International Journal of Advanced Technology and Engineering Exploration, Vol 10(109)
ISSN (Print): 2394-5443 ISSN (Online): 2394-7454
http://dx.doi.org/10.19101/IJATEE.2022.10100187

Research Article

AS-CL IDS: anomaly and signature-based CNN-LSTM intrusion detection system for Internet of Things

Jinsi Jose[1,2*] and Deepa V. Jose[1]
Department of Computer Science, CHRIST University, Bangalore, India[1]
Department of Data Science, St. Joseph's College, Moolamattom, India[2]

Received: 21-September-2022; Revised: 14-December-2023; Accepted: 16-December-2023

Abstract
In recent years, the internet of things (IoT) has had a significant impact on our daily lives, offering various advantages for improving our quality of life. However, it is crucial to prioritize the security of IoT devices and the protection of user's personal data. Intrusion detection systems (IDS) play a critical role in maintaining data privacy and security. An IoT IDS continuously monitors network activity and identifies potential security risks or attacks targeting IoT devices. While traditional IDS solutions exist, intrusion detection heavily relies on artificial intelligence (AI). AI can greatly enhance the capabilities of IoT IDS through real-time monitoring, precise threat identification, and automatic response capabilities. It is essential to develop and utilize these technologies securely and responsibly to mitigate potential risks and safeguard user privacy. A hybrid IDS was proposed for anomaly-based and signature-based intrusions, leveraging convolutional neural network with long short-term memory (CNN-LSTM). The name of the proposed hybrid model is anomaly and signature-based CNN-LSTM intrusion detection system (AS-CL IDS). The AS-CL IDS concentrated on two different IoT IDS detection strategies employing a combination of deep learning techniques. The model includes model training and testing as well as data preprocessing. The CIC-IDS 2018, IoT network intrusion dataset, MQTT-IoT-IDS2020, and BoTNeTIoT-L01 datasets were used to train and test the AS-CL IDS. The overall performance of the proposed model was assessed using accepted assessment metrics. Despite reducing the number of characteristics, the model achieved 99.81% accuracy. Furthermore, a comparison was made between the proposed model and existing alternative models to demonstrate its productivity. As a result, the proposed model proves valuable for predicting IoT attacks. Looking ahead, the deployment strategy of the IoT IDS can anticipate the utilization of real-time datasets for future implementations.

Keywords
Internet of things, Intrusion detection systems, Deep learning, Machine learning, Artificial intelligence, IoT dataset, Hybrid intrusion detection.

1.Introduction
The IoT refers to the network of physical objects or "things" embedded with sensors, software, and other technologies that enable them to connect and exchange data with other devices and systems over the internet. These objects range from everyday items like smart home appliances, wearables, and vehicles, to industrial machinery and infrastructure. Internet of things (IoT) devices are typically designed to collect and share data with other devices or systems, often in real-time, to enable automated decision-making and improve efficiency, productivity, and convenience. For example, a smart thermostat in a home can collect data on temperature and humidity levels and adjust heating and cooling settings accordingly.

At the same time, a sensor-equipped machine on a factory floor can monitor its performance and alert maintenance personnel if a problem is detected. The IoT can potentially revolutionise many aspects of daily life and business but raises concerns about data privacy, security, and the ethical use of personal information [1, 2]. Kevin Ashton claimed the IoT in Massachusetts institute of technology (MIT) Auto-ID labs in 1999 [3]. The first article on IoT in 2004 from MIT was called "Internet 0". The dawning era of the internet has led to the vast evolution of IoT devices in our daily life all over the globe.

Cisco predicts in an annual report of 2022 that the number of internet users, devices, and connections will increase globally between 2018 and 2023. IoT device growth forecasts are represented graphically in *Figures 1(a)* and *1(b)*. By 2023, there will be 8.0

*Author for correspondence

billion people on the planet, up from 7.6 billion in 2018. By 2023, 66 per cent of the population will utilise the Internet, up from 51 per cent in 2018. By 2023, 71 per cent of people will use mobile devices, up from 66 per cent in 2018 [4]. *Figure 1(a)* shows the graphical representation of the growth comparison of internet users and devices between 2018 to 2023. Between 2018 and 2023, networked and mobile-connected devices will expand at a compound annual growth rate (CAGR) of 9.8% and 8.3%, respectively. Globally, the number of devices per person will eventually rise as well. By 2023, there will be 3.6 networked devices and connections worldwide per person, up from 2.4 in 2018, and there will be 1.6 mobile-connected devices worldwide per person, up from 1.2 in 2018 [4]. *Figure 1(b)* shows the growth in usage of devices per capita from 2018 to 2023.

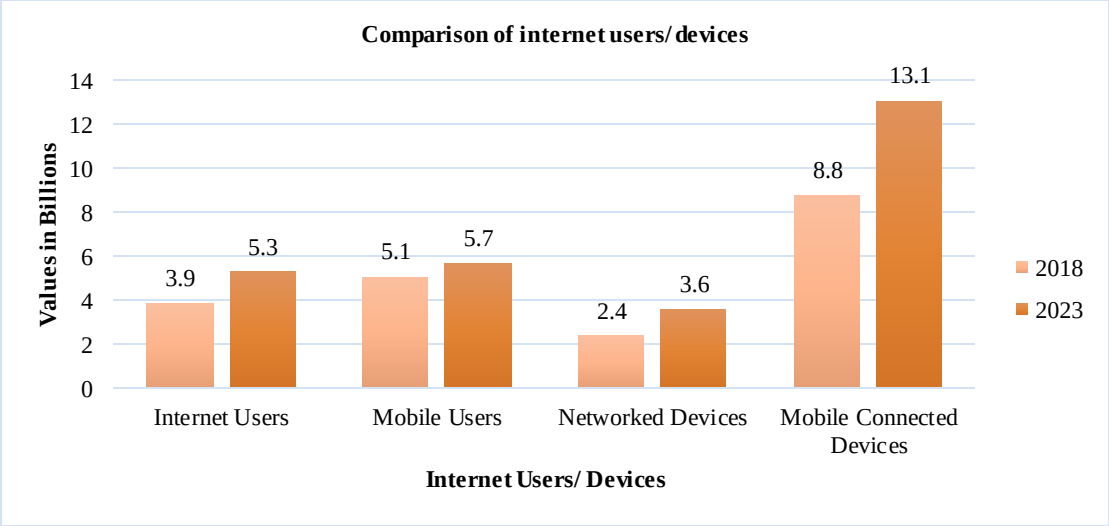

Figure 1(a) Growth comparison of internet users and devices between 2018 to 2023

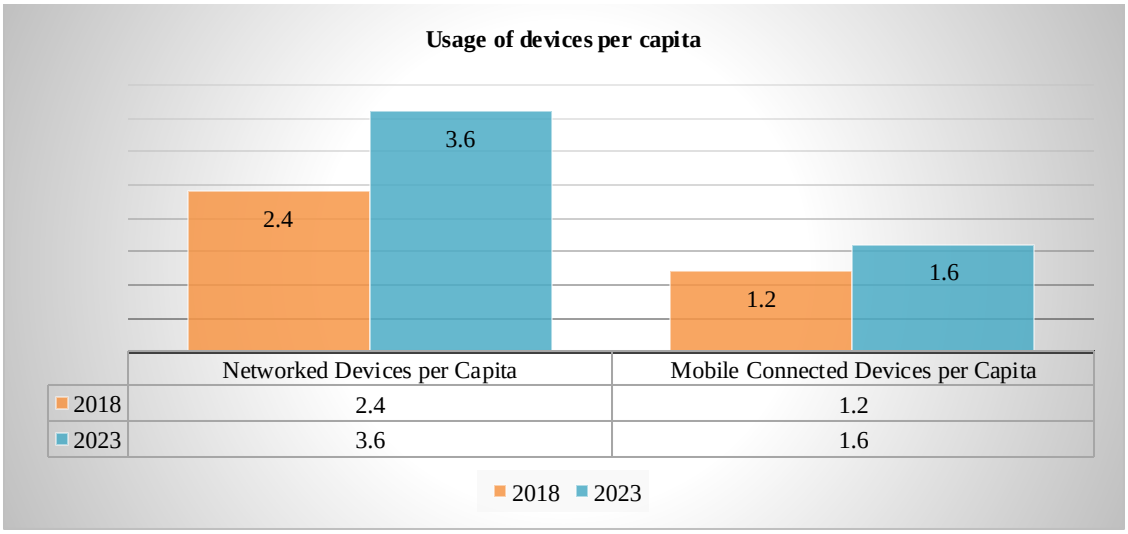

Figure 1(b) Growth in usage of devices per capita from 2018 to 2023

The drastic development of technology leads to every aspect of life-related to various IoT applications. IoT applications can be divided into three main categories: smart health, smart city, and smart industrial. The different categories include various subcategories like medical and health care, personal living, environmental monitoring, home automation, transportation, maritime shipping, smart metering and smart grid, security and emergency, critical infrastructure, the food supply chain, smart retail, agriculture and animal farming, and construction management [5]. Even though IoT is substantial with

numerous and divergent applications, it shares habitual factors. Most IoT applications focus on communications between different devices with less human interference and make the actions easy. So, gathering and keeping track of the data, exchanging data, automation, and collaboration are the standard features of IoT applications [6].

All IoT applications deal with a vast amount of data; as a result, the users are more deliberate about the security of the data. From the beginning, finding out the presence of intrusions in the IoT was a primary concern. IoT intrusion detection system (IDS) is a tool for identifying the intrusion. IoT IDS is a security mechanism designed to monitor and detect unauthorised access or attacks on IoT devices and networks. An IoT IDS works by continuously monitoring the network traffic between devices and analyzing it for any suspicious activity or anomalies that may indicate a security breach [7]. The categorization of IoT IDSs is different based on different researchers. IoT IDS can be divided into different categories based on type, IDS placement strategies, and IDS detection method. Based on types of IDSs can be either host-based intrusion detection system (HIDS) or network-based intrusion detection system (NIDS). The IDS placement strategies are distributed, centralised and hybrid. The detection methods can be anomaly-based, signature-based, and hybrid. [8, 9]. Signature-based detection involves comparing network traffic against a database of known attack patterns, while anomaly-based detection involves monitoring for any unusual or unexpected behavior on the network [10]. The hybrid method uses signature-based and anomaly-based detection methods to identify potential security threats.

An IoT IDS can help prevent security breaches and protect against cyber-attacks by identifying and responding to threats in real time. Some IoT IDS systems can also be configured to automatically respond to potential threats, such as blocking traffic from a specific internet protocol (IP) address or quarantining a compromised device from the network. Many traditional IoT IDS are available; by using data-mining, statistical, payload, and rule-based methods [11]. The main drawback of the conventional method is that it has limitations in identifying new types of attacks. In the last few years, artificial intelligence (AI) advancements have played a significant role in growing IoT IDS. AI can be used to enhance the accuracy and effectiveness of these systems. Here are some ways AI can be used in IoT IDS:

- Anomaly detection: AI can be trained to identify normal behaviour patterns for IoT devices and detect anomalies that could indicate a security breach.
- Machine learning (ML): ML algorithms can analyse data from multiple sources and identify patterns that could indicate a security threat.
- Predictive analytics: AI can predict potential security threats based on past patterns and trends.
- Automated response: AI can trigger automated responses to detected security threats, such as blocking traffic or alerting security personnel.
- Natural language processing (NLP): NLP can analyse human-generated content related to IoT devices, such as user reviews or social media posts, to identify potential security threats [12–15].

Overall, AI can significantly enhance the capabilities of IoT IDS by providing real-time monitoring, accurate threat detection, and automated response capabilities. However, ensuring these systems are designed and implemented securely and responsibly is essential to prevent potential risks and ensure user privacy.

Recently, most researchers have focused on ML and deep learning (DL) techniques in IoT IDS for better results. AI leads to implement the intelligence into the machine as it is human intelligence. ML and DL are subsets of AI [16, 17]. ML is a broad category of algorithms that enable machines to automatically learn patterns in data and make predictions or decisions based on that learning. ML models typically require manual feature engineering, which involves selecting and engineering relevant features from raw data that can be used to make predictions. ML models can also be trained with labelled and unlabeled data, although labelled data is generally required for supervised learning [18]. On the other hand, DL is a specific subset of ML that involves the training of artificial neural networks with many layers. DL models can automatically learn to extract relevant features from raw data, eliminating the need for manual feature engineering. DL has shown tremendous success in areas such as computer vision, NLP and speech recognition, where complex patterns in data need to be learned [19]. In summary, while ML and DL involve training models to make predictions or decisions based on data, DL is a more advanced and specialized form of ML that involves training neural networks with many layers to learn relevant features from raw data automatically.

Researchers have been focused on a hybrid model for intrusion detection in the last few years. This paper

focuses on a hybrid intrusion detection technique for IoT based on a detection method and a combination of two DL techniques. IoT devices are widely used today in many spheres of life. As the number of IoT devices continues to grow, the need for effective IoT IDS systems will become increasingly important to ensure the security and integrity of these devices and the data they collect and transmit. The utmost priority is safeguarding the security and privacy of such data and preventing its exploitation. IDS can be used to help analyse the quantity and types of assaults, and with this knowledge, security systems can be changed, or more effective controls can be put in place for improved security.

The work is motivated by the limitations of the conventional methods in IoT to identify the presence of intrusions in real-time scenarios. The objective of this study is to develop an intrusion detection technique for IoT using a DL approach to increase the accuracy of attack detection. With the help of a convolutional neural network (CNN) with long short-term memory (LSTM), this study suggests a hybrid IDS for anomaly-based and signature-based intrusions. The proposed model is named AS-CL, anomaly and signature-based CNN- LSTM hybrid model. Four datasets—CIC-IDS2018, IoT network intrusion dataset, MQTT-IoT-IDS2020, and BoTNeTIoT-L01—were used to train and evaluate the hybrid model. The overall performance of the proposed model is assessed using accepted assessment metrics.

Our research contributes by developing a hybrid model named as AS-CL IDS, focused on identifying anomaly-based and signature-based attacks rather than particular types of attacks in the IoT environment. Most of the hybrid models in an IoT IDS are just a combination of two models and perform the attack detection, but in the case of AS-CL IDS concentrated on hybrid detection method and security threats in IoT and is also a combination of real-time data analysis models CNN and LSTM. For the analysis of the AS-CL IDS model, the latest four datasets are specially intended for IoT IDS. Each dataset varied the features randomly to verify accuracy, achieving a minimum accuracy of 99.81% and a maximum time of 311 seconds for identifying attacks. The proposed model is well for the identification zero-day attacks. This research worked on several ML and DL models and evaluated with state-of-the-art literature for attack detection and identified the most accurate models for attack detection. This research work provides an improved solution to be cautious about possible attacks and hence helps us to take remedial precautions. The proposed model shows improved accuracy compared to the existing hybrid IoT IDS. Evaluated the effectiveness of the proposed model using relevant parameters against state-of-the-art literature for attack detection.

The structure of the paper is outlined as follows: Section 2 provides the details of related work carried out for this paper. Section 3 covers the methodology and details of the proposed methods of hybrid IoT technique. The dataset used in this paper and different evaluation metrics are discussed in section 4. Experiments and results are discussed in section 5, and the conclusion and future work are discussed in section 6.

2.Literature review

This section proffers to the review of existing hybrid IoT IDS carried out by other researchers. In [20], the proposed hybrid IDS can handle anomaly and specification-based routing attacks. The researchers mainly focus on sinkholes, wormholes, and selective-forwarding attacks. The clustering of the data packets is done by an unsupervised optimum-path forest (OPF) algorithm and MapReduce approaches. The proposed IDS was tested in the smart city environment. The proposed model yielded 76.91% correctness for both sinkhole and selective forward attacks and 96.02% for wormhole attacks, respectively. The main limitation of the work is that MapReduce architecture served as the foundation for the proposed anomaly detection strategy. The researchers themselves suggested that future work will incorporate data mining techniques and computational intelligence-based approaches to enhance the performance of the recommended hybrid IDS system. In another study presented in [21], the researchers have developed a hybrid model combining two models, advanced support vector machine (ASVM) and fuzzy c-means (FCM) clustering, for anomaly detection. The proposed model was tested with the NSL KDD dataset and compared with the other two hybrid methods. The result shows a better detection rate of 99% rather than other methods. The drawback is that the model focused only on the binary classification of the attacks.

The anomaly-based hybrid model presented in [22] consists of a combination of four ML algorithms. The testing of the model used real-time routing-specific attacks in the form of PCAP files and performed data pre-processing and feature extraction. The results showed that the average accuracy of the model is 98.70%. The proposed model outperforms only the known attacks. Another work explained in [23]

proposed a hybrid model based on IDS placement for signature-based intrusion detection. Using the Cooja simulator, these researchers focused on two variants of denial of service (DoS) attacks. Even though the result showed the ability to reduce the false-positive rate of attack detection, the number of different attacks is much less.

Nowadays, the number of IoT devices is increasing, and the amount of data is rising. Most researchers have worked on DL techniques for the hybrid IDS for better results in recent years. In [24], proposed DL models were tested using the most recent CICIDS2017 datasets for distributed denial of service (DDoS) attack detection, which provided the most incredible accuracy of 97.16%. Additionally, proposed models were compared to ML methods. The use of DL algorithms for IoT cyber security is also subject to open research difficulties listed in this study. The proposed CNN-LSTM model specifically for DDoS attack detection was the major limitation. In another study, the authors proposed a hybrid model [25]. This study presents an enhanced genetic algorithm (GA) and deep belief network (DBN)-based intrusion detection model. For the intrusion detection model based on the DBN to attain a high detection rate with a small structure, facing various forms of attacks, the optimal number of hidden layers and neurons in each layer are created adaptively by numerous iterations of the GA. Finally, the model and methods were simulated and evaluated using the NSL KDD dataset. The experimental findings demonstrate that combining the improved intrusion detection model with DBN can significantly increase the rate at which intrusion threats are recognised while lowering the complexity of the neural network's structure. Another study explained in [26] a hybrid model both in model and detection methods. The researchers suggested convolutional-LSTM for anomaly-based and misuse-based attacks. The proposed model was tested with only the ISCX-UNB dataset was considered the main limitation. In [27], researchers proposed a hybrid model using the IoT-Bot dataset for the IoT. The presented model results show that the model has high accuracy and a low false alarm rate for both known and zero-day attacks. But the proposed model was tested with only the NSL KDD dataset.

In [28], the researchers proposed a two-stage hierarchical network intrusion detection approach (H2ID) using a multimodal autoencoder with soft output classifier and validated it with the BoT-IoT dataset. The result shows proposed M2-DAE was acquired better for simple anomaly detection, and the main drawback was that the proposed model was suitable for a particular dataset. In a novel hybrid intrusion detection presented in [29], the researchers implemented a hybrid model by combining two DL algorithms gated recurrent neural network (GRNN) and light convolutional neural network (LCNN) to gain accuracy with a minimum time overhead. With a cloud-based IoT network, no clustering-based anomaly detection was the major limitation of the work. Another method explained in [30] is a hybrid model that yielded 98% accuracy and was validated and compared with the recurrent neural network (RNN) model. The proposed model was focused on the binary classification of the attacks.

A hybrid intrusion detection method is explained in [31] for anomaly-based attacks. The proposed IDS was a combination of CNN and gated recurrent unit (GRU) evaluated with different datasets using accuracy, precision, and recall as the evaluation metrics. However, the proposed model is unable to identify the novel attacks. Another recent study presented in [32] hybrid model with deep random neural network (DRaNN) and multi-layer perceptron (MLP) identified 16 types of cyberattacks. The result has shown that the proposed model achieved high accuracy for both datasets. The researchers were done a comparison of performance metrics with other revolutionary DL models. Sahu et al. [33] proposed identifying malicious in IoT using the CNN-LSTM model. The dataset was used for the evaluation of twenty Raspberry Pi malicious devices. Researchers implemented it in a real-time environment in this experiment and compared other DL models. Another study proposed in [34] a hybrid model for signature and anomaly-based attacks with three stages: traffic filtering, pre-processing, and hybrid IDS. Even though the proposed model gives better results, the model tested with only one dataset.

In [35], a hybrid method with ML and DL algorithms was proposed and achieved 99% accuracy—the model test with four different datasets. The main limitation was that the datasets were not specific to IoT intrusions. Another method explained in [36] proposed a combination of auto-encoder (AE), and LSTM achieved 98% accuracy for the binary classification of the NSL KDD dataset. Another method proposed a hybrid model [37] AE-LSTM model with NSL-KDD dataset with 89%. The proposed model focused only the anomaly detection. The work explained in [38] is a hybrid model for the detection of DDoS attacks by using the CNN-LSTM model. The model yielded 99.2% accuracy. The limitation of the model is the

lack of architecture from serial to parallel and voting technology. Another work in [39] proposed a CNN-LSTM hybrid model for the industrial IoT (IIoT) with two datasets. Even though the model achieved 92.8%, researchers needed more synthetic data. In [40], they proposed a CNN-LSTM model to detect android malware. The results showed 95% accuracy, but the paper lacks the time sequence consequences of the observed behaviors in a dynamic analysis paradigm approach. Another study was carried out in [41] using LSTM and AE for intrusion detection. The model acquired 99.1 % accuracy but could not identify modern attacks.

In summary, although various researchers proposed many hybrid models, they tried to combine any two models and focused on the known attacks. Also, the proposed model is trained and tested with one dataset, outperforming better accuracy only for that particular dataset. Few researchers worked on hybrid methods based on IoT IDS detection methods. Some other works even used the hybrid methods but gave better accuracy for the binary classification. Another notable drawback of the existing models is that most researchers used common network intrusion datasets, not specifically IoT-related datasets. From the inferences of the literature review, to overcome the drawbacks of the existing model, the proposed model focuses on a hybrid intrusion detection technique for IoT. The proposed hybrid model uses a combination of the CNN-LSTM algorithm and a hybrid detection method of attack, both signature- and anomaly-based attacks. The model was tested with four recent datasets specific to IoT attacks. Even though randomly reducing the number of features in each dataset, the proposed model achieved better accuracy than existing models.

3.Methods

This section gives an idea of the methodology and structure of the proposed hybrid model. This proposed model combines two DL models, CNN-LSTM, to discover two kinds of attacks: anomaly-based and signature-based-the entire methodology of the proposed model is shown in *Figure 2*. The methodology includes data pre-processing, model training and testing and result evaluation regarding the accuracy, precision, recall and f1 score.

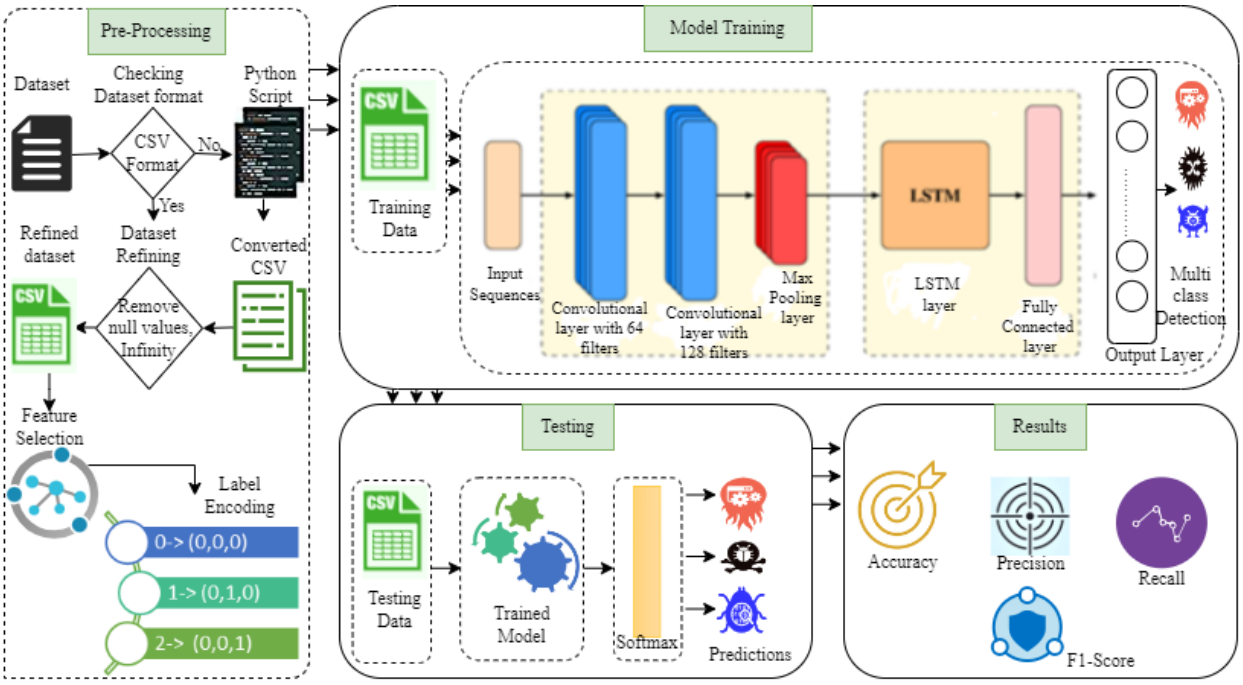

Figure 2 Overview of the proposed methodology

3.1Data pre-processing

Data pre-processing is an inevitable step before feeding the data into the model. Data pre-processing transforms raw data into an understandable format before providing the data. First, check whether the dataset files are in comma separated value (CSV) format in this process. Some datasets are in the PCAP file; such files are converted into CSV format using

Python. Most of the datasets are in several CSV files; after transforming them into CSV files, first appending all the available datasets into a single dataset. In the dataset performed, data pre-processing and data cleaning were performed. In the standardisation of column names, checked whether commas, infinity, null values, and any other special characters exist and removed such values and then generated the description of all the columns with count, mean, standard deviation, minimum values, 25%, 50%, 75%, and maximum values. Here, zeros replaced all the null values and generated the dataset head details. The next focus was on exploratory data analysis (EDA). It is a method to analyse data and frequently summarise its characteristics through the visual approach. To find out the highly correlated data generated pair plots of highly correlated data. Using the principal component analysis (PCA) method to remove the highly correlated data, standardisation and label encoding of the data were done subsequently. The entire datasets are split into 80/20 ratios as train and test datasets.

3.2Model training and testing

The current proposed model is trained and tested with the CNN-LSTM DL algorithm. The CNN-LSTM architecture is a DL model that CNNs and LSTM networks. The CNN component of the architecture is used for feature extraction, which is the process of identifying basic patterns and features in the input data. CNNs are particularly effective at identifying spatial patterns in images and other data types with a grid-like structure. In the context of the CNN-LSTM architecture, the CNN is typically applied to input data as two-dimensional matrices or three-dimensional tensors [42].

The output of the CNN is then fed into the LSTM component of the architecture. LSTM networks are a type of RNN well-suited for modelling sequential data. They are particularly effective at handling long-term dependencies often present in sequential data. In the CNN-LSTM architecture, the LSTM component is used to process the feature maps generated by the CNN. The LSTM can learn the temporal dependencies between the different feature maps, which can be important for understanding the context of the input data. The CNN-LSTM architecture is a powerful DL well suited for processing sequential data. By combining the strengths of CNNs and LSTMs, this architecture can extract features and model long-term dependencies in the input data [43].

Here is an overview of the layers in a typical CNN-LSTM architecture:

1. Convolutional Layers: The first few layers of the network are usually convolutional layers that extract features from the input data. The input to the network is typically an image, video frame, or sequence of audio signals. The convolutional layers use filters to scan the input data and produce feature maps that capture different input aspects.
2. Pooling Layers: After each convolutional layer, a pooling layer is often used to down sample the feature maps and reduce the dimensionality of the data. Max pooling is a common pooling operation that takes the maximum value in each local region of the feature map.
3. LSTM Layers: The output of the convolutional and pooling layers is fed into a set of LSTM layers. LSTM layers are a type of RNN that can process sequential data and capture long-term dependencies. The LSTM layers learn temporal patterns in the input data and generate a fixed-length feature vector summarising the input sequence.
4. Fully Connected Layers: The output of the LSTM layers is then passed through one or more fully connected layers, which transform the feature vector into the desired output format [44–46].

The mathematical formulation for a CNN-LSTM model can be expressed as follows: Let x be the input sequence of length T, where x_t is the t-th element of the sequence, and y be the output sequence of length T, where y_t is the t-th element of the sequence. First, the input sequence is fed into the CNN layers to extract relevant features from the input data. Let f(x_t) be the feature vector for the t-th input element, obtained after passing it through the CNN layers. Next, the feature vectors are fed into the LSTM layers to capture the temporal dependencies in the input sequence. Equation 1, the output of the LSTM layers at time t, denoted as h_t, is calculated as follows:

$$h_t = \text{LSTM}(f(x_t), h_{t-1}) \quad (1)$$

Where LSTM is the LSTM function, f(x_t) is the feature vector at time t, and h_{t-1} is the output of the LSTM layer at the previous time step. Equation 2, the output sequence is obtained by passing the LSTM outputs through a fully connected layer:

$$y_t = W_h h_t + b \quad (2)$$

Where W_h is the weight matrix, and b is the bias vector of the fully connected layer. The CNN-LSTM model can be trained end-to-end using back propagation through time (BPTT) to optimise the model parameters, including the CNN filters, LSTM weights, and fully connected layer weights.

In the proposed model, improving the detection accuracy has reduced the number of false positives (FP). After various experiments discovered the best parametric values, it has 10 epochs with 10 batch sizes. The model consists of convolution and LSTM layers with two max pooling layers, dense and dropout layers. The number of neurons varies in each layer, and used Adam optimiser in the input layer and dropout layer. The activation functions are ReLu for the input and Softmax for the output layer. For the measurement of the loss function used, Categorical Cross- Entropy. The complete description training and testing model is given in *Table 1*. The Keras Python framework with TensorFlow as a backend to implement the proposed model. The graphical processing unit (GPU) is further used to reduce performance lagging.

Table 1 Overall description of the proposed model

Proposed algorithm	Layers	Kernel/ neurons	Optimizer	Activation function	Loss function	Epochs	Batch size
CNN-LSTM	Conv	(128,64,32)	Adam	ReLU	Categorical Cross-Entropy	10	10
	Max Pooling	02	Adam	Softmax			
	LSTM	(128,64,32)					
	Dense	72					
	Dense	128					
	Dropout	0.1					
	Output	10					

3.3Dataset

From the literature review, it's evident that researchers have utilized various publicly available datasets. Similarly, this paper employs four recently released datasets that are publicly accessible. The datasets are CIC IDS 2018, IoT Network Intrusion Dataset, MQTT-IoT-IDS2020, and BoTNeTIoT-L01.

3.3.1CIC IDS 2018

The researchers utilized a publicly available dataset developed through a collaborative project between the Community Security Establishment (CSE) and the Canadian Institute of Cybersecurity (CIC). They employed a specific network topology for attack generation and conducted attacks from one or more external machines outside the target network. The dataset has collected over ten days and contained 16,000,000 instances. The framework includes 50 computers, and the attacking environment has five sections, 30 servers and 420 PCs as terminals. The entire dataset has available in two formats CSV and PCAP files. The CSV files are suitable for AI implementation, and PCAP files are helpful for the extraction of new features. CICFlowMeter-V3 was used for the extraction of PCAP files. The dataset contains seven categories of attacks with 80 features extracted from PCAP files [47, 48]. The dataset comprises seven attacks. The attack categories are shown in *Figure 3*.

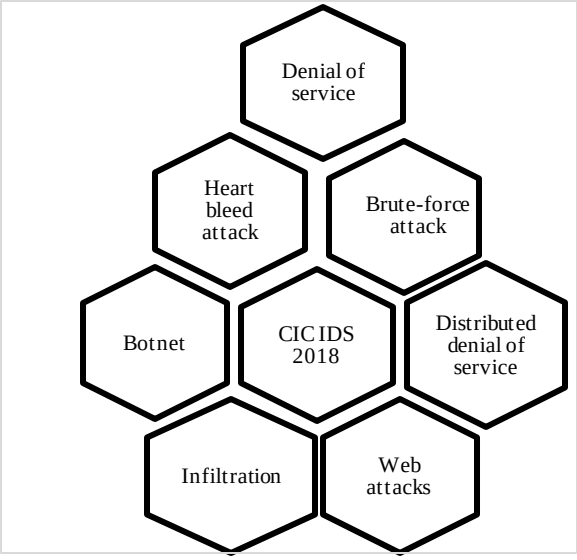

Figure 3 Attack categories of CIC IDS 2018

3.3.2IoT network intrusion dataset

The dataset was released on 27th September 2019 for academic purposes from the IoT environments. The attack taxonomy of the IoT network intrusion dataset is shown in *Figure 4*.

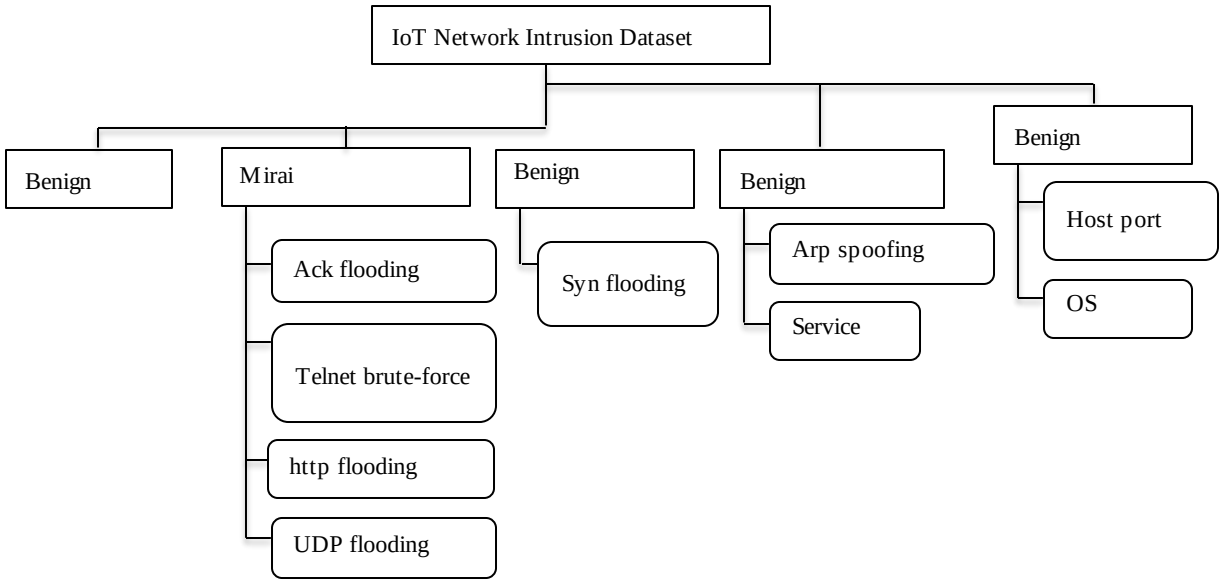

Figure 4 Classification of IoT network intrusion dataset

The dataset contains 42 raw network packet files (PCAP) with different network attacks in IoT. Two typical smart home devices, SKT NUGU (NU 100) and EZVIZ Wi-Fi Camera (C2C Mini O Plus 1080P), were used to create attack environments, including laptops or smartphones in the same wireless network. The packet capturing is done by monitoring the mode of the wireless network adaptor and wireless headers removed by Aircrack-ng. While simulating the attacks, the Nmap tool captured all the attack packets except the Mirai Botnet category. In the Mirai Botnet attack, all the attack packets were generated by laptops and manipulated, and it seemed as if IoT devices generated them. The entire dataset contains 42 PCAP files. The data was presented publicly in IEEE Dataport at https://doi.org/10.227/q70p-q449 [49]. The dataset includes four attacks and eight subcategories rather than Benign (normal) traffic. The PCAP conversion has been carried out using Python and consists of 83 features.

3.3.3MQTT-IoT-IDS2020

It is a publicly available dataset generated by MQTT sensors simulation. The simulation framework contains twelve sensors, a broker, a simulated camera, and an attacker. The dataset was released on 23rd June 2020 and is publically available in PCAP and CSV formats in IEEE Dataport [50]. The researchers created five traffic flows from the simulated attack scenario, including regular traffic and four attack types. The dataset has included five pcap files, such as regular. pcap, sparta.pcap, scan_A.pcap, mqtt_bruteforce.pcap and scan_sU.pcap. The files contain normal, Sparta SSH brute-force, aggressive scan, MQTT brute-force, and user datagram protocol (UDP) scan. The features are extracted from each PCAP file, such as packet features, unidirectional flow features, and bidirectional flow features [50]. *Figure 5* shows the overview of the MQTT IoT IDS 2020 dataset.

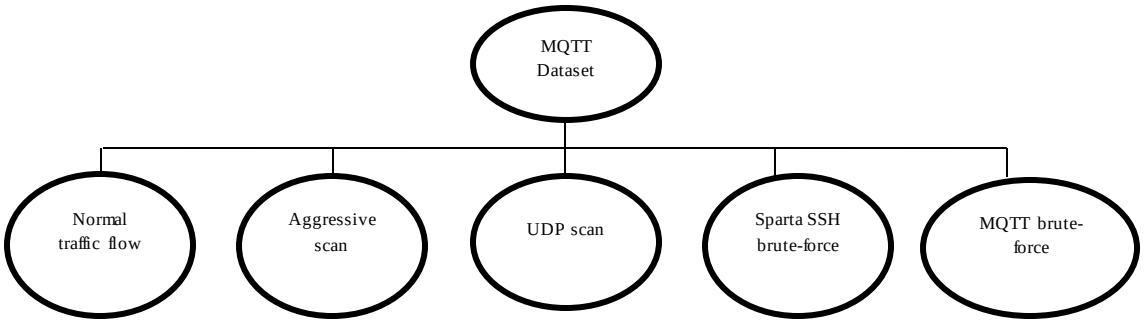

Figure 5 MQTT IoT IDS 2020 dataset classification

3.3.4BoTNeTIoT-L01

This is the most recent dataset released on 28th April 2021 and accessed publicly from the University of New South Wales (UNSW) Sydney website. The dataset created by nine IoT devices traffic was sniffed using a central switch using Wireshark from the local network. The dataset includes two Botnet attacks: Mirai and Gafgyt, and their subcategories. The overall view of the dataset has shown in the figure. The dataset contains 23 statistically engineered features extracted from .PCAP files [51, 52]. *Figure 6* shows the overview of the BoTNeTIoT-L01 dataset.

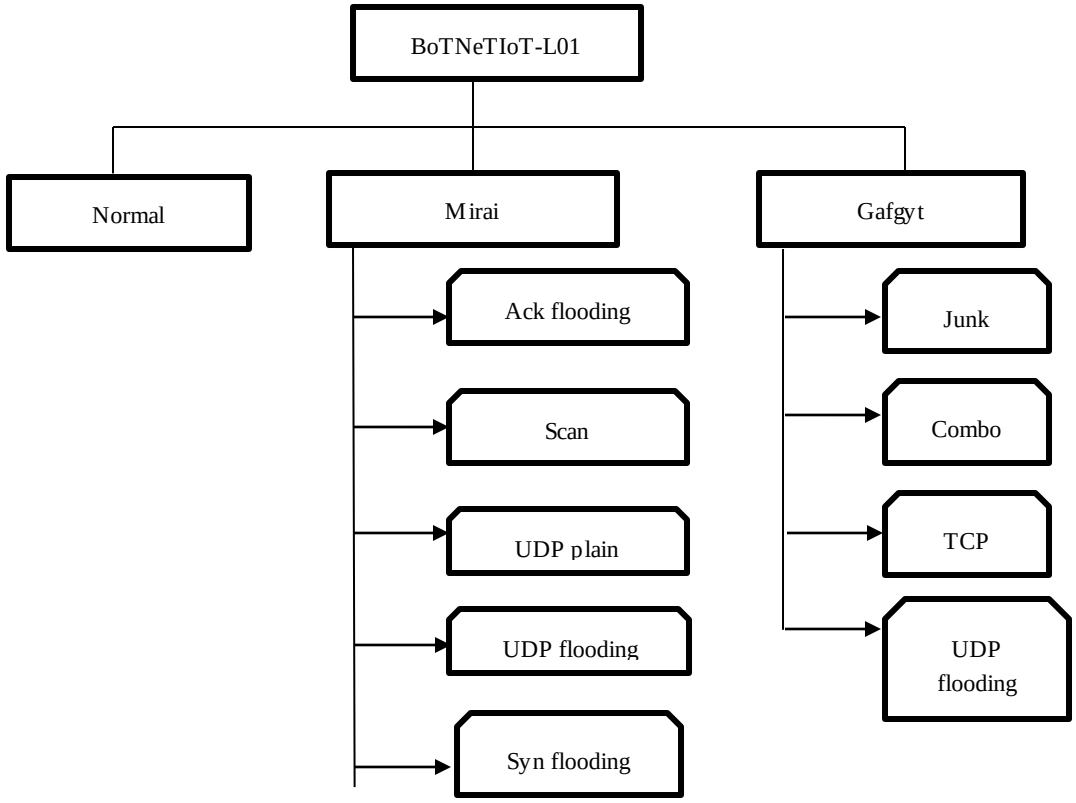

Figure 6 Overview of BoTNeTIoT-L01 dataset

3.3.5 Evaluation metrics

Accuracy, precision, recall, and F1 score are all metrics used to evaluate the performance of a classification model. To calculate the evaluation metrics, the need to identify true positive (TP), true negative (TN), FP, and false negative required (FN) [41, 42, 44, 53]. TP attacks in the dataset are correctly predicted as an attack, while TN is the regular traffic rightly expected as normal traffic. FP are that routine traffic requests are incorrectly predicted as attacks, while FN attacks are incorrectly predicted as regular traffic. Accuracy is the most straightforward metric for classification models; it measures the proportion of correctly predicted labels, both TPs and TNs, out of all the samples in the dataset. Precision measures the proportion of TP predictions out of all the positive predictions made by the model. A high precision score indicates that the model has a low FP rate and correctly identifies positive samples. Recall measures the proportion of TP predictions from all the actual positive samples in the dataset. A high recall score indicates that the model has a low FN rate and correctly identifies all positive samples. The F1 score combines precision and recall, providing a single measure of the model's performance. A high F1 score indicates that the model has a good balance between precision and recall, which means it correctly identifies both positive and negative samples. While accuracy measures the model's overall performance, precision, recall, and F1 score provide more specific insights into the model's ability to identify positive samples correctly. The Equations from 3 to 6 are the mathematical formulation of these metrics.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (4)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (5)$$

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (6)$$

4. Results

This section discusses the results and discussion. The standard evaluation metrics such as accuracy, precision, recall and F1 score are used to attain the model's performance. The proposed model was evaluated with four IoT datasets: CIC IDS 2018, IoT network intrusion dataset, MQTT-IoT-IDS2020, and BoTNeTIoT-L01. Each dataset has a different number of features. To evaluate the model performance, we have done the reduction of features. In DL the model often performs feature selection automatically during the training process. The model learns to extract relevant features from the raw input data and use them to make accurate predictions. This is one of the critical advantages of DL as it can handle high-dimensional and complex data without manual feature engineering. The minimum of features in all four IoT datasets is ten. The feature reduction aims to reduce the computational complexity and identify the model's accuracy if the model has the minimum number of features. *Figures 7* to *10* below show different dataset results with a reduction of several features.

Figure 7 shows the result of the CIC IDS 2018 dataset. The entire dataset has 80 features, reduced to 75 by removing highly correlated data using PCA and the dataset reduced into 50, 30 and 10 randomly. This dataset with 75 features acquired an accuracy of 99.89%, a precision of 99.86%, a recall of 99.78%, and an f1 score of 99.81%. When reduced into 50 features, accuracy is 99.86%, precision 99.85%, recall 99.68%, and f1 score 99.76%. The dataset was reduced into 30 with an accuracy of 99.84%, precision of 99.84%, recall of 99.64% and f1 score of 99.73% and with 10 features yielded accuracy, precision, and recall and f1 score of 99.83%, 99.82%, 99.60%, and 99.70% respectively.

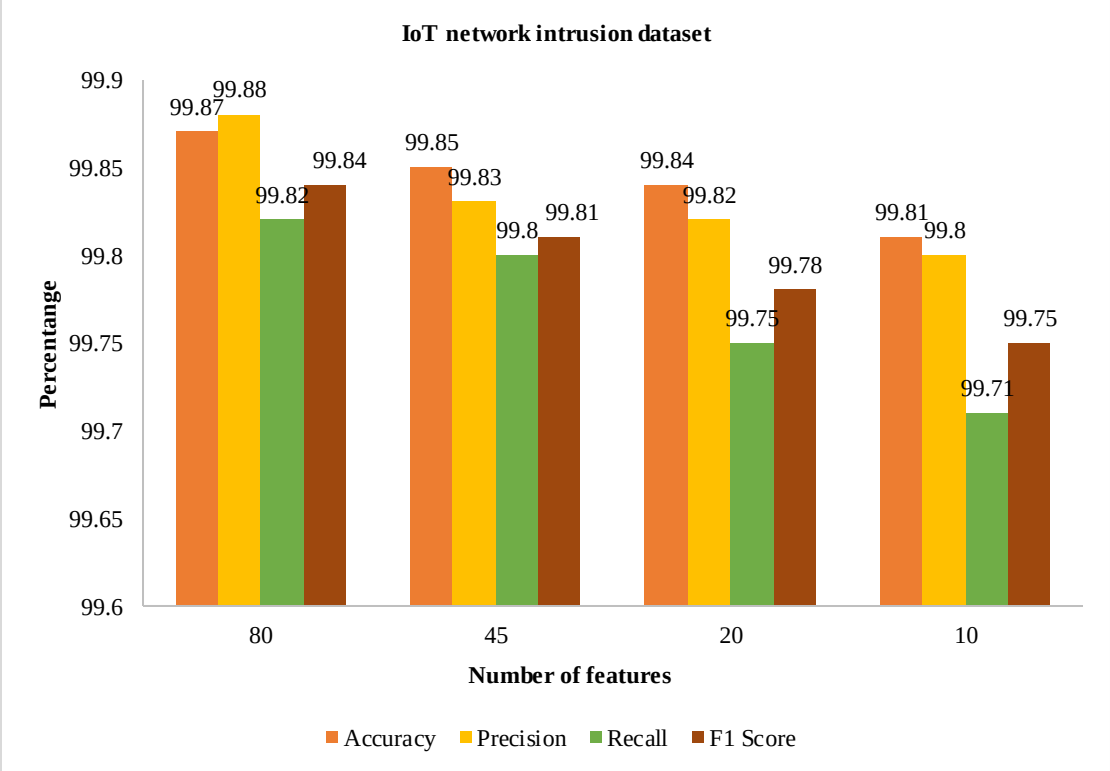

Figure 7 Results of IoT network intrusion dataset

Figure 8 demonstrates the values of evaluation metrics of the IoT network intrusion dataset. The total number of features of the dataset was 83. This total number of features is reduced to 80, 45, 20, and 10, and in the case of 80 features, obtained 99.84% accuracy, 99.88% precision, 99.82% recall and 99.89% F1 score. While reducing to 45 features resulted in an accuracy of 99.81%, precision of 99.83, recall of 99.80%, and 99.81% F1 score. Considering 20 features gained an accuracy of 99.80%, precision of 99.82%, recall of 99.75%, and F1 score of 99.78%. The dataset yielded

99.80% accuracy, 99.79% precision, 99.70% recall, and 99.74% F1 score for the 10 features.

Figure 9 gives the details of the MQTT IoT IDS2020 dataset. The total number of features is 30. After that, the model was trained and tested with full features, reduced to 10. When tested with 30 features resulted in 99.88% accuracy, 99.90% precision, 99.64% recall, and 99.76% F1 score—at the same time, considering 10 features acquired 99.86% accuracy, 99.87% precision, 99.55% recall, and 99.70% F1 score.

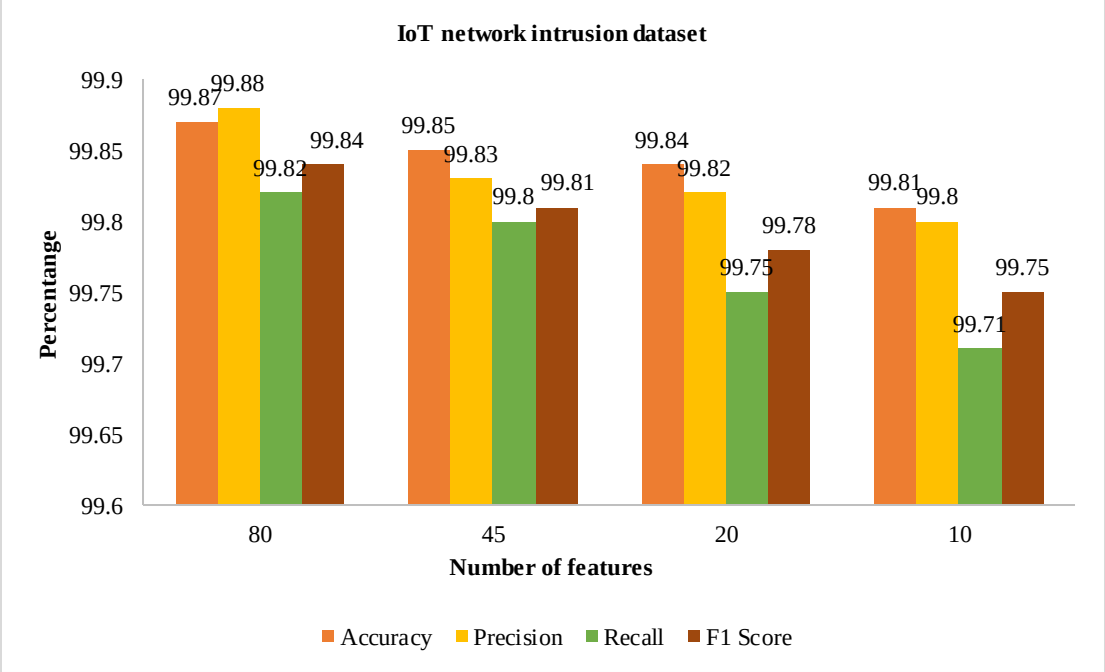

Figure 8 Results of IoT network intrusion dataset

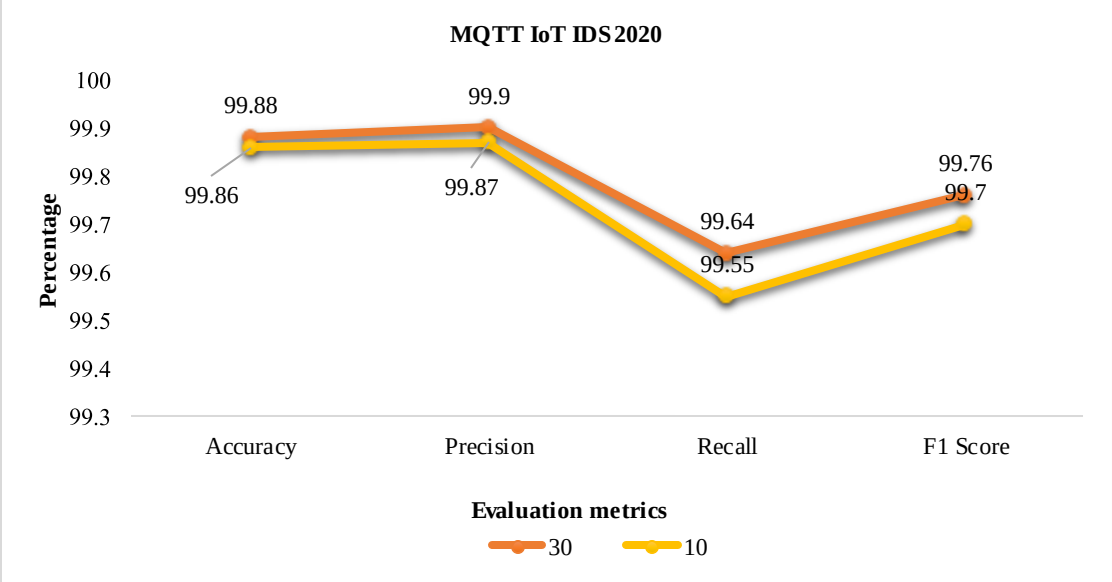

Figure 9 Results of MQTT IoT IDS dataset

Figure 10 illustrates the BoTNeTIoT-L01 dataset with a total number of features are 23 after that, reduced to 10. This is the smallest dataset based on the number of features. The dataset with 23 features yielded 99.86%

accuracy, 99.87% precision, 99.80% recall, and 99.83% F1 score. When the dataset has 10 features obtained, accuracy is 99.85%, precision 99.85%, recall 99.79%, and F1 score 99.81%.

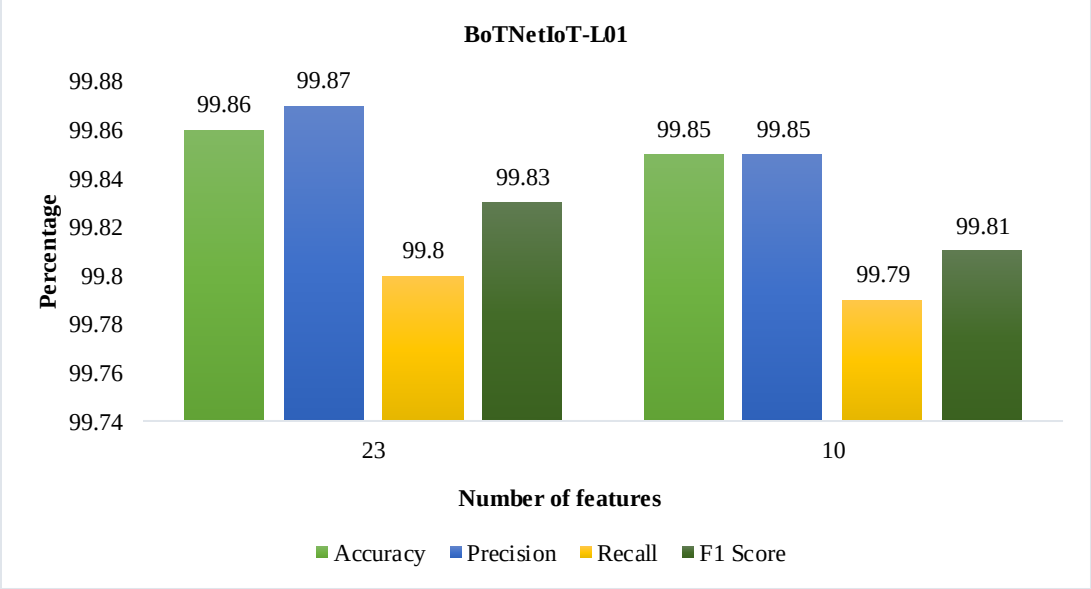

Figure 10 The values of the BoTNeTIoT-L01 dataset

5. Discussion

The effectiveness of the proposed AS-CL approach using relevant parameters against state-of-the-art literature for attack detection is shown in *Table 2*. It shows that the proposed model acquired higher accuracy than other alternative models. In this comparison, *Table 2* gives a detailed review of the last five years carried out in IoT IDS. The table specifies the year, references to the work, algorithms used for the implementation, types of detection methods, name of the dataset used, number of features in each dataset, and the accuracy of the model. The main highlights of the model are:

- This research focused on identifying anomaly-based and signature-based attacks rather than particular types of attacks in the IoT environment.
- Most of the hybrid models in an IoT IDS are just a combination of two models and perform the attack detection, but in the case of AS-CL IDS concentrated on hybrid detection method and security threats in IoT and is also a combination of real-time data analysis models CNN and LSTM.
- For the analysis of the AS-CL IDS model the latest four datasets specially intended for IoT IDS. Each dataset varied the features randomly to check for accuracy and obtained a minimum accuracy of 99.81% and a maximum time of 311Secs for the identification attacks.
- The proposed model is well for the identification zero-day attacks.
- This research, worked on several ML and DL model and evaluated with state-of-the-art literature for attack detection and identified most accurate models for attack detection.
- This research work provides an improved solution to be cautious about the possible attacks and hence helps us to take remedial precautions.
- The proposed model shows improved accuracy compared to the existing hybrid IoT IDS. Evaluated the effectiveness of the proposed model using relevant parameters against state-of-the-art literature for attack detection.

While the proposed model surpassed state-of-the-art counterparts in the literature, it has its limitations. It was tested with the four most recent datasets and achieved better accuracy. Future studies could incorporate real-time data. The proposed hybrid model, which currently utilizes CNN-LSTM, could be enhanced by integrating the advantages of FCRNN and modified CRNN in future iterations.

A complete list of abbreviations is shown in *Appendix I.*

Table 2 Comparison of the proposed model with the existing models

S. No.	Year	References	Algorithm	Detection method	Features	Dataset	Accuracy
1	2019	[26]	Spark ML + Conv-LSTM	*	8	ISCX-UNB	97.29%
2	2019	[25]	GA + DBN	*	41	NSL KDD	≅ 99.00%
3	2019	[24]	CNN + LSTM	Signature-based	**	CIC IDS2017	97.16%
4	2020	[29]	LCNN + GRNN	*	41	NSL KDD	≅ 90.00%
5	2020	[30]	CNN + LSTM	*	**	UNSW-NB15	98.60%
6	2020	[54]	CNN + LSTM	*	**	KF-ISAC(Real-time)	98.07%
						CSIC-2010	91.54%
						CICIDS 2017	93.00%
7	2020	[55]	LSTM + D_MCTS-T	*	80	CICIDS 2017	81.23%
8	2020	[56]	CNN + LSTM	*	80	CICIDS 2017	98.67%
9	2021	[34]	LightNet + Deep Q-learning	Anomaly and Signature-based	41	NSL KDD	96.90%
10	2021	[33]	CNN + LSTM	Signature-based	**	Self-generated	96.00%
11	2021	[32]	DRaNN + MLP	*	13	DS2OS dataset	98.00%
					49	UNSW-NB15	99.00%
12	2021	[57]	CAE+OCSVM	*	41	NSL KDD	91.58%
13	2021	[31]	CNN + GRU	Anomaly-based	**	BoT IoT	99.92%
						IoT Network intrusion	96.77%
						MQTT-IoT-IDS2020	99.91%
						IoT-23	99.88%
14	2022	[36]	AE + LSTM	*	41	NSL KDD	98.88%
15	2022	[37]	AE + LSTM	*	**	NSL KDD	89.00%
16	2021	[57]	RTIDS	*	81	CICIDS 2017	99.35%
					87	CIC DDoS2019	98.58%
17	2022	[58]	CNN + LSTM	*	41	KDD CUP'99	99.95%
					41	NSL KDD	99.53%
18	2022	[59]	Recurrent DL	*	41	KDD CUP'99	99.00%
					49	UNSW-NB15	99.00%
					19	WSN-DS	98.00%
					81	CICIDS 2017	99.00%
19	2023	[38]	CNN + LSTM	Signature-based	41	NSL KDD	99.20%
20	2023	[39]	CNN + LSTM	*	49	UNSW-NB15	92.90%
					68	X-IIoTID	99.80%
21	2023	[42]	CNN + RNN	*	83	CSE-CIC-IDS2018	98.84%
			CNN + LSTM	*	83	CSE-CIC-IDS2018	98.85%
22	2019	[41]	LSTM + AE	*	81	CICIDS 2017	99.99%
					76	CSE-CICIDS 2018	99.10%
23	2023	[40]	CNN + LSTM	*	**	Self-generated	95.30%
24	**This Paper**	***	CNN + LSTM	Anomaly and Signature-based	75	CICIDS 2018	99.89%
					50	CICIDS 2018	99.86%
					30	CICIDS 2018	99.84%
					10	CICIDS 2018	99.83%
					80	IoT Network Intrusion	99.87%
					45	IoT Network Intrusion	99.85%
					20	IoT Network Intrusion	99.84%
					10	IoT Network	99.81%

S. No.	Year	References	Algorithm	Detection method	Features	Dataset		Accuracy
						Intrusion		
					30	MQTT IDS2020	IoT	99.88%
					10	MQTT IDS2020	IoT	99.86%
					23	BoTNetIoT-L01		99.86%
					10	BoTNetIoT-L01		99.85

* Detection method is not specified, ** The number of features is not specified, *** References to this paper

6. Conclusion and future work

The rapid escalation of massive IoT applications has significantly impacted modern society. The foremost concern regarding these applications is security, particularly due to the enormous volume of data generated every second and its utilization. These applications are susceptible to a range of attacks, potentially leading to catastrophic outcomes if not pre-emptively managed and controlled. As the IoT domain expands, concerns about data security threats grow exponentially due to factors like device vulnerability to malware, denial-of-service attacks, and intrusion attempts. To avert such incidents, more robust precautions are necessary, urging system developers and IoT device manufacturers to refine their security mitigation strategies. Recognizing and addressing all potential threats and vulnerabilities specific to IoT infrastructures is crucial. There is a pressing need for extensive research on security attacks to mitigate potential dangers. Identified security challenges must be addressed to prevent them effectively. Future research should focus on addressing security challenges in IoT-based environments, enhancing the reliability of IoT applications for both suppliers and consumers. In this context, AI plays a pivotal role. AI facilitates the integration of machine intelligence, akin to human intelligence. ML and DL, as subsets of AI, offer distinct advantages. ML is suited for small data with feature engineering, training, and validation processes, while DL can automate feature extraction and overcome many limitations of ML. Recent research has been focusing on a hybrid approach to intrusion detection, combining two DL techniques with a detection method for a robust solution. Given the vast connectivity and heterogeneity of IoT devices, which generate large amounts of data and various cyber threats, security must be the primary focus. The proposed model in this research centers on a hybrid IDS with a DL method. Unlike other models that combine two DL models, this one adopts a hybrid detection method specifically for IoT intrusion detection. Beyond model creation and implementation, this work involves data preprocessing and analysis using various methods. It includes working with different datasets and reducing the number of features in each dataset. Looking forward, the research anticipates the use of real-time datasets and the strategic placement of IoT IDS for further implementation and enhancement.

Acknowledgment

None.

Conflicts of interest

The authors have no conflicts of interest to declare.

Author's contribution statement

Jinsi Jose: Conceived and designed the study, conducted experiments, collected and analyzed data, and drafted the manuscript. **Deepa V. Jose:** Guided the research direction, and reviewed the manuscript. All authors have read and approved the final manuscript.

References

[1] Hussain A, Sharif H, Rehman F, Kirn H, Sadiq A, Khan MS, et al. A systematic review of intrusion detection systems in internet of things using ML and DL. In 4th international conference on computing, mathematics and engineering technologies (iCoMET) 2023 (pp. 1-5). IEEE.

[2] Bu T, Huang Z, Zhang K, Wang Y, Song H, Zhou J, et al. Task scheduling in the internet of things: challenges, solutions, and future trends. Cluster Computing. 2023:1-30.

[3] Lu Y, Da XL. Internet of things (IoT) cybersecurity research: a review of current research topics. IEEE Internet of Things Journal. 2018; 6(2):2103-15.

[4] https://www.cisco.com/c/en/us/solutions/executive-perspectives/annual-internet-report/airhighlights.html. Accessed 17 March 2022.

[5] Jose J, Jose DV. The internet of things architectures and use cases. In enterprise digital transformation 2022 (pp. 101-25). Auerbach Publications.

[6] Lohiya R, Thakkar A. Application domains, evaluation data sets, and research challenges of IoT: a systematic review. IEEE Internet of Things Journal. 2020; 8(11):8774-98.

[7] Kaur B, Dadkhah S, Shoeleh F, Neto EC, Xiong P, Iqbal S, et al. Internet of things (IoT) security dataset evolution: challenges and future directions. Internet of Things. 2023:100780.

[8] Aljanabi M, Ismail MA, Ali AH. Intrusion detection systems, issues, challenges, and needs. International Journal of Computational Intelligence Systems. 2021; 14(1):560-71.

[9] Khraisat A, Alazab A. A critical review of intrusion detection systems in the internet of things: techniques, deployment strategy, validation strategy, attacks, public datasets and challenges. Cybersecurity. 2021; 4:1-27.
[10] Thakkar A, Lohiya R. A survey on intrusion detection system: feature selection, model, performance measures, application perspective, challenges, and future research directions. Artificial Intelligence Review. 2022; 55(1):453-563.
[11] Malhotra P, Singh Y, Anand P, Bangotra DK, Singh PK, Hong WC. Internet of things: evolution, concerns and security challenges. Sensors. 2021; 21(5):1-33.
[12] Hanif S, Ilyas T, Zeeshan M. Intrusion detection in IoT using artificial neural networks on UNSW-15 dataset. In 16th international conference on smart cities: improving quality of life using ICT & IoT and AI 2019 (pp. 152-6). IEEE.
[13] Mohamed E. The relation of artificial intelligence with internet of things: a survey. Journal of Cybersecurity and Information Management. 2020; 1(1):30-4.
[14] Kuzlu M, Fair C, Guler O. Role of artificial intelligence in the internet of things (IoT) cybersecurity. Discover Internet of Things. 2021; 1:1-4.
[15] Awotunde JB, Misra S. Feature extraction and artificial intelligence-based intrusion detection model for a secure internet of things networks. In illumination of artificial intelligence in cybersecurity and forensics 2022 (pp. 21-44). Cham: Springer International Publishing.
[16] Al-garadi MA, Mohamed A, Al-ali AK, Du X, Ali I, Guizani M. A survey of machine and deep learning methods for internet of things (IoT) security. IEEE Communications Surveys & Tutorials. 2020; 22(3):1646-85.
[17] Anushiya R, Lavanya VS. A comparative study on intrusion detection systems for secured communication in internet of things. ICTACT Journal on Communication Technology. 2021; 6948:2527-37.
[18] Baich M, Hamim T, Sael N, Chemlal Y. Machine learning for IoT based networks intrusion detection: a comparative study. Procedia Computer Science. 2022; 215:742-51.
[19] Tsimenidis S, Lagkas T, Rantos K. Deep learning in IoT intrusion detection. Journal of Network and Systems Management. 2022; 30:1-40.
[20] Bostani H, Sheikhan M. Hybrid of anomaly-based and specification-based IDS for internet of things using unsupervised OPF based on MapReduce approach. Computer Communications. 2017; 98:52-71.
[21] Kumari VV, Varma PR. A semi-supervised intrusion detection system using active learning SVM and fuzzy c-means clustering. In international conference on I-SMAC (IoT in social, mobile, analytics and cloud) 2017 (pp. 481-5). IEEE.
[22] Bhatt P, Morais A. HADS: hybrid anomaly detection system for IoT environments. In international conference on internet of things, embedded systems and communications 2018 (pp. 191-6). IEEE.
[23] Ioulianou P, Vasilakis V, Moscholios I, Logothetis M. A signature-based intrusion detection system for the internet of things. Information and Communication Technology Form. 2018:1-7.
[24] Roopak M, Tian GY, Chambers J. Deep learning models for cyber security in IoT networks. In 9th annual computing and communication workshop and conference 2019 (pp. 452-7). IEEE.
[25] Zhang Y, Li P, Wang X. Intrusion detection for IoT based on improved genetic algorithm and deep belief network. IEEE Access. 2019; 7:31711-22.
[26] Khan MA, Karim MR, Kim Y. A scalable and hybrid intrusion detection system based on the convolutional-LSTM network. Symmetry. 2019; 11(4):1-14.
[27] Khraisat A, Gondal I, Vamplew P, Kamruzzaman J, Alazab A. A novel ensemble of hybrid intrusion detection system for detecting internet of things attacks. Electronics. 2019; 8(11):1-18.
[28] Bovenzi G, Aceto G, Ciuonzo D, Persico V, Pescapé A. A hierarchical hybrid intrusion detection approach in IoT scenarios. In GLOBECOM global communications conference 2020 (pp. 1-7). IEEE.
[29] Ramadan RA, Yadav K. A novel hybrid intrusion detection system (IDS) for the detection of internet of things (IoT) network attacks. Annals of Emerging Technologies in Computing (AETiC). 2020; 4(5):61-74.
[30] Smys S, Basar A, Wang H. Hybrid intrusion detection system for internet of things (IoT). Journal of ISMAC. 2020; 2(4):190-9.
[31] Ullah I, Ullah A, Sajjad M. Towards a hybrid deep learning model for anomalous activities detection in internet of things networks. IoT. 2021; 2(3):428-48.
[32] Huma ZE, Latif S, Ahmad J, Idrees Z, Ibrar A, Zou Z, et al. A hybrid deep random neural network for cyberattack detection in the industrial internet of things. IEEE Access. 2021; 9:55595-605.
[33] Sahu AK, Sharma S, Tanveer M, Raja R. Internet of things attack detection using hybrid deep learning model. Computer Communications. 2021; 176:146-54.
[34] Otoum Y, Nayak A. As-ids: anomaly and signature based ids for the internet of things. Journal of Network and Systems Management. 2021; 29:1-26.
[35] Ravi V, Chaganti R, Alazab M. Recurrent deep learning-based feature fusion ensemble meta-classifier approach for intelligent network intrusion detection system. Computers and Electrical Engineering. 2022; 102:108156.
[36] Mahmoud M, Kasem M, Abdallah A, Kang HS. Ae-LSTM: autoencoder with LSTM-based intrusion detection in IoT. In international telecommunications conference 2022 (pp. 1-6). IEEE.
[37] Mushtaq E, Zameer A, Umer M, Abbasi AA. A two-stage intrusion detection system with auto-encoder and LSTMs. Applied Soft Computing. 2022; 121:108768.
[38] Issa AS, Albayrak Z. Ddos attack intrusion detection system based on hybridization of CNN and LSTM. Acta Polytechnica Hungarica. 2023; 20(2):1-9.
[39] Altunay HC, Albayrak Z. A hybrid CNN+LSTM based intrusion detection system for industrial IoT networks.

Engineering Science and Technology, an International Journal. 2023; 38:101322.
[40] Calik BE, Koray SO, Dogan B. Deep learning based malware detection for android systems: a comparative analysis. Tehnički Vjesnik. 2023; 30(3):787-96.
[41] Khan FA, Gumaei A, Derhab A, Hussain A. A novel two-stage deep learning model for efficient network intrusion detection. IEEE Access. 2019; 7:30373-85.
[42] Wang YC, Houng YC, Chen HX, Tseng SM. Network anomaly intrusion detection based on deep learning approach. Sensors. 2023; 23(4):1-21.
[43] Xu J, He Z, Zhang Y. CNN-LSTM combined network for IoT enabled fall detection applications. In journal of physics: conference series 2019 (pp. 1-6). IOP Publishing.
[44] Praanna K, Sruthi S, Kalyani K, Tejaswi AS. A CNN-LSTM model for intrusion detection system from high dimensional data. Journal of Information and Computational Science. 2020; 10(3):1362-70.
[45] Alferaidi A, Yadav K, Alharbi Y, Razmjooy N, Viriyasitavat W, Gulati K, et al. Distributed deep CNN-LSTM model for intrusion detection method in IoT-based vehicles. Mathematical Problems in Engineering. 2022; 2022:1-8.
[46] Alkahtani H, Aldhyani TH. Botnet attack detection by using CNN-LSTM model for internet of things applications. Security and Communication Networks. 2021; 2021:1-23.
[47] https://www.unb.ca/cic/datasets/ids-2018.html. Accessed 28 February 2020.
[48] Khan MA. HCRNNIDS: hybrid convolutional recurrent neural network-based network intrusion detection system. Processes. 2021; 9(5):1-14.
[49] https://ieee-dataport.org/open-access/iot-network-intrusion-dataset. Accessed 16 November 2020.
[50] https://ieee-dataport.org/open-access/mqtt-iot-ids2020-mqtt- internet-things-intrusion-detection-dataset. Accessed 16 November 2020.
[51] https://research.unsw.edu.au/projects/bot-iot-dataset. Accessed 21 March 2021.
[52] Alhowaide A, Alsmadi I, Tang J. Towards the design of real-time autonomous IoT NIDS. Cluster Computing. 2021:1-4.
[53] Vujović Ž. Classification model evaluation metrics. International Journal of Advanced Computer Science and Applications. 2021; 12(6):599-606.
[54] Kim A, Park M, Lee DH. AI-IDS: application of deep learning to real-time web intrusion detection. IEEE Access. 2020; 8:70245-61.
[55] Zhang X, Zhou Y, Pei S, Zhuge J, Chen J. Adversarial examples detection for XSS attacks based on generative adversarial networks. IEEE Access. 2020; 8:10989-96.
[56] Sun P, Liu P, Li Q, Liu C, Lu X, Hao R, et al. DL-IDS: extracting features using CNN-LSTM hybrid network for intrusion detection system. Security and Communication Networks. 2020; 2020:1-11.
[57] Binbusayyis A, Vaiyapuri T. Unsupervised deep learning approach for network intrusion detection combining convolutional autoencoder and one-class SVM. Applied Intelligence. 2021; 51(10):7094-108.
[58] Wu Z, Zhang H, Wang P, Sun Z. RTIDS: a robust transformer-based approach for intrusion detection system. IEEE Access. 2022; 10:64375-87.
[59] Umair MB, Iqbal Z, Faraz MA, Khan MA, Zhang YD, Razmjooy N, et al. A network intrusion detection system using hybrid multilayer deep learning model. Big Data. 2022.

Jinsi Jose completed her post-graduation in Master of Computer Applications (MCA) in the year 2017 from CHRIST University, Bangalore, India. Currently, she is a research scholar in the same university working in the domain of Network Security in IoT, Department of Computer Science. Her areas of interest include the Internet of Things, Intrusion Detection, Artificial Intelligence, Machine Learning, and Deep Learning.
Email: jinsi.jose@res.christuniversity.in

Deepa V. Jose holds a PhD in Computer Science from CHRIST University, India. Her area of research interest includes Wireless Sensor Networks, Security in Internet of Things, Block Chain Technology, NLP and Cyber Forensics. She has authored several research papers in national and international levels. She is a reviewer for leading computer science journals such as Peer to Peer Networks.
Email: deepa.v.jose@christuniversity.in

Appendix I

S. No.	Abbreviation	Description
1	AE	Auto-Encoder
2	AI	Artificial Intelligence
3	AS-CL IDS	Anomaly and Signature-based CNN-LSTM Intrusion Detection System
4	ASVM	Advanced Support Vector Machine
5	BPTT	Back Propagation Through Time
6	CAE	Convolutional Auto-Encoder
7	CAGR	Compound Annual Growth Rate
8	CIC	Canadian Institute of Cybersecurity
9	CNN	Convolutional Neural Network
10	CNN-LSTM	Convolutional Neural Network – Long Short Term Memory
11	Conv-LSTM	Convolutional - LSTM
12	CSV	Comma Separated Value
13	CSE	Community Security Establishment
14	DBN	Deep Belief Network
15	DRaNN	Deep Random Neural Network
16	DL	Deep Learning
17	DoS	Denial of Service
18	DDoS	Distributed Denial of Service
19	EDA	Exploratory Data Analysis
20	FCM	Fuzzy C-Means
21	FN	False Negative
22	FP	False Positive
23	GA	Genetic Algorithm

24	GPU	Graphical Processing Unit
25	GRNN	Gated Recurrent Neural Network
26	GRU	Gated Recurrent Unit
27	HIDS	Host-based Intrusion Detection System
28	H2ID	Two-stage Hierarchical Network Intrusion Detection Approach
29	IDS	Intrusion Detection System
30	IIoT	Industrial IoT
31	IP	Internet Protocol
32	IoT	Internet of Things
33	LCNN	Light Convolutional Neural Network
34	LSTM	Long Short-Term Memory
35	MIT	Massachusetts Institute of Technology
36	ML	Machine Learning
37	MLP	Multi-Layer Perceptron
38	NIDS	Network-based Intrusion Detection System
39	NLP	Natural Language Processing
40	OPF	Optimum-Path Forest
41	PCA	Principal Component Analysis
42	RNN	Recurrent Neural Network
43	TN	True Negative
44	TP	True Positive
45	UDP	User Datagram Protocol
46	UNSW	University of New South Wales