International Research Journal of Mathematics, Engineering and IT
ISSN: (2349-0322)
Association of Academic Researchers and Faculties (AARF)
Impact Factor- 7.132 Volume 11, Issue 04, April 2024
Website- www.aarf.asia, Email : editoraarf@gmail.com

ENSEMBLE-ENHANCED THREAT INTELLIGENCE NETWORK (EETIN): A UNIFIED APPROACH FOR IOT ATTACK DETECTION

[*]Ajay Chandra MK

(Ajay Chandra Manukondakrupa)

[*]Department of Computer Science, Swami Ramanand Teerth Marathwada University, Vishnu Puri, Nanded, Maharashtra, 431606.

[*]Corresponding Author Mail id: ajaymanukonda88@gmail.com

ABSTRACT

The procedure of identifying and reacting to unauthorized or malicious behaviours within the IOT system, is called IOT attack detection. There are some several advantages and challenges associated with implementing and maintaining the detection systems. Some of the limitations are Complexity, false positive and negative and scalability. To defeat these restrictions, we suggested a deep-learning based classification model (INTE). The proposed methodology consists of (1) Pre-Processing (2) Customized Dimensionality Reduction (3) Feature Extraction (4) Deep Learning-based Classification via Intelligent Network Threat Ensemble (INTE) Model (5) Evaluating and Testing. In pre-processing, the gathered data are pre-processed via data cleaning and Transformation. The pre-processed data are used to dimensionality reduction using O-PCA. The principal components are optimized via TLBO (Teaching learning-based optimization). Then the dimensionality reduced data are utilized to filter the relevant features using Time and frequency domain features and non-linear features. The extracted features are used to classify the attack detection via INTE model which is a combination of CNN, RNN and transformer-based model. And, the model will be evaluated using the performance metrics. The proposed methodology is executed using MATLAB.

Keywords: *IOT Attack Detection, CNN, RNN, Transformer-based model.*

1. INTRODUCTION

We are currently residing in the epoch of digitalization. The widespread influence of the internet has penetrated every domain of human life internationally, culminating in the genesis of the 'Internet of Things' ideology. To explicate, electronic devices will be able to interact with each other autonomously without requiring human interference (Sivasankari and Kamalakkannan 2022). A multitude of situations can be monitored in real-time through the installation of devices that collect connectivity, process, and other data. These data are then incorporated into a vast network via the Internet. Numerous intelligent devices have the ability to gather data through the utilization of devices, Quick Response Code, barcodes, and RFID technology. The data, which is transmitted via various short-range wireless contact networks, control gateways, home networks, sensor networks, and industrial control networks, is ultimately received at the network layer (Li *et al.* 2022; Pandey and Mishra 2023). Sensors serve as a means to bring the tangible realm in close proximity to the virtual realm, a feat that can be achieved through the utilization of fog computing, in any application of the Internet of Things (Gokhale *et al.* 2018). RFID (Radio Frequency Identification) is recently one of the better-option technologies, having proven successful in identifying elements in a wide range of practical applications such as "animal identification, healthcare, passport control, transportation, supply chain traceability, maritime freight container tracking, protective equipment verification, and toll payments" (Caramés *et al.* 2016).

The term "Internet of Things" (IoT) encompasses strategies that are linked to the net and have the ability to autonomously exchange data. The association of cloud computing with these devices is primarily due to their capacity to produce massive amounts of data that require processing (Habibi *et al.* 2023). The extensive implementation of IoT devices and platforms has resulted in a heightened level of digitization and connectivity across various sectors. The rapid ascent of IoT stages and IoT strategies has resulted in an enhanced stage of computerizing and connecting throughout the entire sector. Numerous thriving IoT industries, including "smart transportation, smart health, and smart energy, are flourishing" (Gao *et al.* 2023; Ma *et al.* 2023). Addressing cyber security challenges in IoT networks has made a key priority, which may be accomplished by developing and deploying an active Intrusion Detection System (IDS) at edge nodes. Machine Learning (ML) and Deep Learning (DL)-based methods are deemed acceptable and produce superior calculation outcomes (Sharma *et al*. 2023).

Because of their dispersed project, gigantic number of things, and directness, IoT schemes have become a popular goal for cyber invaders (Bojarajulu *et al.* 2023). So, it's

important to detect the attack in the early stage, many intrusion detection approaches exist, including those created on numerical investigation, cluster investigation, ANN and DL (Zhang *et al.* 2019). Because of the particular facility necessities of IoT that cannot be met by the central cloud: "low latency, resource restrictions, distribution, scalability, and mobility", attack findings in IoT are drastically dissimilar from prevailing systems (Thamilarasu and Chawla 2019).

The prime contribution of this research work is as follows:

a. To propose O-PCA to reduce the dimensionality of the principal components via TLBO.
b. To propose the Deep-learning based INTE model to classify the intrusion threats accurately.

2. LITERATURE REVIEW

Bhayo *et al.* (2023) demonstrated a ML-based solution for detecting DDoS outbreaks in an SDN-WISE IoT controller. We built a testbed atmosphere to pretend DDoS attack traffic group and added a ML-based detection module into the controller. The traffic is gathered via a logging apparatus built into the SDN-WISE controller, which writes system logs into a log folder before pre-processing and converting them to a data collection. The SDN-WISE controller's ML DDoS detection module classifies SDN-IoT network packets using NB, DT, and SVM methods. We compare the results provided by the ML DDoS finding module to the performance of the proposed context utilizing different traffic simulation situations. The suggested framework achieved NB, SVM, and DT accuracy rates of 97.4%, 96.1%, and 98.1%, respectively.

Hasan *et al.* (2019) investigated Several ML reproductions' predictions of occurrences and irregularities on IoT devices were associated. LR, SVM, DT, RF, and ANN are the ML procedures employed in this study. The measures used to compare presentation are "accuracy, precision, recall, f1 score, and area under the Receiver Operating Characteristic Curve". The classification achieved 99.4% test accuracy for Decision Tree, Random Forest, and ANN. Although these methods have the similar accuracy, other metrics show that Random Forest performs significantly improved.

Gaber *et al.* (2022) introduced an intrusion detection method for detecting injection attacks in IoT applications (such as smart cities) is provided. Two types of feature selection policies (constant removal and recursive feature elimination) were used in this method and verified by a variety of machine learning classifiers "(SVM, Random Forest, and Decision

Tree)". The T-Test was used to measure the effectiveness of the suggested feature selection strategy. The evaluation results using the public dataset AWID revealed that the DT classifier can detect injection assaults with an accuracy of 99% utilizing only 8 features chosen by means of the proposed feature selection strategy.

Regan *et al.* (2022) presented a federated-based technique that leverages a deep autoencoder to detect botnet occurrences utilizing dispersed traffic data collected on-device. Privacy is handled by the proposed federated solution by guaranteeing that the expedient's data is not moved or abandoned the network edge. As a substitute, the ML calculation is transported to the point where the data is created (i.e., the edge layer), with the added advantage of data protection. We show that utilizing our suggested model, we can get up to 98% accuracy in incongruity discovery when training with features such as source "IP, MAC-IP, and destination IP, among others". The total presentation comparison amid our suggested dispersed strategy and a central structure shows a considerable development in attack detection exactness.

Simpson *et al.* (2021) developed trustworthy environment based on fuzzy logic, With the help of Edge Computing, to reduce security concerns in smart cities. The trustworthy environment ensures that hostile entities are identified and cooperative attacks are detected in real time. The computations necessary at the culmination diplomacies can be located to Edge attendants in Edge Computing. This reduces the predicted dormancy and bandwidth consumption for old-style cloud admittance. Protected SEAL employs a fuzzy-based technique to detect and separate rogue nodes in an IoT network. The supposed bulges are re-analysed using the repute value acquired from the reaction-based belief assessment. Finally, the suggested dependable atmosphere's performance was authenticated.

Al-Fayoumi *et al.* (2023) provided a smart trivial exposure scheme that can sense LR-DDoS assaults in a software-defined IoT context using the MQTT protocol. The suggested system compares the presentation of four ML representations on a current dataset (LRDDoS-MQTT-2022) with a minimal feature set (two features only) and a balanced dataset: DTC, MLP, ANN, and NBC. Our exploratory evaluation demonstrates the DTC recognition system's hubris, attaining an accuracy of 99.5% with peak finding speed. Finally, our finest results outperform current algorithms with greater estimate rates.

Rathore and Park (2018) presented a system for fog-based occurrence finding that is based on the fog calculation model and a recently suggested ESFCM method. Fog computing, as a postponement of cloud computing, allows occurrence finding at the network edge and provisions dispersed occurrence finding. The ESFCM approach handles labelled

data with a semi-supervised fuzzy c-means algorithm and an ELM algorithm to deliver strong simplification presentation at a quicker finding rate. The assessment was carried out on the NSL-KDD dataset, and the suggested approach outperformed the central occurrence finding method. It achieved a shorter uncovering time of 11 milliseconds and an accuracy rate of 86.53 percent.

Baig *et al.* (2020) suggested a context for intelligent DoS uncovering that includes components for data gathering, feature position and grouping, training, and testing. The suggested context has been empirically assessed in real-world IoT occurrence circumstances, and the consequences are more accurate than traditional classification algorithms.

Mbarek *et al*. (2021) presented a trust-based finding technique for IoT repetition occurrences, in which a quantity of copy nodes is deliberately put into the network to exam the dependability and retort of witness nodes. Through extensive simulation, we analyse the viability of the suggested uncovering technique and equivalence it to two alternative plans, brute-force and first stayed. The evaluation considers the likelihood of detecting compromised attacks, transaction execution time, and communication failure rate. The imitation consequences demonstration that, while upholding a exposure runtime of 60 seconds on average for up to 1000 nodes, the suggested trust-based plan can meaningfully upsurge the uncovering probability against replication attacks to 90% on average, and thus meaningfully diminish statement disaster. Table 1 shows the Reviews by various authors of research gaps

2.1 Research Gap

Table 1: Reviews by various authors of research gaps

Authors	Aim	Research Gap
Bhayo *et al*, (2023)	To propose a ML based solution for detecting DDoS attacks in an SDN-WISE IOT supervisor.	Didn't include the frame work for large scale and explore the other types of IOT networks
Hasan *et al.* (2019)	To deal with the growing issue about attack and anomaly detection in IOT networks	Didn't explore the blockchain technology and develop a real-time monitoring system for IOT security.
Gaber *et al.* (2022)	To offer an intrusion recognition approach for	The proposed feature selection techniques were not

	sensing inoculation attacks in IoT applications, notably clever capitals.	compared to other better methods in terms of "convergence time, convergence iteration, and their impact on detection accuracy".
Regan *et al.* (2022)	To present a federated-based solution that uses a deep autoencoder to identify botnet attacks in IoT devices using on-device decentralized traffic data.	Didn't use the blockchain technology
Simpson *et al.* (2021)	To offer a Secure method for Secure SEAL that investigates the many possibilities of disturbance in IoT net produced by cooperative occurrences at the smart city environment's edge.	Didn't explore the co-operative occurrences in IOT nets.
Al-Fayoumi *et al.* (2023)	To present a way for capturing low-rate DDoS occurrences in a software-defined IoT atmosphere using the MQTT protocol.	Didn't generate the stimulated data collection casing the varied IOT procedures and occurrences
Rathore and Park (2018)	To present a fog-based occurrence uncovering context capable of detecting novel assaults in IoT and performing attack detection in a distributed manner	Lower in presentation. Due to the accidental task of input bias and weights, this may cause ill-post problems.
Baig *et al.* (2020)	To provide a methodology for detecting DoS attacks in	Service-centric WSNs offer potential for the IoT

	IoT net using averaged dependence estimators.	paradigm's future.
Mbarek *et al.* (2021)	To provide a trust worthy recognition technique for repetition occurrences in IoT networks that is specifically designed to classify and counter cooperated witness nodes in the attendance of testing copy nodes.	Didn't include the dynamic strategy for replica detection.

3. PROPOSED METHODOLOGY

3.1 Overview

The primary target of this paper is to detect the attack in IOT via a Deep learning-based model which is a combination of CNN, RNN and Transformer Based model. The proposed model developed by the following four major phase (i) Data pre-processing (ii) Customized Dimensionality Reduction (iii) Feature extraction (iv) Deep-learning based classification via INTE (v) INTE (vi) Evaluation and Testing. Fig (1) shows the overall architecture of the suggested methodology.

Step 1: Data Preprocessing:

Data cleaning and Transformation: Employ advanced transformations like quantile transformation for data normalization.

Step 2: Customized Dimensionality Reduction:

O-PCA: The principal components are optimized via self-improved TLBO.

Step 3: Feature Extraction

Time and Frequency Domain Features: Extract statistical, time-domain, and frequency-domain features from the IoT network traffic data. Non-linear Features: Utilize nonlinear features like entropy, fractal dimension, and correlation coefficients to capture complex relationships.

Step 4: Deep Learning-based Classification via Intelligent Network Threat Ensemble (INTE) Model:

Convolutional Neural Network (CNN): Design a CNN architecture to capture spatial patterns in network traffic data.

Recurrent Neural Network (RNN):

Implement an RNN (e.g., LSTM) to capture temporal dependencies in the data.

Transformer-based Model: Utilize a transformer architecture to capture long-range dependencies and interactions.

Step 5-Intelligent Network Threat Ensemble (INTE) Model:

Input Layer: Feed pre-processed and feature-extracted data into the ensemble model.

Base Classifiers: Implement three deep learning-based classifiers - CNN, LSTM, and Transformer-based model.

Intermediate Layer: Combine predictions from the three classifiers using a weighted average mechanism.

Fusion Layer: Use a fusion technique (e.g., concatenation) to combine the features learned by the individual classifiers.

Decision Layer: Final classification decision based on the fused features.

Step 6-Evaluation and Testing:

Evaluate the INTE model on the testing set using standard metrics such as "accuracy, precision, recall, F1-score, and ROC-AUC".

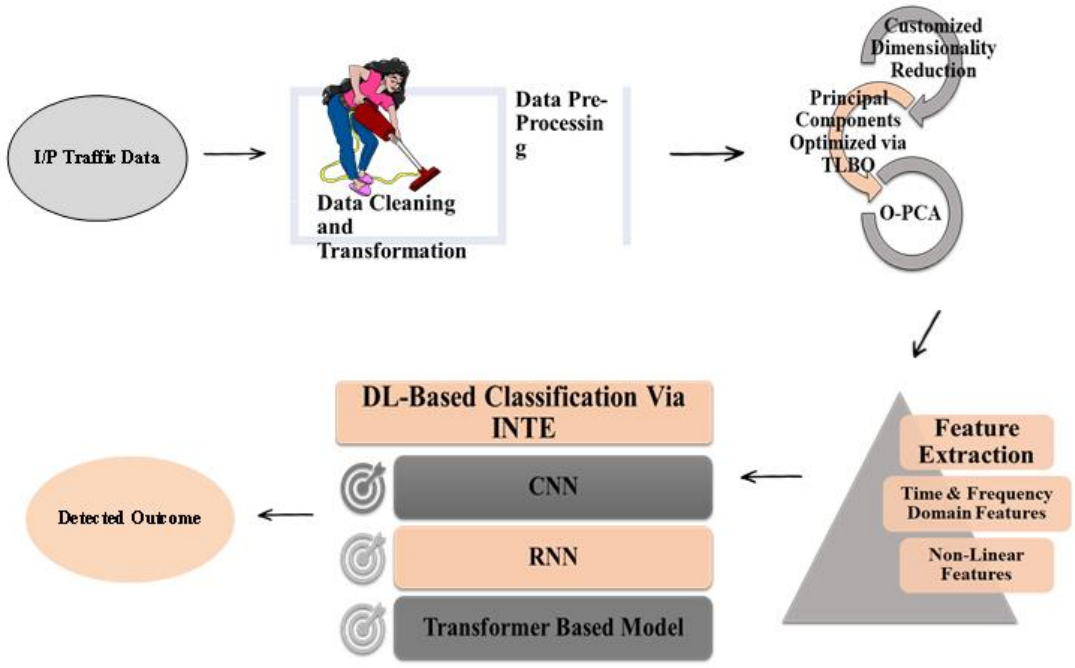

Figure 1: Overall architecture of the proposed model

3.1.1 Data Pre-Processing

The gathered raw data are pre-processed through Data Cleaning and Transformation. Advanced transformations like quantile transformation for data normalization. The act of preparing data involves the conversion of unprocessed data into a structured and effective format. Typically, real-world or newly acquired data is often deficient, prone to manual errors, and lacks uniform formatting. Data pre-processing aims to address these issues, while

simultaneously enhancing the effectiveness and comprehensiveness of datasets utilized for data analysis. Fig (2) shows the Data preprocessing technique which is used in this paper.

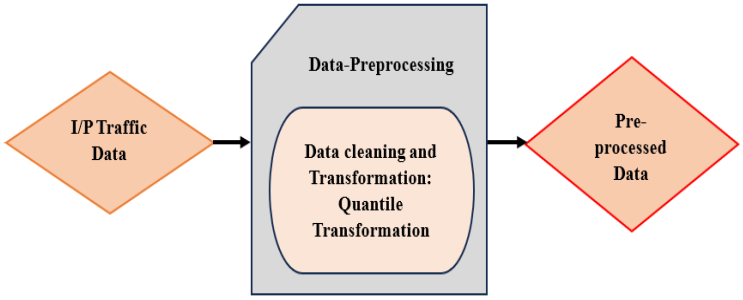

Figure 2: Data preprocessing

3.1.1.1 Quantile Transformation

The process of quantile normalization is a statistical methodology that is frequently utilized for establishing equivalence between two distributions by transforming data into a uniform or Gaussian (normal) distribution. This can be advantageous for various statistical and machine learning algorithms that rely on specific distributional properties.

3.1.2 Customized Dimensionality Reduction

The pre-processed data are used for dimensionality reduction using O-PCA. Dimensionality reduction is an essential component in the arena of data mining as it aids in mitigating the problematic issue of the "curse of dimensionality". This issue arises when datasets possess a large set of features or dimensions, resulting in various kind of challenges and inefficiencies in the analysis of data. Fig 3 shows the Dimensionality reduction phase

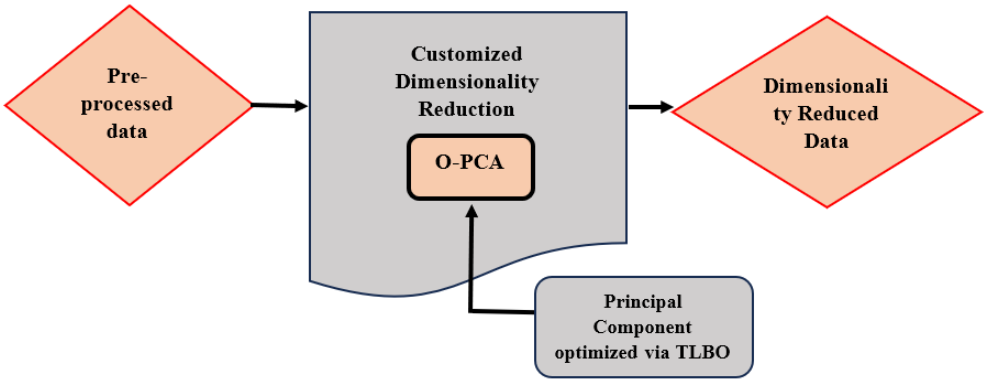

Figure 3: Customized Dimensionality Reduction

1. *PCA*

PCA is the simple and effective approaches for data examination and dimensionality lessening, utilized in big amount of pattern acknowledgement difficulties. The fundamental concept revolves around the substitution of the primary variables with novel variables called

as principal components. These components are acquired by means of linear combination of the initial components, arranged in a particular order and devoid of correlation. The aim is to reduce the loss of the original data while protecting it from higher dimensions to a lower-dimensional space. The principal components are optimized via TLBO.

2. TLBO

TLBO's operation is split into two phases: the "Teacher Phase" and the "Learner Phase" The two stages' functions are described below. The teacher phase in the first section is the procedure by which the students are taught by the teacher. At any iteration i, we assume that there are $'m$' subjects, 'n' learners (population size), and $m_{i,j}$ stands for the results of the learners in a certain subject $'j$' $(j = 1, 2, \ldots, m)$. The following equation serves as the paper's representation of the population (p):

$$p = \begin{bmatrix} x_{1,1} & x_{1,2} & & x_{1,n} \\ x_{2,1} & x_{2,2} & \ldots \ \ldots & x_{2,n} \\ \vdots & \vdots & & \vdots \\ x_{i,1} & x_{i,2} & & x_{i,n} \end{bmatrix}$$

The best learner K_{BEST} output can be interpreted as the best result $X_{TOTAL-KBEST,i}$ for all subjects from all learner populations. For each subject, the difference (ΔX) between the current mean and the associated instructor result is calculated as follows:

$$\Delta X = R_i(X_{j,KBEST,i} - T_f m_{i,j}) \quad (1)$$

Where $R_i$is a random number in the range [0, 1], $X_{j,KBEST,i}$is the outcome for the best learner in topic j, and T_f is the teaching factor, which determines the mean value. T_f's value can either be 1 or 2. The importance of T_f is:

$$T_f = round[1 + range\ (0,1)\{2 - 1\}] \quad (2)$$

The technique uses Eq. (2) to generate a random value forT_f, which is not provided to it as an input. The algorithm works better when the value of T_f is between 1 and 2. In the instructor phase, the answer is modified in accordance with X and the following expression:

$$X'_{j,K,i} = X_{j,K,i} + \Delta X \quad (3)$$

Where $X'_{j,K,i}$ is the updated value of $X_{j,K,i}$ and is acceptable if it gives better function value.

- *Learner phase*

In the second section, the students interact among themselves to further their knowledge, and then they randomly interact with other students to further their knowledge. Two students, p and q , are picked at random from the population of size 'n' as follows:

$$X'_{T-p,i} \neq X'_{T-q,i} \quad (4)$$

Where, after the conclusion of the teacher phase, $X'_{T-p,i}$and $X'_{T-q,i}$represent the updated function values of$X_{T-p,i}$ and$X_{T-q,i}$ of p and q, respectively.

$$X''_{j,p,i} = X'_{j,p,i} + R_i\left(X'_{j,p,i} - X'_{j,q,i}\right),\ if\ X'_{T-p,i} < X'_{T-q,i} \quad (5)$$

$$X''_{j,p,i} = X'_{j,p,i} + R_i\left(X'_{j,q,i} - X'_{j,p,i}\right),\ if\ X'_{T-q,i} < X'_{T-p,i} \quad (6)$$

The equations (5) and (6) are applied to minimization issues. The following equations, Eqs. (7) and (8), are applied to maximizing issues.

$$X''_{j,p,i} = X'_{j,p,i} + R_i\left(X'_{j,p,i} - X'_{j,q,i}\right), if\ X'_{T-q,i} < X'_{T-p,i} \quad (7)$$

$$X''_{j,p,i} = X'_{j,p,i} + R_i\left(X'_{j,q,i} - X'_{j,p,i}\right), if\ X'_{T-p,i} < X'_{T-q,i} \quad (8)$$

The following is a summary of the TLBO implementation guidelines for harm documentation founded on the minimization delinquent.

Rule 1:prepare the optimization parameters, which are:

- Populace extent (number of testes each generation)
- Number of peers
- Number of parameters utilized (number of elements chosen)
- Limits of strategy variables

Rule 2:Hand-picked the best learner for each generation

Rule 3:Assess the alteration between the present mean consequence and finest mean consequence rendering to the neutral occupation.

Rule 4:apprise the learner's acquaintance with the aid of the teacher's acquaintance

Rule 5:apprise the learner's acquaintance by applying the acquaintance of some others learners rendering to Eq (5) and (6), i.e., apprise the dented elements for each iteration.

Rule 6:Recurrence the process from step 2 to 5 until the selected quantity of generations is grasped.

3.1.3 Feature Extraction

Dimensionality reduced Data are move on to the Feature extraction phase. Time as well as Frequency FieldCharacteristics are obtained by statistical features, time-domain, and frequency-domain features also the non-linear features are extracted by entropy, fractal dimension, and correlation coefficients. Feature extraction used to make the process more accurate and it increases the prediction power of the algorithm to select the most relevant features. Figure 4 illustrates the Feature extraction techniques.

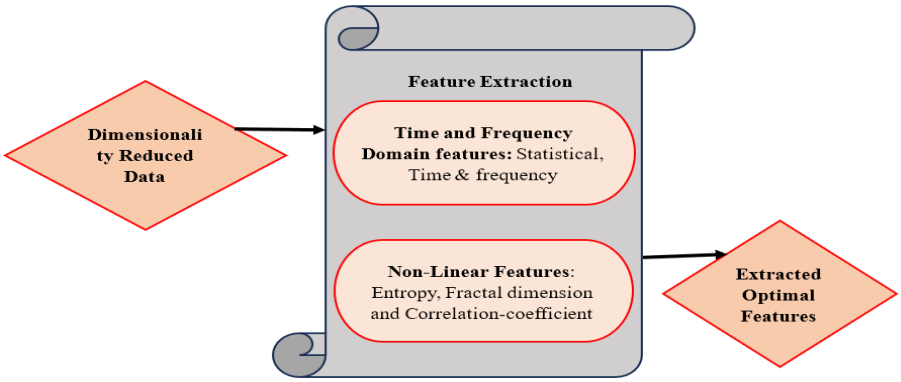

Figure 4: Feature Extraction

3.1.3.1 Time and Frequency Domain Features

In a variety of domains, including data mining, there are two alternative approaches to analyse and display data: in the TD and in the FD. They are frequently used to extract important data and patterns from datasets in signal processing, image analysis, and other data-intensive applications.

- ***Statistical Features***

Statistical features encompass those characteristics of a dataset that can be precisely delineated and evaluated through statistical analysis. This statistical construct is arguably the most commonly employed concept within the realm of data science. Table 2 shows the explanation and formula of the statistical features like mean, median, skewness and kurtosis.

Table 2: mathematical expression of Mean, Median, Skewness and kurtosis.

Metrics	Description	Mathematical Expression
Mean	The mathematical average of the set of two or more numbers	$\boldsymbol{\mu} = \frac{1}{N}\sum_{i=1}^{N} x_i$
Median	Midpoint between the lowest and highest value of the set	$Median = \frac{(N+1)}{2}$
skewness	Measures the symmetry, or the lack of symmetry	$skew = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{(x_i - \bar{x})}{\sigma}\right]^3$
kurtosis	Measures, whether the data are heavy-tailed or light-tailed	$kurt = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{(x_i - \bar{x})}{\sigma}\right]^4$

- Time-Domain

The representation and analysis of data in its original form, when data points are plotted against time or in sequential order is known as time domain. The focus is on how data evolves over time in the time domain, and patterns and trends are frequently seen by directly studying the data points in chronological sequence.

- Frequency-Domain

The transformation of data from the time domain to a domain that highlights the frequency domain, the plotting of data points against frequency can effectively reveal the underlying periodontics and frequency patterns in the data.

3.1.3.2 Non-linear features

Nonlinear features refer to specific attributes or alterations of data that encompass intricate interactions or procedures that are beyond the scope of linear equations. In this paper work the non-linear features are extracted by entropy, fractal dimension and correlation-coefficient.

- **Entropy**

The presence of impurity or randomness in a given dataset is often measured by entropy in the field of ML. This measurement is frequently employed in decision tree algorithms to assess the homogeneity of data at a specific node. A greater value of entropy denotes a more heterogeneous dataset featuring diverse classes, while a lower entropy signifies a purer and more homogeneous subset of data.

$$E = -\sum_{i=1}^{N} p_i \log_2 p_i \qquad (9)$$

Eq (9) shows the mathematical representation of Entropy. N denotes the total number of classes; E is entropy and p_i represents the probability of randomly selecting an example in class i.

- **Fractal Dimension**

A mathematical notion called fractal dimension is used to quantify the complexity, irregularity, and self-similarity of patterns in datasets. Objects with intricate and recursive patterns might have a “roughness” or “texture” that can be described using fractal dimension. In the realm of machine learning (ML), fractal dimensions are valuable as a component of dimensionality reduction, which alters the data processing methods of machine learning systems. Eq (10) shows the formula of Fractal Dimension,

$$d = \frac{\log(n)}{\log(R)} \quad (10)$$

Where, n is the number of boxes that enclosed the part and R is the exaggeration or the opposite of the box size.

- **Correlation Co-efficient**

correlation coefficient is a metric utilized to assess the potency and orientation of the linear correlation between two variables. It is a crucial instrument for detecting correlations between variables and comprehending how modifications in one variable may be linked to modifications in another variable. Eq (11) shows the formula of Correlation Co-efficient,

$$R_{a.b} = \frac{\sum_1^N (A_i - \bar{a})(B_i - \bar{b})}{N\, \alpha_A \alpha_B} \quad (11)$$

N represents the total number of data points, A_i and B_i are individual data points for variable

A and B, where Irange from 1 to N. $\bar{a}$and $\bar{b}$ is the mean average of A and B. α_A, α_B represents the standard deviation of A and B.

3.1.4 Deep Learning-based Classification via Intelligent Network Threat Ensemble (INTE) Model:

The extracted optimal features are move on to the classification stage via INTE model which is a combination of CNN, RNN and transfer based model. Figure 5 shows the DL-Based Classification via Intelligent Network Threat Ensemble (INTE) Model.

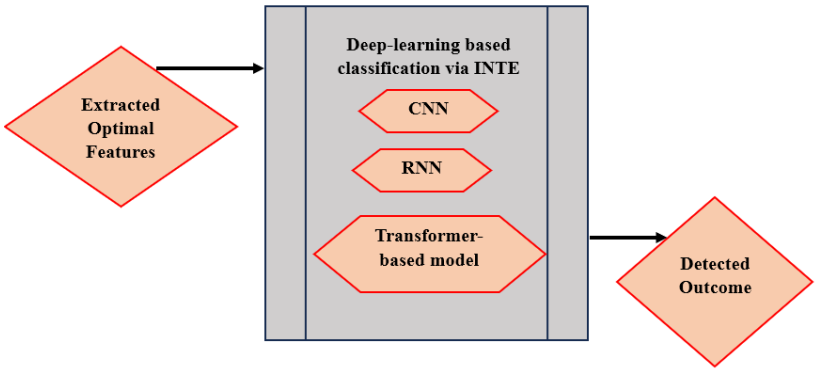

Figure 5: DL-Based Classification via INTE

3.1.4.1 CNN

Using grid-like matrices, convolutional neural networks (CNN), which have evolved from ANN, are mostly used to extract features from datasets. Convolutional, pooling, and fully linked layers are among the layers that make up a CNN in most cases. Convolutional layers perform the work of extracting features from the input data, and the task of down sampling the feature maps by pooling layers reduces their spatial dimensions. The fully

linked layers at the network's end carry out classification or regression tasks based on the learned properties. Eq (12) shows the mathematical representation of CNN,

$$x_{i,j}^{l} = \sum_{a=0}^{m-1} \sum_{b=0}^{m-1} w_{ab}\, y_{(i+a)(j+b)}^{l-1} \quad (12)$$

3.1.4.2 RNN

An ANN architecture called a recurrent neural network (RNN) is made primarily to handle sequential input. RNNs have a recurrence relation that enables them to capture the temporal dependencies contained in sequential data, unlike other neural networks that process data linearly during the feed-forward and back-propagation processes. "Natural language processing, speech recognition, machine translation, time series forecasting, music creation, and other sequential data-related applications have all had success with RNNs". They are effective tools for modelling and predicting data sequences because they can capture temporal dependencies. $H^{(t)}$represents the hidden units, $O^{(t)}$ denoted the outputs and $\hat{Y}^{(t)}$ display the targets.

$$A^{(t)} = B + wH^{(t-1)} + Ux^{(t)} \quad (13)$$

$$H^{(t)} = \tanh\left(A^{(t)}\right) \quad (14)$$

$$O^{(t)} = C + vH^{(t)} \quad (15)$$

$$\hat{Y}^{(t)} = softmax(O^{(t)}) \quad (16)$$

3.1.4.3 Transformer based Model

A transformer model is a form of neural network that acquires an understanding of context and, as a result, meaning by effectively monitoring the connections among sequential data, such as the words found within this sentence. Transformer models employ a dynamic range of mathematical methodologies, which are referred to as attention or self-attention, in order to identify nuanced manners in which even remote elements of data within a series exert an influence on and are reliant upon one another.

3.1.5 INTE model

3.1.5.1 Evaluation and Testing

Finally, the hybrid model is assessed using performance metrics such as "accuracy, precision, F-Measure, Sensitivity, Specificity, NPV, MCC, FPR, FNR". Also, the hybrid model is fine-tuned by adjusting the hyperparameters, architecture, or the optimization process to enhance fault diagnosis accuracy.

3.1.5.2 Performance Metrics

Several metrics are used to measure the performance including "Accuracy, Precision, F-Measure, Sensitivity, Specificity, NPV, MCC, FPR, FNR".

Accuracy

The accuracy is the ratio of properly categorized data to all of the data in the log. The Accuracy is described as,

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (17)$$

Precision

By employing the entire count of examples used in the classification process, precision is the representation of the total number of genuine samples that are appropriately taken into consideration during the classification process.

$$Precision = \frac{TP}{TP+FP} \quad (18)$$

F-Measure

The definition of the F-measure is the consonant mean of recollect rate and accuracy.

$$F_{Measure} = \frac{2\ Precision\ \times Recall}{Precision\ + Recall} \quad (19)$$

Sensitivity

The sensitivity value is obtained by just dividing the total positives by the proportion of true positive predictions.

$$Sensitivity = \frac{TP}{TP+FN} \quad (20)$$

Specificity

Specificity is premeditated by dividing the count of accurately anticipated negative outcomes by the total number of negatives.

$$Specificit\,y\ = \frac{TN}{TN+FP} \quad (21)$$

NPV

NPV determines the effectiveness of an analytical test or another quantifiable metric.

$$NPV = \frac{TN}{TN+FN} \quad (22)$$

MCC

Below is a representation of the two-by-two binary variable association measure known as MCC,

$$MCC = \frac{(TP\times TN-FP\times FN)}{\sqrt{(TP+FN)(TN+FP)(TN+FN)(TP+FP)}} \quad (23)$$

FPR

The count of negative occurrences divided by the count of negative events that were incorrectly ranked as positive yields the false positive rate (false positives).

$$FPR = \frac{FP}{FP+TN} \qquad (24)$$

FNR

The likelihood that an actual positive may be overlooked by the test is known as the false-negative rate, sometimes referred to as the "miss rate".

$$FNR = \frac{FN}{FN+TP} \qquad (25)$$

4. RESULT AND DISCUSSION

This paper presents a INTE model which is a combination of CNN, RNN and transformer-based model. In this part the suggested approach is associated with another present models to know the performance efficiency.

4.1 Comparative Analysis and Discussion

Several metrics such as "accuracy, precision, sensitivity, specificity, F-measure, MCC, NPV, FNR and FPR". To improve the performance of the suggested method, it is compared with CNN, LSTM and TF for each metrics. The assessmentoutcomeis shown in Table 3.

Table 3: Evaluation of metrices for existing and recommended technique

Metrics	CNN	LSTM	TF	Proposed
Accuracy	0.935067	0.9248	0.8812	0.974133
Precision	0.837667	0.812	0.703	0.935333
Sensitivity	0.837667	0.812	0.703	0.935333
Specificity	0.959417	0.953	0.92575	0.983833
F-Measure	0.837667	0.812	0.703	0.935333
MCC	0.797083	0.765	0.62875	0.919167
NPV	0.959417	0.953	0.92575	0.983833
FPR	0.016167	0.040583	0.047	0.07425
FNR	0.064667	0.162333	0.188	0.297

Table 3 displayed that the proposed methodology attained optimal value in all assessed metrics. Consequently, the suggested model exhibits superior performance

compared to all other techniques under comparison. Moreover, the evaluation of each metric is visually presented in the graphical format depicted in Figures 6 to 14.

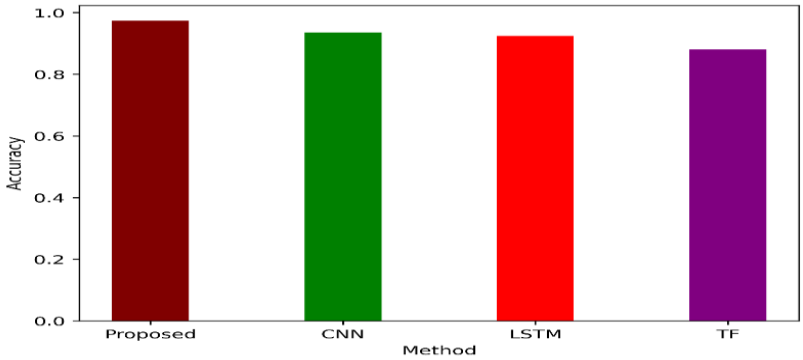

Figure 6: Accuracy analysis of various Algorithm

The graph shown in Figure 6 was made from the values from the Table 3. The accuracy values indicated as 0.935067, 0.9248, 0.8812 and 0.974133 for the methods CNN, LSTM, TF and the proposed methodology in this study. The suggestedsystem has the highest accuracy when associated to additional existing models which are in practice currently.

The precision of recommended method is compared with various methods and the obtained values are graphically represented in below Figure 7.

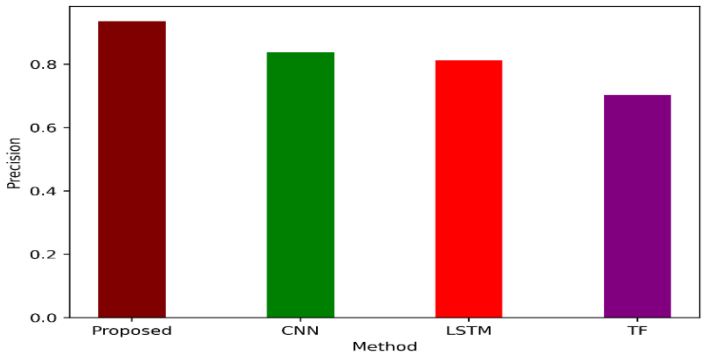

Figure 7: Precision Analysis

The data of the table 3 were used to create the graph that shown in figure 7. The obtained precision values of CNN, LSTM, TF and the suggested model like 0.837667, 0.812, 0.703 and 0.935333. The method that has been put forth exhibits the utmost level of precision when compared to all other available alternatives. When contrasted with the existing models that are currently being utilized, the model that has been recently developed demonstrates a significant degree of precision.

In the visual representation provided in Figure 8 below, the F-measure of a suggested strategy is compared to that of alternative approaches.

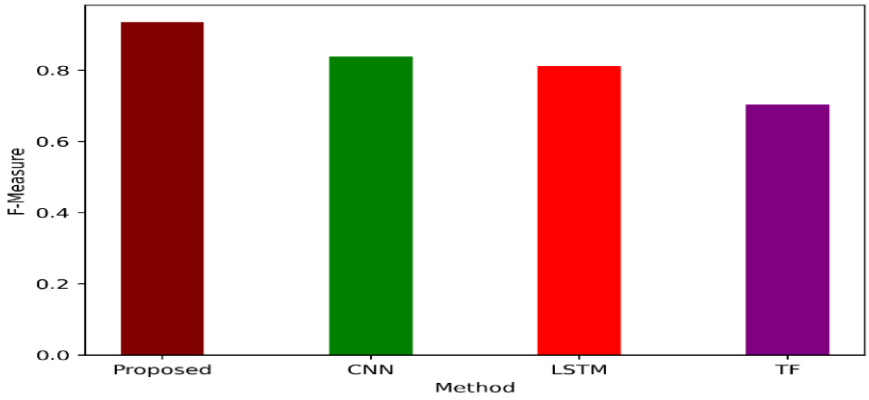

Figure 8: F-Measure Analysis

The graph shown in Figure 8 was made using the information of Table 3. The f-Measure value of CNN is 0.837666667, LSTM is 0.812, TF is 0.703 and the constructed model in this study is 0.935333333. The suggested approach attains the highest value of F-measure. The recently developed model exhibits a superior level of F-measure in comparison to the previous versions that are still being employed.

The outcomes of comparing a MCC of a suggested approach with those of other methodologies are depicted in Figure 9 as shown below.

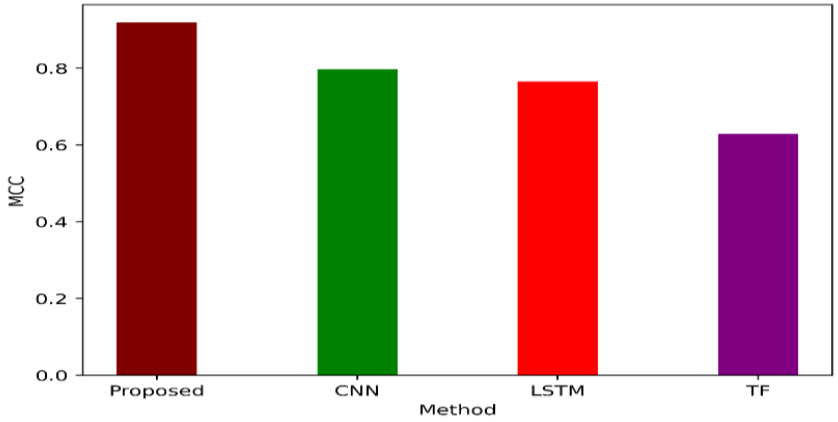

Figure 9: MCC analysis

Utilizing the informatic of table 1, the graph which is shown in figure 9 was created. The MCC values for recommended model, CNN, LSTM and TF are 0.919166667, 0.797083333, 0.765 and 0.62875. The strategy that is recommended yields the highest MCC score. When comparing the model that has been recently produced to the older versions that are still being utilized, it becomes evident that the newly developed model offers the highest rate of MCC.

The figure 10 below shows the outcome of comparing an NPV of a suggested strategy to those of alternative methods.

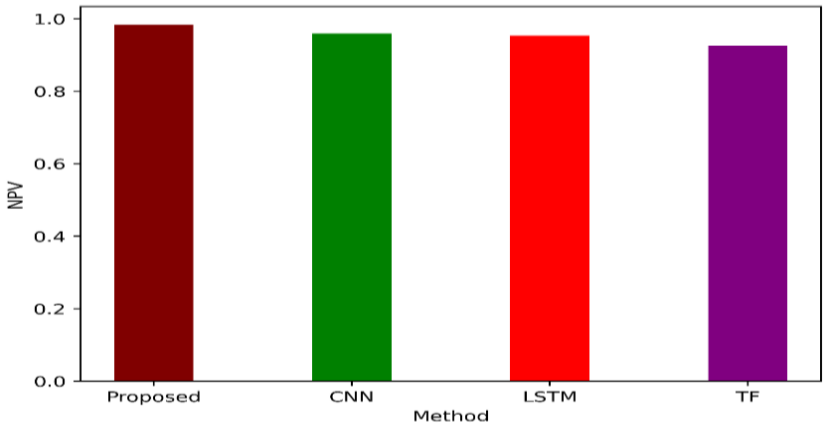

Figure 10: NPV Analysis

The graph shown in figure 10 was made using the data of table 3. CNN, LSTM, TF and suggested model all have NPV values of 0.959416667, 0.953, 0.92575 and 0.983833333.The suggested methodology yields the most favourable NPV score. The developed model employed in this investigation presents the highest NPV rate when compared to the preceding models that are currently in utilization.

Figure 11 below illustrates the results of comparing the FPR of a proposed plan to those of competing strategies.

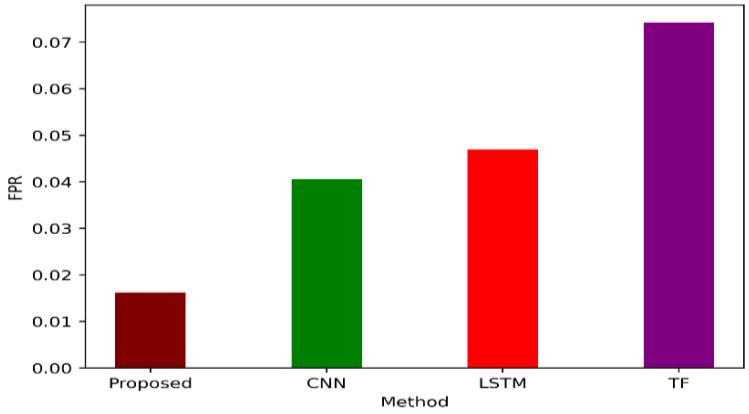

Figure 11: FPR Analysis

The information of Table 1 was utilized to create the graph which is shown in figure 11. The FPR values of CNN, LSTM, TF and the suggested model are 0.040583333, 0.047, 0.07425 and 0.016166667. The method proposed in this study results in the attainment of a value denoted as the lowest FPR value. When compared to the previous models that are currently in use, the model developed and utilized in this research exhibits the lowest FPR rate. The model possessing the minimum FPR value will demonstrate a more efficient performance. Given that the suggested technique possesses the minimum value, it can be considered as the most efficient method among the other models.

The results of comparing the FNR of a suggested scheme to that of rival tactics are illustrated in Figure 12 presented below.

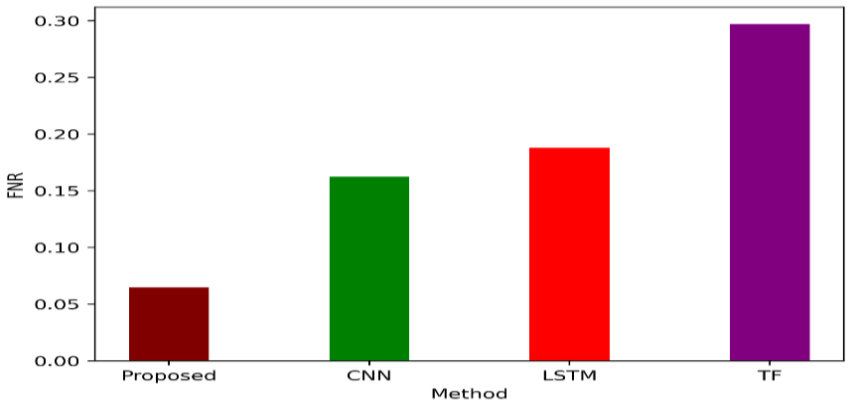

Figure 12: FNR Investigation

The graph displayed in Fig 12 was made from the values from the Table 3. The FNR values of CNN, LSTM, TF and the suggested model are 0.162333333, 0.188, 0.297 and 0.064666667. The model utilized in this investigation exhibited the most minimal FNR compared to the preceding models that are currently being utilized. The model that demonstrates superior performance possesses a diminished FNR magnitude. The suggested approach proves to be more efficacious than alternative models due to its minimal value.

The outcomes of comparing a sensitivity of a suggested approach with those of other methodologies are depicted in Figure 13 as shown below.

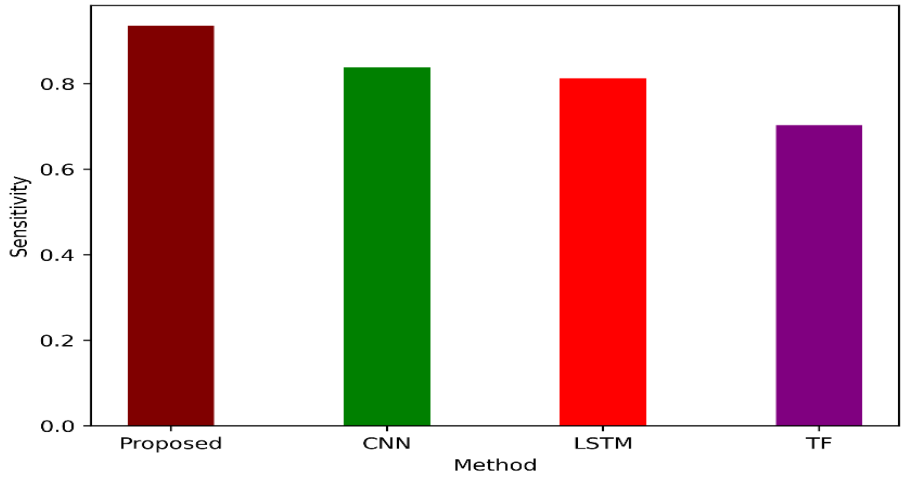

Figure 13: Sensitivity Analysis

The graph in Figure 13 above was made using the numbers from Table 3. The sensitivity values of CNN, LSTM, TF and suggested model are 0.837666667, 0.812, 0.703 and 0.935333333. The recommended method is the utmost level of sensitivity that exists. The recently devised framework provides a significant level of sensitivity in comparison to previous models that are presently being utilized.

The results of comparing a proposed approach's specificity to those of other techniques are shown graphically in Figure 14 below

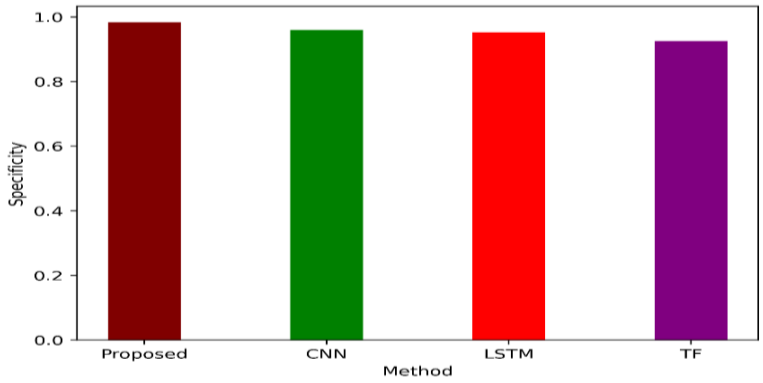

Figure 14: Specificity analysis

The graph displayed in Fig 14 was made from the values from the Table 3. The specificity values of CNN, LSTM, TF and the proposed model are 0.959416667, 0.953, 0.92575 and 0.983833333 respectively. The proposed strategy offers the highest level of specificity. Compared to older versions still in use, the recently designed model offers a high level of specificity.

The method proposed in this research has attained the most optimal value in all the aforementioned metrics. Based on these analyses, the newly devised approach exhibits superior performance compared to all other existing techniques that were examined.

5. CONCLUSION

The paper has developed aDL-basedcataloguing model for attack recognition. The collected data are pre-processed via data cleaning and transformation. From the pre-processed data the dimensions are reduced using O-PCA. The principal components were optimized using TLBO. From the dimensionality reduced data, the features were extracted by time and frequency domain features and non-linear features. Obtained extracted features were applied in the classification model (INTE) which is a combination of CNN, RNN and Transformer based model with the purpose of categorizing the attacks. The proposed model obtained the highest accuracy0.974133333 respectively. The proposed methodology was executed using MATLAB.

DATA AVAILABILITY STATEMENT

All the data is collected from the simulation reports of the software and tools used by the authors. Authors are working on implementing the same using real world data with appropriate permissions.

REFERENCE

[1] Al-Fayoumi, D.M. and Abu Al Haija, Q., Capturing Low-Rate DDoS Attack based on MQTT Protocol in Software Defined-IoT Environment. Capturing Low-Rate Ddos Attack Based on Mqtt Protocol in Software Defined-Iot Environment 2023.

[2] Baig, Z.A., Sanguanpong, S., Firdous, S.N., Nguyen, T.G. and So-In, C., 2020. Averaged dependence estimators for DoS attack detection in IoT networks. Future Generation Computer Systems, 102, pp.198-209.

[3] Bhayo, J., Shah, S.A., Hameed, S., Ahmed, A., Nasir, J. and Draheim, D., 2023. Towards a machine learning-based framework for DDOS attack detection in software-defined IoT (SD-IoT) networks. Engineering Applications of Artificial Intelligence, 123, p.106432.

[4] Bojarajulu, B., Tanwar, S. and Singh, T.P., 2023. Intelligent IoT-BOTNET attack detection model with optimized hybrid classification model. Computers & Security, 126, p.103064.

[5] Diro, A.A. and Chilamkurti, N., 2018. Distributed attack detection scheme using deep learning approach for Internet of Things. Future Generation Computer Systems, 82, pp.761-768.

[6] Fernández-Caramés, T.M., Fraga-Lamas, P., Suárez-Albela, M. and Castedo, L., 2016. Reverse engineering and security evaluation of commercial tags for RFID-based IoT applications. Sensors, 17(1), p.28.

[7] Gaber, T., El-Ghamry, A. and Hassanien, A.E., 2022. Injection attack detection using machine learning for smart IoT applications. Physical Communication, 52, p.101685.

[8] Gao, M., Wu, L., Li, Q. and Chen, W., 2023. Anomaly traffic detection in IoT security using graph neural networks. Journal of Information Security and Applications, 76, p.103532.

[9] Gokhale, P., Bhat, O. and Bhat, S., 2018. Introduction to IOT. International Advanced Research Journal in Science, Engineering and Technology, 5(1), pp.41-44.

[10] Habibi, O., Chemmakha, M. and Lazaar, M., 2023. Imbalanced tabular data modelization using CTGAN and machine learning to improve IoT Botnet attacks detection. Engineering Applications of Artificial Intelligence, 118, p.105669.

[11] Hasan, M., Islam, M.M., Zarif, M.I.I. and Hashem, M.M.A., 2019. Attack and anomaly detection in IoT sensors in IoT sites using machine learning approaches. Internet of Things, 7, p.100059.

[12] Li, K., Ma, W., Duan, H., Xie, H. and Juanxiu, Z.H.U., 2022. Few-shot IoT attack detection based on RFP-CNN and adversarial unsupervised domain-adaptive regularization. Computers & Security, 121, p.102856.

[13] Ma, W., Ma, L., Li, K. and Guo, J., 2023. Few-shot IoT attack detection based on SSDSAE and adaptive loss weighted meta residual network. Information Fusion, 98, p.101853.

[14] Mbarek, B., Ge, M. and Pitner, T., 2021. Proactive trust classification for detection of replication attacks in 6LoWPAN-based IoT. Internet of Things, 16, p.100442.

[15] Pandey, N. and Mishra, P.K., 2023. Performance analysis of entropy variation-based detection of DDoS attacks in IoT. Internet of Things, 23, p.100812.

[16] Rathore, S. and Park, J.H., 2018. Semi-supervised learning based distributed attack detection framework for IoT. Applied Soft Computing, 72, pp.79-89.

[17] Regan, C., Nasajpour, M., Parizi, R.M., Pouriyeh, S., Dehghantanha, A. and Choo, K.K.R., 2022. Federated IoT attack detection using decentralized edge data. Machine Learning with Applications, 8, p.100263.

[18] Sharma, B., Sharma, L., Lal, C. and Roy, S., 2023. Anomaly based network intrusion detection for IoT attacks using deep learning technique. Computers and Electrical Engineering, 107, p.108626.

[19] Simpson, S.V. and Nagarajan, G., 2021. A fuzzy based co-operative blackmailing attack detection scheme for edge computing nodes in MANET-IOT environment. Future Generation Computer Systems, 125, pp.544-563.

[20] Sivasankari, N. and Kamalakkannan, S., 2022. Detection and prevention of man-in-the-middle attack in iot network using regression modeling. Advances in Engineering Software, 169, p.103126.

[21] Thamilarasu, G. and Chawla, S., 2019. Towards deep-learning-driven intrusion detection for the internet of things. Sensors, 19(9), p.1977.

[22] Zhang, Y., Li, P. and Wang, X., 2019. Intrusion detection for IoT based on improved genetic algorithm and deep belief network. IEEE Access, 7, pp.31711-31722.