Journal of Theoretical and Applied Information Technology

15th November 2023. Vol.101. No 21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A NOVEL INTRUSION DETECTION SYSTEM (IDS) FRAMEWORK FOR AGRICULTURAL IOT NETWORKS

SANTOSH KONDE[1], Dr. S.B. DEOSARKAR[2]

[1] Research Scholar, Department of Electronics and Telecommunication Engineering, Dr. Babasaheb Ambedkar Technological University, Lonere Dist. Raigad (M.S.)

[2] Professor. Department of Electronics and Telecommunication Engineering, Dr. Babasaheb Ambedkar Technological University, Lonere Dist. Raigad (M.S.)

E-mail: [1]skonde76@gmail.com, [2]sbdeosarkar@dbatu.ac.in

ABSTRACT

The Tremendous growth of the Agricultural Internet of Things (IOT) applications has required a huge amount of network data and created high computational complexity across various connected devices. IOT devices capture valuable information that enables users to make critical decisions dependent on live streaming. Most of these Agricultural IOT devices have resource limitations such as low CPU, limited memory, and low energy storage. Thus, these devices are vulnerable to attacks due to the lack of capacity to run existing security software. This creates an inherent risk in Agricultural IOT networks. This has resulted in attackers having more incentive to target IOT devices. If the hackers attacking on the networks; the traditional intrusion detection system (IDS) cannot detect threats effectively. Therefore, there is a need to develop effective IDS using machine learning (ML) techniques in the Agricultural IOT networks.

In this paper, we propose IDS, which is a combination of feature selection and classification. We suggest using Pearson's correlation coefficient based feature selection and K-Nearest Number (KNN) based classification method to detect attacks in an Agricultural IOT networks. This will increase the accuracy of the classification and reduce the complexity of the system by extracting only nineteen key features from the original Forty One features in the dataset. The performance assessment of the proposed IDS was conducted using tests conducted on the intrusion benchmark dataset NSL-KDD. In this work, we compared the proposed IDS with other ML models including Support Vector Machine (SVM), Decision Tree, Naive Bayes, Random Forests and KNN. Additionally, we used the most important performance indicators, namely, accuracy, precision, recall and F1 score, to test the effectiveness of proposed IDS. The results obtained show that our proposed IDS can effectively reduce the number of features with higher classification accuracy compared to other ML-based classification methods.

Keywords: *Internet of things (IOT), Machine learning (ML), Intrusion detection system (IDS), Dataset, Algorithm, framework.*

1. INTRODUCTION

IOT is a set of interconnected devices augmented with lightweight processors and network cards that can be managed through web services or other types of interfaces. The Agricultural IOT offers a vision where devices can use sensors to understand the context and connect through network with one-another. Devices in an Agricultural IOT network can be used to collect information based on use cases. These include retail, healthcare, agriculture, and manufacturing industries that use Agricultural IOT devices for tasks such as tracking purchased items, remote patient monitoring, and fully autonomous warehouses. Agricultural IOT is the integration of advanced technologies into existing agricultural operations to improve the quality and productivity of agricultural products. Nowadays, the Agricultural IOT application has been deployed for Agriculture using wireless sensor networks (WSN) such as, Supply chain management, smart monitoring, Smart water and Agrochemicals applications, Disease management, and smart harvesting. As the deployment of thousands of Agricultural IOT-based devices is in the open field, there are possibilities to inject many new threats in Agricultural IOT agriculture. Any new technology

that is widely adopted by the public naturally attracts the interest of attackers to exploit it using various complex hacking techniques such as botnets. Added to this is a lack of standardization in Agricultural IOT systems, as well as the cheap, light and low-power devices that make up many of these systems. When an attacker attempting to penetrate Agricultural IOT network, it use several different approaches such as Distributed Denial of service (DDoS), probing, information theft to disrupt the functioning of the Agricultural IOT devices. For example farmers maintained the soil pH by providing ammonium to soil. This information, an attacker can launch DDoS attacks to disrupt the pH parameters. The cyber attacks can be reduced by secured the pH level of soil data. To protect Agriculture IOT network from destruction, change, unauthorized access, or attack, the use of an IDS along with the authentication, access control, and integrity techniques is necessity. Security attacks against Agricultural IOT networks are categorized into two major groups: Active and Passive. In passive attacks, attackers are typically secret (unseen) and moreover tap the message link to accumulate data; or tear down the performance elements of the network. An adversary essentially affects the operations in the attacked network in active attacks and this may be the reason for the attack and can be detected. The traditional techniques used for detection of attack do not work well while working with large data flows. In order to protect the Agricultural IOT network from intrusion by an adversary, various IDS have been proposed by researchers. The IDS are signature based security dependent on known attack pattern. The other as anomaly based, this anomaly based defense solution is dependent on statically model .the data packets are verify at different time of interval. data packet are classify on the basis of technique as statistical modeling, machine learning(ML),Data mining etc. data packet are normal or anomaly can differentiated by these method.[1, 2, 3,4]

In this work, we propose IDS framework, which is a combination of feature selection and classification. We suggest using Pearson's correlation coefficient based feature selection and KNN based classification method to detect attacks in a Agricultural IOT network. This will increase the accuracy of the classification and reduce the complexity of the system by extracting only 19 key features from the original 41 features in the dataset. The performance assessment of the proposed framework was conducted using tests conducted on the NSL-KDD dataset. In this work, we compared the proposed IDS with other ML models including Random Forest (RF), KNN, Decision Tree (DT), Naïve Bayes (NB), and SVM. The results obtained show that our proposed framework can effectively reduce the number of features with higher classification accuracy compared to other ML-based classification methods.

The rest of this paper is organized as follows: Section 2 gives the details of IDS; Section 3 presents the literature study of recent IDS techniques employed for Agricultural IOT network. Section 4 discusses Proposed IDS Framework, Section 5 discusses the attack detection based on the proposed IDS framework, Section 6 discusses the particulars of data set, implementation, and experimental results. Finally, the concluding remarks of the study are provided in Section 7.

2. INTRUSION DETECTION SYSTEMS (IDS)

Intrusion can be defined as networks or system is not working by given implementation method, also an intrusion is defined as: “any set of actions that attempt to compromise the integrity, confidentiality, or availability of a resource”. IDS are a set of tools to facilitate distinguishing, evaluating, and describing intrusions. IDS is a defense system that can detect abnormal activities [5],[6]. IDS are considered as a defensive wall from the security point. IDS can be deployed along with other security measures such as access control, authentication mechanisms and encryption techniques to make systems more secure against threats. Using patterns of benign traffic or threat, IDS can distinguish between normal and malicious actions [7]. According to Dewa and Maglaras [8], data mining, which is used to describe knowledge discovery, can help implement and deploy IDS with higher accuracy and robust behavior compared to traditional IDS, which may not be as effective against modern sophisticated attacks [9]. A requirement for IDS is, "a low false positive rate and a high true positive rate". Network intruders can be divided into two types: external intruders and internal intruders. **(1) External Intruder :** An unknown who uses various attack methods to enter the network. **(2) Inside intruder :** A compromised node that used to be a network partner. IDS can detect both external and internal intruders, but internal intruders are more difficult to detect. This is because internal attackers have the necessary keying resources to counter any protection taken by the authentication mechanisms. Intrusions can be of any type, such as attempted break, masquerade,

penetration, leakage, DoS, and malicious use. IDS can provide partial detection solutions to these attacks. Perfect IDS that would be able to detect all the intrusions mentioned above [10], [11], [12]. Based on deployment, IDS can be divided into two types: host intrusion detection system (HIDS) and network intrusion detection system (NIDS). HIDS is distributed between the host to detect intrusions such as changes to system files, numerous attempts to access the host, abnormal memory allocation methods, unusual CPU activity or I/O activity. HIDS accomplishes this by monitoring the host's real-time traffic or by examining log files on the host. NIDS can examine the entire packet, payload inside a packet, IP address, or port by either passively or actively listening for network traffic. Based on detection methodologies, IDS can be classified as anomaly-based detection, misuse-based detection, and specification-based detection.

2.1 Misuse detection :

In this case, patterns need to be defined and given to the system. Behavior of nodes is compared to well-known attack patterns. The disadvantage is that this method requires knowledge to create attack patterns to detect new attacks and it also requires an up-to-date database.

2.2 Anomaly detection :

The prime approach describes the known "normal behavior" by using automated training. This method does not look for exact attack patterns but instead checks normal or abnormal behavior. IDS will have high confidence in deciding that a node is malicious if the sensor node is not acting according to the specific specification of a particular protocol. A disadvantage of this method is that the system may illustrate invalid behavior as valid.

2.3 Specification-based detection:

This pays attention to detecting deviations from normal behavior that are not defined by ML techniques or training data. Patterns that describe normal behavior are defined manually, and any action against these specifications is tracked. The disadvantage of this approach is that manual development of all specifications is a time-consuming process for humans and cannot detect malicious behavior that does not violate the defined specifications of the IDS. [13]

3. LITERATURE SURVEY OF ML BASED IDS

Diro et al. [14] presented a distributed IDS based on deep learning (DL) for IOT networks. The author's proposed to deploy this system on a fog computing layer for hosting IDS. Three intrusion datasets namely NSL-KDD, ISCX, and KDD-CUP-99 were used in the performance evaluation, in which the results show up to 97% accuracy.

Muna et al. [15] presented anomaly detection system (ADS) for detecting threats in the Industrial IoT. The ADS used an unsupervised deep auto encoder algorithm for pattern classification. NSL-KDD and UNSW-NB15 datasets were used in the evaluation. The experimental results show 99% accuracy.

Vinayakumar et al. [16] presented IDS against botnet threats based on scanning DNS services in smart city IOT applications. The presented mechanism uses a two-layer environment to monitor DNS logs and search the domain name generated by the domain generation algorithm using DL algorithms to increase accuracy.

Latif et al. [17] presented IDS that uses a lightweight random neural network to detect threats in the Industrial IOT. Compared with traditional ML approaches such as SVM, ANN, and DT, the presented system shows better accuracy. An open-source dataset DS2OS used in experimentation.

Parra et al. [18] presented a distributed architecture using two DL approaches, namely a distributed CNN and an LSTM network. DCNN is used in the IOT micro security plug in, while LSTM is used by the back-end server. The N-BaIOT dataset is used in the performance evaluation, where the results show an accuracy of 98% and 94.30% during the training phase and the testing phase.

Haddad Pajouh et al. [19] presented IDS using a recurrent neural network. The presented mechanism uses three stages; namely data collection, feature extraction and deep threat classifier. Experimental results demonstrate the highest accuracy of 98.18% in 10-fold cross-validation analysis and performance compared to conventional ML classifiers such as Naive Bayes, KNN, RF, and DT.

Koroniotis et al. [20] presented a network forensic scheme called PDF to detect and monitor threat patterns in IOT networks. Identification of anomalous incidents, adaptation of DL parameters and data extraction is three stages used in PDF

scheme. In the second phase, a particle swarm optimization (PSO) algorithm is used, while in the third phase, a deep neural model is used. Experimental results show an accuracy of 99.9% compared to 93.2% with the DT and 72.7% with naive bays.

Selvakumar et al. [21] presented an IDS based on vector convolution DL approach. The authors also presented that the calculations be processed in fog nodes. Internet of medical things(IMOT)find out the threats performed the experiment on the BOT-IOT dataset which shows accuracy up to 99.92%.

Manimurugan et al. [22] presented IDS using a deep belief approach. Port scan threat shows the accuracy of 97.71% and the infiltration threat shows an accuracy of 96.37% when using the CICIDS2017dataset.

Popoola et al. [23] presented a hybrid IDS, called LAE-BLSTM, to detect botnets in IOT networks. The LAE-BLSTM mechanism uses a deep bidirectional long-short-term memory (BLSTM) and a long-term memory auto encoder (LAE). LAE is used for feature dimensionality reduction, while BLSTM is used to identify botnet threat traffic in IOT networks. The Bot-IOT dataset used in the performance evaluation, showing that the LAE-BLSTM mechanism achieved a data size reduction ratio of 91.89%.

Basati et al [24] presented IDS called deep feature extraction. This model is based on CNN. The authors focused mainly on those devices that have low computing power. The authors used the UNSW-NB15, CICIDS2017 and KDD-Cup99 datasets for their experiments. The model was tested for both binary and multiclass classification.

Rashid et al [25] presented a tree-based stacking ensemble approach for intrusion detection in IOT. Two intrusion datasets, NSL-KDD and UNSW-NB15, were used to evaluate the effectiveness of the presented model. The authors also improved efficiency by integrating feature selection strategies to identify the most important features.

Fatani et al [26] presented feature engineering for an IDS system taking advantage of the swarm intelligence (SI) approach. Current IDS mechanism used the popular public dataset as BOT-IOT,NSL-KDD ,KDD99 and CIC2017 for calculating the accuracy.

Alkahtani et al [27] presented three advanced and widely used DL models for intrusion detection. The authors conducted experiments with LSTM, CNN, and a hybrid CNN–LSTM model. To evaluate these DL models, the authors used the IoTID20 dataset.

Keserwani et al [28] presented a method for extracting significant IOT network features for intrusion detection. The presented method consists of a combination of grey wolf optimization and PSO. The authors used the KDDCup99, NSL-KDD and CICIDS-2017 datasets.

Qaddoura et al [29] presented single-layer feed forward neural network to detect threat IoT networks. The authors used data reduction with clustering and SMOTE oversampling approach.

Saba et al [30] presented a two-stage hybrid approach for detecting malicious threats in IOT networks. A genetic algorithm as well as well-known ML approaches such as SVM, ensemble classifier, and DT were used to select relevant features.

Majjed et al. [31] presented DL approach STL-IDS. For dimensionality reduction, the presented system can be used. In this approach, both training and testing time are reduced to achieve higher prediction accuracy of SVM.

Sandhya Peddabachigari et al. [32] evaluated a DT for intrusion detection. DT based intrusion detection was tested on a 1998 DARPA dataset, and the system outperforms traditional models in terms of accuracy. Again, the results show that the training and testing times are better compared to the SVM.

Mrutyunjaya Panda et al. [33] presented a NIDS framework based on Naïve Bayes. For the implementation, KDD Cup 99 is used as the dataset and from the results it is found that the planned system offers higher performance in terms of false positive rate, procedure time and cost.

4. PROPOSED IDS FRAMEWORK

In the field of ML, feature selection is the process of reducing the number of input variables when constructing a hypothetical model. It is desirable to reduce the number of input variables to reduce the computational costs of modelling. Methods of statistical-based feature selection include examining the relationship between individual input variables and target variables using statistics and selecting those input variables that have the strongest correlations with the target variable. Feature selection options can be divided into the filter, wrapper and embedded. Filter-based selection methods use statistical measures to determine the correlation or dependence between input variables that can be filtered to select the most suitable features. Correlation is a commonly used mathematical term meaning how close two variables are to having a linear relationship. Two or more variables linear relationship can find out by correlation method. Correlation can show one variable from the other variable. The idea behind using a correlation in selecting a feature is that the good variables are highly correlated with the target. In addition, variables should be associated with the target but should not be associated with each other. When two variables are linked, we can predict one another. Therefore, when two factors are correlated, the model requires only one of them, as the second does not add additional information. In this work, we used Pearson's correlation coefficient to quantify numerical features. In statistics, Pearson's correlation coefficient is the measure of linear correlation between two sets of data. Consider the ratio of the product of standard deviation and covariance of two variables. The covariance measurement can normalised and always show value in between -1 and +1. Here, we select nineteen of the most important features out of the Forty One found in the NSL-KDD dataset.

As shown in Fig. 1, we have divided our proposed IDS framework for detecting IOT network threats into three steps. The first phase in implementing lightweight IDS is to introduce an initial pre-processing stage for the dataset before training. To achieve this, we used Pearson's correlation coefficient to rank features. By ranking, the features that are strong in determining the output class of the dataset are obtained and nineteen rated features are selected. These selected features represent the most important features of all Forty One features. In the second phase, the training phase, the features selected after pre-processing the NSL-KDD dataset are used to train the proposed IDS to detect possible IOT network attacks. This proposed IDS is used in the IOT network to monitor data from the open field sensor to the cloud server. The final stage, the attack detection phase, is the process by which a non-labelled test dataset is used to test one of the class labels. In this task, we used the KNN classification algorithm to detect the occurrence of an attack. KNN algorithm classifies new objects based on similarity measures. To measure the similarity between different objects mathematical measure Euclidean Distance is used. In KNN algorithm, consider most frequently obtained classes; consider particular class for the test data. Every particular test data point will show the K – nearest training data points. so K-represent training data point and it shows proximity to test data point which can used for consider the particular class.

The steps of KNN algorithm are given below

Step 1: Decide the value of K.

Step 2: distance between query instance and all the training samples are calculated.

Step 3: Sort the distance in ascending order and confirm nearest neighbors supported the Kth minimum distance.

Step 4: nearest neighbors majority of class considered, fix the prediction value of the query instance. [38, 39]

5. ATTACK DETECTION BASED ON THE PROPOSED IDS FRAMEWORK

Fig. 1 shows the proposed IDS for detecting threats as shown below ; The attack detector is installed on 'wireless access points (gateway) or on fog layer nodes', the gateway that connects to the cloud server. In this work, we assume that the gateway is equipped with a better processing subsystem, sensing unit, radio subsystem, and power supply unit. Our proposed IDS are deployed on the gateway for intrusion detection. The gateway will monitor the open field Sensor Nodes (SNs) to detect attacks. Furthermore, the gateway will filter abnormal data and forward all the reliable sensed information to the cloud server for further processing, either directly or via one or more relay nodes. The IDS is deployed only on the gateway to conserve the battery energy of the open field SNs, which in turns prolong the network lifetime and functionality. The attack detector has two components pre-processor and a classifier.

5.1 Pre-processor

This module captures incoming data from open field SNs. Then, it converts captured traffic into a batch of samples. These groups are used in the

classifier to detect attacks. The first step is to extract the useful features and normalization, which scales the feature values in the range [0, 1]. The next step is to address the symbolic features of the data. One-hot encoding is used to convert symbolic features into discrete features. In this case, one feature value is represented as a binary value vector. After this, a feature selection is made, in this work; we use Pearson's correlation coefficient to rank the numerical features. The most relevant features that are strong in determining the output class are ranked and chosen to be used by the KNN algorithm to classify traffic as either normal or anomaly.

5.2 Classifier

The classifier used in the proposed IDS is KNN. It takes sample groups with 19 key features prepared by the pre-processor as input and separates each sample group as either normal or attack. It is a supervised model and requires training with labeled samples before it can be used for attack detection. To train the classifier, the NSL-KDD training dataset is used.

5.2.1. Attack detection

During the attack detection, the samples prepared by the pre-processor are applied to a trained classifier and output is calculated. The output of the model classifies each sample as an attack or normal. When the samples applied to a classifier are classified as an attack, it indicates that the attack was launched in the IOT network. Here, the gateway will reject that attacked data.

6. EXPERIMENTATION

We use NSL-KDD to test the performance of Proposed IDS with other ML-based IDS. Researchers have downloadable files at the disposal which shown in Table no. 1 as shown below. The tests were performed in Google Co laboratory under Python 3 using Tensor Flow and Graphics Processing Unit (GPU).

6.1 Dataset Description

NSL-KDD dataset is proposed by Tavallaee et al. [34] and is recommended to solve some of the inherent problems of the KDD'99 dataset. The researchers have developed many IDS models, analyses the statistical detection accuracy but affect the accuracy due to the internal drawback of KDD cup 99. NSL-KDD data set [35] is a developed version of its predecessor. It include essential records of the complete KDD data set. As consider to the original KDD dataset, the NSL-KDD dataset has the following development: (1) unnecessary records are removed to enable the classifiers to produce impartial result, (2) exact copy of records are removed, (3) the number of selected records is arranged as the percentage of records (e.g. DDTrain+_20Percent.ARFF), and (4) necessary number of records is available in the train and test data sets, which is practically rational and enables to shoe the result of experiments on the complete set (5) The number of selected records from each not easy level of group is inversely proportional to the percentage of records in the original KDD data set [36]. In every particular record 41 attributes are unfolding different features of the flow and The 42nd attribute is a label assigned to each one or the other of two as an attack-type (Probe, DoS, R2L, and U2R) or as normal [34][37]. The particular types of attacks are differentiated into four major categories. Table 2 shows this detail as mention below.

Table 3 shows the distribution of the normal and attack records available in the various NSL-KDD datasets. [36] as mentioned below.

6.2 IDS methodology used in experimentation

The details of the Proposed IDS used in experimentation are illustrated in Fig. 2. Shown below Specifically, the method consists of five stages: (1) datasets stage, (2) pre-processing stage, (3) feature Selection stage, (4) training stage and (5) testing stage.

6.3 Performance Metrics

We used key performance indicators including accuracy, precision, recall and F1 score.

Accuracy: It is a metric used to indicate the proportion of correct classifications on the total records in the test set.

Accuracy = (TP+ TN)/ (TP+ FN+ TN+ FP)

Precision (P): It is a metric that measures the actual performance within the desired response space, i.e. between positions.

P =TP/(TP + FP)

Recall (R): It is a metric that measures how many predicted responses were discarded, or for each correct label, how many other true labels we discarded.

R = TP/ (TP + FN)

F1 (F) score: It is the harmonic mean of two matrices P and R.

F = (2 * P * R)/ (P + R)

Where,

True Positive (TP): cases of anomaly correctly categorized as anomaly.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

False Positive (FP): cases of a normal class miscategorised as an anomaly.

True Negative (TN): cases of a normal class and correctly categorized as normal.

False negative (FN): cases of abnormality misclassified as normal. [7]

6.4 Results and Discussion

In this work, in the first phase, Pearson's correlation coefficient used during the pre-processing stage to select features from the labeled dataset, NSL-KDD. The most relevant features that are strong in determining the output class are ranked and chosen to be used by the KNN algorithm to classify traffic as either normal or anomaly. During the evaluation, we determine the performance of our proposed IDS framework by using an NSL-KDD dataset. The output of the feature selection process is shown in Table 4 below

In the second phase, the training phase, the features selected after pre-processing of the NSL-KDD dataset are used to train the IDS to detect possible attacks in the IOT network. In this work, we used the KNN classification algorithm to detect the occurrence of an attack.

To compare the Proposed IDS, five algorithms of ML were considered, namely SVM, Naive Bayes, Random Forest, Decision Tree, and KNN. For comparison purposes, accuracy, precision, recall, and F1 score were considered, and their comparison results are shown in Table 5 shown below. We can say that the accuracy of the proposed IDS is highest as compared to other approaches.

7. CONCLUSION

In this paper, we have proposed the IDS framework which is a combination of feature selection and classification. Feature Selection is based on the Pearson's correlation coefficient technique, to pre-process dataset before attack classification. The proposed IDS reduce the complexity of the system by selecting important features in the dataset, thus reducing the features from Forty One to Nineteen before classification, using a KNN algorithm. Experimental results obtained show improved performance with a reduced feature set from Forty One to Nineteen.

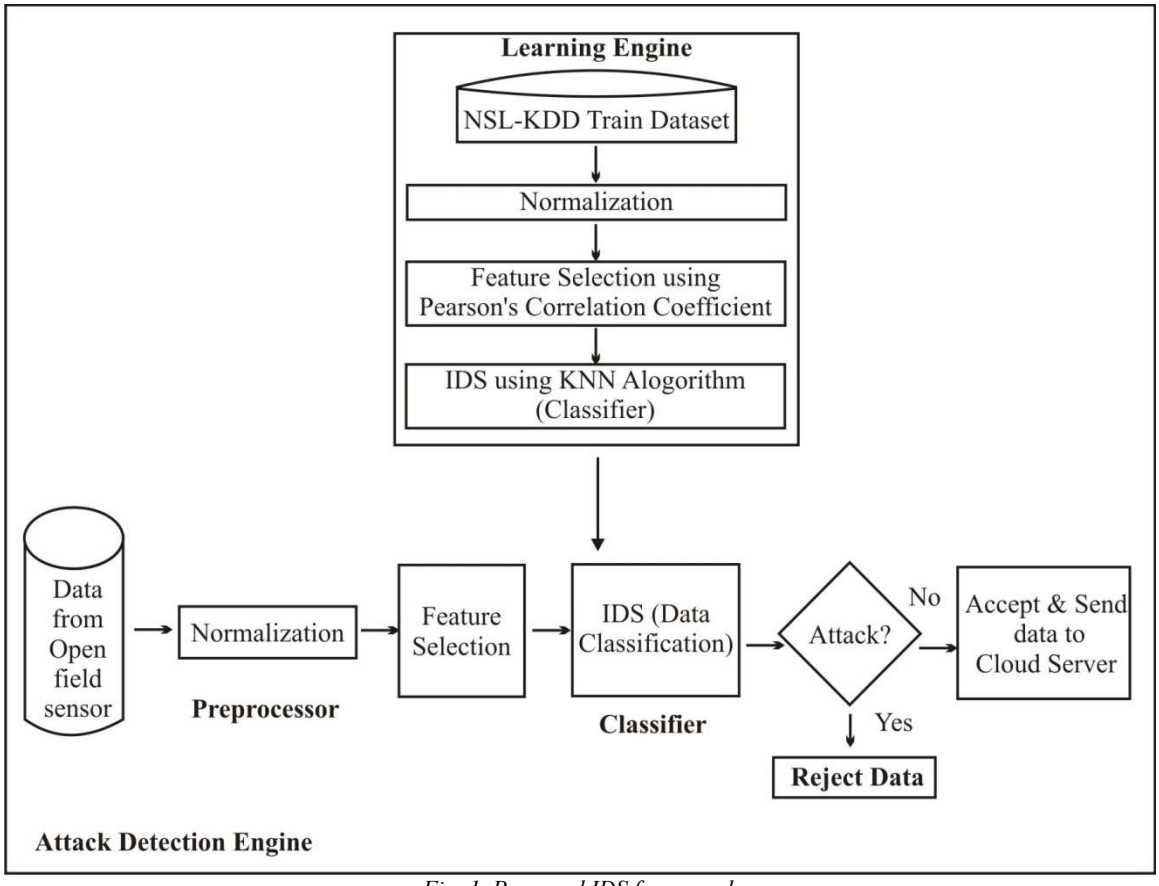

Fig. 1. Proposed IDS framework

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 1: description and list of files in NSL-KDD dataset.

Sr.	Name of the file	Description
1	KDDTrain+.ARFF	The full NSL-KDD train set with binary labels in ARFF format
2	KDDTrain+.TXT	The full NSL-KDD train set including attack-type labels and difficulty level in CSV format
3	KDDTrain+_20Percent.ARFF	A 20% subset of the KDDTrain+.arff file
4	KDDTrain+_20Percent.TXT	A 20% subset of the KDDTrain+.txt file
5	KDDTest+.ARFF	The full NSL-KDD test set with binary labels in ARFF format
6	KDDTest+.TXT	The full NSL-KDD test set including attack-type labels and difficulty level in CSV format
7	KDDTest-21.ARFF	A subset of the KDDTest+.arff file which does not include difficulty level of 21 out of 21of records
8	KDDTest-21.TXT	A subset of the KDDTest+.txt file which does not include Difficulty level of 21 out of 21of records.

Table 2: attack type and attack class mapping

Attack Class	Attack Type
DoS	Land,Back, Neptune, Pod, Smurf,Apache2,Teardrop, Udpstorm, Processtable, Worm
Probe	Satan, Ipsweep, Nmap, Portsweep, Mscan, Saint
R2L	Guess_Password, Ftp_write,Phf, Imap, Multihop, Warezmaster, Spy,Warezclient, Xlock, Xsnoop, Snmpguess, Snmpgetattack, Httptunnel,Named, Sendmail,
U2R	Buffer_overflow, Loadmodule,Perl, Rootkit, Sqlattack,Ps, Xterm,

Table 3: NSL – KDD data set types, details of attack data and normal data

Dataset Type	Total No. of					
	Records	Normal Class	DoS Class	Probe Class	U2R Class	R2L Class
KDD Train+ 20%	25192	13449	9234	2289	11	209
		53.39%	36.65%	9.09%	0.04%	0.83%
KDD Train+	125973	67343	45927	11656	52	995
		53.46%	36.46%	9.25%	0.04%	0.79%
KDD Test+	22544	9711	7458	2421	200	2754
		43.08%	33.08%	10.74%	0.89%	12.22%

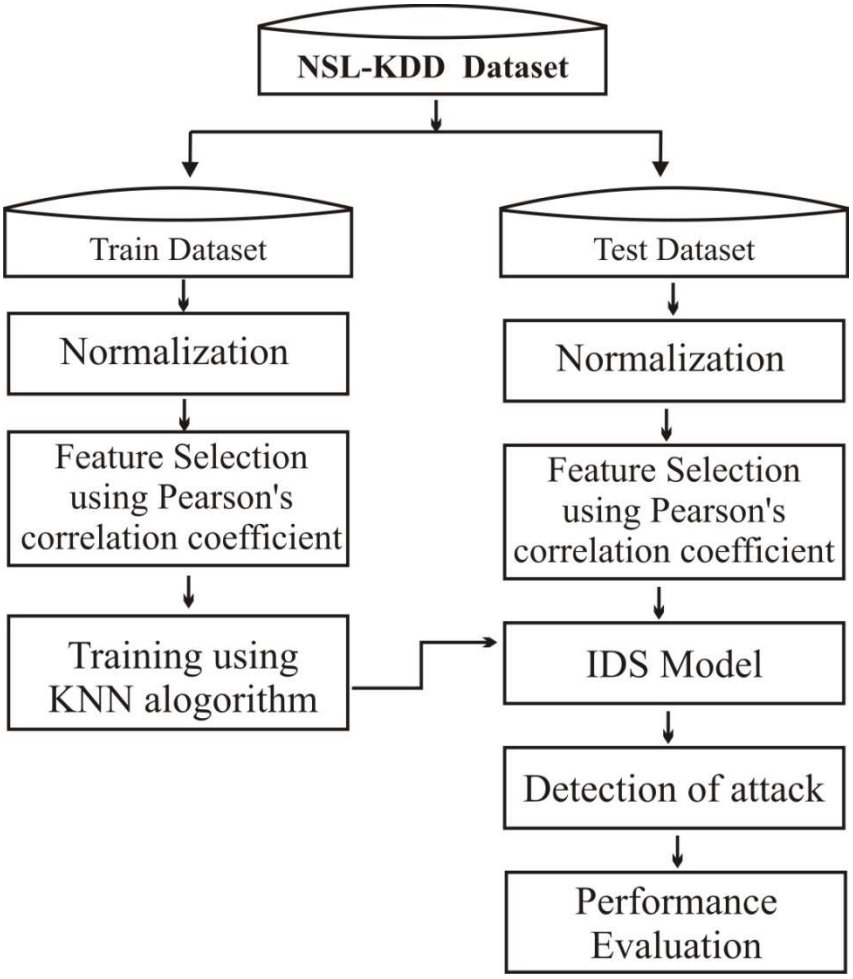

Fig. 2. Flowchart of the proposed IDS framework

Table 4: Output of Pearson's correlation coefficient based Feature Selection Method

Feature selection method	**Feature selected**
Pearson's correlation coefficient	**'dst_bytes','same_srv_rate','src_bytes','dst_host_same_srv_rate','logged_in',' dst_host_srv_count','dst_host_srv_serror_rate','dst_host_serror_rate','serror_ rate','srv_serror_rate','count','dst_host_diff_srv_rate','rerror_rate','dst_host_ srv_rerror_rate','srv_rerror_rate','dst_host_rerror_rate','protocol_type','service',' flag'**

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 5: Comparison of the proposed IDS framework with other ML-based IDS models

Algorithm	Accuracy (overall)	Precision		Recall		F1 Score	
		Attack	Normal	Attack	Normal	Attack	Normal
Proposed IDS Framework	79.24	0.66	0.97	0.97	0.68	0.78	0.80
KNN	77.63	0.63	0.97	0.96	0.66	0.76	0.79
SVM	76.55	0.66	0.90	0.90	0.67	0.76	0.77
Decision Tree	74.17	0.62	0.91	0.90	0.64	0.73	0.75
Random forest	73.24	0.55	0.98	0.97	0.62	0.70	0.76
Naive Bayes	51.17	0.90	0.00	0.54	0.02	0.68	0.00

REFERENCES:

[1] Mohamed Amine Ferrag, Lei Shu, Hamouda Djallel and Kim-Kwang Raymond Choo, *"Deep Learning-Based Intrusion Detection for Distributed Denial of Service Attack in Agriculture 4.0"*, Electronics 2021, 10, 1257. https://doi.org/10.3390/electronics10111257

[2] Safi Ullah, Jawad Ahmad, Muazzam A. Khan, Eman H. Alkhammash, Myriam Hadjouni, Yazeed Yasin Ghadi, Faisal Saeed and Nikolaos Pitropakis, *"A New Intrusion Detection System for the Internet of Things via Deep Convolutional Neural Network and Feature Engineering"*, Sensors 2022, 22, 3607. https://doi.org/10.3390/s22103607.

[3] Ansam Khraisat and Ammar Alazab, *"A critical review of intrusion detection systems in the internet of things: techniques, deployment strategy, validation strategy, attacks, public datasets and challenges"*, Khraisat and Alazab Cybersecurity (2021) 4:18

[4] Eric Gyamfi and Anca Jurcut, *"Intrusion Detection in Internet of Things Systems : A Review on Design Approaches Leveraging Mult i -Access Edge Computing, Machine Learning, and Datasets"*, Sensors 2022 , 22, 3744. https://doi.org/10.3390/s22103744

[5] M. Ngadi, A. H. Abdullah and S. Mandala, *"A survey on MANET intrusion detection"*, International Journal on Computer Science and Security, volume 2, number 1, pages 1-11, 2008.

[6] Y. Zhang, W. Lee and Y.A. Huang, *"Intrusion detection techniques for mobile wireless networks"*, Journal on Wireless Networks, vol. 9, num. 5, pp.545-556, 2003.

[7] Mohamed Amine Ferrag, Leandros Maglaras, Sotiris Moschoyiannis and Helge Janicke, *"Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study"*, Journal of Information Security and Applications, 50 (2020) 102419.

[8] Dewa Z and Maglaras L A, *"Data mining and intrusion detection systems"*, International Journal on Advanced Computer Science Applications 2016; 7(1):62–71.

[9] Stewart B , Rosa L , Maglaras L A , Cruz T J , Ferrag M A and Simões P, *"A novel intrusion detection mechanism for scada systems which automatically adapts to network topology changes"*, EAI Endorsed Trans. Ind. Netw. Intell. Syst. 2017; 4(10):e4.

[10] T. S. Sobh, *"Wired and wireless intrusion detection system: Classifications, good characteristics and state-of-the-art"*, Elsevier Journal on Computer Standards and Interfaces, volume 28, number 6, pages 670-694, 2006.

[11] T. Anantvalee and J. Wu, *"A survey on intrusion detection in mobile ad hoc networks"*, Springer Journal on Wireless Network Security, pages 159-180, 2007.

[12] P. Albers, O. Camp, J. Percher, B. Jouga, L. M. and R. Puttini, *"Security in Ad Hoc Networks: A General Intrusion Detection Architecture Enhancing Trust Based Approaches,"* Proc. 1st International Workshop on Wireless Information Systems (WIS-2002), pp. 1-12, April 2002.

[13] Rakesh Sharma and Vijay Anant Athavale, *"A Survey of Intrusion Detection Techniques and Architectures in Wireless Sensor Networks"*, International Journal on Advanced Networking and Applications, Volume: 10 Issue: 04 Pages: 3925-3937, 2019.

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

[15] Muna A H, Moustafa N and Sitnikova E, *"Identification of malicious activities in industrial internet of things based on deep learning models"*, Journal of Information Security and Applications, 2018, 41, 1–11.

[16] Vinayakumar R, Alazab M, Srinivasan S, Pham QV, Padannayil SK and Simran K A, *"visualized botnet detection system based deep learning for the Internet of Things networks of smart cities"*, IEEE Transactions on Industry Applications, 2020, 56, 4436–4456.

[17] Latif S, Zou Z, Idrees Z and Ahmad J A, *"Novel Attack Detection Scheme for the Industrial Internet of Things Using a Lightweight Random Neural Network"*, IEEE Access, 2020, 8, 89337–89350.

[18] Parra G D L T, Rad P, Choo K K R and Beebe N, *"Detecting Internet of Things attacks using distributed deep learning"*, Journal of Network and Computer Applications, 2020, 163, 102662.

[19] HaddadPajouh H, Dehghantanha A, Khayami R and Choo KKR, *"A deep recurrent neural network based approach for internet of things malware threat hunting"*, Future Generation Computer Systems, 2018, 85, 88–96.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[20] Koroniotis N, Moustafa N and Sitnikova E, *"A new network forensic framework based on deep learning for Internet of Things networks: A particle deep framework"*, Future Generation Computer Systems, 2020, 110, 91–106.

[21] Bhuvaneswari Amma N G and Selvakumar S, *"Anomaly detection framework for Internet of things traffic using vector convolutional deep learning approach in fog environment"* Future Generation Computer Systems, 2020, 113, 255–265.

[22] Manimurugan S, Al-Mutairi S, Aborokbah M M, Chilamkurti N, Ganesan S and Patan R, *"Effective Attack Detection in Internet of Medical Things Smart Environment Using a Deep Belief Neural Network"*, IEEE Access 2020, 8, 77396–77404.

[23] Popoola S I, Adebisi B, Hammoudeh M, Gui G and Gacanin H, *"Hybrid Deep Learning for Botnet Attack Detection in the Internet of Things Networks"*, IEEE Internet Things,2021, 8, 4944–4956.

[24] Basati A and Faghih M M, *"DFE: Efficient IoT network intrusion detection using deep feature extraction"*, Neural Comput Appl 2022,1–21.

[25] Rashid M, Kamruzzaman J, Imam T, Wibowo S, Gordon S, *"A tree-based stacking ensemble technique with feature selection for network intrusion detection"*, Appl Intell 2022, 1–14.

[26] Fatani A, Dahou A, Al-Qaness M A, Lu S, Abd Elaziz M, *"Advanced Feature Extraction and Selection Approach Using Deep Learning and Aquila Optimizer for IoT Intrusion Detection System"*, Sensors 2022, 22, 140.

[27] Alkahtani H and Aldhyani T H, *"Intrusion detection system to advance internet of things infrastructure-based deep learning algorithms"*, Complexity 2021, 2021, 5579851.

[28] Keserwani P K, Govil M C, Pilli E S and Govil P, *"A smart anomaly-based intrusion detection system for the Internet of Things (IoT) network using GWO–PSO–RF model"*, J Reliab Intell Environ 2021, 7, 3–21.

[29] Qaddoura R, Al-Zoubi A, Almomani I and Faris H, *"A multi-stage classification approach for iot intrusion detection based on clustering with oversampling"*, Appl Sci 2021, 11, 3022.

[30] Saba T, Sadad T, Rehman A, Mehmood Z and Javaid Q, *"Intrusion detection system through advance machine learning for the internet of things networks"*, IT Prof 2021, 23, 58–64.

[31] Al-Qatf M, Lasheng Y, Alhabib M & Al-Sabahi K. (2018), *"Deep learning approach combining sparse auto encoder with SVM for network intrusion detection"*, IEEE Access. https:// doi.org/10.1109/ACCESS.2018.2869577.

[32] Peddabachigari S, Abraham A, Thomas J, (2016), *"Intrusion detection systems using decision trees and support vector machines"*, International Journal of Advanced Networking and Applications, 07(04), 2828–2834. ISSN: 0975-0290.

[33] Mahbub M, Gazi M S A, Provar S A A, Islam M S, *"Multi-Access Edge Computing-Aware Internet of Things: MEC-IoT"*, In Proceedings of the 2020 Emerging Technology in Computing, Communication and Electronics (ETCCE), London, UK, 19–20, August 2020; pp. 1–6.

[34] Tavallaee M, Bagheri E, Lu W, Ghorbani A. A. *"A detailed analysis of the kdd cup 99 data set"*, In: 2009 IEEE Symposium on Computational Intelligence for Security and Defence Applications. IEEE; 2009. p. 1–6.

[35] Nsl kdd. https : // www.unb.ca / cic / datasets /nsl.html

[36] L. Dhanabal, and Dr. S.P. Shantharajah, *"A Study on NSL-KDD Dataset for Intrusion Detection System Based on Classification Algorithms"*, international Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 6, June 2015.

[37] Sapna S. Kaushik, Dr. Prof. P. R. Deshmukh, *"Detection of Attacks in an Intrusion Detection System"*, International Journal of Computer Science and Information Technologies, Vol. 2 (3), 2011, 982-986

[38] Opeyemi Osanaiye, Olayinka Ogundile, Folayo Aina, Ayodele Periola, *"FEATURE SELECTION FOR INTRUSION DETECTION SYSTEM IN A CLUSTER-BASED HETEROGENEOUS WIRELESS SENSOR NETWORK"*, FACTA UNIVERSITATIS, Series: Electronics and Energetics Vol. 32, N o 2, June 2019, pp. 315-330, https://doi.org/10.2298/FUEE1902315O.

[39] Manna A, Alkasassbeh M., *"Detecting network anomalies using machine learning and SNMP-MIB dataset with IP group"*, In: 2019 2nd International Conference on new Trends in Computing Sciences (ICTCS). IEEE; 2019. p. 1–5.

[40] Zhou X, Hu Y, Liang W, Ma J and Jin Q, *"Variational LSTM enhanced anomaly detection for industrial big data"* IEEE

Transactions on Industrial Informatics, 2020, 17, 3469–3477.

[41] Niyaz Q, Sun W, Javaid A Y & Alam M, *"A deep learning approach for network intrusion detection system"*, In BICT 2015, New York City, United States.

[42] Xu L, Jurcut A D, Ahmadi H, *"Emerging Challenges and Requirements for internet of things in 5G. In 5G Enabled Internet Things"*, CRC Press: Boca Raton, FL, USA, 2019, pp. 29–48.

[43] Said O, Tolba A, *"Accurate performance prediction of IoT communication systems for smart cities: An efficient deep learning based solution"*, Sustain. Cities Soc. 2021, 69, 102830.

[44] Pawar P and Trivedi A, *"Device-to-device communication based IoT system: Benefits and challenges"*, IETE Tech. Rev. 2019,36, 362–374.

[45] Van N. T, Thinh T. N & Sach L. T., *"An anomaly-based network intrusion detection system using deep learning"*, In 2017 International Conference on System Science and Engineering (ICSSE).

[46] Yang Y, Zheng K, Wu C, Niu X, Yang Y, *"Building an Effective Intrusion Detection System Using the Modified Density Peak Clustering Algorithm and Deep Belief Networks"*, Applications Science 9, 238 (2019).

[47] Michael Riecker, Sebastian Biedermann, Rachid El Bansarkhani and Matthias Hollick, *"Lightweight energy consumption-based intrusion detection system for wireless sensor networks"*, International Journal of Information Security, vol. 14, no. 2, pp. 155-167, 2015.

[48] Mohammad Wazid and Ashok Kumar Das, *"An Efficient Hybrid Anomaly Detection Scheme Using K- Means Clustering for Wireless Sensor Networks"*, Wireless Personal Communications, vol. 90, no. 4, pp. 1971-2000, October 2016.

[49] Ahmed Saeed, Ali Ahmadinia, Abbas Javed and Hadi Larijani, *"Random Neural Network based Intelligent Intrusion Detection for Wireless Sensor Networks"*, In proceedings of International Conference on Computational Science, vol. 80, pp. 2372-2376, 2016.

[50] Yassine Maleh, Abdellah Ezzati, Youssef Qasmaoui and Mohamed Mbida, *"A Global Hybrid Intrusion Detection System for Wireless Sensor Networks"*, The fifth International Symposium on Frontiers in Ambient and Mobile Systems, vol. 52, pp. 1047-1052, 2015.

[51] Jurcut A D, Ranaweera P, Xu L, *"Introduction to IoT security. In IoT Security: Advances in Authentication"*, John Wiley & Sons: Hoboken, NJ, USA, 2020; pp. 27–64..

[52] P. R. Chandre, P. N. Mahalle, and G. R. Shinde, *"Deep Learning and Machine Learning Techniques for Intrusion Detection and Prevention in Wireless Sensor Networks: Comparative Study and Performance Analysis"*, Lecture Notes in Networks and Systems 82, https://doi.org/10.1007/978-981-13-9574-1_5

[53] Nicholas Lee, Shih Yin Ooi and Ying Han Pang, *"A Sequential Approach to Network Intrusion Detection"*, Lecture Notes in Electrical Engineering 603, https : // doi.org /10.1007/978-981-15-0058-9_2

[54] Kishor Kumar Gulla, P. Viswanath, Suresh Babu Veluru, and R. Raja Kumar, *" Machine Learning Based Intrusion Detection Techniques"*, Handbook of Computer Networks and Cyber Security, https://doi.org/10.1007/978-3-030-22277-2_35

[55] Nickolaos Koroniotis, Nour Moustafa , Elena Sitnikova, Benjamin Turnbull, *"Towards the development of realistic botnet dataset in the Internet of Things for network forensic analytics: Bot-IOT dataset"*, Future Generation Computer Systems 100 (2019) 779–796.