Computers, Materials & Continua

Tech Science Press

DOI: 10.32604/cmc.2023.041667

ARTICLE

GMLP-IDS: A Novel Deep Learning-Based Intrusion Detection System for Smart Agriculture

Abdelwahed Berguiga[1,2,*], Ahlem Harchay[1,2], Ayman Massaoudi[1,2], Mossaad Ben Ayed[3] and Hafedh Belmabrouk[4]

[1]Department of Computer Science, College of Science and Arts in Gurayat, Jouf University, Sakakah, Saudi Arabia

[2]Olive Research Center, Jouf University, Sakakah, Saudi Arabia

[3]Department of Electronic Industrial, ENISo, Sousse University, Sousse, Tunisia

[4]Department of Physics, College of Science at Zulfi, Majmaah University, Majmaah, Saudi Arabia

*Corresponding Author: Abdelwahed Berguiga. Email: awberguiga@ju.edu.sa

Received: 01 May 2023 Accepted: 14 August 2023 Published: 31 October 2023

ABSTRACT

Smart Agriculture, also known as Agricultural 5.0, is expected to be an integral part of our human lives to reduce the cost of agricultural inputs, increasing productivity and improving the quality of the final product. Indeed, the safety and ongoing maintenance of Smart Agriculture from cyber-attacks are vitally important. To provide more comprehensive protection against potential cyber-attacks, this paper proposes a new deep learning-based intrusion detection system for securing Smart Agriculture. The proposed Intrusion Detection System IDS, namely GMLP-IDS, combines the feedforward neural network Multilayer Perceptron (MLP) and the Gaussian Mixture Model (GMM) that can better protect the Smart Agriculture system. GMLP-IDS is evaluated with the CIC-DDoS2019 dataset, which contains various Distributed Denial-of-Service (DDoS) attacks. The paper first uses the Pearson's correlation coefficient approach to determine the correlation between the CIC-DDoS2019 dataset characteristics and their corresponding class labels. Then, the CIC-DDoS2019 dataset is divided randomly into two parts, i.e., training and testing. 75% of the data is used for training, and 25% is employed for testing. The performance of the newly proposed IDS has been compared to the traditional MLP model in terms of accuracy rating, loss rating, recall, and F1 score. Comparisons are handled on both binary and multi-class classification problems. The results revealed that the proposed GMLP-IDS system achieved more than 99.99% detection accuracy and a loss of 0.02% compared to traditional MLP. Furthermore, evaluation performance demonstrates that the proposed approach covers a more comprehensive range of security properties for Smart Agriculture and can be a promising solution for detecting unknown DDoS attacks.

KEYWORDS

Drones; DDoS attacks; Internet of Things; deep learning; multilayer perceptron; gaussian mixture model; Industry 5.0; Agricultural 5.0

1 Introduction

The emergence and development of the fourth industrial into the creation of smart and interconnected manufacturing systems that can optimize production, reduce costs, and increase efficiency have created new opportunities for businesses to improve their competitiveness, productivity, and sustainability [1]. Indeed, the Smart Factory concept has been defined as the latest industrial trend, known as Industry 5.0 [2–4]. This concept integrates advanced technologies into production processes, such as big data, artificial intelligence, the Internet of Things (IoT), robotics, and cloud computing.

Smart farming, also called intelligent farming or Agriculture 5.0, represents the adoption of new technologies for strengthening the efficiency of agro-industrial companies. In such regard, Agriculture 5.0 is anticipated to be transformed by Industry 5.0, leading to a radical transformation in such a sector [5]. Industry 5.0 and Agriculture 5.0 are increasingly crucial to our lives since they simplify our daily tasks and interact with the world around us. Also, the application of IoT in Agriculture 5.0 ranges from family farming and plays a vital role in changing the overview of conventional agriculture scenarios. For example, the digitalization of agriculture has been brought about by the Internet of Things (IoT), cloud computing, and Artificial Intelligence (AI), where machines and production systems are connected, monitored, and controlled through a network of sensors and actuators [6]. This can assist farmers in taking appropriate measures at the appropriate moment by monitoring soil moisture, temperature, and other environmental factors, consequently making the production process more scientific regarding fertilization, pest control, and improving crop yields.

Internet of Things, remote sensing, machine learning, and deep learning revolutionize agricultural value chains as they are deployed for advanced data analysis. Additionally, machine learning enables algorithms to collect data from sensors, such as soil moisture sensors, weather stations, and satellite imagery. Collected data are then analyzed to identify and diagnose crop diseases, classify soil types, and monitor weather patterns, among other applications [7]. Also, machine learning algorithms can predict the best time for planting, watering, and harvesting crops, allowing farmers to optimize their operations and increase productivity [8].

Conversely, despite the benefits of implementing IoT in Agriculture 5.0 and Industry 5.0, several challenges could impact the agricultural and food sectors in the coming decades. The large volume of data generated by IoT devices can be overwhelming and require new data management and analysis approaches. In particular, the lack of intrinsic security measures increases the IoT's vulnerability to cyber-attacks and data breaches [9]. These data are subject to DDoS, Sybil attacks, and single-point failure. These attacks can cripple an entire agriculture system, causing significant financial losses and potentially risking lives [10].

Nevertheless, there are many risks of integrating IoT in the agricultural sector. Indeed, IoT devices and sensors generate and transmit sensitive data, such as crop yield, soil moisture, humidity, light, and nutrient levels. However, if this data is not secured correctly, it can be easily targeted for unauthorized access, leading to loss of device connectivity or collecting sensitive personal information. Moreover, IoT devices are vulnerable to malware and ransomware attacks, which can disrupt operations and compromise data integrity. Another type of threat that could be encountered in the integration of IoT in the agricultural sector is Distributed Denial of Service (DDoS) attacks, where attacker overwhelms IoT devices with a tremendous amount of traffic, making the network congested, heavy, or even data loss, specifically at peak times.

Moreover, IoT security-related are significant and must be carefully discussed and addressed appropriately. In this light, Intrusion Detection Systems (IDS) are becoming crucial for detecting and mitigating DDoS attacks in Agriculture 5.0 [11]. By leveraging machine learning algorithms, IDS can quickly detect unusual patterns in network traffic and identify potential DDoS attacks before they cause any harm. In addition, machine learning algorithms can also be adapted to detect new attack patterns, making IDS a powerful tool in ensuring the security of Agriculture 5.0. Overall, integrating IDS and machine learning is essential for protecting the agriculture industry against malicious attacks and ensuring that Agriculture 5.0 continues to provide safe and efficient farming practices.

For example, Al Jouf City, a province in the Northern Region of Saudi Arabia, is the largest plant producing olive oil in the Middle East and has been identified as a potential key player in the production of the finest quality of olive oil. Al Jouf Agricultural Development Company (Saudi Arabia) has made tremendous efforts to modernize olive farming by introducing new technologies such as drip irrigation and precision farming. To achieve such goals, the company deploys drones, sensors, and satellite imagery to help farmers collect data, monitor their crops and make data-driven decisions about planting, harvesting, and irrigation. However, with an area of 7730 hectares and more than 5 million olive trees, it is hard to prevent and detect whether there are cyber-attacks on collected data.

In such cases, machine learning and deep learning can play a more vital and robust role in analyzing vast amounts of data collected from various sources and detecting and preventing if there are cyber-attacks in farming systems [12–14]. Attackers can exploit vulnerabilities in connected devices and other equipment, compromising data security and potentially causing significant harm to farming operations. More crucially, machine learning and deep learning algorithms can aid in analyzing large amounts of data to identify potential threats and detect whether there are cyber-attacks. As a result, farmers can take appropriate measures to prevent or mitigate cyber-attacks' impact on their operations [15].

The remainder of the paper is carried out as follows: Section 2 discusses the relevant scientific literature on IDS models. Then in Section 3, we describe our proposed GMLP-IDS with its main components. The evaluation experiments and results outcomes are described in Section 4. The final part of this paper outlines in Section 5 the conclusion and future works.

2 Related Works

The paper [16] provided a novel approach for detecting assaults in the Internet of Things (IoT) using deep learning. The study focused on the security problems faced by IoT systems due to their distributed design and the enormous number of connected devices. The suggested solution comprises employing a distributed system of edge devices and a centralized server for the training and inference of a deep learning model. The edge devices collect and pre-process data from the IoT devices and transfer it to the centralized server for model training. The learned model is then transmitted back to the edge devices for inference, enabling real-time detection of assaults.

The performance of the suggested strategy is assessed using the KDDCUP'99 dataset and compared to that of more established machine learning and rule-based methods. The findings demonstrate that the proposed deep learning methodology achieves good results for detection accuracy, very low false positive rate, and detection time compared to the alternative approaches. Overall, the study makes a significant contribution to the field of IoT security by outlining a distributed strategy for overcoming the difficulties associated with identifying attacks in complex IoT systems. However, the proposed scheme exchanges immense control messages between edge devices, fog nodes, and the central server, leading to large-scale IoT systems bottlenecks. Moreover, the technique depends on the edge devices' and fog nodes' dependability and security, which may only sometimes be the case in practical use. Finally, the study does not mention the suggested approach's potential for expansion to handle larger datasets or more complex attacks.

Thamilarasu et al. [17] designed an intelligent intrusion detection approach for IoT systems using deep learning algorithms. In their work, authors combined Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) to extract features and classify network traffic anomalies in real-time. The approach is evaluated on the Bot-IoT dataset, and the results show that it can detect various types of attacks with a high detection rate and low false positive rate. However, the authors do not take into account the dynamic nature of IoT networks, where devices can join and leave the network at any time, resulting in changing network traffic patterns. Also, the approach relies on the availability of labeled training data for deep learning algorithms, which is difficult to be applied in practice. Finally, the proposed approach does not consider the impact of network latency and communication overhead on real-time attack detection. Overall, this paper provides a promising approach for detecting intruders in IoT systems, but further research is needed to address these weaknesses and evaluate the proposed approach in real IoT systems is required.

In [18], the authors provided a comprehensive review of the use of Federated Learning (FL) for Intrusion Detection Systems (IDS) in the Internet of Things (IoT) context. They focus on Federated Learning (FL) which is proposed as a potential solution to address the challenges posed by these networks' distributed and heterogeneous nature. Several advantages and limitations of FL-based intrusion detection are discussed in the paper, including the need for secure and efficient communication protocols, the selection of appropriate machine learning models, and the effect of data privacy concerns on FL algorithms' performance. The authors conclude that FL has the potential to improve the effectiveness and efficiency of IDS in IoT, but further research is needed to address the challenges mentioned above and ensure the security and privacy of the data involved.

A novel approach for anomaly detection in Internet of Things (IoT) networks was proposed by Ullah et al. [19]. The authors used Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks instead of a standard RNN for anomaly detection in IoT networks. The proposed approach continuously analyzes the incoming network traffic and compares it with the learned standard patterns to detect anomalies in real-time. Compared to other proposed approaches [20], the output results show that the proposed deep learning model based on the recurrent neural approach outperforms previous deep learning implementations in terms of precision, recall, accuracy, and F1 score, as well as its ability to detect anomalies in real-time makes it a promising approach for intrusion detection in IoT environments. Nonetheless, even if authors explored that the proposed solution is based on a regular pattern of behavior in the IoT network and remains consistent over time, the optimal values for hyper-parameters for the Recurrent Neural Network (RNN) model, such as the number of layers, the number of neurons per layer, and the learning rate may depend on the characteristics of the specific IoT networks. They may require extensive experimentation to find the best configuration.

In [14], the authors addressed how smart farming can improve Agriculture 4.0 and then increase efficiency, productivity, and sustainability. The authors emphasized using cutting-edge technologies, namely, cloud computing, edge computing, AI, IoT, big data, and 6G in agriculture. The authors discussed the benefits of "smart farming", including precision farming, resource optimization, and predictive maintenance. Also, they explore the problems devoted to traditional farming, such as increasing food production, water scarcity, and the need for climate change. According to these characteristics, smart farming can help to solve resource optimization usage and lack of technical expertise. However, a comprehensive analysis of the challenges faced by Agriculture 4.0 technologies is not outlined by the authors. Also, a detailed discussion of the policy implications of implementing Agriculture 4.0 technologies in farming is not present.

In another recent study, Friha et al. [21] developed a novel intrusion detection system (IDS) for the Agricultural Internet of Things (IoT) that uses Federated Learning (FL) as its core technology. The proposed scheme is based on a distributed architecture, where multiple IoT devices collect data and train a machine-learning model using the Federated Learning approach. This latter allows knowledge sharing with privacy maintenance and cost reduction. The model is then used to detect anomalies and intrusions in real-time. The paper also provided a detailed description of the FELIDS architecture and its components, including the data collection module, the FL module, and the IDS module. Consequently, FELIDS is then evaluated using a dataset of real-world attacks on IoT-based agriculture systems, and performance results are compared to other IDS systems. Authors demonstrate that the proposed scheme FELIDS achieves a higher detection rate and lower false positive rate than classic versions of machine learning, making it a promising solution for securing IoT-based agriculture systems.

Gupta et al. [22] have used a hierarchical deep-learning model to analyze network traffic and detect anomalies. The model is trained on an extensive network traffic dataset to learn patterns and identify abnormal behavior. Then, it monitors network traffic in real-time and generates alerts when detecting anomalous activity. The proposed model aims to enhance the cybersecurity of multi-cloud healthcare systems. Authors have also shown that the proposed approach outperforms traditional approaches in terms of accuracy and detection rates. Overall, the proposed approach has the potential to improve the security of healthcare systems and protect sensitive patient information from cyber threats.

Gudla et al. [23] designed a new detection scheme, named deep intelligent DDoS attack detection scheme (DI-ADS), for Distributed Denial of Service (DDoS) attacks in fog-based Internet of Things (IoT) applications. Fog defender utilizes a deep learning-based approach to allow only legitimate requests to be accessed to the cloud. DI-ADS employs a distributed architecture that uses edge devices located at the fog layer of the IoT network to collect and pre-process data before sending it to a central server for analysis. The system utilizes a Convolutional Neural Network (CNN) to classify network traffic as either normal or attack traffic. Traffic attacks are generated using the well-known Kali Linux tool, and the performance of DI-ADS is evaluated using various metrics, including behavior detection time, false positive rate, loss rate, and accuracy.

The authors of [24] proposed a hybrid lightweight system to detect cyber-attacks in the Internet of Medical Things (IoMT) fog environment. The proposed system uses machine learning algorithms to detect network anomalies and abnormal behaviors, and it consists of two main components: a traffic profiler and a hybrid anomaly detection model. The traffic profiler collects network traffic data and extracts relevant features. The hybrid anomaly detection model combines two machine learning algorithms, a one-class Support Vector Machine (SVM) and a decision tree, to detect network anomalies and abnormal behaviors. Experimental results show that the proposed system can detect network anomalies and abnormal behaviors with high accuracy and low computational and memory overhead, making it suitable for use in resource-constrained IoMT environments.

The study conducted by Jiang et al. [25] proposed a dynamic ensemble algorithm for anomaly detection in IoT-imbalanced data streams. The algorithm considers the dynamics, continuity, and data streams generated in IoT systems. To address the data imbalance problem, the proposed algorithm uses a combination of multiple classifiers to improve anomaly detection accuracy. Indeed, the algorithm uses the Borderline-SMOTE approach to generate minority samples in training sets. The Borderline-SMOTE dynamic weighted LightGBM (BSDWLGB) algorithm also uses LightGBM as base classifiers and adopts a chunk-based dynamic weighted ensemble mechanism, enabling classifiers to adapt to data distribution changes and remove the classifier whose accuracy performance is lower than a given threshold. The proposed algorithm is evaluated using several datasets, and the results show that it outperforms several state-of-the-art algorithms in terms of accuracy, precision, recall, and F1 score. Furthermore, the proposed algorithm can effectively detect anomalies in IoT-imbalanced data streams and can be used in various applications, such as intrusion detection, fault diagnosis, and predictive maintenance.

The study conducted by [26] proposed a novel deep learning-based anomaly detection method called FlowADGAN for network intrusion detection systems (NIDS). FlowADGAN is inspired by the successes of Generative Adversarial Networks (GANs) for detecting anomalies in the area of Computer Vision and Images. It consists of two components: a generator and a discriminator. The generator learns to generate normal network traffic data, while the discriminator learns to distinguish between normal and abnormal network traffic data. Furthermore, a comprehensive evaluation of FlowADGAN demonstrates that the proposed scheme can detect known and unknown network intrusions and is a promising deep learning-based anomaly detection method for NIDS.

Wang et al. [27] proposed a new method to classify malicious code families by merging convolutional neural networks (CNNs) with generative adversarial networks (GANs). The authors emphasized that conventional malware detection methods are not effective against the continuous evolution of malicious code variants. To deal with the associated challenges, new methods are crucial to enhance the efficiency and accuracy of malware detection. The proposed framework uses a code visualization method to convert malicious code into images, which can then be fed into a CNN for feature extraction and classification. According to the evaluation performance of the proposed method on a dataset of malicious code families, the authors show that the method achieves a classification accuracy of 97.78% and outperforms the conventional malware detection methods. The summary of deep learning approaches for network intrusion detection for IoT networks is presented in Table 1.

Table 1: Summary of deep learning approaches used for anomaly-based IDSs in IoT networks

	Reference	Published year	Focused domain	Network model	Deep learning techniques	Basic idea	Dataset used	Performance metrics	Type of experiment	Weaknesses
1	[16]	2018	Internet of Things (IoT)	Distributed system of edge devices and a centralized server	Convolutional Neural Network (CNN) with ReLU activation functions and dropout regularization	Distributed approach for detecting attacks in IoT systems using deep learning	KDDCUP'99 and NSL-KDD	Detection accuracy, false alarm rate, and detection time	The experiments involve pre-processing the data, training the deep learning model on the centralized server, distributing the model to the fog nodes and edge devices, and evaluating the model's performance on the test data	Requires a significant amount of communication between the edge devices, fog nodes, and the centralized server, which can be a bottleneck for larger-scale IoT systems. Additionally, the approach assumes that the edge devices and fog nodes are reliable and secure, which may not always be the case in practice
2	[17]	2021	Internet of Things (IoT)	LSTM and CNN	LSTM for classification, CNN for feature extraction	Develop an intelligent intrusion detection approach using deep learning algorithms	Bot-IoT dataset	Detection rate, Recall (TPR), and F1 score	Experimental evaluation on Bot-IoT dataset	Requires labeled training data, does not consider dynamic nature of IoT networks, does not consider network latency and communication overhead
3	[14]	2022	Smart agriculture driven by IoT	N/A	N/A	Integration of IoT, big data analytics, AI, and robotics into the agriculture sector to create a smart and interconnected agricultural system	N/A	N/A	Literature review and case study analysis	A comprehensive analysis of the challenges faced to Agriculture 4.0 technologies is not provided. Also, a detailed discussion is not present of the policy implications of implementing Agriculture 4.0
4	[21]	2022	Intrusion detection system for Agricultural Internet of Things (IoT) using Federated Learning	Distributed architecture	Federated Learning (FL)	Use FL to develop an IDS system for IoT-based agriculture that achieves high detection rate	Real-world attacks on IoT-based agriculture systems	Detection rate, False positive rate	Comparative evaluation of FELIDS with other IDS systems	Limited evaluation with only one dataset; May require significant computational resources to implement FL

(Continued)

Table 1 (continued)

	Reference	Published year	Focused domain	Network model	Deep learning techniques	Basic idea	Dataset used	Performance metrics	Type of experiment	Weaknesses
5	[22]	2022	Multi-cloud healthcare systems	Hierarchical deep learning	Convolutional neural networks, long short-term memory	Enhance cybersecurity of multi-cloud healthcare systems	Proprietary dataset of network traffic	Accuracy, detection rate, false positive rate	Experimental evaluation	Limited discussion of limitations and future work
6	[23]	2022	Fog-based IoT applications and DDoS attack detection	Distributed Fog computing model	Deep neural network (DNN)	A distributed DDoS attack detection scheme	DDoS-SDN dataset (Mendeley Dataset)	Accuracy, detection rate, and false positive rate	Simulation experiments	Limited evaluation of the scalability of the proposed system
7	Our work	/	Smart agriculture driven by IoT	Distributed Fog computing model	Deep neural network (DNN)	A distributed DDoS attack detection scheme using a DNN model deployed on Fog nodes Smart agriculture driven by IoT	CIC-DDoS2019 dataset	Accuracy, detection rate, false positive rate	Simulation experiments	Challenges on time and memory consumption is not discussed at IoT environment

3 Proposed Framework: GMLP-IDS

3.1 Network Model

This research deploys a network model with three layered architectures: The uppermost layer is the cloud layer, the middle layer is referred to as the fog layer, and the lower layer is composed of IoT or smart devices. The three-layered network model is a prevalent architecture in the industry 5.0 paradigm, which has gained significant global traction as a fundamental driver of industrial growth and transformation.

The topmost layer, as shown in Fig. 1, is the cloud layer, consisting of many cloud storage nodes and a cloud data center. The cloud layer provides scalability and resilience for storing and processing large amounts of data. It serves as a centralized storage unit for updated information on the behaviors of various IoT devices. Additionally, the IoT devices in the lower layer receive regular updates from the cloud layer, ensuring that their behaviors are up-to-date, whether normal or under DDoS status.

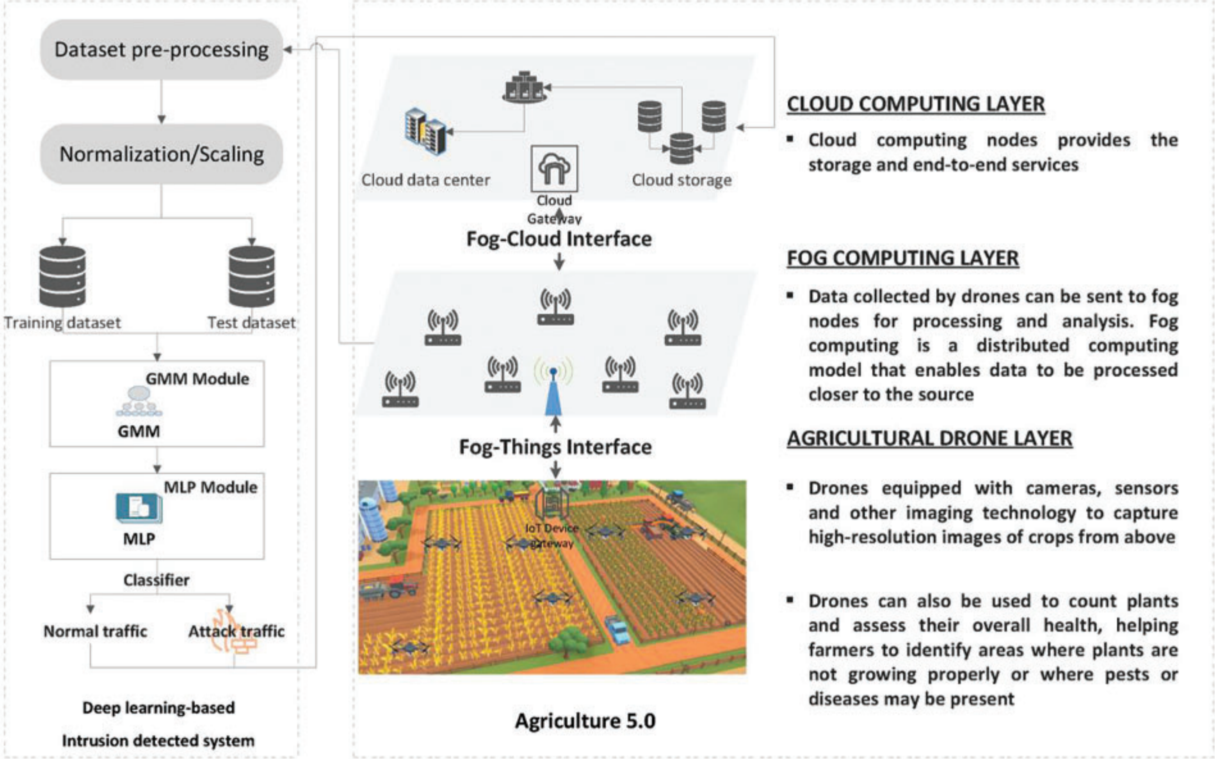

Dataset pre-processing
Normalization/Scaling
Training dataset
Test dataset
GMM Module
GMM
MLP Module
MLP
Classifier
Normal traffic
Attack traffic
Deep learning-based
Intrusion detected system
Cloud data center
Cloud Gateway
Cloud storage
Fog-Cloud Interface
Fog-Things Interface
IoT Device Gateway
Agriculture 5.0
CLOUD COMPUTING LAYER
▪ Cloud computing nodes provides the storage and end-to-end services
FOG COMPUTING LAYER
▪ Data collected by drones can be sent to fog nodes for processing and analysis. Fog computing is a distributed computing model that enables data to be processed closer to the source
AGRICULTURAL DRONE LAYER
▪ Drones equipped with cameras, sensors and other imaging technology to capture high-resolution images of crops from above
▪ Drones can also be used to count plants and assess their overall health, helping farmers to identify areas where plants are not growing properly or where pests or diseases may be present

Figure 1: System architecture for the proposed deep learning-based IDS for DDoS attack in Agriculture 5.0

The intermediate layer is called the fog layer, consisting of fog nodes (FN_1, FN_2, ..., FN_n) for data processing and storage. It is designed to provide low latency and efficient processing power to meet the needs of applications running on the devices in the lower layer. These nodes offer services to nearby IoT devices and maintain timely records of their behavior. This intermediate layer acts as a bridge between different IoT systems and the cloud layer, thus facilitating the interoperability of IoT systems. Communication between fog nodes is carried out via wired or wireless links. The fog layer is connected to the upper layer via cloud gateway (**CG**) and base stations (**BSs**), and communication is carried out using wired or wireless connections. The IoT Device gateway facilitates Communication between the fog and the lower layer.

The lower layer, also known as the IoT or smart devices layer, consists of drones and sensors. In this layer, drones detect changes or events in the environment and process substantial amount of data. By flying over fields, drones can communicate with other IoT devices (e.g., sensors) and be placed directly in the soil to measure soil conditions: temperature, humidity, light, soil moisture, and nutrient levels. Also, drones equipped with cameras and sensors can monitor crop health, detect anomalies, and respond to all end-user requests. As mentioned earlier, these drones communicate with the fog and cloud layers to perform speedy computation and service. However, despite their essential role, IoT devices in this layer have limited storage and computational capacity.

Using our proposed system for DDoS attacks has several potential ethical implications, and the system must handle data privacy issues. In fact, our system handles collected data privately to avoid unauthorized access and ensure that collected data is processed and stored transparently and securely. Secondly, sensitive information is often encrypted to protect it from unauthorized access. Finally, as illustrated in Fig. 1, our IDS ensures that user data is used for its intended purpose and stored at the cloud data center. The cloud data center provides a centralized infrastructure for storing, processing, and managing large amounts of data sent from the agricultural sensors layer in a cloud computing environment.

3.2 DDoS Attack Model

In this section, we examine the system's security issues where the Attacker (Ai) attempts to gain access to a communication channel and control targeting IoT devices such as drones, sensors, etc., through various methods to process a large amount of data.

To achieve this, as shown in Fig. 2, the attacker compromises the IoT devices by replacing their normal functions with malicious code, causing them to behave like malicious nodes. The compromised IoT devices then use various techniques to launch attacks on fog nodes at the intermediate layer, such as sending excessive requests to jam the network and gain control. This study focuses on 11 DDoS attacks, including Network Time Protocol (NTP), Domain Name System (DNS), Lightweight Directory Access Protocol (LDAP), Microsoft SQL Server (MSSQL), Network Basic Input/Output System (NetBIOS), Simple Network Management Protocol (SNMP), Simple Service Discovery Protocol (SSDP), User Datagram Protocol (UDP), the delay that can occur when sending and receiving data over UDP (UDP-Lag), Synchronization (SYN), and Trivial File Transfer Protocol (TFTP), to evaluate our proposed work. To accomplish this, we utilized the CIC-DDoS2019 dataset for training and testing purposes [28]. Indeed, as mentioned earlier, the proposed system model has three layered architectures: The uppermost layer is the cloud layer, the middle layer is referred to as the fog layer, and finally, the agricultural sensors layer comprises various IoT devices and drones that monitor agricultural environment data. The IoT and drones that monitor agricultural environment data make up the agricultural sensors layer. Actuators are activated in this layer, which also features a smart grid design and innovative energy technology for powering IoT devices. In each fog node, a deep learning-based intrusion detection system is placed. The IoT data are transmitted directly to the fog computing layer for analysis and machine learning algorithms. By outperforming nature in detecting DDoS attacks, GMLP-IDS is being implemented in this middle layer, the fog computing layer. The cloud layer provides scalability and resilience for storing and processing large amounts of data. It serves as a centralized storage unit for updated information on the behaviors of various IoT devices.

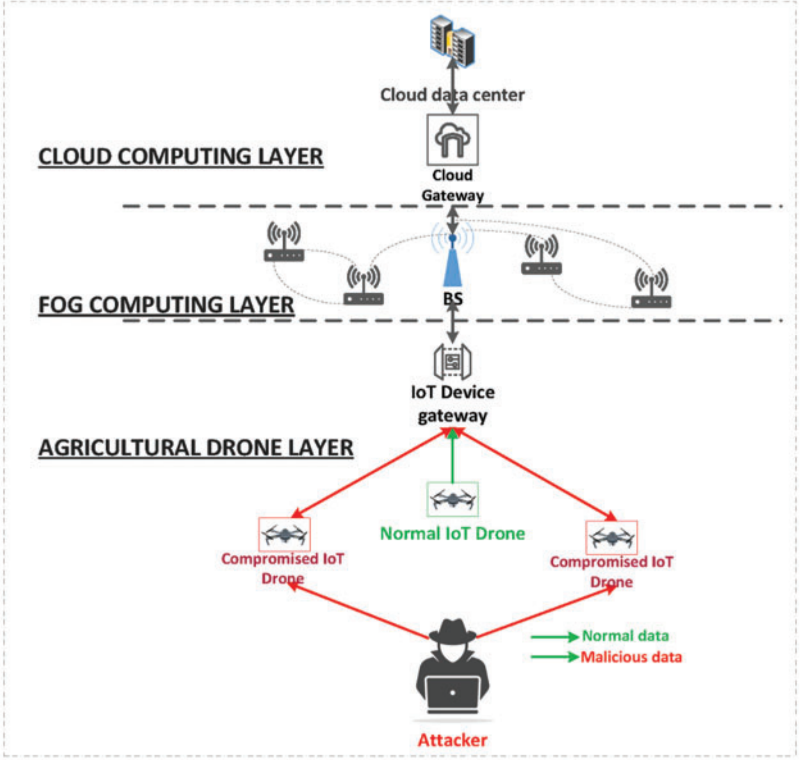


Figure 2: DDoS attack model

3.3 Data Preprocessing

3.3.1 Description of the Dataset and Study Area

In this section, we discuss the effectiveness of our proposed Intrusion Detection System using the CIC-DDoS2019 dataset, which was created as part of a study by the Canadian Institute for Cybersecurity (CIC). It contains a wide range of the most recent and popular Distributed Denial of Service (DDoS) attacks based on real-world information such as LDAP, MSSQL, NetBIOS, PortMap, UDP, SYN, and UDP-Lag, all available in both PCAP and CSV files [28,29]. The CIC-DDoS2019 dataset, shown in Table 2, comprises approximately 50,063,123 records, with 50,006,249 rows dedicated to DDoS attacks and 56,874 rows for benign traffic. Each row has 88 features, and the dataset is divided into two classes: benign and malicious. The performance of our proposed GMLP-IDS model on the CICDDoS2019 dataset is evaluated to detect and classify malicious traffic in real-world scenarios accurately. Moreover, this dataset can be regarded as a valuable resource for researchers and practitioners in the field of network security by providing a suitable basis for evaluating the system's performance against a wide range of DDoS attacks [30].

3.3.2 Binary Pre-Processing

As mentioned earlier, the CIC-DDoS2019 dataset contains a wide range of the most recent and popular DDoS attacks based on real-world information [28]. This dataset was divided into 11 subtypes of attacks to have a better-engineered and more diverse set of attacks to be used for DDoS attack detection. Eighty-eight features describe each record on the dataset, and the attack type is based on binary classification, such as normal or malignant traffic. The features included in the dataset include source and destination IP addresses, port numbers, packet sizes, timestamps, and other features. Each attack generates a different kind of malicious network flow.

Table 2: Summary of attack types in CIC-DDoS2019 dataset

Attack type	Flow count
DNS attack	5074414
NTP attack	1217008
LDAP attack	2181543
MSSQL attack	4524499
NetBIOS attack	4094987
SNMP attack	5161378
SSDP attack	2611375
UDP-lag attack	370606
UDP attack	3136803
SYN attack	1582682
TFTP attack	20107828

To further verify the effectiveness of our GMLP-IDS, CIC-DDoS2019 is used to perform binary classification, so we divided the dataset into two portions: 75% for the training process. In contrast, 25% is allocated for the testing process. Table 3 summarizes the flow count for training and testing processes in the CIC-DDoS2019 dataset.

Table 3: Flow count for training and testing processes in CIC-DDoS2019 dataset

Attack type	Training	Test
DNS attack	3805811	1268604
NTP attack	912756	304252
LDAP attack	1636157	545386
MSSQL attack	3393374	1131125
NetBIOS attack	3071240	1023747
SNMP attack	3871034	1290345
SSDP attack	1958531	652844
UDP-lag attack	277955	92652
UDP attack	2352602	784201
SYN attack	1187012	395671
TFTP attack	15080871	5026957

3.3.3 Multiclass Pre-Processing

The main objective of our proposed model in this section is multiclass classification in the context of DDoS attacks. Each CSV file from CIC-DDoS2019 contains 88 features related to network traffic and is labeled with one of two classes: BENIGN or a specific type of network attack (DNS, MSSQL, UDP, etc.). First, each file belonging to the dataset is unified into a separate information instance (data frame). Next, we extract a random number of rows from each of these data frames. We then append the resulting data frames into a single data frame. Once we combined all the data into a single information instance, we removed specific unwanted attributes (NaN, and infinite) from the selected data.

Additionally, incorporating a data pre-processing phase in the deployment process leads to a more robust training phase and, consequently a more accurate model, as set out by [31]. Table 4 presents the statistics of occurrences that have been selected from each data frame. The final step of data pre-processing is a one-hot encoding process [32]. Indeed, datasets containing categorical information and label classes that cannot be directly processed by machine learning or deep learning models. To enable accurate predictions, the data must be transformed into numerical values using one-hot encoding. This technique involves creating binary vectors to represent each possible category of categorical variables. Also, to further improve the classification process, all fields are normalized by rescaling their dynamic ranges. This normalization step ensures that no single feature dominates the training process [33,34].

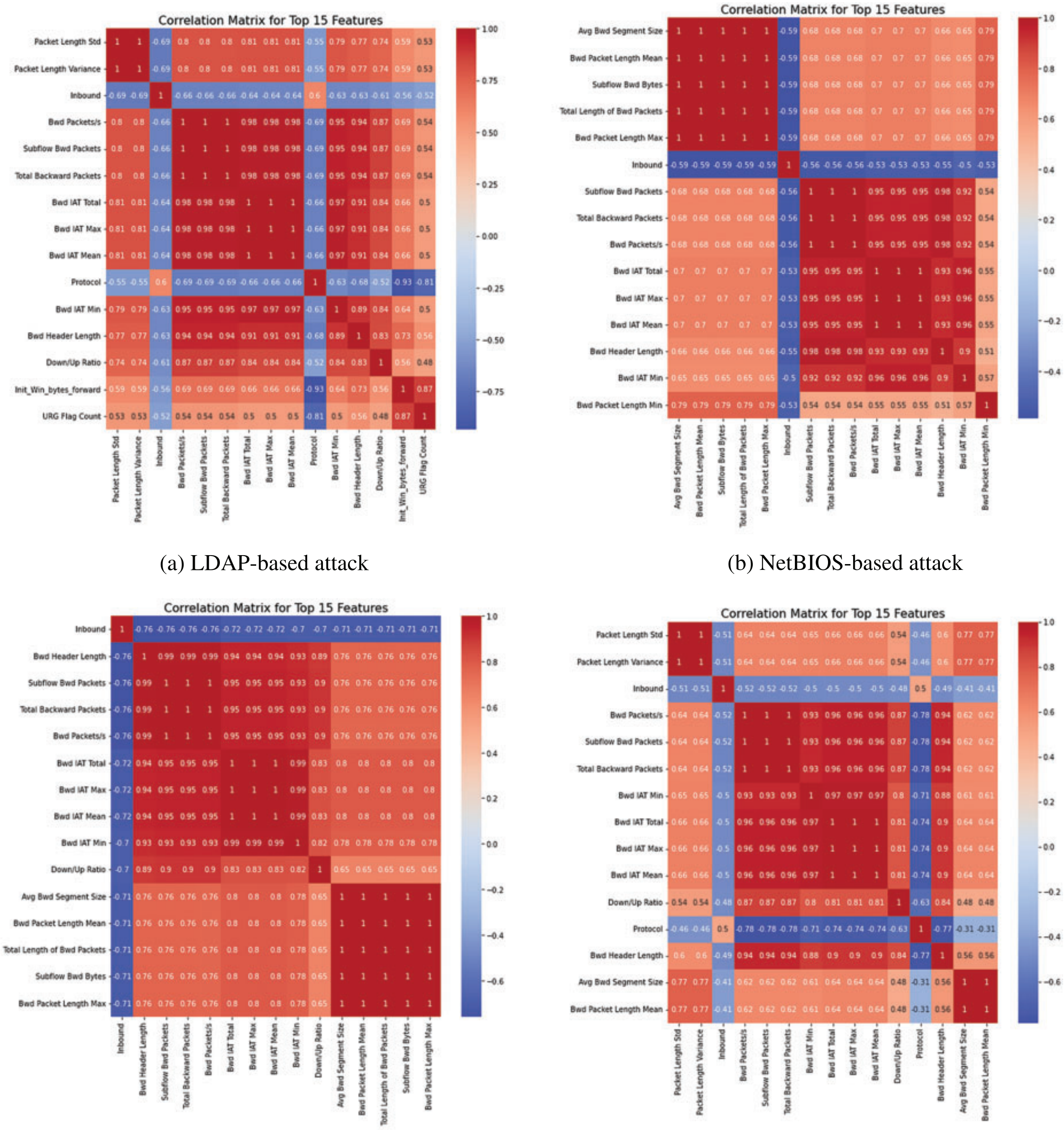
Table 4: Statistic of occurrences selected from CIC-DDoS2019 dataset for multiclass classification

Attack type	Occurrences	Percentage
DNS attack	213726	4.21%
NTP attack	269875	22.18%
LDAP attack	270169	12.38%
MSSQL attack	222855	4.93%
NetBIOS attack	306755	7.49%
SNMP attack	228232	4.42%
SSDP attack	210001	8.04%
UDP-lag attack	347408	93.74%
UDP attack	299385	9.54%
SYN attack	340562	21.52%
TFTP attack	427453	2.13%

3.3.4 Feature Selection

As mentioned earlier, each row on the CIC-DDoS2019 dataset has 88 features and 50,063,123 records. To properly evaluate these huge amounts of data, selecting weighted features with a significant linear relationship with the class label is essential. For this purpose, Pearson's correlation coefficient method was performed in this work to select the 15 features that had the highest correlation with the class label [35]. By selecting only the most important features, the problem's dimensionality can be reduced, and then the model will be simpler and more efficient. Additionally, reducing the number of features can help prevent overfitting, which occurs when the model is too complex. Fig. 3 depicts the correlation matrix for the top 15 features for each DDoS attack classification (DNS, LDAP, MSSQL, NetBIOS, NTP, SNMP, SSDP, UDP, SYN, UDP-Lag, TFTP).

It was found from Fig. 3 that each correlation matrix is characterized by its own set of 15 features identified as the most significant in detecting and classifying DDoS attacks. These matrices visually represent the relationship between features and show how they are correlated to identify patterns and anomalies in network traffic. By analyzing the correlation matrices, we can gain insights into the behavior of network traffic and develop more accurate and effective DDoS attack detection and prevention techniques.



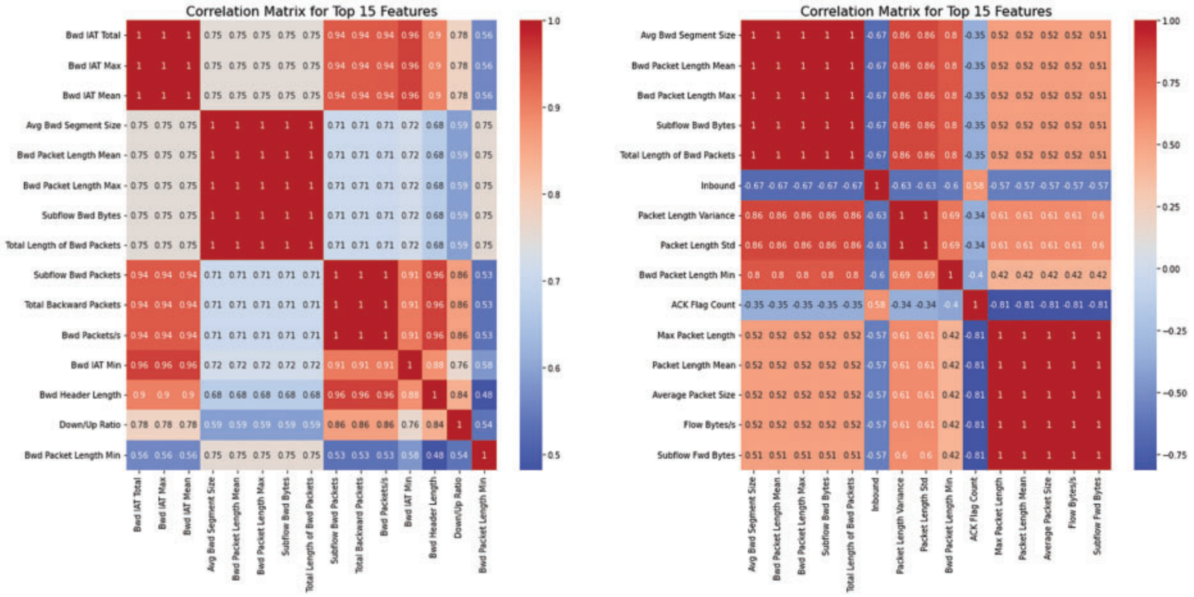
(a) LDAP-based attack

(b) NetBIOS-based attack

(c) NTP-based attack

(d) SNMP-based attack

Figure 3: (Continued)



(e) SSDP-based attack

(f) Syn-based attack

Figure 3: Selecting features by using the Pearson's correlation coefficient method

3.4 GMLP-IDS Module

The main contribution of this research work is to develop a novel intrusion detection system, named GMLP-IDS, to identify cyberattacks in the field of cybersecurity and Agriculture 5.0. Our system utilizes a hybrid approach that combines two different models, namely the Gaussian Mixture Model (GMM) and MLP classifier, to accurately classify input data into one of several possible output classes. GMM is an extension of the single Gaussian model that assumes data points are generated from a mixture of multiple Gaussian distributions. GMM extracts features from input data, which are then passed on to the MLP classifier for classification and, ultimately, for detecting cyberattacks. The functional steps of the proposed GMLP-IDS system are described as follows:

1. **Fitting a GMM to the training data:** In this step, we have used the Expectation-Maximization (EM) algorithm to estimate the underlying probability distribution of the input features. The GMM is configured with a number of Gaussian components equal to 2, covariance type set to "full", and we fix the number of iterations as 100.
2. **Data transformation:** Once the GMM is trained, we transform the input data into a set of feature vectors. This set of feature vectors represents the probability of each input data point belonging to each Gaussian component in the Gaussian Mixture Model.
3. **Training the MLP classifier:** Use the transformed feature vectors as input to train an MLP classifier using the backpropagation technique. This is done using the TensorFlow framework.
4. **Making predictions:** This last step is used to make predictions on new data by first transforming the new data using the same GMM that was used to transform the training data. Then, the MLP classifier is used to predict the class label of the new data.

As shown in Fig. 4, supervised learning is utilized to distinguish between regular traffic and DDoS attacks using a GMLP-IDS. The GMM, based on unsupervised learning, is utilized to model the probability distribution of a dataset as a combination of multiple Gaussian distributions, each representing a subpopulation or cluster within the data. Then, the MLP takes the numeric and normalized values of the selected features prepared in the previous step to predict the class label of the new data. Traffic engineers identify and label each cluster, which is then used to update the GMLP-IDS and enable it to identify previously unseen attacks. Expert analysis is utilized to label and identify unknown traffic or attacks that the Gaussian mixture model has classified. The deep learning model is updated through incremental learning.

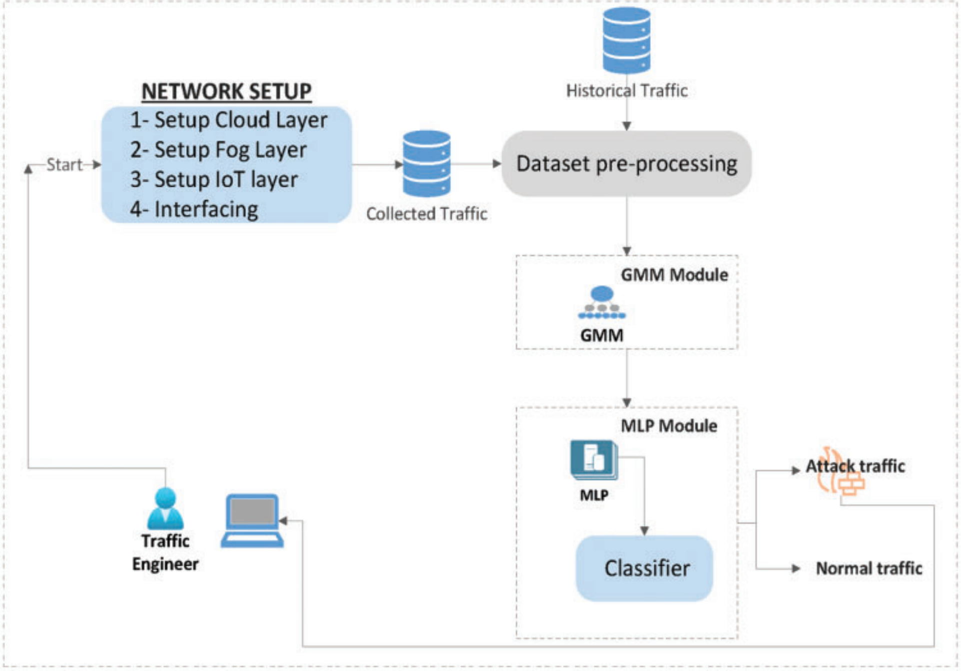


Figure 4: Functional diagram for the proposed deep learning-based IDS for DDoS attack in Agriculture 5.0

4 Experiments

We performed experiments on the Ubuntu 20.6 operating system using 12 GB of memory. To generate various MLP models for comparison purposes, we deployed two popular deep-learning open-source tools, namely Keras and TensorFlow. Also, Pearson's correlation coefficient was utilized to determine the correlation between dataset characteristics and their corresponding class labels.

To achieve high-performance results for detecting attacks on farm systems based on Agriculture 5.0, we selected the top 15 features for further processing using deep learning techniques. Given a large number of features, analyzing each one individually for feature selection is impractical. Therefore, Pearson's correlation coefficient approach is called to reduce the number of features to a more manageable subset. Indeed, by focusing on the features with the highest correlation with the class label, Pearson's correlation coefficient is still the most accurate and foolproof approach to retain the most important predictors of the target variable.

4.1 Performance Metrics

Selecting the appropriate performance criteria is vital for effectively evaluating machine learning and deep learning algorithms. In this section, we examine the overall performance of our IDS approach by testing its impact on prediction against attacks. The proposed GMLP-IDS model is evaluated and compared based on four performance indicators, including precision, F1 score, recall, and accuracy, as specified in Eqs. (1)–(4). Additionally, our study supports the use of confusion matrix analysis for classification accuracy when it makes predictions, as illustrated in Table 5.

$$Accuracy = \frac{TP\_Attack + TN\_BENGN}{TP\_Attack + FN\_Attack + TN\_BENGN + TN\_BENGN} \tag{1}$$

$$Precision = \frac{TP\_Attack}{TP\_Attack + FP\_Attack} \tag{2}$$

$$Recall = \frac{TP\_Attack}{TP\_Attack + FN\_Attack} \tag{3}$$

$$F1\,score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{4}$$

Table 5: Confusion matrix

		Actual	
		Attack	Normal
Predicted	**Attack**	TP (True positive)	FN (False positive)
	Normal	FN (False negative)	TN (True negative)

Where TP, TN, FP, and FN four outcomes are often used to create a confusion matrix. In the context of binary classification, these four outcomes are named true positive (TP), true negative (TN), false positive (FP), and false negative (FN). True positive (TP) occurs when the model correctly identifies a positive sample as positive (attack data that is accurately classified as an attack). The term True negative (TN) refers to a model that correctly identifies a negative sample as negative (benign data that is correctly identified as benign). False positive (FP) occurs when the model incorrectly identifies a negative sample as positive (benign data that was noted for any data that were wrongly identified as an attack). False negative (FN) occurs when the model incorrectly identifies a positive sample as negative (attack data that have been labeled as benign).

4.2 Results and Discussion

The performance results of deep learning techniques in terms of binary classification are shown in Table 6. Indeed, as mentioned earlier, our proposed intrusion detection approach and the traditional MLP are conducted on the CICDDoS2019 dataset to detect binary and multiclass attacks. We use Accuracy, Precision, Recall, and F1 score as performance indicators for each attack type. It can be observed clearly that GMLP-IDS, outperforms the traditional MLP in terms of accuracy and precision scores for most attack types. Table 6 depicts that the proposed classification model, GMLP-IDS, achieved a high accuracy rating of 99.96%, 99.61%, and 99.98% for DNS, NetBIOS, and SYN attacks, respectively. In comparison, the MLP model showed a correct accuracy rating of 95.94% for

DNS attacks, 99.48% for NetBIOS attacks, and reached 99.95% for SYN attacks. The above can be explained by the operation mode of the proposed model on making ethical decisions in diagnosing transformed feature vectors outputted from the Gaussian Mixture Model. Moreover, the GMLP-IDS model obtained a higher precision rating for DNS and SYN attacks, achieving ratings of 98.64% and 99.99%, respectively, compared to the MLP model's ratings of 86.80% and 77.77%. However, for MSSQL attacks, the MLP model achieved a higher precision score, with a score of 98.52% compared to our proposed approach model's rating of 81.66%.

Table 6: Performance results of GMLP-IDS and traditional MLP models to normal and various types of attacks with binary classification

Attack type	Accuracy		Precision		Recall		F1 score	
	GMLP-IDS	MLP	GMLP-IDS	MLP	GMLP-IDS	MLP	GMLP-IDS	MLP
BENIGN	0.9912	0.9774	0.93	0.85	0.99	0.98	0.9591	0.9104
DNS attack	0.9996	0.9594	0.9864	0.868	0.8724	0.5656	0.9259	0.6849
LDAP attack	0.9998	0.9896	0.9998	0.8933	0.9044	0.5342	0.9497	0.6686
MSSQL attack	0.9996	0.9916	0.8166	0.9852	0.718	0.9932	0.7642	0.9892
NTP attack	0.9999	0.9047	0.9999	0.9641	0.7949	0.5523	0.8857	0.7023
NetBIOS attack	0.9961	0.9948	0.8954	0.8884	0.9688	0.9967	0.9306	0.9394
SNMP attack	0.9998	0.9996	0.987	0.9888	0.7607	0.545	0.8592	0.7027
SSDP attack	0.9998	0.9558	0.9921	0.9932	0.6958	0.6667	0.8179	0.7978
UDP attack	0.9998	0.9368	0.8836	0.9065	0.9861	0.9924	0.932	0.9475
SYN attack	0.9998	0.9995	0.9999	0.7777	0.9526	0.9211	0.9757	0.8433
UDP-lag attack	0.9982	0.997	0.9699	0.9413	0.9792	0.8895	0.9745	0.9147
TFTP attack	0.9997	0.9997	0.9887	0.9885	0.9855	0.9889	0.9871	0.9887

Overall, we can affirm that the proposed GMLP-IDS model is more effective in detecting attacks, namely, LDAP, NetBIOS, SYN, etc. These findings demonstrate the importance of choosing an appropriate model and evaluation metrics for detecting different types of attacks. As mentioned earlier, Pearson's correlation coefficient approach deployed with GMM and MLP has retained the most important features of the target variable. Such a solution gives a more accurate and foolproof manageable subset of the trained process.

On the other hand, Table 7 shows accuracy, precision, recall, and F1 score for each of the two compared machine learning models in the case of multi-class classification. The results show that GMLP-IDS gives a higher rating prediction in most attack types. Specifically, the proposed IDS has demonstrated optimistic results in detecting SYN-based attacks with a recall rating of 92.50% and NetBIOS-based attacks with a recall of 96.88%. Compared to the MLP model, which provides detection of SYN-based attacks with recall scores of 90.13% and NetBIOS-based attacks with a recall rating of 94.87%, we can identify that the GMLP-IDS model achieved near-perfect accuracy on multi-class classification and, therefore, improving the accuracy and effectiveness of intrusion detection systems in real-world network environments.

Table 7: Performance results of GMLP-IDS and traditional MLP models to normal and various types of attacks with multi-class classification

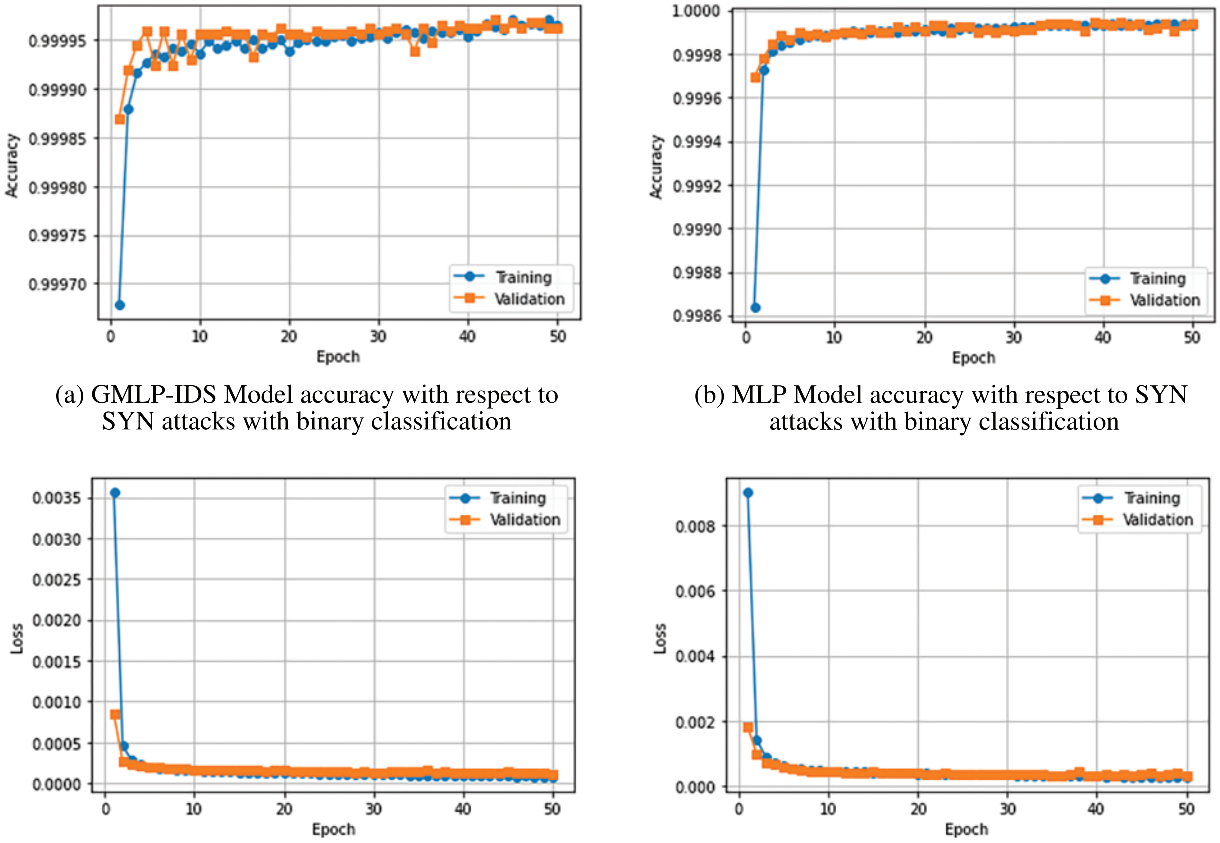
Attack type	Accuracy		Precision		Recall		F1 score	
	GMLP-IDS	MLP	GMLP-IDS	MLP	GMLP-IDS	MLP	GMLP-IDS	MLP
BENIGN	0.9912	0.9334	0.9	0.803	0.96	0.8955	0.929	0.8467
DNS attack	0.9976	0.9594	0.9664	0.8038	0.8234	0.5037	0.8892	0.6193
LDAP attack	0.9908	0.9596	0.9479	0.8373	0.8844	0.5024	0.915	0.6280
MSSQL attack	0.9396	0.9016	0.7566	0.9422	0.708	0.9632	0.7315	0.9526
NTP attack	0.9587	0.8747	0.9129	0.9551	0.7449	0.5426	0.8204	0.6921
NetBIOS attack	0.9961	0.9033	0.8954	0.8364	0.9688	0.9487	0.9306	0.8890
SNMP attack	0.9778	0.9116	0.927	0.9478	0.7507	0.5454	0.8296	0.6924
SSDP attack	0.9038	0.8996	0.9081	0.9882	0.6901	0.6257	0.7842	0.7662
UDP attack	0.9998	0.9368	0.8836	0.8865	0.9464	0.9274	0.9139	0.9065
SYN attack	0.9995	0.9993	0.9934	0.7588	0.925	0.9013	0.958	0.8239
UDP-lag attack	0.9982	0.953	0.9615	0.9111	0.9092	0.8525	0.9346	0.8809
TFTP attack	0.9000	0.8835	0.8787	0.9087	0.9305	0.9365	0.9039	0.9224

Fig. 5 presents the training accuracy and loss convergence concerning epochs for the SYN-based attack. Indeed, a SYN-based attack is a type of network level where the attacker exploits the fact that the TCP protocol uses a three-way handshake to establish a connection between two systems. In such cases, the attacker sends a SYN packet to initiate the handshake with drones but does not complete it by sending the necessary ACK packet. This causes the drone system to keep waiting for the ACK packet, tying up its resources and preventing it from responding to legitimate requests. Moreover, the main disadvantage of IoT and drones is their high energy consumption, which can lead to the rapid depletion of batteries or other power sources.

Drones, in particular, require a significant amount of energy to maintain flight and perform various tasks, such as data communication with soil sensors that continuously measure soil parameters such as moisture, pH levels, and nutrient content. However, if a drone runs out of battery mid-flight, it could crash and cause damage to the crops or other equipment. Furthermore, with such SYN-based attacks, IoT devices such as sensors and smart devices that operate on batteries can also quickly deplete their energy sources, especially when deployed in remote or harsh environments. Additionally, frequent battery replacements can be costly and time-consuming, reducing the efficiency and productivity of IoT and drone applications.

From the profiling perspective of binary classification in Fig. 5, we can see that the training accuracy of GMLP-IDS converges much faster than the traditional MLP model. For GMLP-IDS, the training accuracy curve begins at 99.98% and goes up to 99.99%, and after that, the convergence of training accuracy becomes stable at the 20th epoch. For training loss, Figs. 5c and 5d depict the convergence of the loss with epochs. From Fig. 5c, we can see that the GMLP-IDS achieves the lowest loss of 0.02% at the 25th epoch.

On the other hand, concerning multiclass classification, Fig. 6 illustrates that GMLP-IDS converges must be faster than MLP in the case of training accuracy. Indeed, the training begins at 99.89% and goes up to 99.94% at the 10th epoch. For training loss, Figs. 6c and 6d depict the convergence of the loss with epochs. For multiclass classification, we can see the GMLP-IDS achieves the lowest loss of 0.2% at the 15th epoch, and MLP achieves the lowest loss of 0.25% at the fifth epoch.



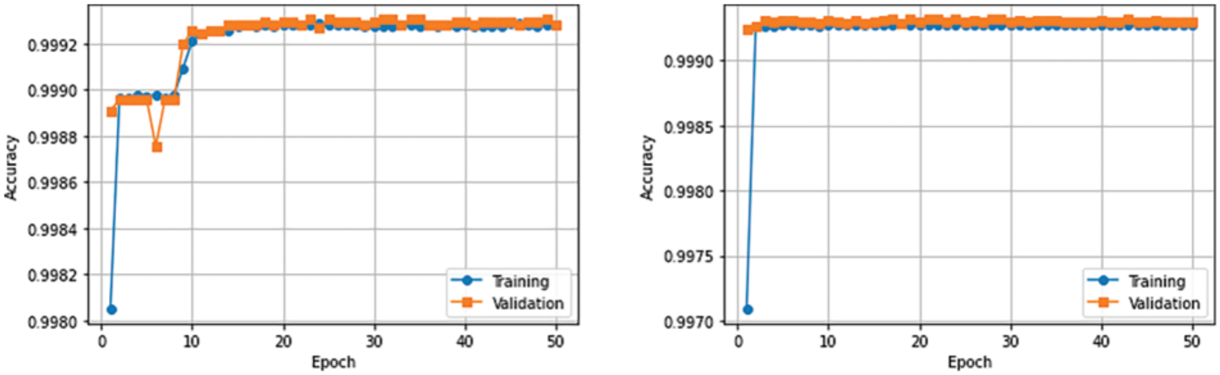
(a) GMLP-IDS Model accuracy with respect to SYN attacks with binary classification

(b) MLP Model accuracy with respect to SYN attacks with binary classification

(c) GMLP-IDS Model loss with respect to SYN attacks with binary classification

(d) MLP Model loss with respect to SYN attacks with binary classification

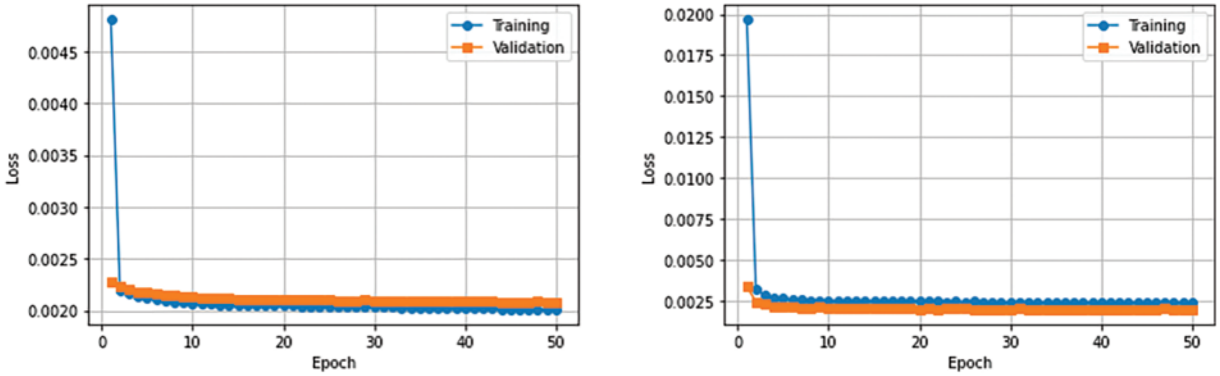
Figure 5: Model accuracy and loss of MLP and GMLP-IDS models with binary classification



(a) GMLP-IDS Model accuracy with respect to SYN attacks with multiclass classification

(b) MLP Model accuracy with respect to SYN attacks with multiclass classification

Figure 6: (Continued)



(c) GMLP-IDS Model loss with respect to SYN attacks with multiclass classification

(d) MLP Model loss with respect to SYN attacks with multiclass classification

Figure 6: Model accuracy and loss of MLP and GMLP-IDS models with multiclass classification

5 Conclusions and Future Guidelines

The present study proposes a novel Deep-intrusion detection system named GMLP-IDS for Smart Agriculture, which leverages a combination of the Gaussian Mixture Model and Multi-layer Perceptron. The CIC-DDoS2019 dataset is employed to train and validate the proposed solution, which is geared towards Agricultural 5.0. This approach encompasses the integration of IoT, drones, and sensors to optimize farming practices. By addressing the drawbacks of traditional agriculture, GMLP-IDS offers an innovative solution to enhance the security of smart agriculture systems. The study compares GMLP-IDS with a conventional MLP approach based on four key performance indicators: precision, F1 score, recall, and accuracy. Furthermore, Pearson's correlation coefficient approach is utilized to reduce the number of features to a more manageable subset. The findings highlight the superior performance of GMLP-IDS in intrusion detection accuracy, thus underscoring its potential to revolutionize the field of smart agriculture security.

In future work, we plan to deploy drones equipped with cameras to capture images of olive leaves to identify different types of infections, including Aculus olearius, olive peacock spot, olive scab, and healthy leaves. Indeed, as mentioned earlier, such open olive farming at Al Jouf, with an area of 7730 hectares and over 5 million olive trees, it is essential a real-time control is achieved in such open-field farming. Additionally, given the massive amount of collected leaves images, data should be protected from unauthorized access or interception. On the other hand, by outperforming nature in detecting DDoS attacks, GMLP-IDS is considered in the proposed framework for being implemented at the fog layer. However, time and memory consumption are crucial considerations due to the limited resources of edge and fog devices. To deal with such issues, balancing the time and memory consumption in DDoS prevention at the fog layer involves optimizing the analysis and response mechanisms, employing memory-efficient strategies, ensuring scalability, and utilizing collaborative defense techniques. Moreover, in our work, we have deployed only one standard dataset, namely CIC-DDoS2019 dataset. The dataset contains a variety of DDoS attacks. However, some network traffic patterns may not be applied in a real agriculture environment, or the dataset might not fully represent the confusions and complexities of real-world network traffic, leading to the proposed IDS being less effective in detecting different types of attacks. Therefore, it is crucial to complement the use of the CIC-DDoS2019 dataset with other relevant datasets and regularly monitor the IDS to decrease false alarms and missed detection rates.

Acknowledgement: The authors extend their appreciation to the Deanship of Scientific Research in Cooperation with Olive Research Center at Jouf University.

Funding Statement: This work was funded by the Deanship of Scientific Research in Cooperation with Olive Research Center at Jouf University under Grant Number (DSR2022-RG-0163).

Author Contributions: The authors confirm contribution to the paper as follows: study conception and design: Abdelwahed Berguiga, Ayman Massaoudi, Ahlem Harchay; data collection: Abdelwahed Berguiga, Mossaad Ben Ayed; analysis and interpretation of results: Abdelwahed Berguiga, Ahlem Harchay, Ayman Massaoudi, Mossaad Ben Ayed, Hafedh Belmabrouk; draft manuscript preparation: Abdelwahed Berguiga, Ahlem Harchay, Ayman Massaoudi, Mossaad Ben Ayed, Hafedh Belmabrouk. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The data used in this paper can be requested from the corresponding author upon request.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] O. Friha, M. A. Ferrag, L. Shu, L. Maglaras and X. Wang, “Internet of Things for the future of smart agriculture: A comprehensive survey of emerging technologies,” *IEEE/CAA Journal of Automatica Sinica*, vol. 8, no. 4, pp. 718–752, 2021.

[2] B. Chen, J. Wan, L. Shu, P. Li, M. Mukherjee *et al.,* “Smart factory of Industry 4.0: Key technologies, application case, and challenges,” *IEEE Access*, vol. 6, pp. 6505–6519, 2018.

[3] Y. Liu, X. Ma, L. Shu, G. P. Hancke and A. M. Abu-Mahfouz, “From Industry 4.0 to Agriculture 4.0: Current status, enabling technologies, and research challenges,” *IEEE Transactions on Industrial Informatics*, vol. 17, no. 6, pp. 4322–4334, 2021. https://doi.org/10.1109/TII.2020.3003910

[4] S. H. Alsamhi, A. V. Shvetsov, S. Kumar, J. Hassan, M. A. Alhartomi *et al.,* “Computing in the sky: A survey on intelligent ubiquitous computing for UAV-assisted 6G networks and Industry 4.0/5.0,” *Drones*, vol. 6, no. 7, pp. 177–206, 2022.

[5] X. Yang, L. Shu, J. Chen, M. A. Ferrag, J. Wu *et al.,* “A survey on smart agriculture: Development modes, technologies, and security and privacy challenges,” *IEEE/CAA Journal of Automatica Sinica*, vol. 8, no. 2, pp. 273–302, 2021.

[6] A. Massaoudi, A. Berguiga and A. Harchay, “Secure irrigation system for olive orchards using Internet of Things,” *Computers, Materials & Continua*, vol. 72, no. 3, pp. 4663–4673, 2022.

[7] A. Scuderi, G. L. Via, G. Timpanaro and L. Sturiale, “The digital applications of Agriculture 4.0: Strategic opportunity for the development of the italian citrus chain,” *Agriculture*, vol. 12, no. 3, pp. 400–413, 2022.

[8] R. Abbasi, P. Martinez and R. Ahmad, “The digitization of agricultural industry–a systematic literature review on Agriculture 4.0,” *Smart Agricultural Technology*, vol. 2, no. 2, pp. 1–24, 2022.

[9] T. H. H. Aldhyani and H. Alkahtani, “Cyber security for detecting distributed denial of service attacks in Agriculture 4.0: Deep learning model,” *Mathematics*, vol. 11, no. 1, pp. 233–252, 2023.

[10] S. Padhy, M. Alowaidi, S. Dash, M. Alshehri, P. P. Malla *et al.,* “Agrisecure: A fog computing-based security framework for Agriculture 4.0 via blockchain,” *Processes*, vol. 11, no. 1, pp. 757–784, 2023.

[11] D. Akgun, S. Hizal and U. Cavusoglu, “A new DDoS attacks intrusion detection model based on deep learning for cybersecurity,” *Computers & Security*, vol. 118, no. 1, pp. 102748, 2022.

[12] M. G. Kavitha, “Deep learning enabled privacy preserving techniques for intrusion detection systems in the industrial Internet of Things,” *Adhoc & Sensor Wireless Networks*, vol. 52, no. 3, pp. 223–247, 2022.

[13] A. Berguiga and A. Harchay, “An IoT-based intrusion detection system approach for TCP SYN attacks,” *Computers, Materials & Continua*, vol. 71, no. 2, pp. 3839–3851, 2022.

[14] M. Javaid, A. Haleem, R. P. Singh and R. Suman, “Enhancing smart farming through the applications of Agriculture 4.0 technologies,” *International Journal of Intelligent Networks*, vol. 3, no. 1, pp. 150–164, 2022.

[15] G. Abbas, A. Mehmood, M. Carsten, G. Epiphaniou and J. Lloret, “Safety, security and privacy in machine learning based Internet of Things,” *Journal of Sensor and Actuator Networks*, vol. 11, no. 3, pp. 38–53, 2022.

[16] A. A. Diro and N. Chilamkurti, “Distributed attack detection scheme using deep learning approach for Internet of Things,” *Future Generation Computer Systems*, vol. 82, no. 6, pp. 761–768, 2018.

[17] G. Thamilarasu and S. Chawla, “Towards deep-learning-driven intrusion detection for the Internet of Things,” *Sensors*, vol. 19, no. 9, pp. 1977, 2019.

[18] E. M. Campos, P. F. Saura, A. González-Vidal, J. L. Hernández-Ramos, J. B. Bernabé *et al.,* “Evaluating federated learning for intrusion detection in Internet of Things: Review and challenges,” *Computer Networks*, vol. 203, no. 3, pp. 108661, 2022.

[19] I. Ullah and Q. H. Mahmoud, “Design and development of RNN anomaly detection model for IoT networks,” *IEEE Access*, vol. 10, pp. 62722–62750, 2022.

[20] Y. Hao, Y. Sheng and J. Wang, “Variant gated recurrent units with encoders to preprocess packets for payload-aware intrusion detection,” *IEEE Access*, vol. 7, pp. 49985–49998, 2019.

[21] O. Friha, M. A. Ferrag, L. Shu, L. Maglaras, K. R. Choo *et al.,* “FELIDS: Federated learning-based intrusion detection system for agricultural Internet of Things,” *Journal of Parallel and Distributed Computing*, vol. 165, no. 15, pp. 17–31, 2022.

[22] L. Gupta, T. Salman, A. Ghubaish, D. Unal, A. K. Al-Ali *et al.,* “Cybersecurity of multi-cloud healthcare systems: A hierarchical deep learning approach,” *Applied Soft Computing*, vol. 118, no. 3, pp. 108439, 2022.

[23] S. P. K. Gudla, S. K. Bhoi, S. R. Nayak and A. Verma, “DI-ADS: A deep intelligent distributed denial of service attack detection scheme for fog-based IoT applications,” *Mathematical Problems in Engineering*, vol. 2022, pp. 3747302, 2022.

[24] S. S. Hameed, A. Selamat, L. AbdulLatiff, S. A. Razak, O. Krejcar *et al.,* “A hybrid lightweight system for early attack detection in the IoMT fog,” *Sensors*, vol. 21, no. 24, pp. 8289, 2021.

[25] J. Jiang, F. Liu, Y. Liu, Q. Tang, B. Wang *et al.,* “A dynamic ensemble algorithm for anomaly detection in IoT-imbalanced data streams,” *Computer Communications*, vol. 194, no. 3, pp. 250–257, 2022.

[26] P. Wang, Z. Li, X. Zhou, C. Su and W. Wang, “FlowADGAN: Adversarial learning for deep anomaly network intrusion detection,” in *Security and Trust Management: 18th Int. Workshop*, Copenhagen, Denmark, pp. 156–174, 2023.

[27] Z. Wang, W. Wang, Y. Yang, Z. Han, D. Xu *et al.,* “CNN-and GAN-based classification of malicious code families: A code visualization approach,” *International Journal of Intelligent Systems*, vol. 37, no. 12, pp. 12472– 12489, 2022.

[28] I. Sharafaldin, A. H. Lashkari, S. Hakak and A. A. Ghorbani, “Developing realistic distributed denial of service (DDOS) attack dataset and taxonomy,” in *Proc. of ICCST*, Chennai, India, pp. 1–8, 2019.

[29] A. Maheshwari, B. Mehraj, M. S. Khan and M. S. Idrisi, “An optimized weighted voting-based ensemble model for DDoS attack detection and mitigation in SDN environment,” *Microprocessors and Microsystems*, vol. 89, no. 1, pp. 104412, 2022. https://doi.org/10.1016/j.micpro.2021.104412

[30] K. B. Dasari and N. Devarakonda, “Evaluation of SVM kernels with multiple uncorrelated feature subsets selected by multiple correlation methods for reflection amplification ddos attacks detection,” in *Proc. of ACSS*, Singapore, pp. 99–111, 2023.

[31] S. B. Kotsiantis, D. Kanellopoulos and P. E. Pintelas, “Data preprocessing for supervised leaning,” *International Journal of Computer Science*, vol. 1, pp. 111–117, 2006.

[32] J. Li, Y. Si, T. Xu and S. Jiang, “Deep convolutional neural network-based ECG classification system using information fusion and one-hot encoding techniques,” *Mathematical Problems in Engineering*, vol. 2018, pp. 1–10, 2018.

[33] M. K. Dahouda and I. Joe, “A deep-learned embedding technique for categorical features encoding,” *IEEE Access*, vol. 9, pp. 114381–114391, 2021.

[34] R. Karthiga, G. Usha, N. Raju and K. Narasimhan, "Transfer learning-based breast cancer classification using one-hot encoding technique," in *Proc. of ICAIS*, Coimbatore, India, pp. 115–120, 2021.

[35] J. Benesty, J. Chen, Y. Huang and I. Cohen, "Pearson correlation coefficient," in *Proc. of Noise Reduction in Speech Processing*, Heidelberg, Germany, pp. 1–4, 2009.