Additional Information: This is an open access article which first appeared in Internet of Things

Data Access Statement: ToN-IoT dataset: https://research.unsw.edu.au/projects/toniot-datasets and BoT-IoT dataset is from here https://research.unsw.edu.au/projects/bot-iot-dataset

Metaverse-IDS: Deep learning-based intrusion detection system for Metaverse-IoT networks

Tarek Gaber [a,b,*], Joseph Bamidele Awotunde [c], Mohamed Torky [d,e],
Sunday A. Ajagbe [f,g], Mohammad Hammoudeh [h], Wei Li [i]

[a] School of Science, Engineering, and Environment, University of Salford, Manchester M5 4WT, UK
[b] Faculty of Computers and Informatics, Suez Canal University, Ismailia 41522, Egypt
[c] Department of Computer Science, Faculty of Information and Communication Sciences, University of Ilorin, Ilorin 240003, Nigeria
[d] Faculty of Artificial Intelligence, Egyptian Russian University, Cairo, Badr City, 11829, Egypt
[e] Scientific Research Group in Egypt (SRGE), Egypt
[f] Department of Computer Science, University of Zululand, Kwadlangezwa, 3886, South Africa
[g] Department of Computer & Industrial Production Engineering, First Technical University Ibadan, 200255, Nigeria
[h] Information and Computer Science Department, King Fahd University of Petroleum & Minerals, Saudi Arabia
[i] College of Computer Science and Technology, Harbin Engineering University, Harbin 150001, China

A R T I C L E   I N F O

A B S T R A C T

Combining the metaverse and the Internet of Things (IoT) will lead to the development of diverse, virtual, and more advanced networks in the future. The integration of IoT networks with the metaverse will enable more meaningful connections between the 'real' and 'virtual' worlds, allowing for real-time data analysis, access, and processing. However, these metaverse-IoT networks will face numerous security and privacy threats. Intrusion Detection Systems (IDS) offer an effective means of early detection for such attacks. Nevertheless, the metaverse generates substantial volumes of data due to its interactive nature and the multitude of user interactions within virtual environments, posing a computational challenge for building an intrusion detection system. To address this challenge, this paper introduces an innovative intrusion detection system model based on deep learning. This model aims to detect most attacks targeting metaverse-IoT communications and combines two techniques: KPCA (Kernel Principal Component Analysis which was used for attack feature extraction and CNN (Convolutional Neural Networks for attack recognition and classification. The efficiency of this proposed IDS model is assessed using two widely recognized benchmark datasets, BoT-IoT and ToN-IoT, which contain various IoT attacks potentially targeting IoT communications. Experimental results confirmed the effectiveness of the proposed IDS model in identifying 12 classes of attacks relevant to metaverse-IoT, achieving a remarkable accuracy of 99.8% and a False Negative Rate FNR less than 0.2. Furthermore, when compared with other models in the literature, our IDS model demonstrates superior performance in attack detection accuracy.

* Corresponding author.
  E-mail address: t.m.a.gaber@salford.ac.uk (T. Gaber).

Introduction

The metaverse represents a groundbreaking online environment for social interaction and communication, driven by cutting-edge technologies. It encompasses a fusion of various technologies, including network communication, the Internet of Things (IoT), Artificial Intelligence (AI), extended reality, Virtual Reality (VR), and blockchain [1] and [2]. Within the metaverse, users adopt digital personas to engage in social activities, leisure, and work. Moreover, it hosts a dynamic marketplace for trading both physical and virtual goods. IoT plays a crucial role in synchronizing the metaverse with the physical world. Mobile devices, VR headsets, and computers seamlessly bridge the gap between the physical realm and the virtual universe, forming the metaverse's digital landscape. The creation platform leverages IoT data for model development, visual representation, and material rendering, while the application ecosystem generates virtual scenarios based on these operations.

Incorporating IoT networks to connect real-world devices will enhance the communication and realism of objects, e.g., avatars, virtual buildings, landscapes, and vehicles, within the metaverse. This integration addresses the complex challenge of blending hybrid data with cloud or digital infrastructures, laying the foundation for advanced systems such as digital twins and virtual simulations. Consequently, IoT facilitates seamless interaction between virtual environments and the physical world, while the metaverse provides the essential 3D user interface for a cluster of IoT devices. This cooperative relationship between IoT and the metaverse holds the potential to revolutionize long-term planning across various sectors including healthcare, energy, and transportation. It empowers us to identify optimized designs and dynamically manage these systems in response to real-world changes. Additionally, the metaverse will incorporate scanned objects capable of engaging with the real world [3]. This amalgamation of IoT and the metaverse heralds a new era of interconnectedness and transformative potential for a wide range of industries.

Consider donning a VR headset, allowing for a complete immersion into a digital environment within the metaverse. Here, IoT-enabled devices seamlessly interact with one's virtual persona. As an illustration, a smartwatch has the potential to offer haptic feedback within the virtual environment, while intelligent appliances can be remotely controlled within the metaverse. Another example of metaverse-IoT networks could is IoT-enabled inventory management systems of virtual retail and commerce where such systems would have the capability to be connected to virtual stores within the metaverse. The users would also have the ability to explore, acquire, and even obtain physical products through their virtual engagements.

In a typical metaverse-IoT network application scenario, IoT technology will manage communications between metaverse objects across various virtual spaces by sensing the real data in the metaverse environment. This would create "metaverse-IoT networks". Such networks would be vulnerable to many security and privacy issues [4] and [5]. For instance, due to the immersive social cyberspace features of metaverse, it is more likely that the exchange of personal information (as in shown in the examples above) would increase which would also lead to increasing privacy and security attacks. Moreover, the advent of the metaverse may give rise to unprecedented challenges and complications. For example, malicious individuals may exploit vulnerabilities within a brain-computer interface that is interconnected with the metaverse, thereby gaining unauthorized access to a person's physical body. Many cybersecurity attacks such as NFTs (Non-Fineable Tokens), fake avatars, dark verse attacks, Cyber-physical attacks, virtual/augmented reality attacks, distributed denial of service, ransomware attacks, scanning, and Cross-Site Scripting (XSS) [4,6]. Hence, it is essential to incorporate extra security measures to protect these devices and the transmitted data within the metaverse-IoT networks. Deep Learning (DL)-driven security solutions, such as anomaly and intrusion detection systems, as well as network traffic monitoring, have the potential to significantly enhance the security of IoT devices.

DL is a machine learning approach that employs artificial neural networks comprising multiple layers to acquire hierarchical representations of data. DL can achieve feature hierarchies by utilizing extensive quantities of unlabelled data, which renders it highly advantageous for the analysis of complex and high-dimensional datasets. DL was used for many intrusion detection systems [7–10], which gave a high performance. Nevertheless, a recent study [11] revealed that the improvement in performance, specifically accuracy, is accompanied by a substantial computational burden that is necessary for the training and administration of deep-learning algorithms. In the metaverse-IoT networks, the rapid growth and utilization of the metaverse will result in data expansion, leading to the establishment of a comprehensive network of large-scale data. Consequently, this will impose significant computational costs on the virtual world [12]. One way of addressing this problem is the application of feature selection techniques [13]. The elimination of irrelevant features will result in a reduction in storage and computational expenses while also preventing substantial information loss or deterioration in learning performance [14–16].

In response to the above problems, this paper proposes a novel deep learning–based (CNN) intrusion detection system to recognize IoT attacks in a metaverse-IoT environment with the support of Kernel Principal Component Analysis (KPCA) for extracting attack features. Using the KPCA would help reduce the computational cost required by the DL algorithm, CNN, in the classification phase.

The main contribution of this paper is as follows:

- Proposing a novel hybrid machine and deep learning-based intrusion detection model that can identify twelve classes of IoT attacks in metaverse IoT networks. The power of the proposed IDS model is using KPCA for selecting discriminative attack features (minimizing processing time) and Convolution Neural Networks (CNN) for performing the attack recognition and classification. The use of KPCA was shown to not only minimize the processing time but also improve the attack detection rate.
- Evaluating the efficacy of the developed deep learning-based intrusion detection model for metaverse-IoT networks. A comprehensive assessment was conducted utilizing eight performance metrics, namely Accuracy, Precision, Sensitivity, F1-score, True Positive (TP) Rate, False Negative Rate (FNR), and ROC curve. This evaluation was performed on two publicly available datasets, the ToN-IoT [17] and BoT-IoT [18]. The proposed model exhibited superior performance compared to the most related work in both datasets.

The rest of this paper is organized as follows. Section 2 presents a discussion and analysis of the related works. Section 3 gives the proposed deep learning IDS model for metaverse IoT networks. Section 3 presents the experimental results and discussion while Section 4 offers a comparison with related. Finally, Section 5 concludes the article and outlines future work avenues.

Literature review

Different security techniques exist in the literature regarding recognizing IoT attack patterns in classical IoT networks [19,20]. However, a few of these studies investigated the detection of IoT attacks in metaverse IoT environments [21].

Convolutional neural network-based IDS models

Yan D et al. [22] introduced an edge computing and data augmentation-based IDS model for detecting traffic anomalies of IoT networks. The detection methodology involves converting the network traffic to images which are then used to train a CNN to classify the network traffic patterns. The experimental results clarified the efficiency of the proposed IDS model in recognizing traffic anomalies with a precision rate reached 96%. Another deep learning-based CNN approach is proposed [7] to determine any possible abnormal traffic behaviour and intrusion. The detection performance of the developed IDS model was tested using two benchmark datasets. The experimental results clarified that the proposed IDS achieved an accuracy of 99.51% and 92.85% of the two used datasets, the NID Dataset and the BoT-IoT dataset, respectively. The focal loss function technique can also be used to develop efficient deep learning-based IDS models to detect intrusion and abnormal traffic in IoT networks [8].

Madhu B et al. [9] proposed a deep learning model called Device-based Intrusion Detection System (DIDS) to predict some patterns of unknown attacks in IoT networks. The experimental results clarified high sensitivity in detecting IoT attacks earlier than other standard algorithms, with detection accuracy reaching 99%. Mahadik S et al. [10] introduced an efficient intrusion detection model called HetIoT-CNN IDS, which utilized CNN to recognize various patterns of DDoS attacks in heterogeneous IoT networks. The experimental results proved that the HetIoT-CNN IDS successfully could identify various DDoS attacks with a detection accuracy of 99.75% for binary classes, 99.95% for 8 classes, and 99.99% for 13 classes of DDoS attacks in the heterogeneous IoT (HetIoT) networks.

Combined generative adversarial network-based IDS models

He Y et al. [23] proposed an access control mechanism called a Wasserstein Distance-based Combined Generative Adversarial Network (WCGAN) for ensuring the communication security of IoT-based intelligent transportation systems in 5 G networks. The validation and verification results clarified that the proposed IoT-based intelligent transport system achieved an attack prediction accuracy of 86.3% and ensured data transmission confidentiality and integrity in Internet of Vehicle (IoV) networks. Another GAN-based IDS system is proposed by Lee J et al. [24] to recognize abnormal traffic data using a deep learning-based autoencoder (AE) model. The performance analysis of the proposed AE-CGAN (auto-encoder-conditional GAN) model using a benchmark dataset (CICIDS-2017) confirmed the efficiency of AE-CGAN model (recall $= 93.29\%$ and F1-Score $= 95.38\%$) compared to other IDS models in the literature.

On the contrary, Zaho S and colleagues [25] introduced an attack model known as AttackGAN, based on Generative Adversarial Networks, which demonstrated remarkable effectiveness in launching attacks against black-box intrusion detection systems, particularly in the context of identifying abnormal traffic data within IoT networks. The analysis of the results made it evident that the proposed AttackGAN outperformed other attack models, including the Fast Gradient Sign Method (FGSM) [26], Project Gradient Descent (PGD) [27], Carlini & Wagner attack (CW) [28], and various GAN-based algorithms [29], when it came to enhancing the success rate of adversarial attacks against black-box IDS models. This highlights the potency of AttackGAN in its ability to exploit vulnerabilities in intrusion detection systems, particularly within IoT environments, where the security landscape is increasingly complex.

Other IDS models

In their work, Yin C et al. [30] introduced an RNN-IDS model based on deep learning, specifically utilizing Recurrent Neural Networks (RNN). The primary objective of this model was to identify four distinct types of cyberattacks: R2L, U2), DoS, and Probing attacks (Probe). The outcomes of their experiments unmistakably demonstrated the superiority of the RNN-IDS over a range of other classification algorithms, including J48, naive Bayes, random forest, multi-layer perceptron, and various others. The numerical results, as presented in their study, shed light on the robust performance of the RNN-IDS model. It achieved an impressive accuracy rate of 81.29% when tested on the KDDTest+ dataset and a notable 64.67% accuracy on the KDDTest−21 dataset. These results notably outperformed the performance of the aforementioned algorithms, signifying the effectiveness of the RNN-IDS in accurately detecting and classifying different types of cyberattacks. This highlights the potential of deep learning, particularly RNNs, in enhancing the capabilities of intrusion detection systems, contributing to more robust cybersecurity practices.

Verma A and Ranga V [31] conducted practical comparison results to assess the performance of five machine learning-based classifiers for developing an optimal intrusion detection system for IoT attacks. The comparison involved AdaBoost (AB), Random Forest (RF), Extreme Gradient Boosting (XGB), Extremely Randomized Trees (ETC), Regression Trees (RT), Multi-Layer Perceptron (MLP), and Gradient Boosted Machine (GBM). The experimental results and statistical analysis showed that classification, regression trees, and XGB classifiers show the best trade-off between prominent metrics and response time. Therefore, both are optimal for
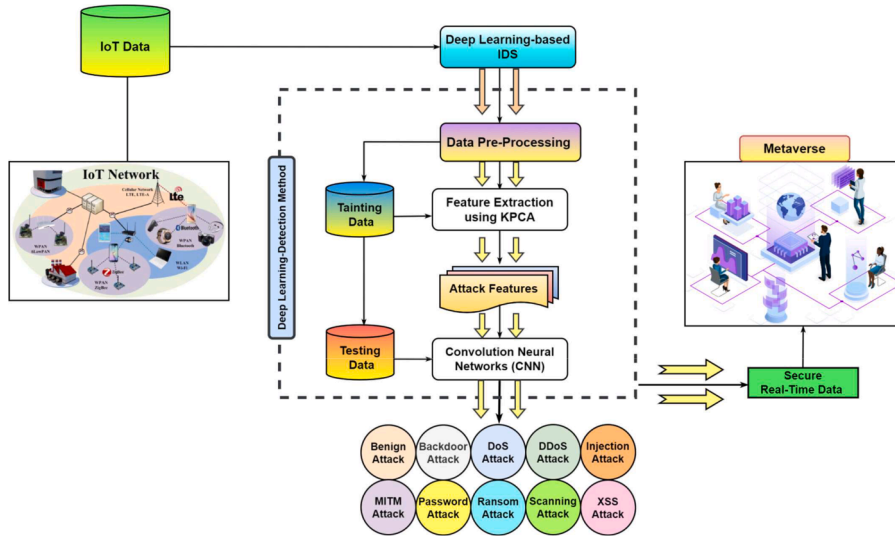
Fig. 1. The proposed CNN-based Metaverse-IDS model.

building efficient IoT–based IDS Models. To enhance the detection methodology of novel IDS models, using semi-supervised learning-based Collaborative Intrusion Detection Systems (CIDSs) may be a promising solution [32–34] for solving the challenge of data labelling compared with traditional supervised classifiers.

From the discussion of the above literature, it was identified that most of the research concentrated on exploring potential proposals related to IoT standards, technologies, architectural designs, IoT security concerns, diverse machine learning methodologies, available datasets, and the tools required for implementation. Nevertheless, to the best of the authors knowledge, no studies yet investigated into the realm of intrusion detection within the future landscape of metaverse-integrated IoT networks, all while taking into account the anticipated computational requirements. This highlights an evident research gap, offering an intriguing avenue for future investigations to explore. Such endeavours could lead to innovative solutions that enhance the security of IoT systems integrated with the metaverse, all the while efficiently managing computational resources. At the same time, identify and detect cyber-attacks from the extensive collected data from the rapid growth and utilization of the metaverse. As described in the section below, a DL-based IDS for metaverse-IoT networks is proposed to address these limitations.

The proposed IDS for metaverse-IoT network

In order to identify and mitigate common IoT attacks within metaverse environments, we present a novel deep learning-driven IDS. The overarching architectural framework of this IDS is illustrated in Fig. 1, which serves as a visual representation of its core components. The detection process employed by this system unfolds in a structured manner, consisting of three fundamental stages. First, the initial stage encompasses a preprocessing step aimed at converting the features within the provided dataset from nominal to numeric representations. This step lays the foundation for subsequent analysis. In the second stage, we leverage the power of Kernel Principal Component Analysis (KPCA) to extract salient attack features. KPCA is helpful in capturing the key characteristics that differentiate malicious activities from benign ones, contributing to the system's efficacy in detecting and addressing threats effectively. Finally, the third and pivotal stage involves the application of a CNN model for the purpose of classifying detected attacks. The CNN model serves as the brain of our IDS, utilizing the rich feature set obtained from the preceding stages to make informed decisions about the nature and severity of any potential security breaches. Detailed explanations of each of these three stages are given in the following subsections.

Pre-processing

The initial phase of pre-processing involves taking the dataset and performing two operations for the purpose of data cleansing, conversion, and normalization. The conversion step transforms categorical attributes into numerical ones. The normalization process aims to bring the attribute values within a more manageable range, reducing the significant differences between them. Specifically, we utilized the minimum-maximum scaling technique, as described in Eq. (1).

$$Y_{norm} = \frac{Y - min(Y)}{max(Y) - min(Y)}$$
(1)

Here, Y represents the value of the feature within the dataset, and it falls within the interval [0, 1].

Feature extraction

After the initial preprocessing phase, the dataset undergoes feature extraction, with KPCA emerging as the favoured technique for reducing data dimensions [49]. However, KPCA does not consider nonlinear data characteristics when dealing with complex structures. By employing KPCA, this limitation can be addressed. Eq. (2), below, represents the mapping of the feature space R within the function P.

$$P \ : \ \emptyset \ \epsilon \ R^m \rightarrow P(\emptyset) \ \epsilon \ R \tag{2}$$

where $\sum_{i=1}^{t} P(\emptyset_i) = 0$, The covariance matrix can be calculated as in Eqs. (3)-(5)

$$Co_{mtx} = \frac{1}{t} \sum_{i=1}^{t} (P(\emptyset_i) - mean)(P(\emptyset_i) - mean)^T \tag{3}$$

$$M_{ean} = \frac{1}{t} \sum_{i=1}^{t} P(\emptyset_i) \tag{4}$$

$$Co_{mtx} = \ \frac{1}{t} \sum_{i=1}^{t} P(\emptyset_i)P(\emptyset_i)^T \tag{5}$$

The terms Eigenvalue and Eigenvector, often used interchangeably, can be computed following the equations provided in 6, 7, and 8:

$$Co_{mtx}I = \ \lambda_i I \tag{6}$$

Combining (5) and (6), we get

$$\frac{1}{t} \sum_{i=1}^{t} P(\emptyset_i)IP(\emptyset_i)^T = \ \lambda_i I \tag{7}$$

The expression for the eigenvector can be reformulated using the equation provided in Eq. (8)

$$I = \ \frac{1}{t} \sum_{i=1}^{t} (\delta_i P(\emptyset_i)) \tag{8}$$

To compute the quotient, a kernel matrix W with dimensions $t \ \times \ t$ is established. The elements of this matrix are computed according to the formula outlined in Eq. (9).

$$W_{ij} = \ P(\emptyset_i)(P(\emptyset_i))^T = \ P(\emptyset_i). \ P(\emptyset_j) = W(\emptyset_i, \ \emptyset_j) \tag{9}$$

When the projected dataset $P(\emptyset_i)$ does not contain a mean

Attacks classification using CNN

In this phase, the convolution layer was employed to train the feature map in the higher layer using a convolution kernel. This process results in a fresh feature map that encompasses multiple feature maps as inputs to the convolution core. By convoluting multiple feature maps, a new output layer is generated for each output feature map. Eq. (10) illustrates the calculation process of the convolution layer.

$$X_j^l = f\left(\sum_{i \in M_j} X_i^{l-1} \times K_{ij}^l + b_j^l\right) \tag{10}$$

In this equation, X_j^l represents the jth feature of layer map l, K_{ij}^l denotes the function of the CNN convolutional kernel, f is the activation function while M_j and b_j^l stand for the input feature and the bias parameter, respectively. Each output feature is generated by combining an input feature along with a bias coefficient. The error signal for the layer, considering the weights of the feature, is determined based on the outcome of the previous step. The constant l is established in the lowermost sample layer as δ. This operation is iterated in the convolution layers to derive the error signal b_j^l for each feature graph j.

$$\delta_j^l = \ \beta_j^{l+1} \ \left(f'\left(u_j^l\right). up\left(\delta_j^{l+1}\right)\right) \tag{11}$$

The layer can be utilized to aggregate the height values as described in Eq. (11). The computation for a sampling operation is detailed in Eq. (12).
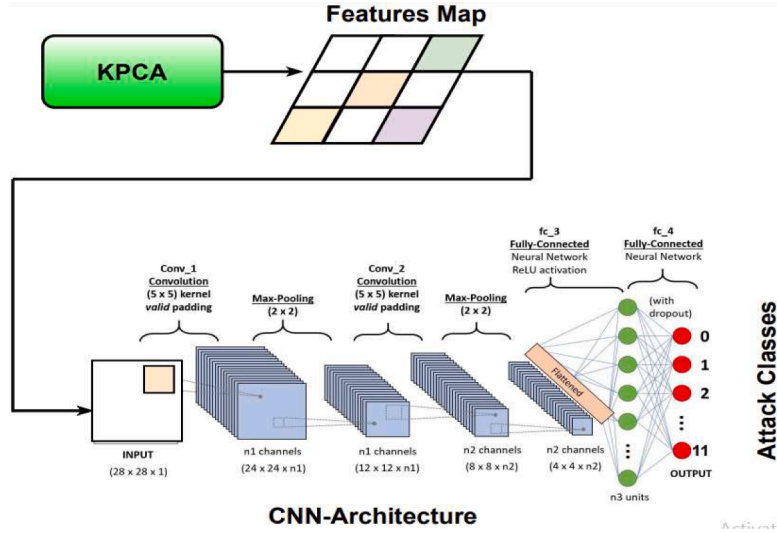
Fig. 2. The CNN architecture for recognizing 12 classes of IoT attacks.

$$\frac{\delta E}{\delta b_j} = \sum_{u,\,v} \left(\delta_j^l\right)_{uv} \tag{12}$$

Ultimately, the weight gradient of the convolution kernel can be computed using the traditional Backpropagation (BP) method in CNNs, involving weighted values with diverse connections. This process begins by generating a gradient for each link linked to a specific weight, followed by combining these gradients as depicted in Eq. (13).

$$\frac{\delta E}{\delta K_{ij}^l} = \sum_{u,\,v} \left(\delta_j^l\right)_{uv} \left(p_i^{l+1}\right)_{uv} \tag{13}$$

Here p_i^{l+1} can be seen as a product of K_{ij}^l involving a small block element within the convolutional layer, where the output feature graph value at position (u, v) is considered. This multiplication can also involve a deconvolution element that results from a small block in the upper (u, v) position. The concept behind the lowest sampling layer relies on the idea that each resulting feature map serves as a compact representation of the convolution layer.

$$X_j^l = f\left(\beta_j^l down\left(X_j^{l-1}\right) + b_j^l\right) \tag{14}$$

To achieve scale invariance, the feature graph is reduced by a factor of n, where down (X_j^{l-1}) represents the sample frame, with a lower resolution by a factor of $n*n$. Each resulting feature graph has its respective scaling offset variable and blending bias variable β.

$$\delta_j^l = f'\left(u_j^i\right).conv2\left(\delta_j^{l+1}, rot180\left(k_j^{l-1}\right), "full"\right) \tag{15}$$

To integrate the convolution function into the overall convolution process, it is necessary to perform a 180-degree rotation of the volume kernel before the computation. In addition, it can manage convolution at the border and handle any missing pixels within complement 0. Subsequently, this process will yield a result denoted as t_0, see the equations below:

$$\frac{\delta E}{\delta b_j} = \sum_{u,\,v} \left(\delta_j^l\right)_{uv} \tag{16}$$

$$\frac{\delta E}{\delta \beta_j} = \sum_{u,\,v} \left(\delta_j^l \cdot down\left(X_j^{l-1}\right)\right)_{uv} \tag{17}$$

In the convolutional neural network, the rise in frequency from time t to time $t+1$ can be likened to the BP approach.

$$w(t+1) = w(t) + \mu\delta(t)x(t) \tag{18}$$

In this context, $\delta(t)$ represents the error term, μ stands for the learning rate, and $x(t)$ represents the neuron's input. Fig. 2 illustrates the structure of the CNN model designed for the identification of 12 categories of IoT attacks.

Table 1

Analysis of the ToN-IoT dataset.

Attack category	Number of instances	Description
Normal Traffic	300,000	Common Non-Malicious Activities
Password Attack	20,000	Employs various methods, such as sniffer and brute-force techniques, with the aim of acquiring login credentials.
Ransomware Attack	20,000	Refers to a type of attack where server data is ciphered, and a ransom is inquired in exchange for the cipher key.
Scanning Attack	20,000	Unauthorized user or malicious software systematically scans a network or system for vulnerabilities, open ports, or potential points of entry.
XSS Attack	20,000	XSS is a technique where a malicious actor injects harmful files into web applications, targeting end users.
MITM Attack	1043	A Person/software in the Middle attack occurs when an attacker intercepts and listens in on communications between a target and the host they are communicating with.
Backdoor Attack	20,000	A malicious actor gains unauthorized access to a computer system, network, or application by exploiting hidden or undocumented vulnerabilities or intentionally created weaknesses in the system's security mechanisms.
DoS Attack	20,000	A deliberate effort to overwhelm the resources of a node (e.g., sensor node, system), disrupting access to its data.
DDoS Attack	20,000	Similar to a Denial of Service (DoS) attack but originates from multiple dispersed sources.
Injection Attack	20,000	Injection attacks, like SQL and Command Injection, in which an attacker inserts or "injects" malicious code or data into an application or system with the intent of manipulating the behaviour of the target system.

Table 2

Analysis of BoT-IoT Dataset.

Attack category	Number of instances	Description
Benign	477	Typical non-malicious patterns
DoS Attack	1650,260	This attack involves a deliberate effort to overwhelm the resources of a node (e.g., sensor node, system), disrupting access to its data.
DDoS Attack	1926,624	Similar to a Denial of Service (DoS) attack but originates from multiple dispersed sources.
Reconnaissance Attack	91,082	Involves an attacker gathering information about a target system, network, or organization to gain insights into its vulnerabilities, weaknesses, and potential points of entry.
Theft Attack	79	Is aimed at stealing sensitive information, assets, or data from individuals, organizations, or systems.

Performance analysis

We employed various performance metrics to assess the outcomes of the proposed algorithm and compared it with the performance of other recent systems utilizing both deep learning and hybrid rule-based techniques. The classification results, comprising correct and incorrect outputs, were aggregated and compared against benchmark results. Commonly utilized matrices in machine learning-based intrusion detection systems include Accuracy, Precision, Specificity, F1-score, True Positive Rate (TPR), and True Negative Rate (TNR). To refine the confusion matrix, numerical metrics such as True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) were calculated using Eqs. (19)-(25).

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{19}$$

$$Precision = \frac{TP}{TP + FP} \tag{20}$$

$$Recall = \frac{TP}{TP + FN} \tag{21}$$

$$Specificity = \frac{TN}{TN + FP} \tag{22}$$

$$F1 - Score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{23}$$

$$TPR = \frac{TP}{TP + FN} \tag{24}$$

$$FPR = \frac{FP}{FP + TN} \tag{25}$$

Experimental results and discussion

Two benchmark datasets were used to validate the efficiency of the proposed deep learning-based IDS model: ToN-IoT [17] and

Fig. 3. Accuracy of IDS model with/without feature selection using the TON-IoT dataset.

Fig. 4. Accuracy of IDS model with/without feature selection using the BoT-IoT dataset.

BoT-IoT [18]. The description of both datasets is summarized in Tables 1 and 2, respectively. The ToN-IoT dataset [17] collects and analyses heterogeneous IoT and IIoT data from connected devices, Windows and Linux system logs, and system network traffic. The ToN-IoT dataset connects multiple virtual machines, cloud layers, blur, edges, and physical systems to test AI-based cybersecurity technologies. It includes simultaneous legal and malicious events in IoT systems. The ToN-IoT dataset categorizes the traffic into normal behaviour and the attack subclass, which consists of nine types of attacks (XSS, DDoS, DoS, password cracking attacks, reconnaissance or verification, MITM, ransomware, backdoors, and injection attacks). A description of these attacks is given [17] and a brief is given in Table 1. In the context of metaverse-IoT applications, these attacks would behave as in the IoT environment (i.e., as described in [17]). For example, in backdoor attacks, malicious users can exploit vulnerabilities or create hidden access points in virtual environments to gain unauthorised access and control over the metaverse platform or users' accounts with persistent access. This form of attack can cause metaverse data breaches, identity theft, and virtual space manipulation. Injection attacks involve injecting code, commands, or data with malicious intent into a virtual environment for the purpose of exploiting vulnerabilities and compromising the security or functionality of the metaverse. These attacks have the potential to target the user's virtual objects/assets, giving attackers the ability to alter or gain control of virtual items, currency, or real estate, thus compromising the user's privacy.

The *BoT-IoT* [18] dataset was created to perform feature selection and accurately identify Bot attacks in the context of IoT networks. The database encompasses data including botnet attacks, amongst others, see Table 2. The development of this database, which includes valuable information features, is conducted in a practical testing environment. This enables accurate monitoring of traffic and the successful implementation of an efficient dataset.

The complexity of the prediction

The complexity of our proposed model is given as follows. The kernel size was 3 x 3, which helped us get fine-grained features. 8, 16, and 32 neurons were tested, but the 16 neurons were found to be the best number to deal with the size of the datasets and to control the problem of overfitting during training and testing. A deeper network was used to help avoid overfitting, which increased the performance of the CNN-based model.

Sigmoid, ReLU and SeLU functions were tested and the ReLU activation function was found the best as it introduced sparsity by setting negative values to zero. The sparsity makes the network more computationally efficient during both training and interpretation.
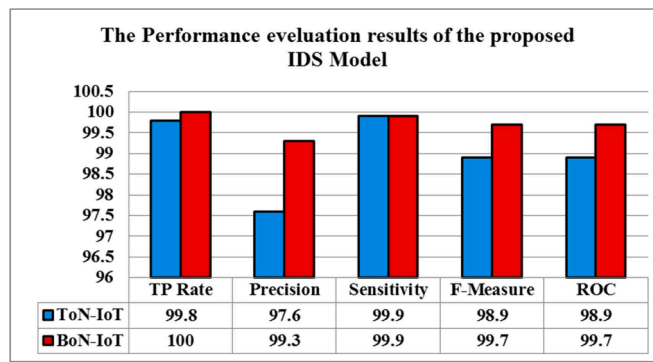
Fig. 5. The performance evaluation results of the proposed Metaverse-IDS model.

This helps reduce the number of operations required to compute and store intermediate results. For the learning rate, 0.01, 0.001, and 0.0001 were tried, and 0.001 learning rate was the best as it helped the model reach the minimum of the loss function, resulting in a more accurate and efficient model.

Multiple epochs (20, 50, 100) were used to help reduce overfitting, a common problem in deep learning. The use of multiple epochs allows the model to learn to discriminate between signal and noise, thus making it less likely to overfit. Also, different batch sizes (8, 16, and 32) were tested, and the 16 batch size performed better to afford intensive memory requirements and fit into the CPU memory. The batch size is generally better when it allows the model to see more diverse examples in each epoch and ensures stability during the training and testing of the proposed model.

Scenario 1: impact of feature selection (KPCA) on the IDS performance

Two experiments were conducted in this scenario which was designed to investigate the impact of employing the feature selection technique, KPCA, in our case, before the CNN classification phase. This investigation was done using two datasets ToN-IoT and BoN-IoT. The results of experiment one (using the ToN-IoT dataset) are summarized in Fig. 3 while the results of the 2nd experiment (using the BoT-IoT dataset) are given in Fig. 4. From Fig. 3, it can be noticed that employing KPCA in the ToN-IoT dataset has improved the detection accuracy for all attack classes compared to the results without using the KPCA. Nearly all attacks can be detected with an accuracy of 97 − 100 except the MITM attack. However, its detection rate was better with applying the KPCA. Similarly, Fig. 4 shows that employing the KPCA to BoN-IoT dataset improved the detection rate of all attack classes. The detection rate reached 100 % in DDoS, Reconnaissance, and Theft attacks. This can be attributed to the fact that the BoN-IoT dataset contains only four types of attacks compared to the NoT-IoT dataset, which contains nine types of attacks. In addition to improving the detection rate, as reported in [35] and [36], using KPCA to select the best set of features for the detection process will result in the elimination of these features, which then leads to a reduction in storage and computational costs, while also preventing substantial information loss or deterioration in learning performance. Such reduction in storage and computational is crucial in an IoT environment with limited resources.

Fig. 3 depicts the performance analysis results of TPR, Precision, Recall/Sensitivity, F1-Score, and ROC metrics. The FNR can be derived from the TPR by subtracting it from 1, i.e., $FNR = 1 − TPR$. The accuracy of intrusion detection is directly affected by FNR which denotes the frequency at which an IDS fails to accurately detect and identify genuine instances of intrusions or attacks. A high FNR indicates a significant proportion of actual attacks remain undetected, hence raising concerns over the adequacy of security measures in place. Based on the findings presented in Fig. 3, it is evident that the FNR of our proposed IDS model is below 0.2. This indicates that our IDS model exhibits a high level of effectiveness in immediately detecting and addressing legitimate threats, hence mitigating the potential risks associated with unauthorised access, data breaches, and system compromises.

Scenario 2: evaluating the proposed IDS performance

The results of Scenario 1 suggested that applying KPCA would improve the detection rate of our proposed IDS. Then, Scenario 2, was aimed to further evaluate the quality of the proposed model under the most common intrusion detection metrics, i.e., Precision, Sensitivity, F1-score, and TPR.

In this scenario, two main experiments were conducted: ToN-IoT and BoT-IoT datasets. Then a comparison between the results of each experiment was recorded, as given in Fig 5. From this figure, it can be noticed that the obtained results confirm the efficiency of the proposed deep learning-based IDS model in detecting the 12 classes of IoT attacks in the metaverse in both datasets, ToN-IoT and BoN-IoT. It can also be seen that the results in all metrics of the proposed IDS model using the BoT-IoT dataset are better than those using the IoN-IoT dataset. As explained above, this could be because the BoN-IoT dataset contains only four types of attacks compared to the NoT-IoT dataset, which contains nine types of attacks, i.e., the latter is more complex than the former.

Table 3
Comparison results of IoT attacks-detection accuracy.

| Model	Detection Performance	
	Data	AVG of Detection Accuracy
Wrapper + neuro tree [35]	KDDCUP 99 data set [41]	98.38%
SVM+EML+K-Means [36]	benchmark data set [36]	95.75%
GA +SVM [37]	KDD Cup 1999 data [41]	97.3%
CNN+LSTM [38]	NSL-KDD dataset [42]	94.12%
Modified KNN [39]	Malimg, Malheur, VirusShare, and Microsoft Kaggle datasets [39]	98.7%
GA+RuleEval+ANN [40]	NSL-KDD and UNSW-NB15 [40]	98.8%
CNN-based IDS [7]	NID Dataset and BoT-IoT datasets [7]	96.15%
DIDS [9]	Bench Mark datasets [9]	99%
HetIoT-CNN IDS [10]	CICDDoS2019 [43]	99.8%
Proposed Model	ToN-IoT and BoT-IoT [17,18]	99.8%
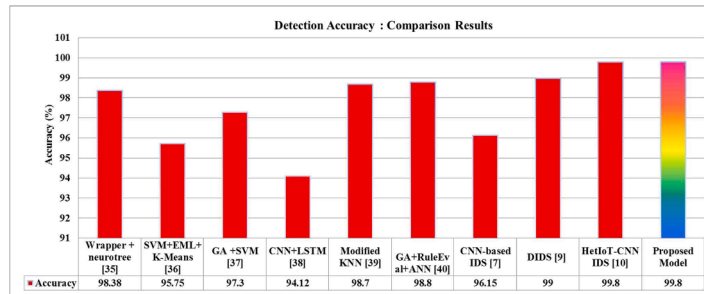
Fig. 6. Comparison results of attack detection accuracy of the proposed IDS model against other techniques in the literature.

Table 4
The proposed CNN-based Metaverse-IDS Vs GAN -based IDS method.

	Dataset	Feature Extraction	Attack Classification	Attack Types
GAN-based IDS model [44]	One dataset: SDN intrusion dataset [45]	Deep auto-encoder algorithm	GAN + Random Forest classifier	Dos, DDos, Web Attacks, R2L, Malware, Probe and U2R attack
The proposed CNN–based Metaverse-IDS model	Two datasets: ToN-IoT [17] & BoT-IoT [18].	KPCA	CNN	Normal, MITM, Password, Injection, Backdoor, Ransomware, DoS, DDoS, Scanning, Reconnaissance, XSS, and Theft attack
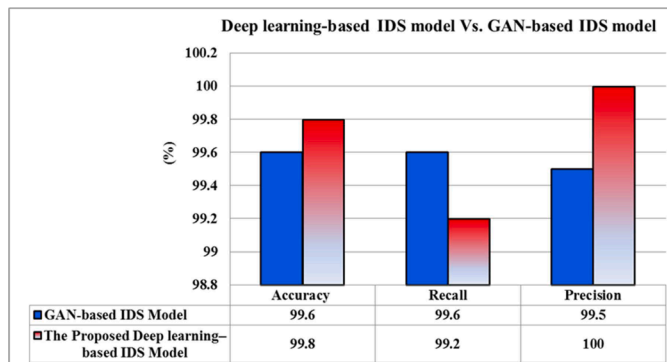
Fig. 7. Comparison results of multiple attack detection performance of the proposed CNN-based Metaverse IDS model against the GAN-based IDS model [44].

Comparison with the literature

To assess the effectiveness of the suggested IDS model, we conducted a comprehensive comparative analysis. This evaluation aimed to scrutinize the detection capabilities of our proposed IDS model when pitted against other well-established IDS techniques

documented in the literature. The summarized findings of this comparative analysis are presented in Table 3, which clearly highlights the superior attack detection accuracy achieved by our proposed IDS model. For a visual representation of the performance comparison, please refer to Fig. 6, which graphically illustrates how the detection accuracy of our proposed IDS outperforms that of other techniques found in existing literature. Although the accuracy of the model in HetIoT-CNN IDS [10] has the same accuracy as our proposed model, the latter is more reliable for two reasons. Firstly, HetIoT-CNN IDS [10] used a generic dataset CICDDoS2019 which not only contains IoT traffic. Secondly, our model used two datasets, ToN-IoT [17] and BoT-IoT [18], which are specifically designed for IoT environments. So, our model is more reliable than HetIoT-CNN IDS [10].

Before writing this study, there was only one research attempt to investigate the possibility of developing intrusion detection techniques for 5 G networks to establish metaverse-real-time communication [44]. Although the authors of this study used a different dataset [45] to validate their proposed IDS model, an interesting finding clarified better performance of our proposed deep learning-based IDS model compared to the use of a Generative Adversarial Network (GAN) for the IDS model in [44]. Table 4 summarizes the comparison results between both two IDS models' methodologies. Fig. 7 compares the detection performance results between both two IDS models. Although the GAN-based IDS model [44] achieved better true positive rate (i.e., recall) results than our proposed IDS model, the proposed IDS model performed the best attack detection accuracy and precision results. In addition, our proposed model was evaluated using two recent datasets collected from the IoT environment, i.e., more realistic for the evaluation of the IDS model for Metaverse-IoT neConverselyther side, the work in [44] used datasets generated from traditional SDN environments. IoT specifications were not considered in such an environment. Furthermore, as can be seen from Table 4, our proposed IDS model can successfully detect a higher number of attacks (12 classes) than the model in [44] (8 classes only). This proves that our proposed model is more efficient than most related work.

One of the limitations of the proposed model is that its results must be interpreted with caution because the proposed IDS model was tested and evaluated using a benchmark IoT dataset which would be similar to that dataset which would be collected from the real metaverse-IoT network. The benchmark IoT datasets were used because there was no any public datasets from the integration of metaverse and IoT data. However, as explained above, the used datasets (ToN-IoT and BoN-IoT) contain attacks which would target the metaverse-IoT networks. In addition, using novel encryption techniques such as visual chaotic image/avatar encryption may represent the promising base for building more robust and secure IDS models in Metaverse networks [46–48].

Conclusion

This study aimed to investigate the efficiency of a proposed deep learning-based intrusion detection system to recognise twelve classes of IoT attacks in the metaverse. The power of the proposed IDS model is based on two essential techniques: KPCA for performing the attack feature extraction step and CNN for performing the attack recognition and classification step. Two benchmark datasets containing 12 types of IoT attacks, that will target IoT communications in metaverse spaces, were used to test how well the proposed IDS model works. The results of this investigation showed that the proposed IDS model is an efficient detection model for recognising IoT attacks in the metaverse, where it achieved a detection accuracy of 99.8%, a precision of 100%, a recall of 99.2%, and the FNR of 0.2. The low FNR is a crucial feature of the IDS model, as it is necessary for the early identification of attacks. The comparison results of the attack detection accuracy of the proposed IDS model against other techniques in the literature also confirmed the superiority of the proposed IDS model compared to eight other IDS techniques. These findings suggest a role for deep learning technology in developing an intelligent IDS model for recognising various classes of IoT attacks in metaverse communications. Further research efforts might explore the efficiency of novel deep learning models for identifying other types of attacks in the metaverse, such as NFTs (Non-Fineable Tokens) attacks, fake avatar attacks, dark verse attacks, cyber-physical attacks, and virtual, augmented, or extended reality attacks. Another important future work could be the determination of whether to do the predictions in the cloud or at the edge. Both cloud computing and edge computing include specific advantages and disadvantages, requiring careful consideration of individual requirements when selecting between the two options. These requirements would include maintenance, scalability, latency and cost.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT in order to improve language and readability. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

ToN-IoT dataset: https://research.unsw.edu.au/projects/toniot-datasets and BoT-IoT dataset is from here https://research.unsw.edu.au/projects/bot-iot-dataset

References

[1] H. Ning, H. Wang, Y. Lin, W. Wang, S. Dhelim, F. Farha, M. Daneshmand, A survey on the Metaverse: the State-of-the-Art, technologies, applications, and challenges, IEEE IoT J. (2023).

[2] O. Nnamonu, M. Hammoudeh, T. Dargahi, Digital forensic investigation of web-based virtual reality worlds: decentraland as a case study, IEEE Commun. Mag. 61 (9) (2023) 72–78.

[3] Z. Lv, S. Xie, Y. Li, M.S. Hossain, A. El Saddik, Building the Metaverse by digital twins at all scales, state, relation, Virtual Real. Intell. Hardware 4 (6) (2022 Dec 1) 459–470.

[4] Y. Huang, Y.J. Li, Z. Cai, Security and privacy in metaverse: a comprehensive survey, Big Data Min. Anal. 6 (2) (2023) 234–247.

[5] I. Yaqoob, K. Salah, R. Jayaraman, M. Omar, Metaverse applications in smart cities: enabling technologies, opportunities, challenges, and future directions, IoT (2023), 100884.

[6] S. Singh, P.K. Sharma, S.Y. Moon, D. Moon, J.H. Park, A comprehensive study on APT attacks and countermeasures for future networks and communications: challenges and solutions, J. Supercomput. 75 (8) (2019 Aug) 4543–4574.

[7] T. Saba, A. Rehman, T. Sadad, H. Kolivand, S.A. Bahaj, Anomaly-based intrusion detection system for IoT networks through deep learning model, Comput. Electr. Eng. 99 (2022 Apr 1), 107810.

[8] A.S. Dina, A.B. Siddique, D. Manivannan, A deep learning approach for intrusion detection in Internet of Things using focal loss function, IoT (2023 Jan 16), 100699.

[9] B. Madhu, M.V. Chari, R. Vankdothu, A.K. Silivery, V. Aerranagula, Intrusion detection models for IOT networks via deep learning approaches, Measurement: Sensors 25 (2023 Feb 1), 100641.

[10] S. Mahadik, P.M. Pawar, R. Muthalagu, Efficient intelligent intrusion detection system for heterogeneous Internet of Things (HetIoT), J. Network Syst. Manag. 31 (1) (2023 Mar) 2.

[11] N.C. Thompson, K. Greenewald, K. Lee, G.F. Manso, Deep learning's diminishing returns: the cost of improvement is becoming unsustainable, IEEE Spectr 58 (10) (2021) 50–55.

[12] Sun, J., Gan, W., Chen, Z., Li, J., & Yu, P.S. (2022). Big data meets metaverse: a survey. arXiv preprint http://arxiv.org/abs/arXiv:2210.16282.

[13] T. Gaber, A. Tharwat, V. Snasel, A.E. Hassanien, Plant identification: two dimensional-based vs. one dimensional-based feature extraction methods, in: 10th International Conference on soft computing models in Industrial and environmental applications, Springer International Publishing, 2015, pp. 375–385.

[14] J. Li, K. Cheng, S. Wang, F. Morstatter, R.P. Trevino, J. Tang, H. Liu, Feature selection: a data perspective, ACM Comput. Surv. (CSUR) 50 (6) (2017) 1–45.

[15] M.A. Ali, M. Meselhy Eltoukhy, F. Rajeena PP, T. Gaber, Efficient thermal face recognition method using optimized curvelet features for biometric authentication, PLoS ONE 18 (6) (2023), e0287349.

[16] T. Gaber, A. El-Ghamry, A.E. Hassanien, Injection attack detection using machine learning for smart IoT applications, Phys. Commun. 52 (2022), 101685.

[17] N. Koroniotis, N. Moustafa, E. Sitnikova, B. Turnbull, Towards the development of realistic botnet dataset in the internet of things for network forensic analytics: bot-iot dataset, Future Generat. Comput. Syst. 100 (2019 Nov 1) 779–796.

[18] N. Moustafa, A new distributed architecture for evaluating AI-based security systems at the edge: network TON_IoT datasets, Sustain. Cities Soc. 72 (2021 Sep 1), 102994.

[19] K. Tsiknas, D. Taketzis, K. Demertzis, C. Skianis, Cyber threats to industrial IoT: a survey on attacks and countermeasures, IoT 2 (1) (2021 Mar) 163–186.

[20] N. Abosata, S. Al-Rubaye, G. Inalhan, C. Emmanouilidis, Internet of things for system integrity: a comprehensive survey on security, attacks, and countermeasures for industrial applications, Sensors 21 (11) (2021 May 24) 3654.

[21] S. Ding, L. Kou, T. Wu, A GAN-based intrusion detection model for 5G enabled future metaverse, Mobile Netw. Appl. (2023 Jan 10) 1–5.

[22] D. Yuan, K. Ota, M. Dong, X. Zhu, T. Wu, L. Zhang, J. Ma, Intrusion detection for smart home security based on data augmentation with edge computing, in: InICC 2020-2020 IEEE International Conference on Communications (ICC), IEEE, 2020 Jun 7, pp. 1–6.

[23] Y. He, M. Kong, C. Du, D. Yao, M. Yu, Communication security analysis of intelligent transportation system using 5G internet of things from the perspective of big data, IEEE Trans. Intell. Transp. Syst. (2022 Mar 23).

[24] J. Lee, K. Park, AE-CGAN model-based high-performance network intrusion detection system, Appl. Sci. 9 (20) (2019 Oct 10) 4221.

[25] S. Zhao, J. Li, J. Wang, Z. Zhang, L. Zhu, Y. Zhang, attackgan: adversarial attack against black-box ids using generative adversarial networks, Procedia Comput. Sci. 187 (2021 Jan 1) 128–133.

[26] Y. Liu, S. Mao, X. Mei, T. Yang, X. Zhao, Sensitivity of adversarial perturbation in fast gradient sign method, in: 2019 IEEE symposium series on computational intelligence (SSCI), IEEE, 2019 Dec 6, pp. 433–436.

[27] Y. Deng, L.J. Karam, Universal adversarial attack via enhanced projected gradient descent, in: In2020 IEEE International Conference on Image Processing (ICIP), IEEE, 2020 Oct 25, pp. 1241–1245.

[28] N. Carlini, D. Wagner, Towards evaluating the robustness of neural networks, in: In2017 IEEE Symposium on security and privacy (sp), IEEE, 2017 May 22, pp. 39–57.

[29] M. Usama, M. Asim, S. Latif, J. Qadir, Generative adversarial networks for launching and thwarting adversarial attacks on network intrusion detection systems, in: 2019 15th international wireless communications & mobile computing conference (IWCMC), IEEE, 2019 Jun 24, pp. 78–83.

[30] C. Yin, Y. Zhu, J. Fei, X. He, A deep learning approach for intrusion detection using recurrent neural networks, in: Ieee Access, 5, 2017 Oct 12, pp. 21954–21961.

[31] A. Verma, V. Ranga, Machine learning based intrusion detection systems for IoT applications, Wirel. Pers. Commun. 111 (2020 Apr) 2287–2310.

[32] W. Li, W. Meng, M.H. Au, Enhancing collaborative intrusion detection via disagreement-based semi-supervised learning in IoT environments, J. Netw. Comput. Appl. 161 (2020 Jul 1), 102631.

[33] N. Alexopoulos, E. Vasilomanolakis, N.R. Ivánkó, M. Mühlhäuser, Towards blockchain-based collaborative intrusion detection systems, in: InCritical Information Infrastructures Security: 12th International Conference, CRITIS 2017, Lucca, Italy, October 8-13, 2017, Springer International Publishing, 2018, pp. 107–118. Revised Selected Papers 12.

[34] T.V. Khoa, Y.M. Saputra, D.T. Hoang, N.L. Trung, D. Nguyen, N.V. Ha, E. Dutkiewicz, Collaborative learning model for cyberattack detection systems in iot industry 4.0, in: 2020 IEEE Wireless Communications and Networking Conference (WCNC), IEEE, 2020 May 25, pp. 1–6.

[35] G. Stein, B. Chen, A.S. Wu, K.A. Hua, Decision tree classifier for network intrusion detection with GA-based feature selection, in: Proceedings of the 43rd annual Southeast Regional Conference 2, 2005 Mar 18, pp. 136–141. Volume.

[36] G. Goswami, A.K. Das, A. Chakraborty, B. Chakraborty, A feature cluster taxonomy-based feature selection technique, Expert Syst. Appl. 79 (2017 Aug 15) 76–89.

[37] B.M. Aslahi-Shahri, R. Rahmani, M. Chizari, A. Maralani, M. Eslami, M.J. Golkar, A. Ebrahimi, A hybrid method consisting of GA and SVM for the intrusion detection system, Neural. Comput. Appl. 27 (2016 Aug) 1669–1676.

[38] C.M. Hsu, H.Y. Hsieh, S.W. Prakosa, M.Z. Azhari, J.S. Leu, Using long-short-term memory-based convolutional neural networks for network intrusion detection, in: InWireless Internet: 11th EAI International Conference, WiCON 2018, Taipei, Taiwan, October 15-16, 2018, Springer International Publishing, 2019, pp. 86–94. Proceedings 11.

[39] H. Naeem, B. Guo, M.R. Naeem, F. Ullah, H. Aldabbas, M.S. Javed, Identification of malicious code variants based on image visualization, Comput. Electric. Eng. 76 (2019 Jun 1) 225–237.

[40] J.B. Awotunde, C. Chakraborty, A.E. Adeniyi, Intrusion detection in the industrial Internet of things network based on deep learning model with rule-based feature selection, Wirel. Commun. mobile Comput. 2021 (2021 Sep 2) 1–7.

[41] K. Raghuveer, et al., Performance evaluation of data clustering techniques using KDD Cup-99 Intrusion detection data set, Int. J. Inf. Netw. Secur. (IJINS) 1 (4) (2012) 294–305.

[42] L. Dhanabal, S.P. Shantharajah, A study on NSL-KDD dataset for intrusion detection system based on classification algorithms, Int. J. Adv. Res. Comput. Commun. Eng. 4 (6) (2015 Jun 6) 446–452.

[43] UNB website, DDoS evaluation dataset CICDDoS2019. (Online), available: https://www.unb.ca/cic/datasets/ddos-2019.html (2019). Accessed 10 Jun 2020.

[44] S. Ding, L. Kou, T. Wu, A GAN-based intrusion detection model for 5G enabled future metaverse, Mobile Netw. Appl. 27 (6) (2022 Dec) 2596–2610.

[45] M.S. Elsayed, N.A. Le-Khac, A.D. Jurcut, InSDN: A Novel SDN Intrusion Dataset, 8, IEEE Access, 2020 Sep 8, pp. 165263–165284.

[46] J.S. Khan, J. Ahmad, S.S. Ahmed, H.A. Siddiqa, S.F. Abbasi, S.K. Kayhan, DNA key based visual chaotic image encryption, J. Intell. Fuzzy Syst. 37 (2) (2019) 2549–2561.

[47] S.F. Abbasi, J. Ahmad, J.S. Khan, M.A. Khan, S.A. Sheikh, Visual meaningful encryption scheme using intertwining logistic map, in: Intelligent Computing: Proceedings of the 2018 Computing Conference, Volume 2, Springer International Publishing, 2019, pp. 764–773.

[48] J.S. Khan, J. Ahmad, S.F. Abbasi, S.K. Kayhan, DNA sequence based medical image encryption scheme. In 2018 10th Computer Science and Electronic Engineering (CEEC), IEEE, 2018, pp. 24–29.

[49] J.B. Awotunde, T. Gaber, L.N. Prasad, S.O. Folorunso, V.L. Lalitha, Privacy and security enhancement of smart cities using hybrid deep learning-enabled blockchain, Scalable Comput.: Pract. Exp. 24 (3) (2023) 561–584.