IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 4 January 2023, accepted 7 February 2023, date of publication 22 February 2023, date of current version 28 February 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3246660

RESEARCH ARTICLE

Deep Learning Based Anomaly Detection for Fog-Assisted IoVs Network

SHUMAYLA YAQOOB[1], ASAD HUSSAIN[2], FAZLI SUBHAN[2], GIUSEPPINA PAPPALARDO[3], AND MUHAMMAD AWAIS[4]

[1]Department of Electrical Electronic Computer and Telecommunication Engineering, University of Catania, 95124 Catania, Italy
[2]Faculty of Engineering and Computer Science, National University of Modern Languages, Islamabad 44000, Pakistan
[3]Department of Civil Engineering and Architecture, University of Catania, 95124 Catania, Italy
[4]Department of Computer Science, Edge Hill University Lancashire, L39 4QP Ormskirk, U.K.

Corresponding author: Shumayla Yaqoob (shumayla.yaqoob@phd.unict.it)

This work was supported by the SAFE DEMON SAFE Driving by E-Health Monitoring, Sicilian Region, Italy, under Grant G29J18000720007 and Grant PON FESR 2014/2020-Action 1.1.5.

ABSTRACT Internet of vehicles (IoVs) allows millions of vehicles to be connected and share information for various purposes. The main applications of IoVs are traffic management, emergency messages delivery, E-health, traffic, and temperature monitoring. On the other hand, IoVs lack in location awareness and geographic distribution, which is critical for some IoVs applications such as smart traffic lights and information sharing in vehicles. To support these topographies, fog computing was proposed as an appealing and novel term, which was integrated with IoVs to extend storage, computation, and networking. Unfortunately, it is also challenged with various security and privacy hazards, which is a serious concern of smart cities. Therefore, we can formulate that Fog-assisted IoVs (Fa-IoVs), are challenged by security threats during information dissemination among mobile nodes. These security threats of Fa-IoVs are considered as anomalies which is a serious concern that needs to be addressed for smooth Fa-IoVs network communication. Here, smooth communication refers to less risk of important data loss, delay, communication overhead, etc. This research work aims to identify research gaps in the Fa-IoVs network and present a deep learning-based dynamic scheme named CAaDet (Convolutional autoencoder Aided anomaly detection) to detect anomalies. CAaDet exploits convolutional layers with a customized autoencoder for useful feature extraction and anomaly detection. Performance evaluation of the proposed scheme is done by using the F1-score metric where experiments are carried out by exploiting a benchmark dataset named NSL-KDD. CAaDet also observes the behavior of fog nodes and hidden neurons and selects the best match to reduce false alarms and improve F1-score. The proposed scheme achieved significant improvement over existing schemes for anomaly detection. Identified research gaps in Fa-IoVs can give future directions to researchers and attract more attention to this new era.

INDEX TERMS Fog computing, smooth communication, Internet of Vehicles, anomaly detection, fog-assisted IoVs.

I. INTRODUCTION

The Internet of vehicles (IoVs) is a network that permits mobile communication among vehicles. A conventional VANET is rising into IoVs to meet the upcoming requirements of an intelligent transportation system in the future [1], [2]. It is an emerging era that is the base for traffic management, efficient traffic monitoring, mobile sensing, parking alerts, audio/video streaming in vehicles, and accident reporting [3]. Internet of vehicles is supported in both areas, urban and rural. IoVs network allows communication among vehicle-to-vehicle (V2V), vehicle-to-grid (V2G), vehicle-to-device (V2D), Vehicle to infrastructure (V2I), and vice versa. IoVs are also used to provide E-health applications

The associate editor coordinating the review of this manuscript and approving it for publication was Jenny Mahoney.

as mobile hospitals [2]. IoVs is an important constituent of Intelligent Transportation Systems (ITS) to manage road safety and related transportation services by saving data at central online repositories for better decision-making [3], [4]. Moreover, IoVs help to manage huge data transmission, computation, and storage for users and owners [5], [6].

With the frequently growing rate of IoVs network, security is a serious research concern of this era [3], [7]. Many preventive measures have been taken to some extent to halt these attacks. Security threat progression results in poor network communication. In the IoVs network, message congestion and its security during information dissemination is a major issue in Vehicle communication [4], [8]. To deal with the IoVs concerns, fog computing was introduced for smooth communication [4], [9], [10]. Here, smooth communication refers to no message congestion and limited security threats.

The term ''fog computing'' was invented by Cisco organization in 2012 [12], [13]. Fog computing is the middle layer between clients and the cloud [12]. It is not a replacement for the cloud, but it is an enhancement of cloud computing which is like the cloud and gives data computation, data storage, and networking between end users and cloud servers. Cloud computing follows centralized servers. When a network or application needs a quick response from a central controller, Cloud computing lack to provide immediate attention with a short response time. The high distance between the cloud and end devices increases the latency rate, which is directly proportional to increase response time [14]. Furthermore, cloud computing has limited mobility options and a high workload due to a central server. High workload, latency, and response time are considered limitations of cloud computing. Furthermore, data congestion is also one serious concern of cloud computing, the central nature of cloud computing may lead to congestion and data loss [4], [15]. Fog computing is an emerging domain that may enhance cloud features and aims to resolve issues such as latency, mobility, and location tracking and provide services to users [9], [12], [16]. Fog computing computes data at the network edges by sharing computation workload [14], [17]. Fog nodes act as local servers where it computes given tasks. Fog servers are also capable to process data at local servers and help to take timely actions [15], [18]. Furthermore, Fog computing reduces computation processes and user-related services and reduces the burden of traditional cloud computing data centers [19]. Some basic characteristics of cloud and fog are stated in Table 1.

Besides the fair factors of fog computing, it has also some limitations that need to be explored. A major concern of fog computing is security, due to local servers information security is not guaranteed [12], [19], [20]. Fog nodes also face security threats due to data computation on edges and the mobile nature of nodes [13], [14], [21]. Specifically, security constraint includes account hijacking, Daniel of Service (DDOS), data breaches, and data loss [8], [14], [22], [23], [24]. In DDOS users are prevented to use system resources

TABLE 1. Comparison of cloud and fog computing.

Features	Cloud Computing	Fog Computing
Geographic Nature	centralized	decentralized
Response Time	high	Low
Mobility	limited	high
Workload	high	low
Latency	high	low
Computation Cost	high	low
Security	low	high
Congestion	High	Low

while important information may have lost in data breaches and data loss by natural disasters or any malware attack. Therefore, need to improve the failure of local servers in fog computing for smooth communication. Thousands of people had been died due to traffic accidents [25], [26]. Secure Fog computing can be used to connect vehicles for smooth and secure communication and data transmission [6], [27]. This research work will provide secure information dissemination among vehicles by handling the limitations of fog computing.

Fog-assisted IoVs (Fa-IoVs) refer to various local services over the vehicular network for the sake of load balancing and cost reduction [17], [28]. Fa-IoVs involve many vehicles and fog nodes to develop a network for smooth communication. Fog nodes reduce congestion and IoVs security attack to score smooth communication [4], [7]. Fog computing works in a distributed manner on network edges [12]. For Fa-IoVs, Fog computing is a central layer between IoVs and cloud works through roadside units (RSU) [4], [28]. Fog computing map fog nodes locally that help in load balancing, resource management, and computation cost reduction in parallel to cloud computing or only RSU's involvement. Unfortunately, local fog nodes on IoVs edges are threatened with security issues, which is a critical challenge. Security issues of Fog-assisted IoVs are considered as anomalies.

Anomaly is an outlier or deviated action over normal behavior [29], [30]. Anomaly detection (AD) is a recent way to detect security challenges or threats [31]. Basically, AD is a way that can identify known and unknown threats by observing a change in normal behavior. Anomaly detection of fog-assisted IoVs network refers to ''security concerns'' of fog computing [31], [32]. Security concerns of fog computing may include authentication, data integrity, access control, DDOS, and malware attacks [14], [33]. Before ML and DL anomaly detection methods, researchers detect fog computing security challenges by giving pattern-matching algorithms to detect known threats [33], [34]. Existing pattern matching or signature-based schemes were limited to detecting known threats efficiently with the limitation of high latency and limited accuracy [21], [32]. Deep learning-based

anomaly detection is considered a better approach over existing signature-based and shallow schemes [35], [36]. There are several machine learning and deep learning-based algorithms introduced for anomaly detection for various domains and datasets. Although ML and DL-based schemes improved anomaly detection of fog computing threats but still accuracy is limited for real networks such as the Fa-IoVs network. Some other limitations need to be improved such as a high rate of false alarms, and high execution time due to the combination of various models [34], [37]. Therefore, it is important to improve the anomaly detection rate and reduce false alarms for the sake of smooth communication. Here, smooth communication refers to the limited risk of important data loss, delay, communication overhead, etc.

We present a deep learning-based scheme that significantly improves anomaly detection for fog-assisted IoVs networks. The main contributions of this work are the following:

- We explore the literature on anomaly or threats detection in fog, IoVs, and Fog-assisted IoVs networks where the focus is on the comparison of each existing scheme based on the addressed problem, the accuracy level of anomaly detection, limitation, and remarks of each scheme.
- From the literature and comparison of the existing schemes, we identify the main research gaps to formulate the problem.
- We present a deep learning-based scheme named convolutional autoencoder-aided anomaly detection (CAaDet) that follows convolutional layers with a customized encoder-decoder structure for automatic feature extraction and then anomaly detection.
- In the comparison of existing schemes, CAaDet reduces false alarms and improves the F1-score rate significantly for anomaly detection. Indeed, less rate of false alarms and anomaly detection improvement leads to smooth communication.
- To improve the performance of the proposed scheme, we also observe the behavior of fog nodes and hidden neurons and chose the best match to get a better F1-score.

The rest of the paper consists of six more sections. Section II. reviewed existing related work and placed a ground for problem formulation. Fog-assisted IoVs challenges and Problem formulation is stated in Section III. Section IV explained the preliminaries before the explanation of the proposed methodology and the dataset. Result evaluation is carried out in Section V, explaining how CAaDet outperforms the existing approaches. Finally, the conclusion of this research is stated in Section VI along with some future directions.

II. LITERATURE REVIEW

With the growing rate of population and vehicles, the Internet of vehicles (IoVs) has become the most attention gained topic for researchers. It aims to attain safety via intelligent transportation systems by using different types of information. In the IoVs scenario, vehicles exchange information directly for communication [2], [4]. The main problem, in this case, is the distance between the two vehicles. If it is shorter than the communication range, it provides a successful connection. Network connectivity is a key issue in enabling information transmission for communication [6]. To deal with the increasing number of connected vehicles, fog computing smartly manages computing, storage, and networking resources by working on the edges [17], [28]. Thus researchers go to the Fog-assisted vehicular network, which takes moving vehicles as communication nodes to establish better network connectivity [4]. Later, it is highlighted that fog computing has issues with security and failure of servers which needed to recover.

A. SECURITY THREATS IN IOVS NETWORK AND AI-BASED INTRUSION DETECTION

In [49], authors performed a survey about security threats to IoVs network and highlighted various security threats such as Distributed Denial of Service (DDoS), black hole attacks, gray hole, sinkhole, spamming, and man-in-the-middle attacks. In the recent era, the IoVs network is an attractive and highly explorable research domain. In [8], the author detects the Distributed Denial Of service (DDOS) attack in the IoVs network and provides a good path for safe V2V communication. Its objective is to reduce communication delays. The model creates the controlled clusters and applies several limits-based analyses to identify an attack, but still needing to ensure secure communication over the network by controlling repeated transmission, may lead to less communication loss. In [23], authors proposed a deep neural network-based scheme to identify DDOS attacks by implementing LSTM and autoencoder methods. This scheme improves the accuracy rate of attack detection but has some limitations that can be addressed in the future such as the need to implement it in large real networks such as IoVs networks. In [50], authors proposed an interesting way to improve DDOS attack detection by implementing SVM and KNN with the combination of C4.5 classifier but it increases time cost that can be addressed in the future.

Authors in [40] presented five machine learning models: Stochastic gradient drop Classifier (SGD) Classifier, Ridge Classifier, Decision Tree Classifier, Random Forest Classifier, and Extra Tree Classifier to examine the intrusion detection system to learn the best model to cope with network attack detection. NSL-KDD data set was utilized for result evaluation. The experimental findings reveal that both the Random Forest Classifier and the Additional Tree Classifier performed well, with the extra tree model retaining excellent stability and accuracy while dealing with challenging tasks. The limitation of [40] is its limited anomaly detection accuracy rate which can be improved using dynamic algorithms.

Authors, in [41], proposed a novel framework for anomaly detection based on a five-layer autoencoder used for

networks. A new data pre-processing approach is adopted that filters out the damaged outliers from the input and removes the skewness of the dataset. Afterward, an error function reconstruction-based model is employed for anomaly detection in a network traffic domain. Limited anomaly detection accuracy can be improved in the future.

In [43], research introduces a one-stage intrusion detection strategy that merges a one-dimensional convolutional autoencoder (1D CAE) and a one-class support vector machine (OCSVM) as a classifier together into a concurrent optimization framework. A. Binbusayyis et al. shown the result that the suggested strategy has the potential to serve as a foundation for developing a successful Intrusion detection system, but with limitation of high execution time.

In another study [44] a new approach is proposed for IDS based on Principal Component Analysis (PCA) and Fuzzy Clustering with K-Nearest Neighbour. The model consists of two sections, one is responsible for the classification and the other to check the robustness of the model that is evaluated on an NSL-KDD dataset. Generally, the model prediction is limited due to the high rate of false classification. In [51], authors proposed a machine learning model based on SVM to detect DDOS attacks. The performance of this scheme is evaluated by using the NSL-KDD benchmark dataset with the limitation of implementation in a real scenario such as the IoVs network.

The study [45] proposed autoencoder-based feature reconstruction for anomaly detection. The result presents a medium level of anomaly detection with a high rate of computational cost.

This study [46] proposed feature extraction utilizing a basic autoencoder and SVM to characterize attacks on intrusion detection systems. Moreover, performed experiments achieved a medium level of detection for attacks. Authors in [47], proposed that with various settings of hyperparameters of stacked autoencoder algorithms, high accuracy can be achieved for network intrusion detection systems. Therefore, [46] and [47] require exploring dynamic algorithms for real-time anomaly detection.

B. ANOMALY DETECTION IN FOG AND FOG-TO-THINGS ENVIRONMENT

Cloud Security Alliance [14] has recognized twelve serious security issues. These issues directly impact the distributed, shared, and on-demand nature of cloud computing. Indeed, distributed sharing is a key feature of fog computing. Therefore, the Fog platform can also be affected by the same threats. However, fog computing cannot be deemed to be secure, since it still inherits various security risks from cloud computing [9]. Unlike Cloud systems, there are no standard security certifications and measures defined for Fog computing. In addition, it could also be stated that a Fog platform:

- Has relatively smaller computing resources due to this nature, it would be difficult to execute a full suite of security solutions that can detect and prevent sophisticated, targeted, and distributed attacks.
- Is an attractive target for cyber-criminals due to high volumes of data throughput and the likelihood of being able to acquire sensitive data from vehicles.

In [38], authors proposed a machine learning method named One-Class Support Vector Machine (SVM) based scheme to detect anomalies in a framework containing a fog-enabled infrastructure. The fog-enabled infrastructure offers improved computing resources for selecting the best learning model and sample ratio. The results revealed that the proposed optimal learning model attained medium-level detection accuracy, but with the limitation of many false alarms that may lead to misleading classification. In [39], the authors presented a Genetic Algorithm Wrapper-Based feature selection and Nave Bayes for Anomaly Detection Model (GANBADM) in a Fog Environment that removes immaterial features. GANBADM used the NSL-KDD dataset for result evaluation. The limitation of GANBADM includes high execution time cost due to the integration of two methods. High execution time costs could be improved using machine learning algorithms.

The study [21] developed a distributed deep learning approach for detecting cyber-attacks in fog-to-things computing based on stacked autoencoders. Deep models surpass shallow models in terms of detection performance with 99.27%, a false alarm rate of 0.85%, and scalability. The specific domain and the targeted dataset are its limitations.

Authors in [34], proposed an LSTM network for cyber-attack detection in fog-to-things communication for the distributed use case. The data set used for experiments were ISCX and AWID for the detection of cyber-attacks. The results reveal that the employed deep learning method surpasses the traditional machine learning methods.

Another study [37] presented a method (Auto-IF) for intrusion detection in the fog environment that is based on a deep learning approach employing Autoencoder (AE) and Isolation Forest (IF). Because fog devices are primarily concerned with distinguishing attacks from regular packets. The limitation of [37] is that the technique only focuses on the binary classification of incoming packets with high execution times.

In [32], the authors presented a network intrusion detection system based on the Exact Greedy Boosting ensemble approach to safeguard critical infrastructure from hazardous activity detection that is rapid and precise in the fog-to-things network. The developed scheme investigates traffic flow monitoring in innovative IoT Intrusion Dataset 2020(IoTID20) network traffic by recognizing and categorizing attack types based on deviations from normal behavior.

In [52], the authors proposed the $\boldsymbol{ES^2A}$ (Efficient and Secure Service-Oriented Authentication) Scheme for the security threat of fog-assisted 5G architecture in IoTs environment that results in confidential data loss. $\boldsymbol{ES^2A}$ is based on the slice/service types of accessing services. The privacy-preserving slice selection mechanism is introduced to

preserve data. ES^2A guarantees secure access to service data in fog cache and remote servers with low latency. But still has limitations like the network is divided into three regions which consume high energy and increases communication costs. Fog slice still has network security threats due to on edges.

In [10] the author presents a Privacy-preserving data reporting and requesting (PARE) scheme for serious security threats to fog computing of user privacy (e.g., data content, preference). PARE is constructed by leveraging one-way hash chains, marked mix-nets, and groups of fog nodes to avoid security attacks. PARE reduces computational costs and communication overhead. PARE has limitations such as the hotspot area being considered. A group of fog nodes may cause high delays and communication costs. Therefore, it needs a smart solution to avoid network attacks of fog computing and improve connection orientation.

C. FOG COMPUTING FOR ANOMALY DETECTION IN IOTS SYSTEM

The study in [11] used the fog computing idea for DDoS mitigation by distributing traffic monitoring and analysis work locally. The presented technique is evaluated in an industrial control system test platform, with the tests evaluating the latency and rate for two types of DDoS attacks. Here [11], false alarms are considered as limitations that can be improved using customized algorithms. Future work can be anomaly detection in the Fog- to -IoT network.

Another study [31] proposed a fog computing-based hybrid anomaly mitigation system for IoT to enable quicker and more accurate anomaly detection. The model was composed of two modules, one used a signature-based and the other employed anomaly-based detection approaches. The signature-based module used a database of attack sources (blacklisted IP addresses) to ensure faster detection of attacks launched from a blacklisted IP address, whereas the anomaly-based module employed an extreme gradient boosting algorithm to accurately classify network traffic flow as normal or abnormal. Using an IoT-based dataset, the study assessed the performance of both modules in terms of reaction time for the signature-based module and accuracy in binary and multiclass classification for the anomaly-based module. The result showed better performance for anomaly detection. Finding anomaly detection for the fog-to-things in the environment can be future work.

In this study [42] authors proposed a convolutional neural (CNN) network-based approach for intrusion detection in fog nodes for malicious users' attacks on the network. CNN model is based on sliding windows for the incoming traffic on the nodes. The outcome was assessed using the NSL-KDD dataset. The proposed technique of [42] can be explored in a fog-to-things environment.

A data-driven intrusion detection scheme based on RSU is proposed [28], [48] to evaluate the link load behaviors of the Roadside Unit (RSU) in the IoV against different threats that cause unusual variations in traffic flow but it takes more time to identify and inform cloud. Therefore, dynamic algorithms can reduce the computational cost of external units such as RSU in the future.

The fog-based identity authentication (FBIA) technique proposed in [7] is divided into two layers: the security authentication layer for cars outside the fog and the security monitoring layer for the remaining vehicles. To perform real-time security in the IoV, two-way authentication based on the vehicle's identification was implemented. In the IoV, the FBIA scheme has higher authentication accuracy and better adaptation. Here, needs to explore Fog-assisted IoVs in the future.

III. EXISTING RESEARCH GAPS

This section concludes the literature review by stating the existing Fa-IoVs challenges, to formulate the problem.

A. FA-IoVs CHALLENGES

Fog-assisted IoVs is a novel term and provides several opportunities but still, there are some limitations, a few of which have been observed in the above lines. Some further limitations are discussed below that need to be explored.

1) STRUCTURAL ISSUES

In Fog-assisted IoVs, edges based networks are used as computing infrastructure. These components consist of various processors that are general purpose, being challenged [53], [54]. Here, the selection of nodes (vehicles) and resource configuration needs to be focused on, such as normal vehicles looking for smart vehicles.

2) RESOURCE MANAGEMENT

In IoVs fog computing needs to manage resources to utilize computation nodes and storage resources [13]. Nodes can be vehicles, base stations, RSUs, and routers. Computation nodes should be cheaper in cost for V2V communication to send messages to nearby vehicles. Several protocols and algorithms must be planned to detect idle resources and utilize them efficiently among vehicles. However, future research needs to explore new algorithms and techniques to share resources among nodes efficiently.

3) DEPLOYMENT

Fog-assisted IoVs deployment for dynamic situations is quite challenging because it is based on several fog nodes or components. Fog nodes or components include servers, routers, bridges, base stations, RSUs, and vehicles [28].

4) MOBILITY

Fog computing introduced mobility among vehicles by using geography distribution concepts for known patterns of vehicles. Mobility is a major challenge in vehicles and other wireless networks [4], [28]. Mobility for random patterns needs to be explored in future research work as it is formulated.

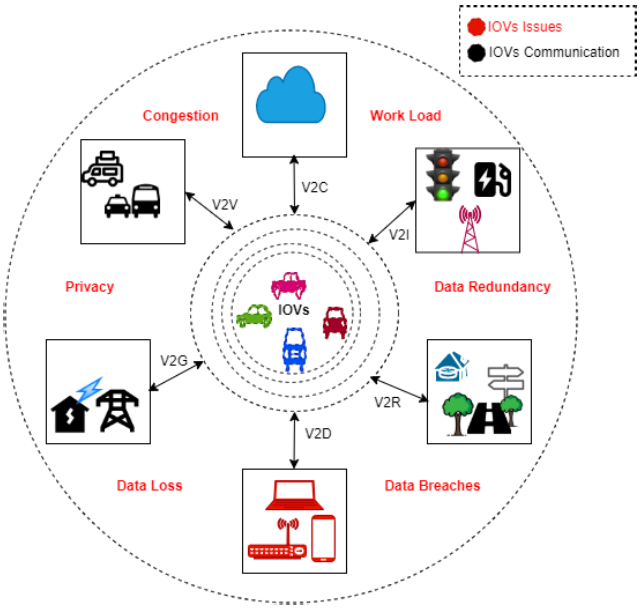

FIGURE 1. Issues of internet of vehicles (IoVs).

5) SECURITY ASPECTS

Fog-assisted IoVs are considered weak in security aspects [27]. Fog is based on a distributed concept, therefore it's difficult to manage security for several servers, to authenticate the data on different gateways [7], [52], Mobile information-centric-based techniques are required to avoid security issues of vehicles [12]. Security implementation is directly proportional to the QoS of fog, it affects the services of fog [55]. So, it's difficult to deal with real-time applications like communication among IoVs. Fog computing cannot decrypt the data while to maintain privacy applications need encrypted data. In IoVs fog computing has various security issues such as regarding data, virtualization, network, virus, and monitoring [12], [53]. With time, it is highlighted that fog computing is also not sufficient for information dissemination among IoVs. It is needed to resolve the security constraint of fog computing for IoVs.

B. PROBLEM FORMULATION

IoVs communication includes V2V, V2I, V2G, and V2D, etc., where each node sends various types of information packets to each other for various purposes as shown in Fig. 1. IoVs communication purposes can be traffic monitoring, accident reporting, crash prevention, and parking, etc. A bulk of information packets at the same time can lead to congestion that can result in packet loss and information duplication. Similarly, IoV communication faces privacy hazards that can lead to data leakage and packet loss.

In the last few years, IoV communication remains a serious concern of the research community where fog computing was proposed to resolve IoVs issues such as congestion, workload, and privacy as illustrated in Fig. 1.

The term Fog computing with the internet of vehicle networks gives the concept of Fog-assisted IoVs (Fa-IoVs). Fog-assisted IoVs are a distributed network of IoVs communication. Unfortunately, fog nodes are also facing various security hazards like DDOS attacks, malware, and authentication attacks that can lead to fog node crashes, delay in packet delivery and data loss as shown in Fig. 2. Security of Fa-IoVs is an appealing research domain in this era.

From the above literature [5], [6], [7], [8], [9] it is concluded that it is hard to disclose sensitive data during transmission due to the Network security threats in Fog-assisted IoVs (Fa-IoVs). Researchers have been working to limit security threats for smooth communication. In this way, data breaches and data loss can be avoided, and reduce road accidents by transmitting emergency messages timely.

Suppose that there is an accident on the road-side, where an accidental car is an abnormal vehicle $v_A \leftrightarrow V$ surrounded by 5 nearby vehicles $\boldsymbol{V} = \{v_1, v_2, v_3, v_4, v_5\} \forall v_j \in \boldsymbol{V}$ transmitting 100 of emergency packets $P = \{p_1, p_2, p_3, \ldots, p_m\} \overset{m}{\underset{k=1}{p}} \in \boldsymbol{P} wherem = 1, 2, 3 \ldots 100$ among 4 communication networks such as V2D, V2V, V2G, V2I $N = \{D, V, G, I, R\}$. There are three hops $= \{H_1, H2, H_3\} \forall H_i \in \boldsymbol{H}$ *where* $i = 1, 2, 3$. In the first hop, the transmitted number of packets is one hop $[100 \times (5 \times 4)]$ while the transmitted number of packets increases in 2^{nd} and 3^{rd} hop respectively as $[100 \times (5 \times 4)^2]$ and $[100 \times (5 \times 4)^3]$ which can lead to data congestion and data loss, delay, and message overhead. When there is a fog server between IoVs' communication, it will play a medium role. When (5×4) packets reach the fog server no duplication will be raised because the fog server is responsible for packet delivery to nearby vehicles and servers by reducing congestion, packet duplication, and workload.

From Table 2. Some limitations of Fa-IoVs security have been identified such as:

- High rate of false alarms or misleading.
- High execution time
- Delay in communication
- Communication overhead

Unfortunately, when fog nodes are hacked or attacked by malicious attackers it may lead to packet loss, redundancy, delay, and communication overhead. Therefore, it is important to detect attacks or anomalies in the Fa-IoVs network.

IV. MODEL SYNTHESIS

This research work follows various steps such as data preprocessing, labeling, modeling testing, and result validation as shown in Fig. 3. Proposed model CAaDet exploits a standard collection of security threats database referred to as the NSL-KDD dataset. CAaDet is an integration of convolutional layers and autoencoder. CAaDet can perform 2 tasks such as (i) Feature extraction (ii) Anomaly identification and limiting false alarms. First task feature extraction is used to extract meaningful data for model training and testing instead of useless extensive data. The second module aims to detect anomalies by passing a dataset to a convolutional autoencoder.

TABLE 2. Comparison between existing schemes.

Method	Type	Problem addressed	A	Limitation	Remarks
One-class Support Vector Machine [38]	ML	Anomaly detection in Fog Environment	M	High false alarm rate.	Needs to improve the rate of misleading false alarms.
Genetic Algorithm + Nave Bayes [39]	SS	Anomaly discovery in Fog environment	M	High execution time than ML	Using ML computational time could be saved.
DDOS mitigation scheme [11]	SS	DDoS attacks detection in Fog environment	M	Limited threat detection and high false alarms.	Requires DDoS mitigation in the Fog- IoT network
Stacked autoencoder-Softmax [21]	DL	Attack detection in the fog-to-things environment	H	Results are limited to a specific dataset.	Evaluate different datasets and conditions
Random forest [40]	ML	Network intrusion detection	M	Limited unknown threat detection	Dynamic algorithms are required.
LSTM [34]	ML	Distributed attack detection in Fog-to things	H	High rate of misclassification	Needs more customized algorithms for scalability.
Autoencoder + Isolation Forest [37]	DL	cyber-attack detection in Fog Environment	M	high execution time	Needs an algorithm to improve the accuracy
Autoencoder-based network anomaly detection [41]	DL	Network anomaly detection	M	The limited anomaly detection rate	Need to customize features and parameters
CNN [42]	DL	Network intrusion detection in Fog	H	The specific type of threats is targeted	Need to explore algorithm for fog-to things
one-class support vector machine [43]	ML	Network intrusion detection	L	Limited rate of anomaly detection	Need to explore joint optimization framework
PCA + KNN [44]	ML	Network intrusion detection	M	Limited accuracy	Dynamic approaches are required for better accuracy
Conditional Variational Autoencoder [45]	DL	Network security threats detection	M	Limited accuracy is achieved	Improve imbalance class classification
Autoencoder [46]	ML	Network anomaly detection	M	Less experimental data is used	Explore algorithms for large and real datasets
Stacked-Autoencoder [47]	DL	Network Intrusion Detection	H	Limited dataset size	Need to explore algorithm on a real or large dataset.
CNN [48]	DL	anomaly Detection for Intelligent IoVs	H	Involvement of external devices e.g RSU	Need dynamic algorithm to reduce the computational cost
Fog-Based Identity Authentication [7]	DL	security authentication for IoVs using fog computing	L	High rate of false alarms	Need to explore optimizing the algorithm

Note. A represents Accuracy while ML, DL, L, M, and H represent Machine Learning, Deep Learning, Low, Medium, and High respectively.

A. DATASET

The proposed scheme employs the benchmark dataset named NSL-KDD, which is online available and can be easily downloaded from the Kaggle site. This dataset is a reduced version of the KDD cup'99 dataset. The KDD cup dataset was collected at MIT Lincoln Lab in 1998; it has a large amount of redundant data that might result in biased outcomes [23], [51]. The NSL-KDD dataset is close to a real environment, which is why very effective and used for various research purposes [40], [41], [51]. NSL-KDD consists of 125,973 training and 22543 testing samples. Each packet of NSL-KDD consists of 41 features and label classes to categorize the network traffic into normal or attack. The given dataset consists of two major categories: normal and abnormal data packets as shown in Table 3 along with training and testing vectors. Abnormal data packets refer to attacks or malicious data. Furthermore, the given dataset is divided into two sets for training and testing as 70 and 30 percent of the entire dataset respectively. The training dataset consists of 50622 attacks and 53340 normal cases while the testing dataset includes 26120 attacks and 18435 normal data as illustrated in Table 3. In each case, labels and 41 attributes are defined to identify the attack and normal data packet. The feature consists of symbolic, continuous, and discrete variables, which is not appropriate for the proposed model as CAadet is considering all dataset

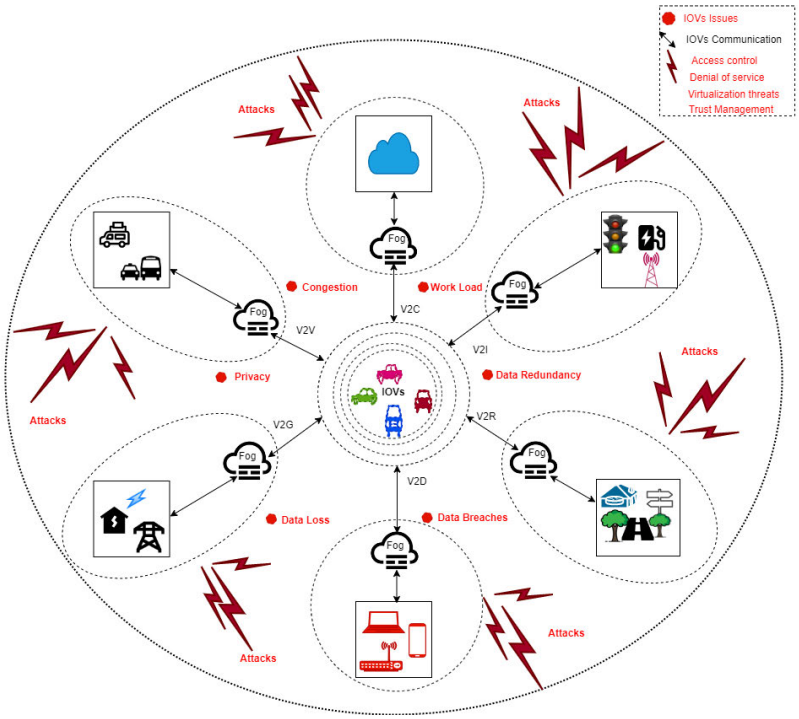

FIGURE 2. Security attacks in fog-assisted IoVs environment.

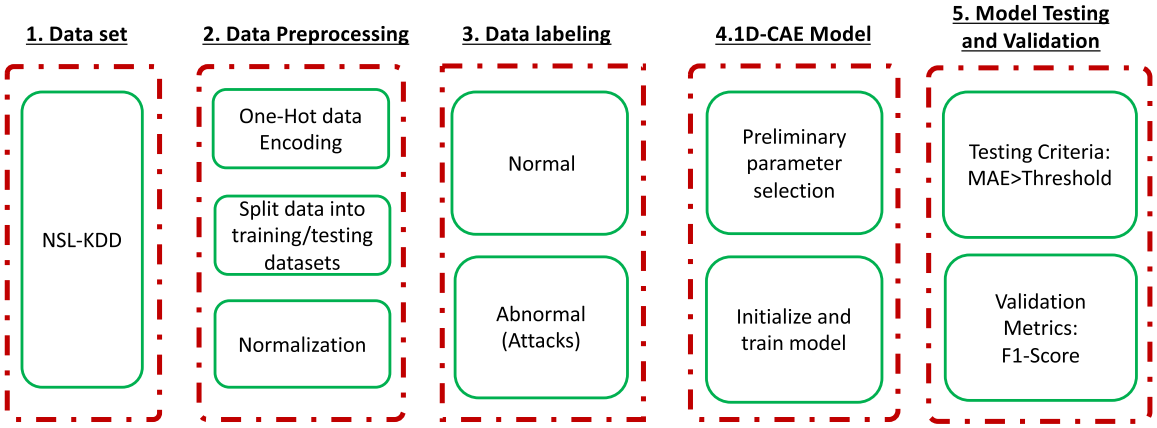

FIGURE 3. Flow diagram of proposed work.

features. Therefore, each feature has equal importance for the model.

1) DATASET PREPROCESSING

To reduce the misclassification error of the proposed model, the given dataset requires initial preprocessing before model training.

Preprocessing of a given dataset consists of three steps. Step 1: Identify and handle missing values of the dataset. Step 2: Data Encoding, One-hot encoder applied to categorical data such as integer's representation to extract binary variables for each integer value. Step 3: Data Normalization, after data encoding data normalization is applied to training and testing datasets to achieve scaled data. Scaled data

TABLE 3. Dataset categories and number of data packets in training and testing datasets-70/30.

Dataset Category	Training Dataset	Testing Dataset
Normal	53340	18435
Abnormal (Attacks)	50622	26120
Total	103962	44555

consists of all scaled features. Here, we performed normalization by using 0 and 1 for mean and standard deviation respectively.

Pseudo Code 1: Data Preprocessing

Input: *Dataset DT*
Output: *Preprocessed Dataset = D*
Step 1: DTU ←FindMissingValues (DT)
Step 2: DTU_M ←ReplaceMissingValues (DTU, Mean)
Step 3: ED ←Transforamtion(DTU_M, OneHotEncoder)
Step 4: TrainData ← Split(ED,80%),
TestData ←Split(ED,20%)
Step 5: D ← Feature Scaling (Normalization (TrainData, TestData))

B. PRELIMINARIES

1) AUTOENCODER

Autoencoder (AE) is an Artificial Neural Network (ANN), that consists of two main parts i.e. encoder and decoder [35] [41], [47]. The encoder part of AE is responsible for compressing input into a latent vector. The latent vector is the least representation of data. A latent vector is also known as a bottleneck. On the bottleneck layer, the Encoder compresses input until it reaches the smallest feature representation. The decoder's job is to reconstruct output from a latent vector and observe if it is like the input. To measure the decoder performance, reconstruction error or reconstruction loss is required to be calculated by using the constructed output and original input [41].

AE follows backpropagation for training like other feed-forward neural networks. The purpose of backpropagation during training is to minimize *Re.* Minimum reconstruction loss is an achievement of AE. The encoder function E compresses the input data INP into X stated as $X = E(INP)$. The decoder D is responsible for regenerating the given input as $X' = D(E(INP))$. The purpose of AE is to compute reconstruction errors. AE produces the output similar to the input with little variation, this change is known as reconstruction error.

There are several ways to compute *Re* such as mean squared error, mean absolute error, smooth absolute error, etc. Mean squared error (MSE) is one of the most useful methods to compute the reconstruction error. MSE measures the distance of a fitted line toward data points, it is calculated as illustrated in (1).

$$MSE = \frac{1}{k}\sum_{i=1}^{k}\left(X_i - X_i'\right)^2 \tag{1}$$

where
MSE = Mean Squared Error
K = Total number of data points
X_i = Input values
X_i' = Predicted input

2) CONVOLUTIONAL NEURAL NETWORK

Convolutional Neural Network (CNN) is a deep learning method used to analyze and process grid pattern-based datasets [35]. Applications of CNN include image processing, automatic feature extraction, and anomaly detection [35], [36], [56]. CNN architecture consists of three basic layers named input, hidden, and output layers. Typically, the hidden layer consists of convolutional layers, pooling, and connected layers. Convolutional layers are responsible to transform input into feature maps and pass input to the next layer. Convolutional layers extract features and learn patterns efficiently. The job of pooling layers is to minimize the data dimensions by integrating neuron outputs of the previous cluster into a single neuron of the next layer. The last layer is fully connected. Connected layers connect each neuron of one layer to the neurons of another layer. Flattened input passes through the connection layer to classify the data.

CNN is a layered neural network made of neurons [56]. Each neuron is calculated as a function f with input 'Y' and output 'Z. The function f is expressed as in (2).

$$Z^{(k)} = f\left(Y^{(k-1)} * w^{(K-1)} + c^{(k-1)}\right) \tag{2}$$

where 'w' is the weight vector and c is the bias of the neuron. A weight vector is a nonzero space of weights. Weight is a parameter that represents the connection strength of each neuron. In a neural network, each input is associated with a weight parameter. Weight improves the sharpness of the activation function. The weight parameter decides how the quick activation function will initiate. 'c' is a bias of the neuron, and act as a constant used to delay the triggering of an activation function. In equation (1) function *f(-)* is recognized as an activation function. There are various activation function choices such as Rectified Linear Unit (ReLU), sigmoid function, and hyperbolic tangent function, etc. In a neural network, the activation function is responsible for transforming weighted input into output.

Training of neural networks investigates the parameters (i.e weight, bias) to develop a relationship between inputs and outcomes. The neural network training phase follows the backpropagation algorithm by using an already-defined loss function. Mostly, the loss function for unique training data packet $\left(Y^{(n)}, P^n\right)$ is calculated in (3).

$$L\left(w, c, Y^{(n)}, P^n\right) = \frac{1}{2}\left\|j_{w,c}Y^{(n)} - P^n\right\|^2 \tag{3}$$

where $j_{w,c}Y^{(n)}$ is output. For T number of data packets loss function is calculated as in (4).

$$L(w, c) = \frac{1}{n}\sum_{n=1}^{T} L\left(w, c, Y^{(n)}, P^n\right) \tag{4}$$

CNN's backpropagation algorithm finds out the optimal outcome by reducing the loss function by using a gradient descent approach. Gradient descent follows the partial derivative which is denoted as $\partial Lo\partial Z_{(k,h)}^{(l)}$. Here, $Z_{(k,h)}^{(l)}$ expresses the weights among the layers of the neural network.

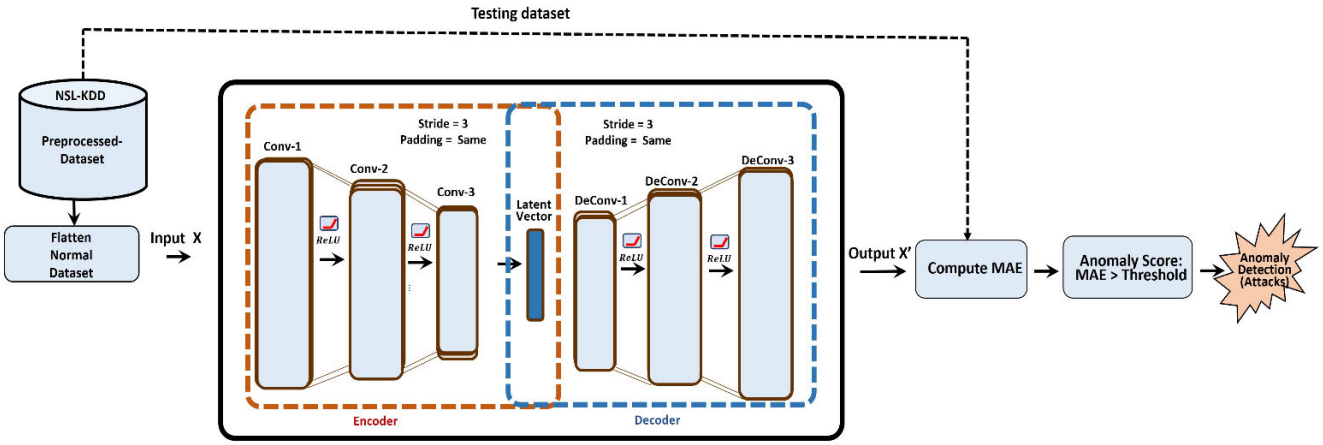

FIGURE 4. The architecture of proposed model.

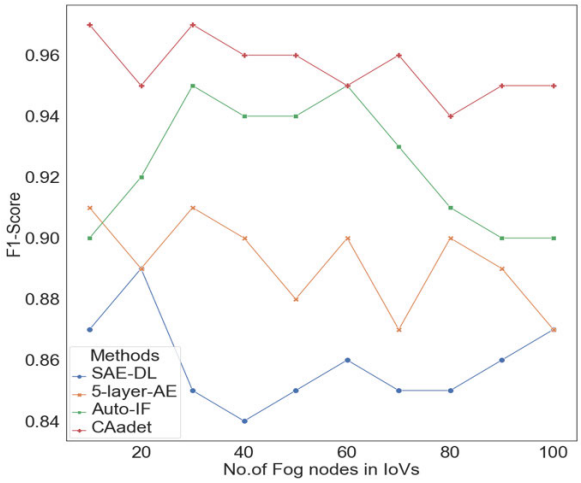

FIGURE 5. Comparison of proposed and existing schemes based on F1-Score concerning No. of Fog Nodes.

The parameter for each layer is updated as in (5).

$$Z^{(l)}_{(k,h)} = Z^{(l)}_{(k,h)} - \mathfrak{y}\frac{\partial Lo}{Z^{(l)}_{(k,h)}} \tag{5}$$

where $\mathfrak{y}$ = learning rate

C. PROPOSED SCHEME CAadet FOR FA-IoVs

This research aims to propose a deep learning-based anomaly detection scheme for Fog-assisted internet of vehicles (Fa-IoVs). We call the proposed scheme **C**onvolutional autoencoder **A**ided **a**nomaly **det**ection (CAadet) to detect attacks for Fa-IoVs environment as it is a major concern for network security. In this section we present data on our proposed scheme, CAadet consists of a convolutional neural network and autoencoder.

Figure 4 illustrates the complete architecture of the proposed scheme. CAadet follows a convolutional layers-based autoencoder for automatic feature extraction and then anomaly detection. Autoencoder architecture follows the encoder and decoder respectively. The encoder includes three convolutional layers. Each convolutional layer is responsible to minimize the input dimension concerning stride size. As an outcome, the convolutional layers turn into a latent vector. To avoid the overfitting issue, we included dropout layers along the input and convolutional layers. Dropout layers are used to handle regularization and prevent copying input as output. In fact, during training dropout layers drop some random neurons to avoid model overfitting. The CAadet decoder consists of three deconvolutional layers to reconstruct output from the latent vectors. The decoder outcome is compared with the input by computing the mean absolute error (MAE) as in (6).

$$MAE = \frac{\sum_{i=1}^{k} |X'_i - T_i|}{K} \tag{6}$$

where

$$\begin{aligned} \text{MAE} &= \text{Mean Absolute Error} \\ \text{X}'_\text{i} &= \text{Reconstructed output} \\ \text{T}_\text{i} &= \text{testing dataset input,} \\ \text{K} &= \text{Total instances in the testing dataset} \end{aligned}$$

To label an instance as an ''attack'', we used a threshold value. If an instance has a reconstruction error greater than the threshold, we label it as an ''attack'', otherwise as ''normal''. We arrived at this threshold value based on the model loss over training data and not on validation data.

To label a packet as ''abnormal'' we used a threshold value. When this MAE is greater than a given threshold value an anomaly count is increased in an otherwise normal data packet. Observe that the threshold value is a critical parameter. To achieve threshold value, MAE loss is observed concerning real anomalous packets' detection. The given dataset consists of data packets as illustrated in (7).

For a given dataset

$$D_p = \{(a_1, b_1), (a_2, b_1) \ldots (a_n, b_1)\} \tag{7}$$

where

$$D_p = Data\ Packets$$
$$a_n = total\ number\ of\ packets$$
$$b_1 = \{0, 1\}$$

Each packet of D_p may carry some important information. Data packets of a given dataset are labeled into binary classification. Normal packets of the dataset are labeled as '0' while '1' represents malicious packets. CAadet trains only normal data packets, which is more practical for real-time. For Fa-IoVs, it is important to identify normal or malicious data packets in real time.

V. RESULT

A. EXPERIMENTAL SETUP

Several public datasets are available to evaluate security threats. We opt NSL-KDD dataset for research evaluation. NSL-KDD is a benchmark dataset that is famous for cybersecurity research. The experiment is conducted by a Deep learning-based approach named convolutional autoencoder. The proposed model is chosen for automatic feature extraction and to detect anomalies by using a decided anomaly score which is explained below in this section. In this work, we target the security threats of Fog-assisted IoVs networks to ensure secure communication. Binary classification of normal and abnormal packets is done.

B. PERFORMANCE METRICS

Result evaluation metrics follow four classifications of results True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). Suppose we have a dataset consisting of normal and abnormal data packets. True and false follows the correct and wrong classification of packets respectively. TP means the model detected attacks (abnormal packets successfully, FP means there was no abnormal packet, but the model classified normal as abnormal, TN presents normal packets are detected as normal while FN means there were some abnormal packets that are misclassified as normal packets.

We concluded this research by using well-renowned evaluation parameters such as recall, precision, and F1-score [31]. These metrics are described as follows:

1) RECALL (R)

Recall is a performance metric for machine learning classification and pattern recognition-based problems. It is used to measure model sensitivity. The recall represents the relevant proportion of detected anomalous packets over the total number of packets. R can be calculated as in (8).

$$Recall(R) = \frac{TP}{(TP + FN)} \tag{8}$$

2) PRECISION (P)

Precision is also a performance metric like a recall to measure the performance for pattern recognition and classification-based problems. Precision represents the number of original packets among detected packets. A precision metric is directly proportional to measuring the rate of false alarms. False positive (FP) represents the rate of false alarms over correct anomaly detection. False alarms are misleading detections, which means when there is no threat, but the system detects a threat and misleads the user. P can be obtained as in (9).

$$Precision\ (P) = \frac{TP}{(TP + FP)} \tag{9}$$

Pseudo Code 2

Input: *Matrix* D^U, ***output:*** θ
params (w,qx, qh) ***where:*** *w : Weight between layers, qx Encoder's params, qh Decoder's Params*
initial Variables
h ← null // vector for the hidden layer
D′ ← null // Reconstructed D
L ← null // vector for Loss Function
l ← batch number
i ←0
loop statement
While *i< l do*
for *l=1: number of layers [3]*
for *ks=1: kernalsize [4]*
for *f=1: No.of filters [16]*
// Encoder function maps an input D^U *to hidden representation h :*

$$h = f(p[i].w + p[i] + qx)$$

//Decoder function maps hidden representation h back to a reconstruction D′ :

$$D' = g(p[i].w^T + p[i] + qx)$$

//For nonlinear reconstruction, the reconstruction loss is general from cross-entropy

$$L = -Sum\left(x \cdot log\left(D'\right) + (1 - x) \cdot log\left(1 - D'\right)\right)$$

/ For linear reconstruction, the reconstruction loss is generally from the squared error: */*

$$L = Sum\left(D - D'\right)^2$$
$$\theta\,[i] = L\left(D - D'\right)$$

End
End
End
End While

Pseudocode 3

Input: *D= {NDP, ADP} where NDP= {np_1, np_2, …, np_x} ADP= {ap_1, ap_2, …, ap_y} and $x \gg Y$*
Step 1: $D^U \leftarrow$ *Flatten (NDP)*
Step 2: *CovAutoMod* $\leftarrow D^U$
Output: ***from*** *step 2* θ
Step 3: *AC*$\leftarrow 0$
If *(Loss> Th)* ***then***
AD$\leftarrow$*Yes*
Increment AC
Else
AD $\leftarrow$ *No*
Endif
Out Put: AC

TABLE 4. Performance of the proposed model.

Metrics	Result %
Recall	99
Precision	96
F1-Score	97.4

3) F1-SCORE: THE

F1-score is also a common performance metric. The F1-score metric is a weighted average between recall and precision that can be calculated as in (10).

$$F = \frac{2 \times (Precision \times Recall)}{(Precision + Recall)} \tag{10}$$

C. PERFORMANCE EVALUATION

Performance evaluation is carried out based on a comparison between the proposed scheme and existing schemes. From existing schemes, two categories are targeted such as shallow-based schemes and AI-based approaches.

The proposed scheme (CAaDet) follows the NSL-KDD dataset as input and predicts results using a convolutional autoencoder. Convolutional autoencoders first extract useful features and then predict output against given input. To evaluate the result, an anomaly score is selected. Anomaly score is a criterion to consider an anomaly i.e., when the Mean absolute error (MAE) is greater than the threshold means an anomaly is detected successfully as shown in (11). MAE is calculated by using input and constructed output as shown in (1). Threshold selection is done by performing multiple experiments and observing the threshold on-ground truth. Best performance is achieved by using the 0.19 threshold.

$$Anomaly_{Score} = MAE > Threshold \tag{11}$$

where MAE = Mean absolute error

Table 4. depicted the performance of the proposed scheme to detect anomalies for Fog-assisted IoVs network. Recall above 0.8 is considered excellent performance.

TABLE 5. Comparison between proposed and existing standard algorithms.

Serial No.	Method	F1-Score
1.	SVM	0.71
2.	Random Forest	0.93
3.	Decision Tree	0.92
4.	Genetic Algorithm + Naïve bayes	0.61
5.	Variational Autoencoder	0.86
	Proposed Scheme	**97.4**

By using the above-mentioned evaluation metrics, we evaluated the proposed model and achieved better results in terms of recall, precision, and F1-score as shown in Table 4. For Fa-IoVs anomaly detection, the recall, the precision, and the F1-score of the proposed method are 99%, 96%, and 97.4% respectively.

D. COMPARISON WITH OTHER ANOMALY DETECTION METHODS

Anomaly detection has become the most trending and hot topic in the recent era. Researchers are exploring machine and deep learning algorithms for anomaly detection in various fields such as anomaly detection for IoTs-based networks is a serious concern. In this section, we compare the proposed methods with other anomaly detection methods using same the NSL-KDD dataset, where the evaluation metric is accuracy. Table 5 is adapted from [39] and [43] to compare with the proposed method.

We have compared the performance metric F1-score of the proposed method with other methods (extracted from [39] and [43]) using the same dataset to detect security threats as shown in Table 5. The findings show that the proposed scheme beats the other algorithms. Our deep learning-based proposed model was evaluated using extracted deep patterns from the training dataset. The scalability of the proposed distributed anomaly detection scheme has been evaluated by using many active nodes. A larger number of active nodes results in larger distribution and help to improve the F1-score rate as shown in Fig.5. This improvement in F1-score could be the result of model and parameters sharing between fog nodes. Being inference, parallel model training leads to master node level aggregation that can help for better learning in distributed environments such as Fog-assisted IoVs network.

Figure 6 depicts the relationships between the number of neurons and reconstruction error. The number of neurons in hidden layers plays a vital role in building reconstruction error. We observed that as the number of neurons increases less reconstruction error is calculated which results in low accuracy and F1-score.

Figure 7 shows the superior results of the proposed model against similar existing approaches based on autoencoders.

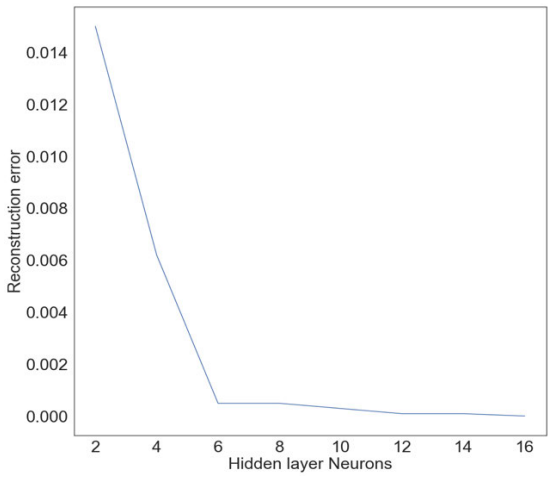

FIGURE 6. Reconstruction Error concerning hidden layer neurons.

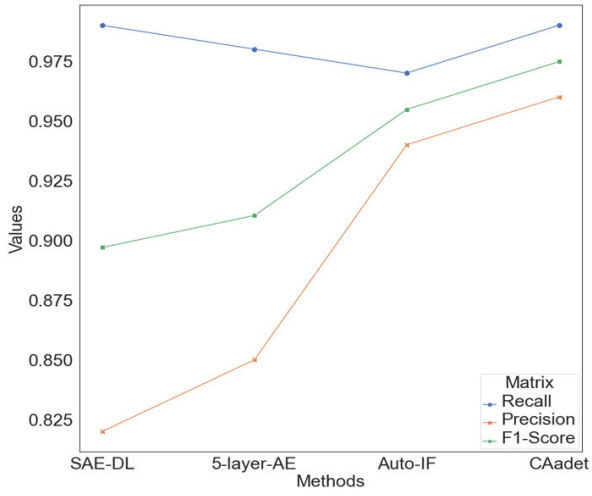

FIGURE 7. Performance comparison between proposed and existing schemes.

Three autoencoder-based anomaly detection methods with the same dataset named NSL-KDD have opted for comparison. First method SAE-DL [21] detect anomalies of Fog-assisted IoTs, it follows a stacked autoencoder approach while other two methods 5- layer-AE [37] and Auto-IF [41] also detect security threats as intrusion detection. The findings show that the proposed model outperforms the other methods with F1-score of 97.4% while SAE-DL [21], Auto-IF [37] 5- layer-AE [41] indicated an F1-score of 89%, 95%, and 91% respectively.

To evaluate the performance of CAadet for Fa-IoVs, an experiment is carried out to implement the proposed model and two others selected methods on the same machine. The two selected methods are based on a similar approach to the proposed method. Then, the proposed method is also compared with the existing results of other schemes using the same dataset and having similar objectives.

VI. CONCLUSION

The Internet of vehicles (IoVs) is a more popular network of this era. Over time, IoVs challenges have become a focal point of research. Fog computing was one of the suitable solutions for IoVs challenges to achieve smooth communication by reducing the risk of congestion, security, etc. Unfortunately, fog nodes are also challenged by security threats due to their local and on-edge nature. We can refer Fog and IoVs combination as Fog-assisted IoVs (Fa-IoVs) which is a distributed network that allows communication among vehicles such as V2V, V2I, V2D, etc. Applications of Fa-IoVs include a safe navigation system, traffic guidance, crash prevention, and intelligent vehicle control. Like fog computing, Fa-IoVs are challenged by security threats which is a serious concern for the smooth communication of the Fa-IoVs network. We consider security threats as anomalies and proposed a deep learning model for Fa-IoVs anomaly detection by using a benchmark dataset named NSL-KDD. Convolutional autoencoder has been deployed using customize parameter settings against convolutional layers for feature engineering and anomaly identification. Anomaly identification is done by setting a threshold against reconstructed error. To evaluate the performance of the proposed model "CAadet," three major comparisons are carried out: (a) Comparison with similar existing schemes using the same dataset for anomaly detection rate, (b) comparison with statistical and machine learning models (extracted from existing work [39], [43]), (c) Comparison of fog nodes and reconstruction error to achieve better F1-score. For comparison among proposed and other schemes, performance metrics such as recall, precision, and F1-score have been used. The experiments prove not only the competence of the deep learning-based proposed model over existing models but also highlighted the efficiency of distributed learning optimization over fog nodes by scoring a high F1-score. We concluded that a deep learning-based attack detection scheme for distributed Fa-IoVs networks supported by fog nodes can help to improve the accuracy rate of cyber-attack detection. Furthermore, efficient anomaly detection leads to smooth communication by reducing delay, data loss, and communication overhead. In the future, we will investigate proposed approaches in different IoTs domains by using different datasets and other deep learning models.

REFERENCES

[1] O. Kaiwartya, A. H. Abdullah, Y. Cao, A. Altameem, M. Prasad, C. T. Lin, and X. Liu, "Internet of Vehicles: Motivation, layered architecture, network model, challenges, and future aspects," *IEEE Access*, vol. 4, pp. 5356–5373, 2016, doi: 10.1109/ACCESS.2016.2603219.

[2] W. Xu, H. Zhou, N. Cheng, F. Lyu, W. Shi, J. Chen, and X. Shen, "Internet of Vehicles in big data era," *IEEE/CAA J. Autom. Sinica*, vol. 5, no. 1, pp. 19–35, Jan. 2018, doi: 10.1109/JAS.2017.7510736.

[3] J. Contreras-Castillo, S. Zeadally, and J. A. Guerrero-Ibañez, "Internet of Vehicles: Architecture, protocols, and security," *IEEE Internet Things J.*, vol. 5, no. 5, pp. 3701–3709, Oct. 2018, doi: 10.1109/JIOT.2017.2690902.

[4] S. Yaqoob, A. Ullah, M. Akbar, M. Imran, and M. Shoaib, "Congestion avoidance through fog computing in Internet of Vehicles," *J. Ambient Intell. Humanized Comput.*, vol. 10, no. 10, pp. 3863–3877, Oct. 2019, doi: 10.1007/s12652-019-01253-x.

[5] A. Ullah, S. Yaqoob, M. Imran, and H. Ning, "Emergency message dissemination schemes based on congestion avoidance in VANET and vehicular FoG computing," *IEEE Access*, vol. 7, pp. 1570–1585, 2019, doi: 10.1109/ACCESS.2018.2887075.

[6] W. Zhang and G. Li, "An efficient and secure data transmission mechanism for Internet of Vehicles considering privacy protection in fog computing environment," *IEEE Access*, vol. 8, pp. 64461–64474, 2020, doi: 10.1109/access.2020.2983994.

[7] L. Song, G. Sun, H. Yu, X. Du, and M. Guizani, "FBIA: A fog-based identity authentication scheme for privacy preservation in Internet of Vehicles," *IEEE Trans. Veh. Technol.*, vol. 69, no. 5, pp. 5403–5415, Mar. 2020, doi: 10.1109/TVT.2020.2977829.

[8] A. J. Siddiqui and A. Boukerche, "On the impact of DDoS attacks on software-defined Internet-of-Vehicles control plane," in *Proc. 14th Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Jun. 2018, pp. 1284–1289, doi: 10.1109/IWCMC.2018.8450433.

[9] J. Ni, K. Zhang, X. Lin, and X. S. Shen, "Securing fog computing for Internet of Things applications: Challenges and solutions," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 1, pp. 601–628, 1st Quart., 2018, doi: 10.1109/COMST.2017.2762345.

[10] L. Zhu, M. Li, and Z. Zhang, "Secure fog-assisted crowdsensing with collusion resistance: From data reporting to data requesting," *IEEE Internet Things J.*, vol. 6, no. 3, pp. 5473–5484, Jun. 2019, doi: 10.1109/JIOT.2019.2902459.

[11] L. Zhou, H. Guo, and G. Deng, "A fog computing based approach to DDoS mitigation in IIoT systems," *Comput. Secur.*, vol. 85, pp. 51–62, Aug. 2019, doi: 10.1016/j.cose.2019.04.017.

[12] R. Mahmud, R. Kotagiri, and R. Buyya, "Fog computing: A taxonomy, survey and future directions," *Internet of Everything*. Cham, Switzerland: Springer, 2018, pp. 103–130, doi: 10.1007/978-981-10-5861-5_5.

[13] S. Yi, C. Li, and Q. Li, "A survey of fog computing: Concepts, applications and issues," in *Proc. Workshop Mobile Big Data*, Jun. 2015, pp. 37–42, doi: 10.1145/2757384.2757397.

[14] S. Khan, S. Parkinson, and Y. Qin, "Fog computing security: A review of current applications and security solutions," *J. Cloud Comput.*, vol. 6, no. 1, pp. 1–22, Dec. 2017, doi: 10.1186/s13677-017-0090-3.

[15] S. Yaqoob, A. Ullah, M. Awais, I. Katib, A. Albeshri, R. Mehmood, M. Raza, S. Ul Islam, and J. J. P. C. Rodrigues, "Novel congestion avoidance scheme for Internet of Drones," *Comput. Commun.*, vol. 169, pp. 202–210, Mar. 2021, doi: 10.1016/j.comcom.2021.01.008.

[16] L. M. Vaquero and L. Rodero-Merino, "Finding your way in the fog: Towards a comprehensive definition of fog computing," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 44, no. 5, pp. 27–32, Oct. 2014, doi: 10.1145/2677046.2677052.

[17] X. Hou, Y. Li, M. Chen, D. Wu, D. Jin, and S. Chen, "Vehicular fog computing: A viewpoint of vehicles as the infrastructures," *IEEE Trans. Veh. Technol.*, vol. 65, no. 6, pp. 3860–3873, Jun. 2016, doi: 10.1109/TVT.2016.2532863.

[18] S. Yaqoob, A. Ullah, M. Akbar, M. Imran, and M. Guizani, "Fog-assisted congestion avoidance scheme for Internet of Vehicles," in *Proc. 14th Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Jun. 2018, pp. 618–622, doi: 10.1109/IWCMC.2018.8450402.

[19] A. Yousefpour, C. Fung, T. Nguyen, K. Kadiyala, F. Jalali, A. Niakanlahiji, J. Kong, and J. P. Jue, "All one needs to know about fog computing and related edge computing paradigms: A complete survey," *J. Syst. Archit.*, vol. 98, pp. 289–330, Sep. 2019, doi: 10.1016/j.sysarc.2019.02.009.

[20] S. Tu, M. Waqas, S. U. Rehman, M. Aamir, O. U. Rehman, Z. Jianbiao, and C.-C. Chang, "Security in fog computing: A novel technique to tackle an impersonation attack," *IEEE Access*, vol. 6, pp. 74993–75001, 2018, doi: 10.1109/ACCESS.2018.2884672.

[21] A. Abeshu and N. Chilamkurti, "Deep learning: The frontier for distributed attack detection in fog-to-things computing," *IEEE Commun. Mag.*, vol. 56, no. 2, pp. 169–175, Feb. 2018, doi: 10.1109/MCOM.2018.1700332.

[22] S. T. Zargar, J. Joshi, and D. Tipper, "A survey of defense mechanisms against distributed denial of service (DDoS) flooding attacks," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 4, pp. 2046–2069, Mar. 2013, doi: 10.1109/SURV.2013.031413.00127.

[23] S. Sumathi, R. Rajesh, and S. Lim, "Recurrent and deep learning neural network models for DDoS attack detection," *J. Sensors*, vol. 2022, pp. 1–21, Sep. 2022, doi: 10.1155/2022/8530312.

[24] S. Sumathi and R. Rajesh, "Comparative study on TCP SYN flood DDoS attack detection: A machine learning algorithm based approach," *Wseas Trans. Syst. Control*, vol. 16, pp. 584–591, Nov. 2021, doi: 10.37394/23203.2021.16.54.

[25] Y. Xu, W. Yang, X. Yu, H. Li, T. Cheng, X. Lu, and Z. L. Wang, "Real-time monitoring system of automobile driver status and intelligent fatigue warning based on triboelectric nanogenerator," *ACS Nano*, vol. 15, no. 4, pp. 7271–7278, Apr. 2021, doi: 10.1021/acsnano.1c00536.

[26] G. Albertus, M. Meiring, and H. C. Myburgh, "A review of intelligent driving style analysis systems and related artificial intelligence algorithms," *Sensors*, vol. 15, no. 12, pp. 30653–30682, 2015, doi: 10.3390/s151229822.

[27] J. Kang, R. Yu, X. Huang, and Y. Zhang, "Privacy-preserved pseudonym scheme for fog computing supported Internet of Vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 8, pp. 2627–2637, Aug. 2018, doi: 10.1109/TITS.2017.2764095.

[28] X. Wang, Z. Ning, and L. Wang, "Offloading in Internet of Vehicles: A fog-enabled real-time traffic management system," *IEEE Trans. Ind. Informat.*, vol. 14, no. 10, pp. 4568–4578, Oct. 2018, doi: 10.1109/TII.2018.2816590.

[29] H. Zhang, Y. Luo, Q. Yu, L. Sun, X. Li, and Z. Sun, "A framework of abnormal behavior detection and classification based on big trajectory data for mobile networks," *Secur. Commun. Netw.*, vol. 2020, pp. 1–15, Dec. 2020, doi: 10.1155/2020/8858444.

[30] A. A. Cook, G. Misirli, and Z. Fan, "Anomaly detection for IoT time-series data: A survey," *IEEE Internet Things J.*, vol. 7, no. 7, pp. 6481–6494, Jul. 2020, doi: 10.1109/JIOT.2019.2958185.

[31] M. A. Lawal, R. A. Shaikh, and S. R. Hassan, "An anomaly mitigation framework for IoT using fog computing," *Electronics*, vol. 9, no. 10, pp. 1–24, 2020, doi: 10.3390/electronics9101565.

[32] D. K. K. Reddy, H. S. Behera, J. Nayak, B. Naik, U. Ghosh, and P. K. Sharma, "Exact greedy algorithm based split finding approach for intrusion detection in fog-enabled IoT environment," *J. Inf. Secur. Appl.*, vol. 60, Aug. 2021, Art. no. 102866, doi: 10.1016/j.jisa.2021.102866.

[33] J. Yakubu, S. M. Abdulhamid, H. A. Christopher, H. Chiroma, and M. Abdullahi, "Security challenges in fog-computing environment: A systematic appraisal of current developments," *J. Reliable Intell. Environ.*, vol. 5, no. 4, pp. 209–233, Dec. 2019, doi: 10.1007/s40860-019-00081-2.

[34] A. Diro and N. Chilamkurti, "Leveraging LSTM networks for attack detection in fog-to-things communications," *IEEE Commun. Mag.*, vol. 56, no. 9, pp. 124–130, Sep. 2018, doi: 10.1109/MCOM.2018.1701270.

[35] C. Yin, S. Zhang, J. Wang, and N. N. Xiong, "Anomaly detection based on convolutional recurrent autoencoder for IoT time series," *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 52, no. 1, pp. 112–122, Jan. 2022, doi: 10.1109/tsmc.2020.2968516.

[36] M. Munir, S. A. Siddiqui, A. Dengel, and S. Ahmed, "DeepAnT: A deep learning approach for unsupervised anomaly detection in time series," *IEEE Access*, vol. 7, pp. 1991–2005, 2019, doi: 10.1109/ACCESS.2018.2886457.

[37] K. Sadaf and J. Sultana, "Intrusion detection based on autoencoder and isolation forest in fog computing," *IEEE Access*, vol. 8, pp. 167059–167068, 2020, doi: 10.1109/ACCESS.2020.3022855.

[38] S. Xu, Y. Qian, and R. Q. Hu, "A semi-supervised learning approach for network anomaly detection in fog computing," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2019, pp. 1–6, doi: 10.1109/ICC.2019.8761459.

[39] J. O. Onah, S. M. Abdulhamid, M. Abdullahi, I. H. Hassan, and A. Al-Ghusham, "Genetic algorithm based feature selection and Naïve Bayes for anomaly detection in fog computing environment," *Mach. Learn. Appl.*, vol. 6, Dec. 2021, Art. no. 100156, doi: 10.1016/j.mlwa.2021.100156.

[40] H. Ao, "Using machine learning models to detect different intrusion on NSL-KDD," in *Proc. IEEE Int. Conf. Comput. Sci., Artif. Intell. Electron. Eng. (CSAIEE)*, Aug. 2021, pp. 166–177, doi: 10.1109/CSAIEE54046.2021.9543241.

[41] W. Xu, J. Jang-Jaccard, A. Singh, Y. Wei, and F. Sabrina, "Improving performance of autoencoder-based network anomaly detection on NSL-KDD dataset," *IEEE Access*, vol. 9, pp. 140136–140146, 2021, doi: 10.1109/ACCESS.2021.3116612.

[42] H. O. M. Omar, S. B. Goyal, and V. Varadarajan, "Application of sliding window deep learning for intrusion detection in fog computing," in *Proc. Emerg. Trends Ind. 4.0*, May 2021, pp. 2–7, doi: 10.1109/ETI4.051663.2021.9619421.

[43] A. Binbusayyis and T. Vaiyapuri, "Unsupervised deep learning approach for network intrusion detection combining convolutional autoencoder and one-class SVM," *Int. J. Speech Technol.*, vol. 51, no. 10, pp. 7094–7108, Oct. 2021, doi: 10.1007/s10489-021-02205-9.
[44] H. Benaddi, K. Ibrahimi, and A. Benslimane, "Improving the intrusion detection system for NSL-KDD dataset based on PCA-fuzzy clustering-KNN," in *Proc. 6th Int. Conf. Wireless Netw. Mobile Commun. (WINCOM)*, Oct. 2018, pp. 1–6, doi: 10.1109/WINCOM.2018.8629718.
[45] R. F. Lova, R. M. Fifaliana, and W. P. De Silva, "Intrusion detection toward feature reconstruction using Huber conditional variational AutoEncoder," in *Proc. Int. Conf. Inf. Netw. (ICOIN)*, Jan. 2022, pp. 13–17, doi: 10.1109/ICOIN53446.2022.9687135.
[46] Y. N. Kunang, S. Nurmaini, D. Stiawan, A. Zarkasi, and Jasmir, "Automatic features extraction using autoencoder in intrusion detection system," in *Proc. Int. Conf. Electr. Eng. Comput. Sci. (ICECOS)*, Oct. 2018, pp. 219–224, doi: 10.1109/ICECOS.2018.8605181.
[47] Y. Song, S. Hyun, and Y. G. Cheong, "Analysis of autoencoders for network intrusion detection?" *Sensors*, vol. 21, no. 13, pp. 1–23, 2021, doi: 10.3390/s21134294.
[48] L. Nie, Z. Ning, X. Wang, X. Hu, J. Cheng, and Y. Li, "Data-driven intrusion detection for intelligent Internet of Vehicles: A deep convolutional neural network-based method," *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 4, pp. 2219–2230, Oct. 2020, doi: 10.1109/TNSE.2020.2990984.
[49] S. Sharma and B. Kaushik, "A survey on Internet of Vehicles: Applications, security issues & solutions," *Veh. Commun.*, vol. 20, Dec. 2019, Art. no. 100182, doi: 10.1016/j.vehcom.2019.100182.
[50] S. Sumathi, R. Rajesh, and N. Karthikeyan, "DDoS attack detection using hybrid machine learning based IDS models," *J. Sci. Ind. Res.*, vol. 81, no. 3, pp. 276–286, 2022, doi: 10.56042/jsir.v81i03.58451.
[51] S. Sokkalingam and R. Ramakrishnan, "An intelligent intrusion detection system for distributed denial of service attacks: A support vector machine with hybrid optimization algorithm based approach," *Concurrency Comput., Pract. Exper.*, vol. 34, no. 27, pp. 1–18, Dec. 2022, doi: 10.1002/cpe.7334.
[52] J. Ni, S. Member, X. Lin, and X. S. Shen, "Efficient and secure service-oriented authentication supporting network slicing for 5G-enabled IoT," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 3, pp. 644–657, Mar. 2018, doi: 10.1109/JSAC.2018.2815418.
[53] L. Wang and X. Liu, "NOTSA: Novel OBU with three-level security architecture for Internet of Vehicles," *IEEE Internet Things J.*, vol. 5, no. 5, pp. 3548–3558, Oct. 2018, doi: 10.1109/JIOT.2018.2800281.
[54] C. Fang, H. Yao, Z. Wang, W. Wu, X. Jin, and F. R. Yu, "A survey of mobile information-centric networking: Research issues and challenges," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 2353–2371, 3rd Quart., 2018, doi: 10.1109/COMST.2018.2809670.
[55] Y. Yang, L. Wu, G. Yin, L. Li, and H. Zhao, "A survey on security and privacy issues in Internet-of-Things," *IEEE Internet Things J.*, vol. 4, no. 5, pp. 1250–1258, Oct. 2017, doi: 10.1109/JIOT.2017.2694844.
[56] Y. Liu, S. Garg, J. Nie, Y. Zhang, Z. Xiong, J. Kang, and M. S. Hossain, "Deep anomaly detection for time-series data in industrial IoT: A communication-efficient on-device federated learning approach," *IEEE Internet Things J.*, vol. 8, no. 8, pp. 6348–6358, Apr. 2020, doi: 10.1109/JIOT.2020.3011726.

SHUMAYLA YAQOOB received the M.S.C.S. degree (Hons.) from the Department of Computer Science, NUML, Pakistan, in 2018. She is currently pursuing the Ph.D. degree with the Department of Electrical Electronic Computer and Telecommunication Engineering, University of Catania, Italy. She was a participant in the 3MT Presentation and won two national-level software competitions. Her research background includes international journal publications. Her research interests include vehicular communication, road safety, and the IoT.

ASAD HUSSAIN received the M.Sc. degree in electronics from Quaid-i-Azam University, Islamabad, Pakistan, in 2008, and the M.S. degree in electronics engineering from ISRA University, Islamabad, in 2016. He is currently pursuing the Ph.D. degree with the Department of Engineering and Applied Sciences, University of Bergamo, Italy. He has been a Lecturer with the Department of Engineering, National University of Modern Languages. His research interests include automatic modulation classification, signal processing, and machine learning application of heuristic computation techniques in engineering problems.

FAZLI SUBHAN received the M.Sc. degree in computer science from the University of Peshawar, Khyber Pakhtunkhwa, Pakistan, in 2003, and the Ph.D. degree in information technology from Universiti Teknologi PETRONAS, Malaysia, in 2012. He is currently an Assistant Professor with the Department of Computer Science, National University of Modern Languages (NUML), Islamabad, Pakistan. His research interests include navigation and indoor positioning systems, machine learning, and bioinformatics.

GIUSEPPINA PAPPALARDO received the Ph.D. degree in engineering of road infrastructures from the University of Catania, in 2004. Since 2004, she has been a Postdoctoral Researcher with the Department of Civil Engineering and Architecture, where she is currently a Research Fellow. Her current research interests include road safety, vulnerable users, and road asset management. Her research products include more than 60 published papers in international journals and conference proceedings. She is a member of national and international committees (such as PIARC and TRB).

MUHAMMAD AWAIS is currently a Senior Lecturer of biomedical and electrical engineer by backboard with Edge Hill University, U.K. Previously, he worked as a Research Fellow in data analytics and AI with the University of Hull, U.K., and signal processing and machine learning with the University of Leeds, U.K. His research interests include signal processing, applied machine learning, deep learning to develop ICT-based systems for the Internet of Things, industry 4.0, smart healthcare, and emergency and disastrous scenarios.

Open Access funding provided by 'Università degli Studi di Catania' within the CRUI CARE Agreement