time of crisis but also retains a negative pressure at the exit from the lungs, assisting the lung evacuation process and thus reducing barotrauma.

2.1. Constructional and working principle

In the case of COVID-19, the corona virus is both highly virulent and highly infectious. Under such circumstances, exhaled air needs to be properly treated. The air leaving the lungs of COVID-19 patients contains fine droplets capable of transmitting diseases over both short and long distances [6]. With the highly infectious nature of this disease, it is urgent to handle the exhaled air separately. The constructional and working principles of the proposed design are illustrated in Figure 4.

The proposed ventilator works on the principle of negative suction pressure at the end of the expiratory tube. There are two cylinders with a piston arrangement, driven by a servo motor. The inspiration cylinder is connected to the inspiration tube; it also has an inlet valve for taking in fresh air. The inspiration tube connects the cylinder opening and the ETT through an inspiration valve. The second cylinder is the expiration cylinder, and its opening is connected to the expiration tube through a valve to the ETT. The ETT connects the trachea of the patient to the inspiration and expiration tubes. All valves are non-return by nature, permitting flow in only one direction. In the case of extreme emergency, the ETT can be replaced with a mask.

The motion of the piston from outer dead centre (ODC) to inner dead centre (IDC) is called the compression stroke and that from IDC to ODC is called the expansion stroke. Initially, both pistons execute a compression stroke, starting from the ODC, corresponding to the minimum pressure inside the cylinder (-15 cm H_2O).

The inspiration cylinder delivers the air into the inspiration tube, and the expiration cylinder flushes the air out through the expiration outlet, marking the completion of the inspiration process in the lungs and the initiation of expiration. During the next stroke inside the cylinders, as the inspiration piston displaces from IDC to ODC, it admits a fresh volume of air while the expiration cylinder retains a negative pressure with no fluid inside. Both the pistons are kept on hold
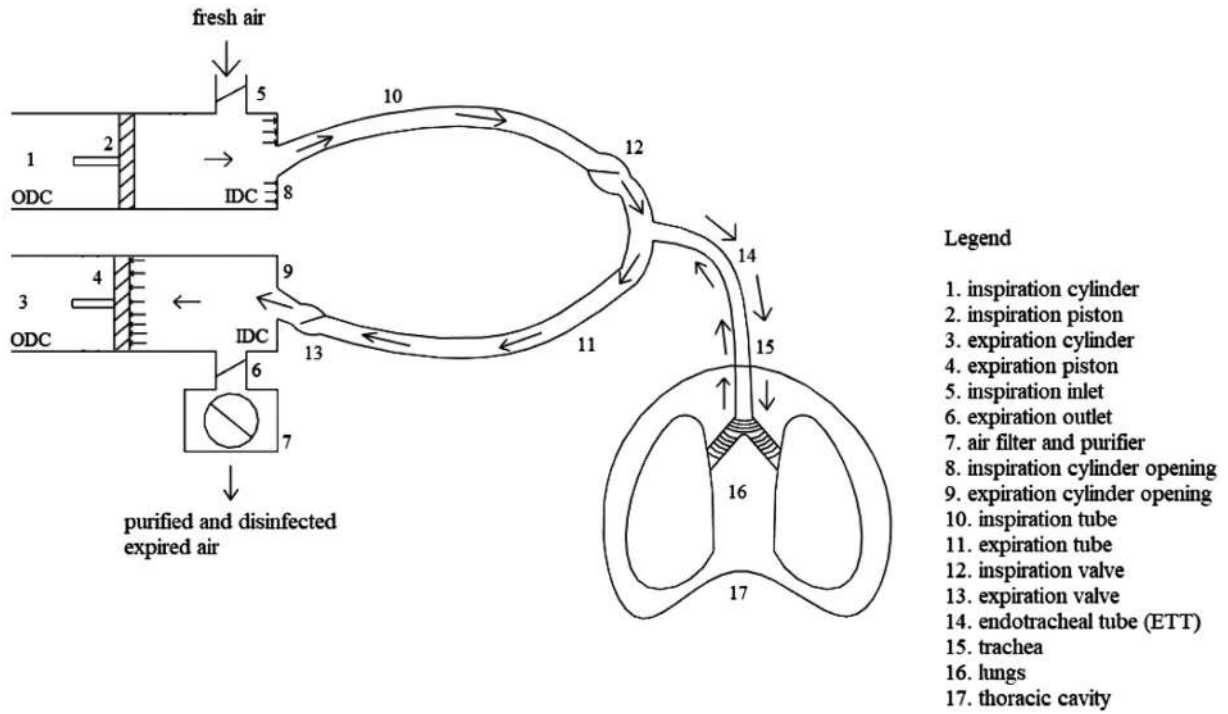
Figure 4. Emergency ventilator model for COVID-19 patients.

till the expiration is completed, allowing the prolonged time for the expiration process. The expired air is admitted into the expiration cylinder through the expiration valve; the pressure inside the expiration cylinder at the end of the expiration process is negative and aids in faster lung evacuation. As the next cycle begins, the inspiration piston displaces compressed air into the inspiration tube and the expiration piston drives the exhaled air of the previous cycle through the expiration outlet. Exhaled air is treated and purified and then let out into the surroundings. The cycle then repeats.

The calculations needed to arrive at the design specifications of the ventilators are shown in the subsequent sections.

2.2. Estimation of pressure range during respiration

Emergency ventilators are a type of CMV with a provision to set the volume and control the pressure. The pressure ranges in which the device works are of paramount importance here and these ranges are interpreted from the physiological lung pressures presented in Section 2. The changes in the lung condition owing to dyspnoea or pneumonia are also accounted for and the calculations carried out and presented as in Gattinoni et al. [11].

The inspiration pressure is given by the following equation:

$$P_{LO} = P_{CF} + P_{ST} + P_{CW} \qquad (4)$$

where P_{CF} is the compressive force applied to the ventilator, P_{LO} is the lung opening pressure ($+45$ cm H_2O), P_{ST} is the surface tension in the airway ($+20$ cm H_2O) and P_{CW} is the chest wall pressure (10 cm H_2O) and can be calculated as $+15$ cm H_2O.

The lung compliance is an indication of the condition of the lungs and is expressed by the following equation:

$$C = \frac{1}{E} = \frac{\Delta V}{\Delta P} = \frac{V_T}{P_P - P_{Tot}} \qquad (5)$$

where C is the compliance of the distressed lung (0.2 l/cm H_2O), E is the elastance of the lung, V_T is the TV (2.2 l), P_P is the plateau pressure (32 cm H_2O) and P_{Tot} is the total PEEP of the ventilator, calculated as 20 cm H_2O from Equation (5). P_{Tot} is, however, the sum of internal and external PEEP and is expressed by the following equation:

$$P_{Tot} = P_I + P_E \qquad (6)$$

where P_I is the intrinsic/auto PEEP of the lungs, usually 5 cm H_2O, and P_E is the external PEEP of the ventilator, the pressure available at the end of the ETT, calculated as 15 cm H_2O from Equation (6); the same negative pressure is retained at the exit of

Table 1. Lung activity corresponding to piston positions.

Tr (s)	Cylinders	Piston positions ODC IDC	Activity of lungs
0	IC		Inspiration commences following a reduction in pressure inside the lungs from P_{atm} to P_T (negative).
	EC		
1.25	IC		Inspiration process completes at P_I in the inspiration and the lungs prepare for expiration. The pressure inside the lungs increases from P_{atm} to P_L (positive).
	EC		
2.5	IC		The expiration process is in progress with negative pressure, P_E established in the expiration cylinder.
	EC		
3.75	IC		The expiration completes at the exit negative cylinder pressure, P_E and the lungs are at auto PEEP condition.
	EC		

the expiration tube in the present ventilator. Hence, the working pressure of the device is fixed in the range (−15, +15) cm H_2O. Lung activities corresponding to piston positions are summarised in Table 1.

3. Design of the cylinders

3.1. Pressure of the ventilator

The ventilator proposed and presented here has the same working principle as that of a reciprocating air compressor operating with maximum pressure +15 cm of H_2O, and minimum pressure −15 cm of H_2O (vacuum). The pressure–volume (P–V) diagram for the cylinder is as shown in Figure 5.

In the P–V diagram represented by Figure 5, the process 1–2–3–4 represents the typical working cycle of the ventilator on the principles of an air compressor, while 1′–2′–3′–4 is the actual working cycle of the inspiration cylinder. Here, 1′–2 is the compression; 2–3′ is the delivery to the inspiration tube; 3′–4 is the expansion and 4–1′ is the suction. A similar process happens in the expiration tube with the difference
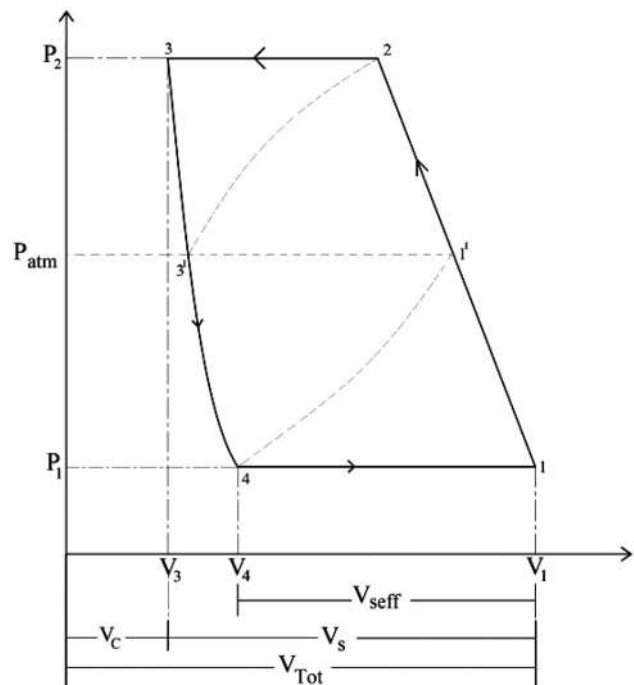

Figure 5. P–V diagram of the ventilator.

being that during suction, expired air is admitted, and during delivery the exhaled air is driven out through the expiration outlet.

The volume of air delivered to the lung is decided by the volume of the cylinder, termed as swept volume (V_s), which is the volume swept by the pistons during its compression and expansion strokes. The P–V diagram of a normal compressor demonstrates that the effective swept volume (V_{seff}) will always be less than V_s. The ventilator is intended to deliver the sufficient TV V_T of air to the patient. Hence, V_{seff} of the cylinder should be made equal to the required V_T. From the P–V diagram given in Figure 5, V_{seff} is expressed as in the following equation:

$$V_{seff} = V_1 - V_4 \tag{7}$$

This volume is identified as the TV, V_T, and is expressed as in the following equation:

$$V_1 - V_4 = V_T \tag{8}$$

The volumetric efficiency (η_{vol}) of an air compressor is defined as presented in the following equation:

$$\eta_{vol} = \frac{V_{seff}}{V_s} = \frac{V_1 - V_4}{V_1 - V_3} \tag{9}$$

η_{vol} of an air compressor is normally in the range 60–85% [12]. The V_s can be obtained as presented in the following equation:

$$V_1 - V_3 = \frac{V_1 - V_4}{\eta_{vol}} = \frac{V_T}{\eta_{vol}} \tag{10}$$

In the present design, $V_T = 2.2\,l$ and $\eta_{vol} = 70\%$.

For the estimation of the optimum design parameters of the ventilator cylinders, such as the stoke length of the cylinder, L, and the bore diameter, D, the displacement volume, V_D, of the compressor is kept the same as that of V_S. Hence V_D can be obtained as in the following equation:

$$V_D = \frac{\pi D^2}{4} L = V_s = V_1 - V_3 \tag{11}$$

Since both inspiration and expiration cylinders are to operate within the same pressure ranges and also with the same volumes, these cylinders are identical in terms of all the parameters estimated above.

3.2. Estimation of operating characteristics of cylinders

The frequency of reciprocating motions (f_{rec}) of the piston is estimated, taking into consideration the normal RR required for keeping a patient stable as shown in the following equation:

$$f_{rec} = \frac{1}{T_r} \tag{12}$$

T_r is the total time for one complete respiration. The variation in the time required for inspiration and expiration is also taken into account. The RR is expressed as n breaths/minute and varies according to sex and age [13]. Time of inspiration (T_i), which is defined as the time required to complete one single process of inhalation, is calculated as shown in the following equation:

$$T_i = \frac{60}{3n} \tag{13}$$

Time of expiration (T_e) is defined as the time required to complete one single process of exhalation. For $T_i{:}T_e{=}1{:}r$, T_e is given by the following equation:

$$T_e = r\left(\frac{60}{3n}\right) \tag{14}$$

The present ventilator is designed to deliver 2.2 l of air to the lungs in $T_i = 1.25\,s$ and to exhale almost the same volume from the lungs in $T_e = 2.5\,s$ as per Equations (13) and (14), corresponding to $n = 16$ (for adults on a gender-based average) and $r = 2$. Even though the inspiration and expiration time periods are different, the ventilator cylinders are designed to operate with the same time period of reciprocating motion. The time lag in the exhalation process is accommodated by holding both inspiration and expiration cylinders for an additional 1.25 s after the expansion stroke. The time period of piston reciprocating motion for both cylinders is fixed at $T_p = 1.25\,s$, which is the same as the inspiration time.

Volume flow rate for the inspiration cylinder Q_{IC}, and expiration cylinder Q_{EC}, can be obtained from Equations (15) and (16), respectively.

$$Q_{IC} = \frac{V_s}{T_i} \tag{15}$$

$$Q_{EC} = \frac{V_s}{T_e} \tag{16}$$

Volume flow rate through the ventilator can also be represented by the following equation:

$$Q = \frac{\pi D^2 L}{4} \frac{N}{60} \tag{17}$$

N is the speed in rpm. Equation (17) can also be expressed in terms of V_D as in the following equation:

$$Q_{IC} = V_D \frac{N}{60} \tag{18}$$

From Equations (16)–(19), the speed (N) of the cylinder is estimated.

Fixing a suitable value for the bore diameter (D), stroke length (L) can be obtained from Equation (18). The cylinders are designed to be geometrically

identical. The speed of the piston (U_P) during reciprocating motion for a single acting piston is obtained from the following equation [14]:

$$U_P = LN \tag{19}$$

Assuming that the entire volume of air in the cylinder is delivered at the end of an inspiration (compression) stroke, the clearance volume (V_C) of the cylinder becomes negligible, and the total volume (V_{Tot}) will be equal to V_s. Hence, the volume, V_1 as given in the P–V diagram (Figure 5) can be estimated as in the following equation:

$$V_1 = V_{Tot} = V_s = V_1 - V_3 \tag{20}$$

3.3. Estimation of ventilator power requirement

Work done (W) by the compressor is given by the following equation [14]:

$$W = \frac{\gamma}{\gamma - 1} P_1 V_1 \left[\left(\frac{P_2}{P_1} \right)^{\frac{\gamma-1}{\gamma}} - 1 \right] \tag{21}$$

P_1 and P_2 are the maximum and minimum pressures of the compression process, respectively. P_1 and P_2 are estimated in Section 2.2. Considering the polytrophic process, γ is the polytrophic constant for air which is taken as 1.35 [14]. Work done by both cylinders will be the same. Power required (P) for operating the cylinder in kW can be obtained from the following equation:

$$P = \frac{WN}{1000} \tag{22}$$

4. Results

4.1. Validation of lung evacuation

Based on the operating characteristics obtained from the design analysis, a pressure–time graph is plotted for the ventilator; pressure inside the cylinder is calculated over the complete respiratory cycle and is plotted against time. The graph is presented in Figure 6. It can be observed that the pressure plot of the cylinder is in phase with the pressure time plot of the lungs. The comparison is made in Figure 6.

A model of the ventilator expiratory cylinder is fabricated. The effect of negative pressure is verified by connecting the expiratory cylinder to an AMBU bag filled with 500 ml of air using a pump. The volumetric strain on the AMBU bag, $\in_V$, can be calculated analytically using the following equation:

$$K = \frac{-dP_A}{\in_V} \tag{23}$$

where K is the isothermal bulk modulus of air taken as 101 kPa, and dP_A is the change in pressure on the AMBU bag for a given volumetric strain, $\in_V$. The pressure corresponding to total PEEP, P_{Tot} pressure (20 cm H_2O), is achieved in the AMBU bag by subjecting it to 5% volumetric strain, $\in_V$. The time required to evacuate the AMBU bag for various pressure outlet conditions is recorded as shown in Figure 7.

Design specifications of the proposed mechanical ventilator, estimated from Equations (1)–(23) are presented in Table 2.
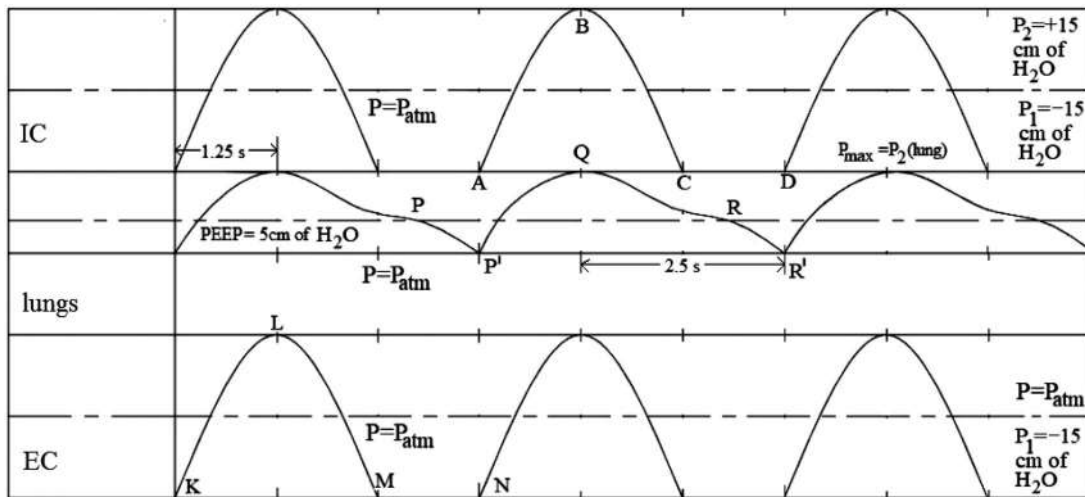

Figure 6. Pressure time histories of the ventilator model: IC and EC represent inhalation and exhalation cylinders, respectively. AB: IC compression; BC: IC expansion; CD: hold for facilitating complete expiration; KL: EC compression; LM: EC expansion; MN: hold for facilitating negative exit pressure; P'Q: lung inspiration; QR: lung expiration.

5. Discussion

In our study, we observed that the proposed automated mechanical ventilator can facilitate respiratory support to patients, eliminating the need for external aid. The cylinder capacity and stroke lengths are so designed as to ensure delivery of the required volume of air at the required RR. The proposed design also creates a negative pressure at the outlet during expiration, making expiration effortless. For the optimum respiratory activity, pressure inside the cylinder is kept below PEEP by 10 cm of H_2O to ensure complete evacuation of the lung and thereby preventing pressure build up. The proposed design is compact, with cylinder diameter 0.1 m and length 0.4 m. From the
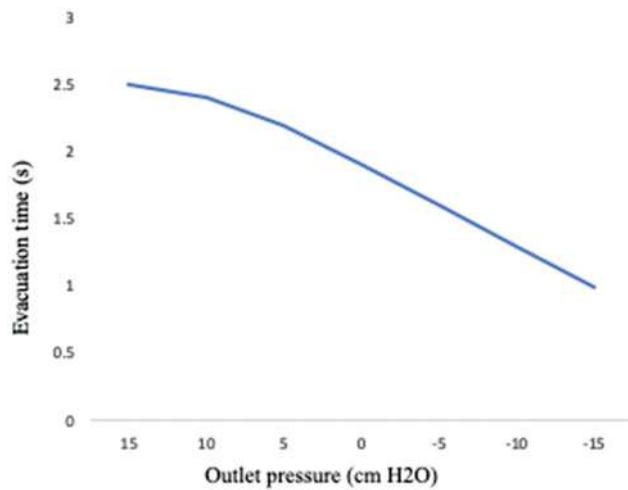

Figure 7. Variation of time of evacuation with outlet pressure.

study, it is observed that the required time for evacuation decreases with increasing negative pressure at the cylinder outlet during expiration. 15 cm of H_2O vacuum is observed to be the optimum outlet pressure showing an evacuation time of 1.25 s, which matches the expiration time for a normal lung.

The principles of design and prototyping of low cost ventilators for mass casualty cases and for resource-deprived environments are a much-researched topic of societal importance, as explained in Al Husseini et al. and Garmendia et al. [12,15]. Recent designs have also included chained wooden planks for compressing the AMBU bag, with provision also for controlling the breath as well as displaying the respiratory cycle [16]. There have been many models of ventilators fabricated recently in order to respond to the ever-growing demand. Nevertheless, in most of them, the pumping rate is constant and does not allow the prolonged time for expiration for patients unable to breathe on their own due to ARDS. As a result, the lungs undergo over-distention leading to barotrauma [17]. Barotrauma is caused by the inability of the lungs to maintain slightly higher pressure than atmospheric pressure; it is due to this pressure difference between the lungs and the atmosphere that expiration takes place. The pressure maintained in the lungs during this phase of respiration is called intrinsic or auto PEEP [18]. The intrinsic PEEP is likely to lead to the collapse of the alveoli, the soft tissues lining the lungs. This problem is addressed in this paper by creating a negative pressure

Table 2. Specifications of the ventilator model.

Parameter	Inspiration cylinder	Expiration cylinder	Ventilator	Unit
TV, V_T	–	–	2.2	l
RR	–	–	16	per minute
PEEP	–	–	5	cm of H_2O (G)
Maximum pressure (P_2)	15	15	–	cm of H_2O (G)
Minimum pressure (P_1)	−15	−15	–	cm of H_2O (Vacuum Gauge)
Effective swept volume, V_{seff}	2.2	2.2	–	l
Swept volume, V_s	3.14	3.14	–	l
Total volume, V_{Tot}	3.14	3.14	–	l
Clearance volume, V_C	0	0	–	l
Volumetric efficiency, η_{vol}	70%	70%	–	
Volume flow rate, Q	2.51	2.51	–	l/s
Bore diameter, D	0.1	0.1	–	m
Stroke length of the cylinder, L	0.4	0.4	–	m
Ratio of inspiration time to expiration time, T_i:T_r	–	–	1:2	–
Inspiration time, T_i	–	–	1.25	s
Expiration time, T_e	–	–	2.5	s
Time period of piston reciprocation, T_p	1.25	1.25	–	s
Time for compression stroke	1.25	1.25	–	s
Time for expansion stroke	1.25	1.25	–	s
Time period of hold	1.25	1.25	–	s
Number of strokes	16	16	–	per minute
Speed of the motor, N	48	48	–	rpm
Piston speed	0.32	0.32	–	m/s
Work done, W	72.8	72.8	–	J
Power required, P	58.24	58.24	116.48	W
Nature of power supply	Intermittent	Intermittent	–	

externally during exhalation at the exit of the expiration tube.

In the condition of dyspnoea, due to medical conditions such as ARDS and pneumonia, the lungs may not be able to function on their own. To sustain the process of respiration, mechanically supplying air to the lungs is the only solution as mouth-to-mouth respiration would be unhygienic. A ventilator is thus used to artificially support or replace the spontaneous breathing of a distressed patient. The evolution of mechanical ventilators is presented in Kacmarek [19]. The ventilators of the nineteenth century were based on negative pressures; the positive pressure ICU devices have come a long way over four generations. Assisting both inhalation and expiration externally is important especially in the case of children using home ventilators [20] for prolonged periods, to avoid the risk of barotrauma.

There are various modes of ventilation such as assist control (AC) or CMV, synchronised intermittent mandatory ventilation (SIMV), etc., where pressure or volume can be set and controlled. The above-mentioned components are vital in the design of a mechanical ventilator. However, an emergency ventilator does not have a mechanism to measure, record and display the input and output parameters. Hence, ventilator-related complications are not uncommon even after extensive research into the development of this life saving machine [21].

The proposed design in this paper provides a compact ventilator unit with a surface area of 0.8×0.4 m^2 with a minimal power requirement of 116.48 W, less than that of an average LCD projector. Production is expected to be economical as the cylinders are to be made of polypropylene and the piston heads using polyisoprene. A lithium–polymer (Li–Po) battery will be used to supply AC power to the servo motor. The overall cost per ventilator can be approximated to $250 and is expected to reduce further with mass-production. The proposed unit is easy to manufacture owing to very few linkages and joints.

The two-cylinder arrangement discussed in this paper is believed to offer safe operating conditions as the inspired and expired air are handled separately. The expired air is purified before it is let out into the atmosphere to prevent the spread of infection through droplets. As the total PEEP, P_{Tot}, (20 cm H$_2$O) in ventilated patients is higher than atmospheric pressure, there is net evacuation even when the outlet pressure is 15 cm H$_2$O. However, due to the absence of constricted spaces inside an AMBU bag, unlike human lungs, the evacuation seems to be slightly faster in this case. Negative pressure (–15 cm H$_2$O) has helped the evacuation rate increase by 50% over when the expiration takes place at an outlet at atmospheric pressure (0 cm H$_2$O). Hence with constricted spaces, the evacuation process is expected to improve by at least 20–25% when negative exit pressure prevails at the expiratory outlet. Hence, it can be considered as an ideal standalone unit for catering for emergency requirements of COVID patients without causing any side-effects to them.

5.1. Limitations

The design presented will be of benefit in the situation where the number of ventilated patients is increasing significantly every day. Nevertheless, there are certain limitations at this stage of design, the first being the lack of provision to measure and display the discharge and pressure of air. The second limitation is the invasive process involved in ETT, which can be carried out only in the presence of an experienced medical practitioner. Replacing ETT with a mask can result in loss of TV during inspiration. Another limitation is the possibility of slip due to the reciprocating motion of the piston in the inspiration and expiration cylinders. Finally, there is the limitation of the lack of clinical testing and validation of the present design. The model once developed will be first tested on artificial lungs and then extended to animal lungs to validate the device to serve for a medical emergency.

6. Conclusions

The ventilator design proposed here is not only compact, lightweight, power efficient and portable, but also uses a negative pressure at the exit of the expiration tube to help in easy lung evacuation during expiration. The experiment carried out in the present work shows faster lung evacuation due to the presence of a pressure (–15 cm H$_2$O) that is lower than that of atmospheric pressure at the exit. In human lungs where airway resistance and constricted spaces are quite common, the expiration takes more time than what is depicted in the experiment. However, the presence of negative pressure will ensure the lungs are cleared of the inspired air to a greater extent during expiration before the next cycle of respiration commences. Hence, the possibility of barotrauma can be minimised. The range of operating pressures of the ventilator chosen in this work (–15 to +15 cm H$_2$O) is well within the respiratory range of human lungs. By automating the entire mechanism, the risk of health

workers contracting the infection is also eliminated. To accomplish the operating conditions of the ventilator, a thorough engineering analysis of flow in and out of the ventilator has been carried out to arrive at its specifications. The ventilator is also found to be low cost, economic and affordable.

7. Scope for future work

In order to understand the practical implications of the ventilator design presented here, the work will be extended to prototyping. The prototype so developed will be initially used to carry out experiments using lung simulators, and later to treat patients in distress under medical supervision and guidance.

Author contributions

Conceptualisation and ideation, S.J.; data curation, V.C.; formal analysis, V.C. and S.J.; investigation, V.C.; methodology, S.J.; project administration, S.J. and R.R.; resources, S.J. and R.R.; supervision, S.J.; validation, V.C. and R.R.; visualisation, V.C.; writing – original draft, V.C. and S.J.; writing – review and editing, S.J. and R.R.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

[1] Faiza A. Development and Validation of Corona Virus Anxiety Scale (CVAS). Walailak J Sci Technol. 2020; 17(9):958–966.

[2] Wilson F. Artificial ventilation. In: Essential accident and emergency care. Dordrecht: Springer; 1981. p. 221–231.

[3] Carter C, Osborn M, Agagah G, et al. COVID-19 disease: invasive ventilation. Clin Integr Care. 2020;1: 100004.

[4] Jayaweera M, Perera H, Gunawardana B, et al. Transmission of COVID-19 virus by droplets and aerosols: a critical review on the unresolved dichotomy. Environ Res. 2020;188:109819.

[5] Gama De Abreu M, Rocco PRM, Pelosi P. Pros and cons of assisted mechanical ventilation in acute lung injury. In: Annual update in intensive care and emergency medicine. Berlin, Heidelberg: Springer; 2011. p. 159–173.

[6] Halek G. Physics of breathing. BMJ. 1951;1(4700): 246–246.

[7] Naimark A, Cherniack RM. Compliance of the respiratory system and its components in health and obesity. J Appl Physiol. 1960;15(3):377–382.

[8] Papineni RS, Rosenthal FS. The size distribution of droplets in the exhaled breath of healthy human subjects. J Aerosol Med. 1997;10(2):105–116.

[9] Weiler N, Heinrichs W, Dick W. Assessment of pulmonary mechanics and gastric inflation pressure during mask ventilation. Prehosp Disaster Med. 1995; 10(2):101–105.

[10] Gattinoni L, Chiumello D, Rossi S. COVID-19 pneumonia: ARDS or not? Crit Care. 2020;24:1–3.

[11] Gattinoni L, Collino F, Maiolo G, et al. Positive end-expiratory pressure: how to set it at the individual level. Ann Transl Med. 2017;5(14):288.

[12] Al Husseini AM, Lee HJ, Negrete J, et al. Design and prototyping of a low-cost portable mechanical ventilator. J Med Dev. 2010;4(2):027514; Naimark A, Cherniack RM. Compliance of the respiratory system and its components in health and obesity. J Appl Physiol. 1960;15(3):377–382.

[13] Wallis LA, Healy M, Undy MB, et al. Age related reference ranges for respiration rate and heart rate from 4 to 16 years. Arch Dis Child. 2005;90(11):1117–1121.

[14] Ganeshan V. Internal combustion engine. 2nd ed. Noida: Tata McGraw Hill Publishing Company Limited; 2012. p. 606–670.

[15] Garmendia O, Rodríguez-Lazaro MA, Otero J, et al. Low-cost, easy-to-build non-invasive pressure support ventilator for under-resourced regions: open source hardware description, performance and feasibility testing. Eur Respir J. 2020;55(6):2000846.

[16] Shahid M. Prototyping of artificial respiration machine using AMBU bag compression. 2019 International Conference on Electronics, Information, and Communication (ICEIC). IEEE; 2019. p. 1–6.

[17] Eisner MD, Thompson BT, Schoenfeld D, et al. Airway pressures and early barotrauma in patients with acute lung injury and acute respiratory distress syndrome. Am J Respir Crit Care Med. 2002;165(7):978–982.

[18] Blanch L, Bernabé F, Lucangelo U. Measurement of air trapping, intrinsic positive end-expiratory pressure, and dynamic hyperinflation in mechanically ventilated patients. Respir Care. 2005;50(1):110–124.

[19] Kacmarek RM. The mechanical ventilator: past, present, and future. Respir Care. 2011;56(8):1170–1180.

[20] Mohsin H. Home ventilation for children in Oman, are we prepared for this new reality? Oman Med J. 2019; 34(5):389–390.

[21] Hashemi SH, Hashemi N, Esna-Ashari F, et al. Clinical features and antimicrobial resistance of bacterial agents of ventilator-associated tracheobronchitis in Hamedan, Iran. Oman Med J. 2017;32(5):403–408.

❏ 559

Detection of Bundle Branch Blocks using Machine Learning Techniques

Praveena S Kammath[1], Vrinda V Gopal[2], Jerry Kuriakose[3]

[1,2,3]Department of Electronics and Communication Engineering, SCMS School of Engineering and Technology, India

Article Info

ABSTRACT

The most effective method used for the diagnosis of heart diseases is the Electrocardiogram (ECG). The shape of the ECG signal and the time interval between its various components gives useful details about any underlying heart disease. Any dysfunction of the heart is called as cardiac arrhythmia. The electrical impulses of the heart are blocked due to the cardiac arrhythmia called Bundle Branch Block (BBB) which can be observed as an irregular ECG wave. The BBB beats can indicate serious heart disease. The precise and quick detection of cardiac arrhythmias from the ECG signal can save lives and can also reduce the diagnostics cost. This study presents a machine learning technique for the automatic detection of BBB. In this method both morphological and statistical features were calculated from the ECG signals available in the standard MIT BIH database to classify them as normal, Left Bundle Branch Block (LBBB) and Right Bundle Branch Block (RBBB). ECG records in the MIT- BIH arrhythmia database containing Normal sinus rhythm, RBBB, and LBBB were used in the study. The suitability of the features extracted was evaluated using three classifiers, support vector machine, k-nearest neighbours and linear discriminant analysis. The accuracy of the technique is highly promising for all the three classifiers with k-nearest neighbours giving the highest accuracy of 98.2%. Since the ECG waveforms of patients with the same cardiac disorder is similar in shape, the proposed method is subject independent. The proposed technique is thus a reliable and simple method involving less computational complexity for the automatic detection of bundle branch block. This system can reduce the effort of cardiologists thereby enabling them to concentrate more on treatment of the patients.

Corresponding Author:

Praveena S. Kammath,
Department of Electronics and Communication Engineering,
SCMS School of Engineering and Technology,
Karukutty, Kerala, India.
Email: praveenakamath@scmsgroup.org

1. INTRODUCTION

The Electrocardiogram indicates the electrical activity of the heart recorded non-invasively from various locations of a person's body. It is the most commonly used biological signal to study the functioning of the heart. ECG consists of several deflections named P wave, QRS complex, and T wave. The electrical conduction system of the heart regulates the heart's pumping action and coordinates the contraction of the various chambers of the heart. An important part of the cardiac system is the Bundle branches. Bundle Branch block is a common cardiac disorder in which there is a blockage in the pathway of the electrical impulses of the heart which leads to an irregular heartbeat. The LBBB and the RBBB are the two types of BBBs which are caused by the delayed activation of the left and right ventricles respectively. LBBB is often associated with serious heart disease and can be the result of extensive strain, myocardial injury, or hypertrophy. RBBB can be caused due to heart attack, blood clot in the lungs, high blood pressure etc. The presence of BBB causes an

Performance evaluation of deep neural network on malware detection: visual feature approach

V. Anandhi[1,2] · P. Vinod[3] · Varun G. Menon[1,2] ⓒ · Korankotte Manoj Aditya[1,2]

Abstract

Nowadays, several malicious applications target computers and mobile users. So, malware detection plays a vital role on the internet so that the device is secure without any malicious activity affecting or gathering the useful content of the user. Researches indicate that the vulnerability of adversarial attacks is more in deep neural networks. When there is a malicious sample in a family, there will not be many changes in the variants, but there will be more signatures. So, a deep learning model, DenseNet was used for detection. The adversarial samples are created by other types of noise, including the Gaussian noise. We added this noise to a subset of malware samples and observed that for Malimg, the modified samples were precisely identified by the DenseNet, and the attack cannot be done. But for BIG2015, we found that there was some marginal decrease in the performance of the classifier, which shows that the model performs well. Further, experiments on the Fast Gradient Sign Method (FGSM) were conducted, and it was observed that a significant decrease in classification accuracy was detected for both datasets. We understand that deep learning models should be robust to adversarial attacks.

Keywords Deep neural network · Deep learning · Adversarial attack · DenseNet · Noise · Fast gradient sign method

1 Introduction

Malware causes catastrophic damage to a computer or network of computers. The malicious software is designed to cause hardware failure, slow down the system speed, restrict access to our own files, pop up unwanted programs, send spam messages from the affected system to any network, etc [1].

At the end of March 2020, when COVID-19 started spreading worldwide, people started working from home and students started attending virtual classrooms. Internet usage has increased drastically. Due to this, security has become a real issue as the network area is widespread. As

seen in [2], it is clear that malware threats have increased day by day and reached a maximum in the mid of march. In [3], the Lab report specifies 12,755,097 COVID-19 related malicious files were detected from June 2020 to June 2021. The attackers would like to leverage this pandemic situation as an opportunity to attack organizations and the computers used by professionals at home too. Many systems are connected online. This gives a chance to the attacker to infiltrate the information.

Machine learning is a major application of artificial intelligence. Machine learning is best used in medical diagnosis, image processing applications, image prediction, image classification, etc. The computational systems built on machine learning algorithms can learn from past experience or previous data. Deep learning is a subset of machine learning that works on different layers to create a deep neural network that can learn and make decisions intelligently by itself.

Classical methods for the detection of malware are slow [4]. Traditionally malware detection is performed through signature-based methods, where signatures are unique string/bytes used to identify malicious payload. In this approach a signature repository of known malware is

✉ Varun G. Menon
varunmenon@ieee.org

1 Department of Computer Science and Engineering, SCMS School of Engineering and Technology, Karukutty, Kerala, India

2 APJ Abdul Kalam Technological University, Thiruvananthapuram, Kerala, India

3 Department of Computer Applications, Cochin University of Science and Technology, Cochin, Kerala, India

Ananya Rajesh [1], Geetika Gopinath [1], Tejas C J [1], Amal Prasannan [1], Dr. Parvathy M [2]

Automated Power Depiction System Using Iot Platform

Ananya Rajesh [1], Geetika Gopinath [1], Tejas C J [1], Amal Prasannan [1], Dr. Parvathy M [2]

[1] B. Tech students,
SCMS School of Engineering and Technology

[2] Associate Professor, Department of ECE
SCMS School of Engineering and Technology
meetparu.parvathy@gmail.com

Abstract—

Precise and well-timed expertise of electricity consumption is an imperative requirement for implementing any power saving strategy. The primary function of energy management systems is to measure the electricity utilization in the homes. Monitoring energy usage in homes indirectly offers knowledge about consumer preferences, types of activities carried out by residents, and forecasting potential energy consumption, energy saving methods and lots of other proposes. Electricity consumers are oblivious to their daily unit consumption of electricity. They know about their power consumption only when they receive the electricity bill on a cycle of 2 months to 3 Months which give consumers no possible option to control the usage of electricity as there are only limited number of methods for the consumers to know about how much units of electricity they have consumed. So, we are going to come up with a solution for the same so as to develop a method which would easily replicate the Electricity Meter reading and would also provide us with the power consumption of each individual appliance on to your phone. This helps us to compare the energy consumption and also helpful for the users to understand if their electricity usage is more or less on a need basis and can compensate or reduce electricity usage in turn reducing electricity bill.)

Keywords—Internet of things (IOT), Android smart phone, NodeMCU.

I. INTRODUCTION

The advancement of the IOT (Internet of Things) have become an additional standard as a result of its contribution to economical solutions for several real time applications. It conjointly revolves the association between M2M that are embedded with electronics, software, sensors etc., which assist users in observing and controlling devices with efficiency. In an IoT system, objects and living beings are provided with unique

identifiers with the ability to transfer the data. Nowadays, IoT is being applied in several areas like gas, water, electricity etc. to make our lives automated. Electricity is a crucial invention and its demand is also increasing at a relentless rate and is being utilized for numerous purposes such as agriculture, industries, hospitals etc. So, it is becoming additionally sophisticated to handle the electricity requirements and maintenance. Clearly there ought to be a necessity for measuring the consumed electricity. Therefore, it is indispensable to execute a technique of taking energy meter readings automatically, which can realize the power consumption management to the customers to be adjustable and manageable to save the electrical energy. Also, to take the reading of the meter, a human operator has to go turn by turn to every resident & commercial building, hence this will increase the work and potency. Therefore, the operating hours also increases to realize the complete area data reading. So, to achieve an efficient energy meter reading, reduce billing error, and operation cost, we need a system that can read the meter reading automatically at each time interval.

The right way to cope up with the problem is to understand the situation thoroughly. At first, we need to realize the quantity of energy that we tend to consume so that we begin to limit it. So, as a solution to it, we intended to construct an application using IOT, which would display the overall power consumed by the household together with the power consumed by individual appliances and conjointly enables the user to set a threshold which upon surpassing will notify the user with an alert message thus helping the user to monitor the power consumption in their house. Thus, this device uses IoT to automate the purpose of measuring consumption of power in homes, allowing for web access and digital technology.