AN INTELLIGENT CYBERSECURITY MODEL FOR IOT NETWORKS USING ADVERSARIALLY REGULARIZED PARALLEL DEEP TRANSFER NETWORK

LAVANYA VEMULAPALLI [1*], P.CHANDRA SEKHAR [2]

[1*]Research Scholar, Department of Computer Science and Engineering
Gitam University, Visakhapatnam, A.P, INDIA
[2]Associate Professor, Department of Computer Science and Engineering
Gitam University, Visakhapatnam, A.P, INDIA
[1]avanyavemulapalli@gmail.com[1], chandoo.potala@gmail.com[2]

ABSTRACT

The increasing number of technology and information has led to significant security problems, which have increased the significance of the creation of sophisticated intrusion detection systems (IDS). Big data may be handled through deep learning, which has demonstrated excellent performance in several disciplines. As a result, security professionals want to use deep learning in intrusion detection systems. This subject has been the subject of a great deal of research, which has produced a wide variety of methods. To categorise network traffic, the majority of these methods employ predetermined characteristics that have been retrieved by specialists.However most of the methods are incapable of effectively classifying the intrusion. To solve this problem, this research introduced a novel deep learning method to effectively classify the intrusion with lesser delay. Using a multi-level clustering strategy that includes the k-nearest neighbour, reverse k-nearest neighbour and k-means clustering approach, the undesirable information and outliers in the data may be eliminated. The Adversarially Regularised Parallel Deep Transfer Network (AR-PDTN) is then used to classify the IoT data intrusion in order to identify any network irregularities. The parameters of the proposed neural network model will then be appropriately adjusted using Alpine skiing optimisation. Five independent datasets like CICIDS2017, NSL-KDD, KDDCup99, UNSW-NB15 and BOT-IOT were utilised to categorise the incursion. In the results section, the accuracy, precision, recall, and f1-score of the proposed model are contrasted with those of several other current models.The proposed model achieved 99.3% accuracy in UNSW-NB15 dataset, 99.7% in CICIDS2017, 99.1% in NSL-KDD dataset, 99.8% using BoT-IoT dataset and 94.5% in KDDCup99 dataset, respectively.The performance of the proposed framework was the best when compared with each dataset. In this research, an Adversarially Regularized parallel deep transfer Network (AR-PDTN) is introduced to classify the intrusion from IoT data. The unwanted information and outliers data could be reduced by using multi-level clustering approach concerning k-nearest neighbour, reverse k-nearest neighbour and k-means clustering approach

Keywords: *Linear Time Complexity, Adjacent Neighbours, High Density Points, Wasserstein Auto Encoder, Critic Unit, Adversary Auto Encoder, Down Sampling, Convolutional Filter Tube*

1. INTRODUCTION

The expansion of internet facilitated the way for online monitoring or data gathering technology through Internet of Things (IoT) [1]. Due to the increased advantage of IoT devices, the usage of the particular technology is increasing. However, the data collected by those technology increases proportional to the IoT usage. IoT is made of number of components [2] such as sensors, protocols, and actuators for data transmission; it transmit the data through layering process. IoT device are mostly used in highly confidential data based application such as military data acquisitions, wearable IoT for healthcare, information regarding smart city etc., [3-6] which need to be protected for future use. Since the IoT devices are connected with internet the IoT devices are vulnerable to security threats. In some application such as cloud computing [7], the data collected by IoT devices are stored in cloud server from which the data user request the data. In this case, the third party will alter the data or interrupt the data acquired by the user.

Cybersecurity is the major need in IoT infrastructure, which has been studied by many researchers for years, since the threats in IoT increasing day-by-day. The comprehensive assessment of machine learning techniques for the detection of intrusion in IoT network has been conducted in [8-10]. The IoT security has been detected by using decision tree, gradient decent [11], Logistic Regression, Random Forest and Support Vector Machine [12]. These models are analysed by incorporating different methods for pre-processing and sampling stages. This results in highly accurate output however, the problem of timely detection of intruded data in the network is also a major concern. Moreover, timely detection make the system to find any solution to security issue while, the machine learning methods could be rained and validated with long duration. Hence, the researchers searched for alternative solution, which resulted in usage of deep leaning methods.

Owing to the advancement of deep learning techniques, more precise security threat in IoT network is detected. For the detection of industrial security threats, multi-cascaded convolutional neural network [13] is utilized. The adopted deep learning model differentiate different types of attack encountered in the data. The unauthorized access of data, in IoT infrastructure is detected by using long short term memory based markov chain [14]. That extensively reduce the outliers in data for detection process. However, the destructive cyber attack types are not classified with accurate features, hence multi-dimensional deep learning architectures [15] are used such as 1D, 2D and 3D. Most of the detection techniques are tested on benchmark dataset for intrusion detection such as NSL-KDD, KDD Cup99, BoT-IoT, UNSW-NB15 [16, 17] etc., for effective multiple intrusion activity detection. However, the usage of irrelevant features and absence of proper pre-processing methods, the precise and timely detection of anomaly activity in IoT environment is still in research.

1.1 Motivation

The development of an intelligent cyber security model for Internet of Things (IoT) networks is the utmost importance in today's digital landscape. The vast number of connected devices and the lack of standardization in security protocols make IoT vulnerable to cyber-attacks. The complexity of IoT systems and the difficulty in updating firmware and software also pose significant security risks. The motive behind the proposed mechanism for the new deep learning based approach.This model aims to enhance the security of IoT networks by leveraging the power of deep learning and transfer learning techniques. To ensuring the safety and integrity of IoT networks, a novel AR-PDTN model is designed to detect cyber-attacks. This innovative approach has the potential to revolutionize the field of cyber security and provide a robust defence against emerging threats. The major contributions are as follows:

- To detect the malicious activity in IoT environment using hybrid deep adversarial network.
- To develop a pre-processing method to remove the unwanted data from IoT devices.
- To examine the proposed system with existing state-of-art techniques with respect to datasets.

The rest of the paper is organized as follows: section 2 represents the related works, section 3 represented the proposed methodology, section 4 denotes the results and discussion, section 5 indicates the conclusion and future scope.

2. LITERATURE SURVEY

For the enhancement of secure communication in IoT systems, Saba et al. [18] employed deep neural network for intrusion detection in smart village. In that model, the attack in the data was categorized by using convolutional neural network. The adopted neural network suffers from losses such as binary cross-entropy and categorical cross-entropy, which reduce the precise estimation of different attacks in IoT network. Hence to cope with that Adam optimizer is used, which regulate the weight of hidden layers. Finally, the adopted model was tested on BoT-IoT and NID dataset.

As number of attacks in the IoT platform is increasing, the solution for cybersceurity issues were provided by Ullah et al. [19] using recurrent neural network architectures. For the detection of anomaly the author investigated the performance of long short term memory, bilateral long short term memory and gated recurrent unit. Then the feature from NSLKDD, BoT-IoT, IoT-NI, IoT-23, MQTT, MQTT set, and IoT-DS2 datasets are learned by using convolutional neural network and recurrent neural network. Finally, to classify attacks in the network binary classification model is deployed.

The intrusion detection using parallel auto-encoders was reported by Basati and Faghih

[20] for secure IoT network. The reported lightweight parallel deep auto-encoder model would take the feature vector based on local and surrounding information. The model make used of 2 dimensional feature vector for further processing with the dilated convolution function. The input feature vectors are applied to convolutional function from which a split channel is applied. Then the encoder model is processed with feature extractor tube at one channel and local feature extractor at other channel. Then features are merged and shuffled for extracting global features, which is again splited in to two channels. That process extract the latent features and decoded in next stage; thus making parallel coding. The model is examined for KDDCup99, CICIDS2017, and UNSW-NB15 datasets.

By considering a multi-objective function, the anomaly of the IoT data is detected by Asgharzadeh et al. [21] using optimized convolutional neural network. The features of IoT data was extracted by using feature extraction convolutional neural network that extract both high-level and low level features. These features are extracted by using convolutional neural network, which are selected for particular feature using binary multi-objective enhanced capuchin search algorithm. The obtained features from NSL-KDD and TON-IoT dataset are classified by using random forest classifier.

To deal with the issue of cyber security in IoT network Sharma et al. [22] proposed a filter based deep learning model. In that model, a filter based feature selection method is adopted that evaluate the score of data through pearson's correlation coefficient. Depending on score and threshold value, the feature is selected. However, before feature selection pre-processing and data augmentation has been carried out so that number of data for training and testing is improved. Then for the detection of affected data, the selected feature from pearson's correlation coefficient was given as input to generative adversarial neural network. The utilized network model has number of dense layer with varied neurons at each layer. However, it uses leaky rectified linear unit for activation with sigmoid function at final stage. The adopted model is tested on UNSW-NB15 dataset.

2.1 Problem statement

Ensuring the security of Internet of Things (IoT) devices is crucial in protecting against cyber threats. Robust security measures must be implemented to safeguard against potential breaches and protect sensitive data. There are several research that are to ensure the security in the iot system, but there are several limitations that are presented. The inaccuracy, time consuming and complex structure are some of the limitations present in the existing models. To resolve all the challenges,a novel deep learning based intrusion detection model is present in this research.

3. PROPOSED METHODOLOGY

The IoT devices make use of public data transfer medium (internet) for the data transfer, hence the data is prone to security issues. Moreover, the data transmitted by IoT devices does not use proper encryption standards thus the intruder can easily access the credentials and other important data send by IoT devices. However, it is necessary to detect whether the data has been interrupted by intruder so that the data received could be efficiently used. This malicious activity detection for IoT data provides alert signal to the network operator, thereby remedy for intrusion activity is made. To detect the intrusion, many machine learning methods are adopted, but the algorithm process slowly due to high training time. Followed by that deep learning architectures are used, which need separate feature extraction mechanism and then classify the output. However, the naive adoption of deep learning may lead to misleading design choices and leads to several drawbacks, such as slow detection and reaction time. Hence to reduce these difficulties, here a novel hybrid classifier is designed.

The dataset has many unwanted fields of data, which is not relevant for detection process. Hence to reduce that a pre-processing model using clustering approach is performed. The outliers and unwanted information in the data could be reduced by using multi-levelclustering approach concerning k-nearest neighbour, reverse k-nearest neighbour and k-means clustering approach. Then the intrusion of IoT data is classified by using Adversarially Regularized parallel deep transfer Network (AR-PDTN) for network anomaly detection. The proposed network make use of convolutional auto encoder to convert the input data into latent representations. Adversial learning will be used to promote the category-level feature discriminability. Then the parameters of proposed neural network model will be optimally tuned by Alpine skiing optimization. The utilized tuning model would reduce the error function caused in proposed neural network. Figure 1 represents the block diagram of the proposed method.

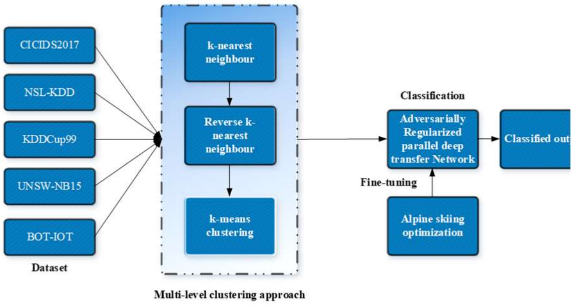

CICIDS2017
NSL-KDD
KDDCup99
UNSW-NB15
BOT-IOT
Dataset
k-nearest neighbour
Reverse k-nearest neighbour
k-means clustering
Multi-level clustering approach
Classification
Adversarially Regularized parallel deep transfer Network
Classified out
Fine-tuning
Alpine skiing optimization

Figure 1: Block Diagram Of The Proposed Method

3.1 Multi-level clustering approach

A multi-level clustering approach is a powerful tool for uncovering complex relationships within large datasets. It simultaneously performs hierarchical clustering and creates clusters of different granularities — from the most general, overarching clusters of data points that share a similar trend, to the most specific, tightly-knit groups that share a more precise set of characteristics. This multi-level analysis can reveal rich, meaningful patterns within a large dataset that may remain undetected with more traditional clustering approaches. Furthermore, the simultaneous clustering of data points at multiple levels of granularity can help to avoid the over-simplification of data that may occur when clustering using a single parameter. Additionally, a multi-level clustering approach can speed up processing times and reduce the amount of computational resources needed to analyze large datasets. This approach is particularly useful when dealing with data sets containing many dimensions, and can lead to a better understanding of the underlying structure of large datasets by uncovering both global and local patterns. The multi-level clustering approach is a powerful tool for data mining and knowledge discovery, and can be invaluable for helping to identify valuable insights from data.

3.1.1 K-nearest Neighbors Algorithm

KNN is a non-parametric classification and regression algorithm. In both cases, the input is the K closest preparation examples in the feature space and the output be contingent on whether K-NN is used in organization or regression. The output of KNN classification is the class membership. An object is classified by the popular of its neighbors. The article is assigned to the class that is most common among its K nearest neighbors (k being a positive integer, usually small). If k = 1, the object is just allotted to that class of that sole nearest neighbor.

3.1.2 Reverse k-nearest neighbour

The reverse K-nearest neighbour (also known as backward k-nearest neighbour) algorithm is a data-mining tool that can be used to identify near neighbours of a given goal object in a great dataset. This algorithm is especially useful when the dataset contains a large number of objects due to its relatively low computational difficulty. In contrast to standard adjoining neighbour pointed algorithms, reverse k-nearest neighbour follows a backward method by preliminary from the target and increasing outwards in search for the k-nearest neighbours. In its place of scanning the entire dataset, the backward approach only requires a partial number of judgments in order to effectively locate the closest objects in the dataset. The backward k-nearest neighbour is particularly useful in the field of database indexing where it can advance the presentation of query processing. The capability of the reverse k-nearest neighbour algorithm to rapidly and excellently trace adjacent neighbours makes it an appreciated data-mining tool for several responsibilities such as object classification, data clustering and pattern recognition. The reverse k-nearest neighbour of j is as follows as in equation 1.

$$N_{rk}(j) = \{ o \mid o \in G, j \in N_k(o) \} \tag{1}$$

The higher number of neighbours of another point in the sample, the higher the concentration of that point in the sample. This point is not isolated while high density points are typically found in data points with more reverse neighbours. The K-nearest neighbor matrix for each data point, and the number of Reverse K-nearest Neighbours for each data point. It is define as:

$$rKnn[j], j = 0,...,n. \tag{2}$$

Where n is the number of model points.

If $j \in N_k(o), o \in G, N_k(o)$ is the k-nearest neighbour set of o, thus. Where K is the loop number. The global average number of neighbours that are k-nearest to each other is called the " avgrknn ".

$$\text{avgrknn} = \frac{1}{n}\sum_{j=0}^{n} rknn[j] \tag{3}$$

3.1.3 K-means Algorithm

One of the easiest ways to solve the problem of clustering is by using k-means cluster. The main purpose of k-means clusters is to divide and group the data into normal and attack instances. The method works on the dataset D, which contains n objects and divides these objects into k clusters. The method begins by choosing n objects from the dataset D and then calculates the cluster centre based on the average value of each object in each cluster. The K-means clusters start with k arbitrary cluster centers in space and divides the collection of giving objects into k subsets based on the distance metric. The clusters of clusters are updated iteratively according to the optimization of an objective function. K-means clusters are one of the most widely used clustering techniques because it is easy to implement very efficiently with a linear time complexity. The main purpose of using the K-Means clustering scheme is to divide the collection of normal data and attack data that behave similar to each other into several partitions, which are called cluster centroids (K-th cluster). In simple terms, K-means estimates a fixed number of k, the optimal cluster centroid representing data with similar behavior.

Using Adversarially Regularised Parallel Deep Transfer Network (AR-PDTN) for network anomaly detection, the incursion of IoT data is then categorised.

3.2 Adversarially Regularized parallel deep transfer Network (AR-PDTN)

The intrusion of IoT data is classified by using Adversarially Regularized parallel deep transfer Network (AR-PDTN) for network anomaly detection. The proposed network make use of convolutional auto encoder to convert the input data into latent representations. Adversial learning will be used to promote the category-level feature discriminability

A GAN-regularized latent structure and a discrete auto-encoding are combined in Adversarially Regularized Autoencoder (ARAE). A learnt distribution across the space P is produced by the whole model. It seems sense that this technique would offer discrete series with a flexible previous a better hidden encoding. Thefollowing section demonstrates how this straightforward network may be formally interpreted as a latent variable model using the Wasserstein auto encoder framework.

The framework includes an auto encoder regularised with a previous sharing,

$$\min_{\theta,\zeta} \lambda_{rec}(\theta,\zeta) + \gamma^{(1)} A\left(C_q, C_s\right) \tag{4}$$

Here, A stands for the Wasserstein distance among the previous distribution C_s and the encoder model's C_q distribution (i.e., $enc_\theta(y)$, where $y \sim C_*$). The A component is calculated as before with an embedded critic function that is optimised in opposition to the encoder and generator.

The algorithm has been trained using coordinate descent over the decoder and encoder to minimise reconstruction, the critic unit to approximately determine the A term and the encoder in opposition to the critic to reduce A:

$$\min_{\theta,\zeta} \lambda_{rec}(\theta,\zeta) = E_{y\sim C_*}\left[-\log p_\zeta\left(Y \mid enc_\theta(Y)\right)\right] \tag{5}$$

$$\max_{m\in M} \lambda_{cri}(m) = E_{y\sim C_*}\left[f_m\left(enc_\theta(Y)\right)\right] - E_{x\sim C_x}\left[f_m\left(\hat{X}\right)\right] \tag{6}$$

$$\max_{\theta} \lambda_{rec}(m) = E_{y\sim C_*}\left[f_m\left(enc_\theta(Y)\right)\right] - E_{x\sim C_x}\left[f_m\left(\hat{X}\right)\right] \tag{7}$$

In practise, the model's effectiveness was significantly influenced by the previous distribution C_X that was used. The adversary auto encoder (A2E) is produced discretely when a fixed distribution, including a Gaussian $G(0,J)$ is used. The most descriptive features for classification may be extracted while removing misleading information from the input vector that misled the classifier. But in NIDS, traditional deep auto-encoders have severe limits. In order to obtain sufficient efficiency in identifying key characteristics using traditional convolution filters. The model should be piled together in numbers. This leads to intricate deep networks that are unsuitable for IoT devices. Therefore, the proposed method makes use of the Parallel Deep Auto-Encoder (PDAE), an innovative auto-encoder.

A PDAE utilises two auto-encoders: an ordinary auto-encoder which employs regular convolutional filters and a dilated auto-encoder which employs dilated convolutional filters. The 2D illustration of the input vector generated after using the pre-processing method serves as the input to a PDAE. A PDAE also includes a transfer layer in addition to encoder, latent feature and decoder being its other three main components. Eight regular convolution filters in the transfer layer send the input to eight channels for representation. The encoder was then fed by splitting the channels into two equally sized pieces.

The encoder has two tubes: a tube for extracting features from the immediate area and another for extracting features locally. A 3×3-size dilated convolutional filter auto-encoder's encoder component is located in the surrounding feature extractor channel. In an auto-encoder using traditional 3×3 convolutional filters, the local feature extractor unit also serves as the encoder component. Three consecutive convolutional filter layers are included in each tube and when combined with down sampling, they capture certain lower dimension characteristics from the source data and transmit them on to the subsequent layer. Reduced features with smaller dimensions are found at the end of each tube; these features serve as the inputs for the latent features, which come next.

The latent features component mixes and shuffles the outputs from two encoder tubes. Following that, it uses an average-pooling layer to separate the global characteristics from the shuffled characteristics. The next component is the decoder, which receives each of the two channels that were created from these characteristics. A dilated convolutional filter unit and a normal convolutional filter tube compose the decoder component, which also has two tubes. Each decoder tube has a layering architecture that is similar to that of the matching encoder tube, with the exception that the layers are arranged in the opposite order. The latent characteristics are entered into each channel, which then transmits them to higher

Up-sampling and all three layers of filters are used to create dimension characteristics. The following procedure is to place an average-pooling layer over the combined and shuffled results from the two decoder tubes to get the PDAE result. Training a PDAE's goal is to approach the identity operation, just as training regular auto-encoders. A trained PDAE's latent features can also be employed, like those of regular auto-encoders, to represent its inputs within a reduced and reduced-dimensional form.

However, PDAEs are more resilient than standard auto-encoders and may effectively construct the lower dimension model of their input with fewer layers and parameters. This is the key distinction between PDAEs and typical auto-encoders. Additionally, PDAE extracts the related spatial information of its input parameters more effectively than standard auto-encoders, making it a better choice for NIDS. Figure 2 represents the block diagram of proposed classifier.

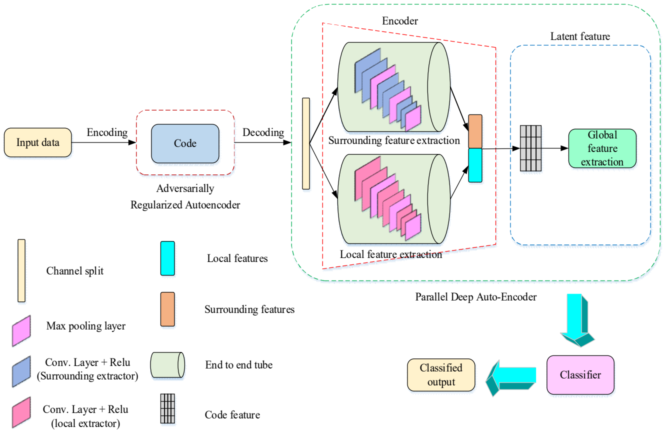


Figure 2: Block Diagram Of Proposed Classifier

Convolutional auto encoder is used by the proposed network to transform the input data into latent representations. The promotion of the category-level feature discriminability will make use of adversarial learning. The proposed neural network model's parameters will then be modified using Alpine skiing optimisation. The proposed neural network's error function would be reduced by the tuning model in use.

3.3 Parameter updation using Alpine skiing optimization

Alpine skiing is a sport full of possibility for optimization and improvement. With the right technique, skiers are propelled downhill with improved speed and control. To obtain maximum efficiency, skiers should strive for a proper position on the skis through good knee and ankle positioning. A proper stance will result in improved contact between the skis and the snow which will increase speed and control. This will also create more power for initial acceleration out of turns. Balance is key in alpine skiing so a skier must find the perfect balance of flexibility and tension in the body. This will help to center the skier's weight over the ski and give them better control. Furthermore, weight transfer is also important during ski turns and requires a skier to quickly move their weight from one foot to the other. Muscle memory is imperative in skiing and must be developed properly for consistency and optimization. The Involved movements that focus on mastering proper technique and muscle memory will quicken the process of learning and enhancement.

➢ *Initialization*

The result space is determined by the ASO algorithm. The skiers will find their position, which is randomly assigned at this stage. Once the parameters have been determined, the ASO makes an original population to define the capability parameters. The total number of skier is kept endless.

➢ *Exploration*

The ASO iterations are achieved. Each skier changes from their current place to a new position with some frequency. It is vital to note that skiers notice their standing and change closer to the leading position. The skier's fitness values are designed and related to their prior results. The best solutions are retained in the present iteration. The psychological change of the skier is based on the two ultimate issues: first place and avoiding falling. The distance between each skier in the ASO can be calculated as follow in the equation (8).

$$S_k(j) = \left\| y_k(j) - y_a(j) \right\| \quad (8)$$

Where $y_k(j)$ = skier's point in iteration number j and $y_a(j)$ = main place in iteration number j. During the competition, the skiers will utilize more skiing skills to take the prime in the competition. As the game advances, the skiers' physical strength will change as well, which can be resolute by equation (9).

$$a_k(j) = \frac{L}{1 + e^{-n\left(y_k(j) - y_0\right)}} \quad (9)$$

Where y = 0.5; nn = Logistic Growth Level;

L = Maximum Value y_0 = sigmoid's center.

The step and the position of the skiers can be updated using the equation (10) and (11).

$$\Delta K(j) = q.X(j).Y(j) \quad (10)$$

$$K(j+1) = K(j) + \Delta K(j) \quad (11)$$

Where the $X(j)$ and $Y(j)$ are denoted here as the distance matrix. 'q' Denotes here as the random number in (0,1), $K(j+1)$ denoted the places of the skiers in the iteration numeral $(j+1)$. The "fall down" in the ASO means that a skier's corporeal strength changes significantly in comparison to their physical strength at the previous moment. When there is a difference between the two values of 15% or more, a skier is considered to have "falled down". When there is a difference of 0.5 or more between the two values, the other skier's positions are updated using a "Levy flight strategy". The other skier's location can be calculated using the equation (12).

$$K(j+1) = K(j) + Levy(k(j)) \times K(j) \quad (12)$$

$Levy(k)$ is indicates the measurement of the skiers' positions. The Levy flight is considered using the equation (13).

$$Levy(k) = 0.01 \times \frac{q_1 \times \eta}{|q_2|^{\frac{1}{\alpha}}} \quad (13)$$

Where q_1 and q_2 are the random number that are obtainable in the [0,1], α is a constant (α = 1.5), η can be obtained using the equation (14).

$$\eta = \left(\frac{\Gamma(1+\alpha)\times \sin\left(\frac{\pi\alpha}{2}\right)}{\Gamma(1+\alpha)\times\alpha\times 2^{\left(\frac{\alpha-1}{2}\right)}} \right)^{\frac{1}{\alpha}} \qquad (14)$$

Where Γ(k) = (k − 1)!.

- *Exploitation*

The leading skier will proceed along the present route. The extra skiers will confirm the new skiing movement by joining the first skier's forward route with their present point and alignment. When the convergence criterion is not met, the iteration will end. The ASO stopping criterion is set to the highest number of iterations with no development.

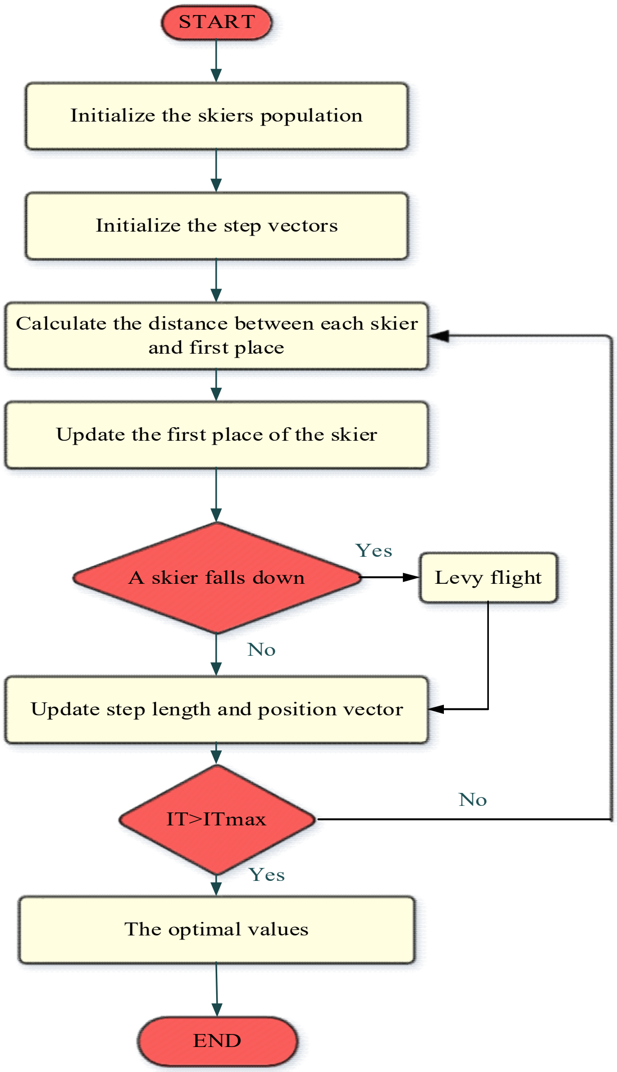

START
Initialize the skiers population
Initialize the step vectors
Calculate the distance between each skier and first place
Update the first place of the skier
A skier falls down
Yes
Levy flight
No
Update step length and position vector
IT>ITmax
No
Yes
The optimal values
END

Figure 3: ASO Flowchart

4. RESULTS AND DISCUSSION

The research model is implemented on PYTHON platform. the system specifications of Intel(R) Core (TM) i5-4670 CPU @ 3.40GHz with installed memory of 8 GB with an operating system of 64- bit without using pen or touch input.The proposed model is evaluated using several existing methods like Firefly algorithm (FFA), Grey wolf optimiser (GWO), Bat algorithm (BAT), Multiverse optimisation algorithm (MVO), transient search optimization-differential evolution (TSODE), CNN-LSTM (Convolutional neural network- Long short term memory), CNN-Recurrent Neural Network (CNN-RNN) and CNN-Gated Recurrent Unit (CNN-GRU) in terms of accuracy, precision, recall and f1-score. The proposed model usesCICIDS2017, NSL-KDD, KDDCup99, UNSW-NB15, BOT-IOT dataset to train the proposed model. The dataset description is given below.

*KDDCup-99:*The information was compiled based on the MIT Lincon laboratory's submission to the 1998 DARPA intrusion detection difficulty. The network traffic data was collected using a set of 1000 UNIX workstations and 100 clients. For the purpose of creating the KDD Cup 1998 dataset, the recorded data was saved in tcpdump format. On the analysed tcpdump data, employing audit information mining for automatic models for the ID (MADMAID) structure, extraction of features has been done. In research, 10% of the entire KDDCup-99 dataset is utilised and the connection records were normalised. 41 characteristics and 5 attack types are included in KDDCup-99.The characteristics are divided into three primary groups: fundamental characteristics, which include packet capture (Pcap) files, content features, which include the whole payload of TCP/IP packet information and time-dependent traffic features with a 2 second overlapping frame.

*NSL-KDD:*After duplicate connection records were removed, a better version of KDDCup-99 was created. The dataset's 41 characteristics and five assault types are also available in CSV format. The dataset's comprehensive statistics are shown in Table 1.

Table 1: Comprehensive Statistics

Attack type	KDDCup-99		NSL-KDD	
	Train	Test	Train	Test
Normal	97278	60593	67343	9710
DoS	391458	229853	45927	7458
Probe	4107	4166	11656	2422
R2L	1126	16189	995	2887
U2R	52	228	52	67
Total	**494021**	**311029**	**125973**	**22544**

Bot-IoT: The Bot-IoT dataset was created at the Cyber Range Lab of the UNSW Canberra Cyber Centre using IIoT traffic samples collected from industrial IoT (IIoT) smart home equipment. Thermostats, motion-activated lighting, remote-controlled garages, refrigerators, freezers, and weather monitoring systems are just a few examples of smart IIoT equipment. The information is provided in two versions: the complete version, which has more than 72 million entries, and the 10% version, which has around 3.6 million records. We make the decision to test the suggested model using a collection of the top ten features on 5% of the whole dataset.

*CICIDS-2017:*At the University of New Brunswick's Canadian Institute for Cyber security (CIC), the attack detection dataset CICIDS-2017 was developed. The CICIDS-2017 has almost 1.5 million records that simulate actual real-world data (PCAPs). The dataset includes information on a variety of attack methods, such as brute force, DoS, DDoS, infiltration, heart bleed, bot, and web-based. Utilising the results of CIC Flow Meter's analysis of network traffic, PCAP traffic files from the dataset were utilised to create CSV files. 25 user behaviours are examined by the CIC Flow Meter programme using several connection protocols (HTTP, FTP, SSH and email protocols). There are 80 network traffic characteristics and flows labels in the entire dataset, which was stored in CSV files.

*UNSW-NB15:*More than 2.5 million network packets are simulated for the UNSW-NB15 data set. Nine different types of attacks (Exploits, Reconnaissance, DoS, Generic, Shellcode, Fuzzers, Backdoors, Worms and Analysis) are included in this data set, along with non-anomalous packets. The data set is severely unbalanced since more than 87% of the packets are non-anomalous in nature.

4.1 Performance evaluation

Different performance metrics are used to evaluate the effectiveness of the proposed techniques. Parameters such as accuracy, precision and F1 measure are evaluated. The total number of normal cases that were accurately identified is known as True Negative (T_n). The False Negative (F_n) refers to the amount of normal data that was incorrectly assessed. True Positive (T_p) refers to the quantity of attacks that were accurately categorised. The number of attack samples that are incorrectly classified into normal sections is known as a false positive (F_p).

Accuracy: The accuracy metric measures the detection of cyberbullying activities accurately from the input data to the total dataset sample. The assessment of accuracy is presented as follows

$$Accuracy \quad = \frac{T_p + T_p}{T_p + T_n + F_p + F_n} \tag{15}$$

Precision: The precision measure calculates the number of accurately categorized cyberbullying incidents from the total number of true positives in a dataset. Precision is defined as

$$\Pr ecision = \frac{T_p}{T_p + F_p} (16)$$

Recall: Recall is similar to Precision, but it targets False Positives instead of False Positives. Once again, True Negatives are ignored.Recallis evaluated as

$$\operatorname{Re} call = \frac{TP}{TP + FP} (18)$$

F measure: The value of F measure is equal to the sum of precision and sensitivity. Precision and sensitivity are calculated as 'one' in F measure and the value of F-measure becomes 'zero' when sensitivity or precision reaches zero. F measure is evaluated as

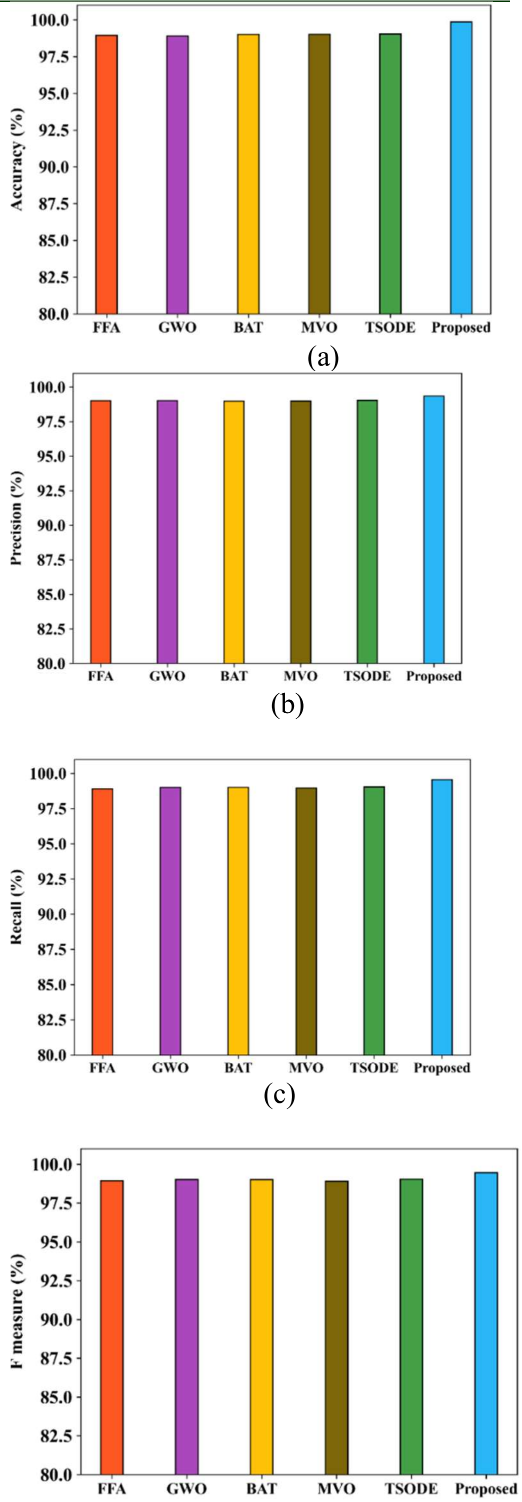
$$F - measure = 2 \times \frac{\Pr ecision \times Sensitivity}{\Pr ecision + Sensitivity}$$
(17)

4.2 Performance evaluation using BoT-IoT dataset

Figure 4 represents the performance evaluation of proposed and existing methods by using BOTIOT dataset. The class accuracy of the existing methods and proposed model is represented in the figure 3(a). The accuracy is limited in the existing system due to over fitting, computational complexity misclassification issue andtraining time. The proposed study uses improved techniques to overcome these issues and achieves higher optimization accuracy. The proposed model achieved the accuracy of 99.8%. The figure 3(b) represent the precision comparison with the existing system from the resultant figure gives the higher precision compared to the existing system. The proposed model have the precision value of 99.3%. The figure 3(c) represented the recall comparison of proposed with the existing system using BoT-IoT dataset. The proposed model achieved the recall value of 99.5%. The figure represents that, the proposed system has the best recall compared to the all other existing techniques. The figure 3(d) represented the f-measures of the existing system and the proposed system. The proposed model achieved the value of 99.4% which is the best value by comparing existing models. Table 2 represents the comparison of proposed and existing models using BoT-IoT dataset.

Table 2: Comparison Of Proposed Over Existing In Bot-Iot

Performance matrix	Proposed	TSODE	MVO	BAT	GWO	FFA
Accuracy	0.998593	99.042	99.03	99.01	98.9	98.95
Precision	0.99355	99.042	98.96	99.02	99.004	98.9
Recall	0.995696	99.042	99	98.98	99.013	99.007
F1Score	0.994622	99.042	98.9	99.01	99.02	98.94



(a)

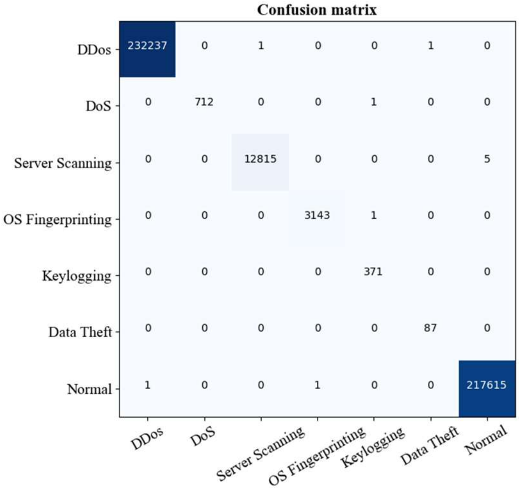
(b)

(c)

(d)

Figure 4: Performance Evaluation Using Bot-Iot Dataset

The confusion matrix illustrations the effectiveness of the proposed categorizing period. The confusion matrix is calculated by comparing the actual labels and the projected labels. Figure 5 represents the confusion matrix of the BoT-IoT dataset.



	DDos	DoS	Server Scanning	OS Fingerprinting	Keylogging	Data Theft	Normal
DDos	232237	0	1	0	0	1	0
DoS	0	712	0	0	1	0	0
Server Scanning	0	0	12815	0	0	0	5
OS Fingerprinting	0	0	0	3143	1	0	0
Keylogging	0	0	0	0	371	0	0
Data Theft	0	0	0	0	0	87	0
Normal	1	0	0	1	0	0	217615

Figure 5: Confusion Matrix Of The Bot-Iot Dataset

4.3 Performance evaluation using KDDCup99 dataset

The performance evaluation of proposed and existing model using KDDCup99 dataset is presented in figure 6. Figure 6 (a) represents the accuracy of the proposed and existing model. The proposed model achieved 94.5% of accuracy which is the best among all other existing model using KDDCup99 dataset. Figure 6 (b) compares the precision of the proposed system to that of the existing system and the proposedmodel has more precision than the existing model. The proposed model has a 98.6% precision rate which is the best using KDDCup99 dataset. The recall comparison of the proposed and current systems using the KDDCup99 was shown in figure 6(c). The proposed model successfully attained a 92.9% recall rate. The figure shows that, when compared to all existing method, the proposed system has the best recall. The f-measures of the proposed system and the existingsystem were depicted in figure 6(d). When compared to other models, the proposed model's value of 93.6% was the highest. Table 2 represents the performance value of the proposed and existing model. confusion matrix of the KDDCup99 dataset is represented in figure 7.

Table 2: Performance Value Of The Proposed And Existing

Performance evaluation	Proposed	TSODE	MVO	BAT	GWO	FFA
Accuracy	0.945462	92.06	91.61	90.347	87.61	84.318
Precision	0.986994	92.06	84.93	90.58	84.48	85.69
Recall	0.929994	89.94	86.64	89.13	84.13	91.609
F1Score	0.93693	90.07	84.48	90.04	88.53	84.28

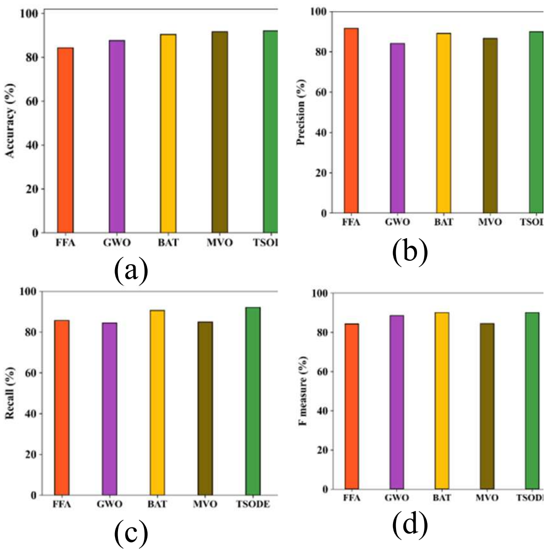
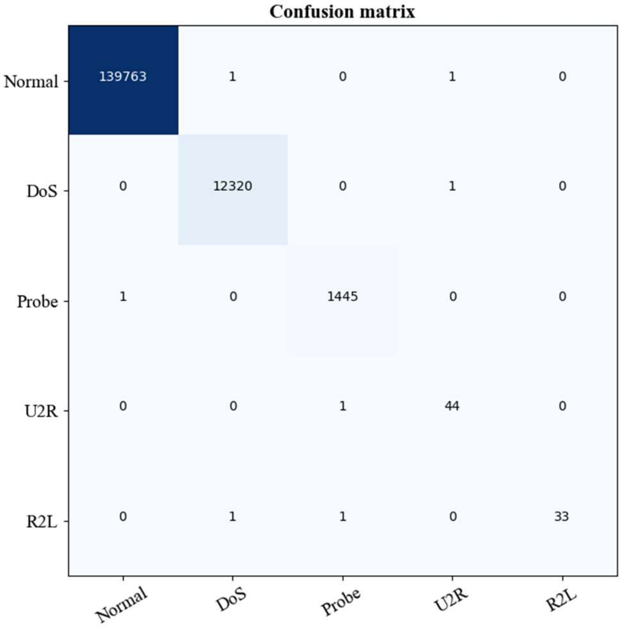


Figure 6: Performance Evaluation Using Kddcup99 Dataset



	Normal	DoS	Probe	U2R	R2L
Normal	139763	1	0	1	0
DoS	0	12320	0	1	0
Probe	1	0	1445	0	0
U2R	0	0	1	44	0
R2L	0	1	1	0	33

Figure 7: Confusion Matrix Of Kddcup99 Dataset

4.4 Performance evaluation using CICIDS2017 dataset

The figure 8 shows the accuracy, precisio recall and f measures of the proposed syste CICIDS2017 dataset with the existing data s such as CNN-GRU, CNN-RNN and the CNN LSTM. The proposed system that shows the be performance compared to the existing system The figure 8(a) shows that the accuracy of tl proposed and the existing system. The proposed model achieved the accuracy of 99.7% which is the best among all other existing models. The figure 8(b) shows that the precision of the existing and the proposed system. The proposed system has the best precision compared to the existing system. The proposed model achieved the value of 99.8%. The figure 8(c) represented the recall of the existing and the proposed system. The recall of the proposed system has the better performance among existing models while achieving the value of 99.8%. The figure 8(d) is presented here as the f1 score. The f1 score that is used here as to analysis of the proposed system and the existing system. The proposed model achieved the F1-score of 99.8% which is the best among all other existing models. Table 3 represent the comparison of proposed and existing model using CICIDS2017 dataset.

Table 3: Comparison Of Proposed And Existing Model Using CICIDS2017 Dataset

Parameters	Accuracy	Precision	Recall	F1score
Proposed	0.997985	0.998933	0.998282	0.998608
CNN-LSTM	99.1575	99.25	99.25	99
CNN-RNN	97.2475	97.5	97.25	96.5
CNN-GRU	96.125	96.25	96.25	95.5

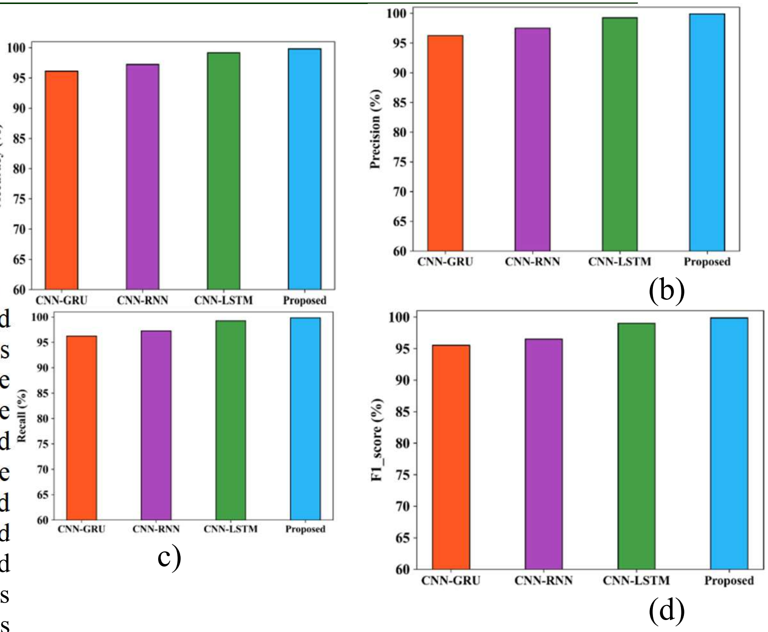
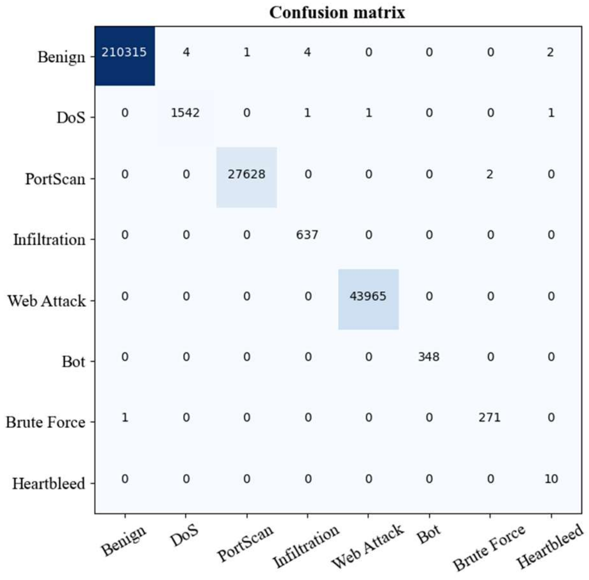


Figure 8: Performance Evaluation Using CICIDS2017 Dataset

The confusion matrix shows the usefulness of the proposed categorizing stage. The confusion matrix is calculated by associating the real labels and predicted labels.Figure 9 represents the confusion matrix of the proposed model.



	Benign	DoS	PortScan	Infiltration	Web Attack	Bot	Brute Force	Heartbleed
Benign	210315	4	1	4	0	0	0	2
DoS	0	1542	0	1	1	0	0	1
PortScan	0	0	27628	0	0	0	2	0
Infiltration	0	0	0	637	0	0	0	0
Web Attack	0	0	0	0	43965	0	0	0
Bot	0	0	0	0	0	348	0	0
Brute Force	1	0	0	0	0	0	271	0
Heartbleed	0	0	0	0	0	0	0	10

Figure 9: Confusion Matrix Usingcicids2017 Dataset

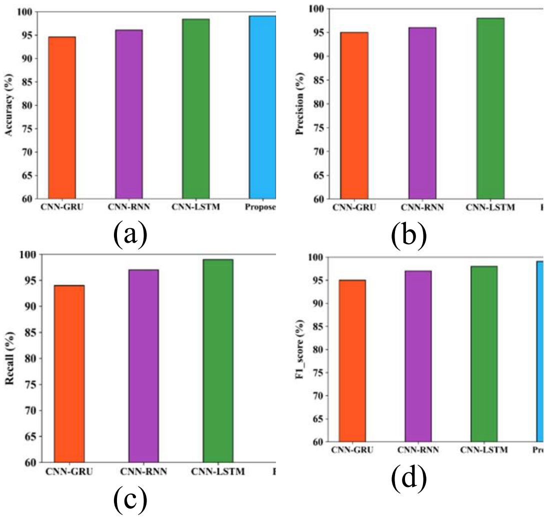
4.5 Performance evaluation using NSL-KDD dataset

Figure 10 compares the proposed system's accuracy, precision, recall and f metrics using

NSL-KDD dataset. Figure 10(a) illustrates theaccuracy of the proposed and existing systems. The proposed model, which outperformed existing models have an accuracy rate of 99.1%. Figure 10(b) illustrates the precision of the proposed and existingsystems. When compared to the existing system, the proposed system has the highest level of precision. 98.5% was obtained using the suggested model. Figure 10(c) depicted the recall of the proposed and existingsystems.When compared to existing models, the recall of the proposed system performs better and achieves a value of 99.5%. The f1 score is shown here as the figure 10(d). The examination of the proposed system and the existing system is done using the f1 score in this case. The proposed model outperformed all other known models with an F1-score of 99%. The proposed system that outperforms existing systems in terms of performance. Using the NSL-KDD dataset, Table 4 compares the proposed and current models. Figure 11 represents the confusion matrix of the NSL-KDD dataset.

Table 4: Performance Evaluation Using Nsl-Kdd Dataset

Methods	**Accuracy**	**Precision**	**Recall**	**F1-score**
Proposed	0.9913	0.985889	0.995792	0.990816
CNNLSTM	98.17	98	98	97.75
CNNRNN	95.745	96.25	96	95.75
CNNGRU	94.475	95	94	94.25



(a) (b)

(c) (d)

Figure 10: Performance Evaluation Using Nsl-Kdd Dataset

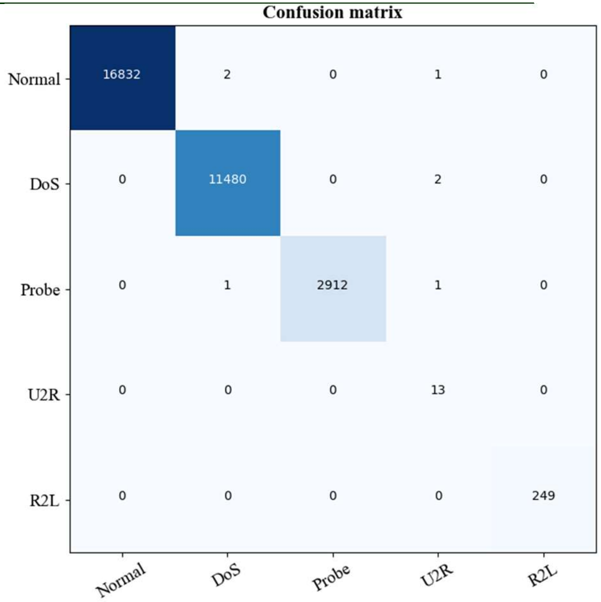


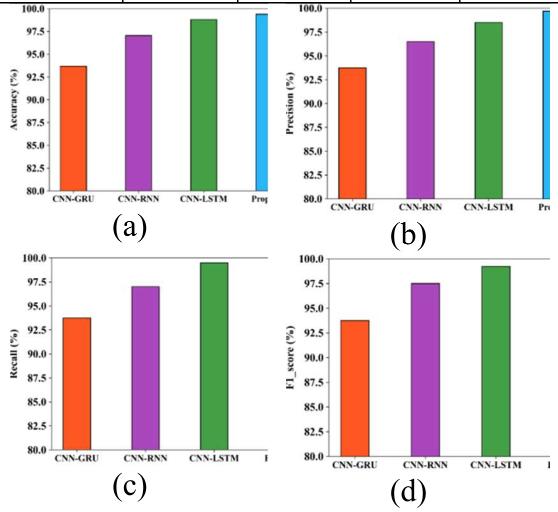
Figure 11: Confusion Matrix Of Nsl-Kdd Dataset

4.6 Performance evaluation using UNSW-NB15dataset

Using the UNSW-NB15 dataset, Figure 12 compares the proposed system's accuracy, precision, recall, and f metrics. The accuracy of the proposed and current systems is shown in Figure 12(a). The suggested model, which outperforms current models, had a 99.3% accuracy rate. The precision of the proposed and existing methods is shown in Figure 12(b). The proposed system offers the maximum level of precision when compared to the existing methods. Using the proposed model, 99.6% of precision was attained. The recall of the proposed and existing systems was shown in Figure 12(c). The recall of the proposed system works better and obtains a value of 99.5% when compared to existing models. Figure 12(d) represents the f1 score in this instance.In this scenario, the f1 score is used to compare the proposed system to the existing system. With an F1-score of 99.6%, the proposed model performed better than all other existing models. Table 5 compares the proposed and current models using the NSL-KDD dataset. The confusion matrix from the UNSW-NB15 dataset is shown in Figure 13.

Table 5:Performance Evaluation Using Unsw-Nb15 Dataset

Algorithm	Accuracy	Precision	Recall	F1-score
Proposed	0.9938	0.996811	0.995991	0.996401
CNNLSTM	98.81	98.5	99.5	99.25
CNNRNN	97.05	96.5	97	97.5
CNNGRU	93.67	93.75	93.75	93.75



(a) (b)

(c) (d)

Figure 12: Performance Evaluation Using Unsw-Nb15 Dataset

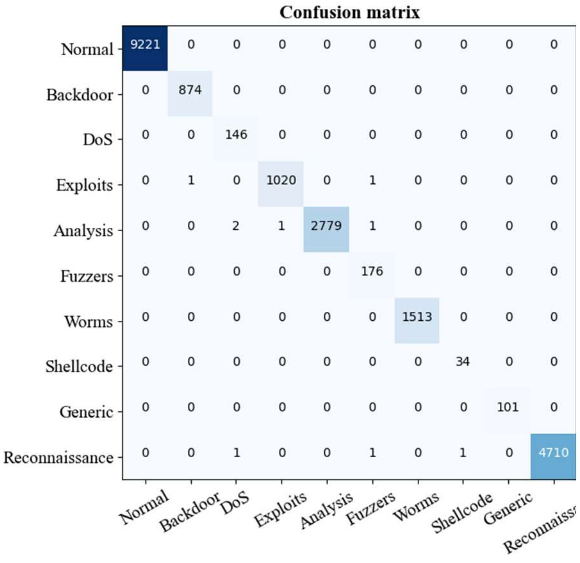


Figure 13:Confusion Matrix Of Unsw-Nb15 Dataset

Figure 14 represents the comparison among datasets in terms of time in second. The BoT-IoT dataset achieved the time of 0.12 second while CICIDS2017 dataset achieved the output at 0.1 second. The KDDCup99 and NSL-KDD have the time duration of 0.09 and 0.072 second, respectively. The UNSW-NB15 dataset achieved the output in 0.063 second which is the best among all other dataset.

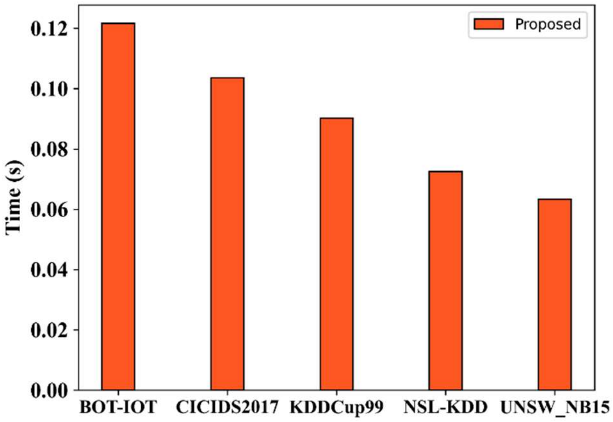


Figure 14: Comparison Between Different Dataset

5. CONCLUSION

In this research, an Adversarially Regularized parallel deep transfer Network (AR-PDTN) is introduced to classify the intrusion from IoT data. The unwanted information and outliersdata could be reduced by using multi-level clustering approach concerning k-nearest neighbour, reverse k-nearest neighbour and k-means clustering approach.The classification of the IoT data incursion is then done using the Adversarially Regularised Parallel Deep Transfer Network (AR-PDTN) to find network anomalies.After that, Alpine skiing optimisation will be used to optimally adjust the parameters of the proposed neural network model.CICIDS2017, NSL-KDD, KDDCup99, UNSW-NB15, BOT-IOT are the 5 different dataset used to classify the intrusion. In result section, the proposed model is compared with different existing models in terms of accuracy, precision, recall and f1-score. The suggested model's accuracy rates were 99.3% for the UNSW-NB15 dataset, 99.7% for CICIDS2017, 99.1% for NSL-KDD, 99.8% for BoT-IoT, and 94.5% for KDDCup99, respectively. When compared to each dataset, the proposed structure performed the best.In future, this research likes to improve the detection method and also intent to compare the proposed model with more recent method to evaluate the performance.

REFERENCE

[1] Castaño, Fernando, StanisławStrzelczak, Alberto Villalonga, Rodolfo E. Haber, and Joanna Kossakowska. "Sensor reliability in cyber-physical systems using internet-of-things data: A review and case study." *Remote sensing* 11, no. 19 (2019): 2252.

[2] Kuppusamy, Palanivel. "Smart education using internet of things technology." In *Emerging Technologies and Applications in Data Processing and Management*, pp. 385-412. IGI Global, 2019.

[3] Donghao, Cui, Zhang Bohua, Ou Chaomin, and Chen Zhiyu. "Research on Military Internet of Things Technology Application in the Context of National Security." In *2021 2nd International Conference on Electronics, Communications and Information Technology (CECIT)*, pp. 992-998. IEEE, 2021.

[4] Sengupta, Jayasree, SushmitaRuj, and Sipra Das Bit. "A comprehensive survey on attacks, security issues and blockchain solutions for IoT and IIoT." *Journal of Network and Computer Applications* 149 (2020): 102481.

[5] Ghubaish, Ali, Tara Salman, MaedeZolanvari, DevrimUnal, Abdulla Al-Ali, and Raj Jain. "Recent advances in the internet-of-medical-things (IoMT) systems security." *IEEE Internet of Things Journal* 8, no. 11 (2020): 8707-8718.

[6] Al-Turjman, Fadi, HadiZahmatkesh, and RamizShahroze. "An overview of security and privacy in smart cities' IoT communications." *Transactions on Emerging Telecommunications Technologies* 33, no. 3 (2022): e3677.

[7] Sadeeq, Mohammed Mohammed, Nasiba M. Abdulkareem, Subhi RM Zeebaree, DindarMikaeel Ahmed, Ahmed Saifullah Sami, and Rizgar R. Zebari. "IoT and Cloud computing issues, challenges and opportunities: A review." *Qubahan Academic Journal* 1, no. 2 (2021): 1-7.

[8] Verma, Abhishek, and VirenderRanga. "Machine learning based intrusion detection systems for IoT applications." *Wireless Personal Communications* 111 (2020): 2287-2310.

[9] Da Costa, Kelton AP, João P. Papa, Celso O. Lisboa, Roberto Munoz, and Victor Hugo C. de Albuquerque. "Internet of Things: A survey on machine learning-based intrusion detection approaches." *Computer Networks* 151 (2019): 147-157.

[10] Saranya, T., S. Sridevi, C. Deisy, Tran Duc Chung, and MKA Ahamed Khan. "Performance analysis of machine learning algorithms in intrusion detection system: A review." *Procedia Computer Science* 171 (2020): 1251-1260.

[11] Douiba, Maryam, Said Benkirane, AzidineGuezzaz, and MouradeAzrour. "An improved anomaly detection model for IoT security using decision tree and gradient boosting." *The Journal of Supercomputing* 79, no. 3 (2023): 3392-3411.

[12] Hasan, Mahmudul, MdMilon Islam, MdIshrak Islam Zarif, and M. M. A. Hashem. "Attack and anomaly detection in IoT sensors in IoT sites using machine learning approaches." *Internet of Things* 7 (2019): 100059.

[13] Sankaran, K. Sakthidasan, and Bong-Hyun Kim. "Deep learning based energy efficient optimal RMC-CNN model for secured data transmission and anomaly detection in industrial IOT." *Sustainable Energy Technologies and Assessments* 56 (2023): 102983.

[14] Shanmuganathan, V., and A. Suresh. "LSTM-Markov based efficient anomaly detection algorithm for IoT environment." *Applied Soft Computing* 136 (2023): 110054.

[15] Ullah, Imtiaz, and Qusay H. Mahmoud. "Design and development of a deep learning-based model for anomaly detection in IoT networks." *IEEE Access* 9 (2021): 103906-103926.

[16] Kilincer, IlhanFirat, FatihErtam, and AbdulkadirSengur. "A comprehensive intrusion detection framework using boosting algorithms." *Computers and Electrical Engineering* 100 (2022): 107869.

[17] AbuAlghanam, Orieb, HadeelAlazzam, Esra'A. Alhenawi, Mohammad Qatawneh, and Omar Adwan. "Fusion-based anomaly detection system using modified isolation forest for internet of things." *Journal of Ambient Intelligence and Humanized Computing* 14, no. 1 (2023): 131-145.

[18] Saba, Tanzila, AmjadRehman, Tariq Sadad, HoshangKolivand, and Saeed Ali Bahaj. "Anomaly-based intrusion detection system for IoT networks through deep learning

model." *Computers and Electrical Engineering* 99 (2022): 107810.

[19]Ullah, Imtiaz, and Qusay H. Mahmoud. "Design and development of RNN anomaly detection model for IoT networks." *IEEE Access* 10 (2022): 62722-62750.

[20] Basati, Amir, and Mohammad Mehdi Faghih. "PDAE: Efficient network intrusion detection in IoT using parallel deep auto-encoders." *Information Sciences* 598 (2022): 57-74.

[21] Asgharzadeh, Hossein, Ali Ghaffari, Mohammad Masdari, and FarhadSoleimanianGharehchopogh. "Anomaly-based Intrusion Detection System in the Internet of Things using a Convolutional Neural Network and Multi-Objective Enhanced Capuchin Search Algorithm." *Journal of Parallel and Distributed Computing* (2023).

[22] Sharma, Bhawana, Lokesh Sharma, ChhaganLal, and Satyabrata Roy. "Anomaly based network intrusion detection for IoT attacks using deep learning technique." *Computers and Electrical Engineering* 107 (2023): 108626.