Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.7762179

IOT_SLINK: ATTACK DETECTION IN IOT NETWORK USING BIO-INSPIRED OPTIMIZATION AND ENSEMBLE MACHINE LEARNING TECHNIQUES

P.Golda Jeyasheeli[1, a], G. Priyanka[1, b], K. Chithra Mala[1, c], Purnima Murali Mohan[2,d], Member IEEE, Zurina Mohd. Hanapi[3,e]
[1]Mepco Schlenk Engineering College (Autonomous), Sivakasi – 626005, India
[2] Singapore Institute of Technology, Singapore
[3]Universiti Putra Malaysia, Malaysia

Abstract—Internet of things (IoT) and Machine to Machine (M2M) technology has become indispensable in everyday life, finding real-time applications in smart tracking/monitoring services, wearable technologies, telemetry, autonomous vehicles, agriculture, healthcare, transportation, etc. It employs millions of sensors and devices connected to automate the tasks to be accomplished by collecting data, activating sensors and communicating between the devices either through the internet or point-to-point connections. Consequently, the smart device usage and the connected devices in the IoT network is increasing exponentially with limitless potential. However, with the increasing connectivity comes the increasing attack surface with attack entry points from different sources in the network. Hence security related risk and network anomalies need to be detected earlier in-order to prevent the devices in the IoT network from bringing down the whole network. In our proposed system, two-layers IoT Stacked Ensemble Learning Network (IoT_SLINK) is designed for attack detection in the IoT network at the source with feature selection using machine learning algorithms. This work focuses on detecting attacks originating from OSI layer 1 to layer 6based on the Random Forest (RF), Decision Tree (DT), Navie Bayes (NB) and K-Nearest Neighbour (KNN)machine learning algorithms. The feature selection algorithms are used to improve the model's performance in terms of accuracy, reduce overfitting and reduce the training time. The prominent features from the datasets are selected by using Particle Swarm Optimization (PSO) algorithm and Improved Salp Swarm Optimization Algorithm (ISSA) feature selection technique. The performance of the proposed system is evaluated with three IoT benchmark datasets IoT23, BoT-IoT and Distributed Smart Space Orientation System (DS2OS).The proposed system outperforms state-of-art techniques with more than 99% of accuracy with very low false positive rate.

Index Terms—Internet of Things (IoT), Machine to Machine (M2M) technology, attack detection, Machine Learning (ML), Ensemble Learning.

I. INTRODUCTION

NTERNETof Things (IoT) connects large number of physical objects such as sensors, mobiles, computers, home appliances, cars, agricultural gadgets, health monitoring devices and so on. These devices communicate with each other and the information is shared across the network without human interaction through the internet or direct point-to-point (P2P) connectivity. Such IoT devices can be controlled and monitored remotely and the data can be stored and processed to make decisions. But the increase in connected devices and data sharing makes the IoT network vulnerable and the hackers may exploit the network by injecting malicious viruses

and attacks such as DDoS [4], Torii [35], Okiru [36] and mirai[20] attacks. So, it is essential to keep the network safe and secure from the external attacks and zero-day attacks by mitigating security related risks.

Machine Learning techniques are being used in the literature to tackle the security issues in IoT network [5-9]by learning from the past experience to produce a decision on the real-time traffic characteristics. Most of the existing works used supervised Machine Learning (ML) techniques for malware detection anticipating the future events based on the historical data [37]. These algorithms continuously learn and develop patterns by analysing the traffic data used for detecting malwares in the network. ML techniques in security system can detect new zero-day attacks as well as similar kinds of attacks by analysing and learning the traffic patterns of known attacks and respond to the future attacks [38].In machine learning, classification/predictions are done by a single MLmodel, with this the high-performance rate cannot be achieved and result can be biased. In order to achieve better attack detection performance, Ensemble Learning based approach has been proposed in this work.

Ensemble learning is a machine learning technique which produces the optimal and efficient predictive model by combining a set of machine learning classifiers. In the previous works, different machine learning algorithms and deep learning models are used for detecting different kinds of attacks such as Distributed Denial of Service Attack (DDoS), Spying, Scan, Data Probing, web attacks etc. In this proposed work, the ensemble machine learning model is built with the Random Forest (RF), Decision Tree (DT), Naive Bayes (NB), K-Nearest Neighbor (KNN) as base classifiers to detect attacks in the IoT network. The attacks such as port scan attack, okiru, DDoS, C&C attack, HeartBeat, data theft, etc., from each OSI layers are considered. The open source IoT23, BoT-IoT, DS2OS datasets has been used for machine learning model evaluation. For the efficient classification, the features selection techniques such as Particle Swarm Optimization (PSO) and Improved Salp Swarm Optimization Algorithms(ISSA) are used which select the most important features among all the features in the datasets. The main contribution in this work are as follows:

Imbalanced target classes are handled using Synthetic Minority Oversampling Technique (SMOTE).

Novel Bio-inspired Feature selection approach is applied to all the three datasets to select the prominent features among the entire features in the datasets with different evaluator algorithms.

Ensemble machine learning model is built to exhibit efficient classification outcomeby combining all the weak base classifier to form a strong classifier for the final prediction.

II. RELATED WORKS

A.Motivation

The rapid growth in IoT devices and network in various fields generate huge amount of data which in turn demand security[27] because the unsecured devices connected in the network will be vulnerable to attacks. Hacker can explore the network and inject devices with malwares, attacks, and perform other operations in the system remotely which may harm the entire network. Artificial Intelligence (AI) is one of the propitious approaches for addressing cyber-attack and providing security. Here, the related works are studied based on the attack detection in IoT network using machine learning and deep learning techniques.

B. Existing Works

Ömer KASIM [18] (2020) proposed a deep classification model by combining the Auto-Encoder model and the Support Vector Machine (SVM) for classifying anomalies in the network. Autoencoder model was used for feature learning and dimensionality reduction and the SVM classifier were used to identify the normal and DDoS attack traffic. The dataset used in this work was CICIDS dataset and the virtually generated DDoS traffic.

C. Malathi et al., [14] (2021) work focused on providing security for the IoT network by multiple machine learning algorithms (ML) which are mostly used to identify the interrelated network assaults immediately. In this work, seven different machine learning algorithms were evaluated against BoT-IoT dataset. The performance of the model is evaluated by using precision, recall, accuracy and f-measure.

H.T. Manjula et al., [9] work focused on detecting DDoS attack using ML classification algorithms. In this proposed work, Loic attacking tools was used to launch the DDoS attacks the dataset was transferred to Hadoop framework Apache spark for detection analysis.For the dataset classification the Apache spark machine learning algorithms (MLib) and other classification algorithms such as Naive Bayes, KNN and RF algorithms were used for classifying the normal traffic and attack traffic.

In Georgios Tertytchny et al., [8] (2020) work, the internal communication environment of an Energy Aware Smart Home (EASH) system was studied. Their proposed framework was evaluated by using both simulation and the real-time testbed environment. The abnormality classes were analysed based on the similarity of the features that each classes shares and the MLP algorithm was used to perform the feature-based analysis.

Eirini Anthi et al., [4] (2021) proposed a rule-based approach for detecting Denial of Service attack in IoT smart home network by generating AML attack samples. In this data points, the DoS packets and the benign packets were selected for AML experiments. In the DoS packets, the important features were extracted by using Information gain filter and InfoGain ratio attribute evaluation. The data are then classified by using the classifiers such as J48 Decision Tree, Random Forest, Bayesian network and SVM classifiers.

Amiya Kumar Sahu et al., [1] (2021) proposed a deep learning (DL) based approach for attack detection in the IoT network. In this work, Convolutional neural network (CNN) was used to select the accurate features from the dataset. These selected features were used for model classification by using Long Short-Term Memory (LSTM) model. In this, IoT-23 dataset was used which contains the malicious network traffic data were collected from twenty

Zhida Li et al., [25] (2021) proposed a recurrent neural network (Long Short-Term Memory and Gated Recurrent Unit) and Broad Learning System to classify the known intrusion. Border Gateway Protocol routing records along with the NSL-KDD and CICIDS2017 and CSE-CIC-IDS2018 datasets were used to trained and tested the model. In incremental learning, in each step the features are selected and ranked accordingly.

Swathi Sambangi et al., [23] (2020) proposed a machine learning model by applying the multiple regression analysis to predict the DoS attack and BoT attack in the cloud environment. The CICIDS 2017 dataset is used and the Information Gain (IG) method was applied to choose the most important feature. Then the selected features were used for multiple regression analysis. ANOVA model has been used to feature selection and the multiple linear regression model was used for classification.

Zhihong Tian et., [26] (2020) proposed a web attack detection system by analysing URLs based on distributed deep learning technique which was implemented on edge devices. In this work, they have implemented two concurrent deep models and the result was compared with the existing system with HTTP Dataset CSIC 2010, FWAF and HttpParams datasets. The performance of the model was evaluated by using accuracy, recall, FP and precision metrics.

A novel SaaS framework was proposed by Reddy SaiSindhuTheja et al., [19] (2020) for mitigating attack nodesin the cloud platform by transferred the control to lightweight bait approach by using Deep Belief Network and MFoSLO algorithm was used to fine tuning weight and activation function. The performance of the proposed model was evaluated based on accuracy, sensitivity, specificity, precision, false positive and negative rate, false detection and omission rate metrics.

Asmaa A. Elsaeidy et al., [2] (2021) developed a hybrid RBM (Restricted Boltzmann Machine) deep learning model to detect the replay attack that breaches the authentication conditions in smart city infrastructure and DDoS attack. The real-life smart city datasets such as environmental, smart river and smart soil datasets were used with simulated DDoS and replay attack for experimental evaluation.

The nature of adversarial problem happens in Network Intrusion Detection System was studied in this work by Elie Alhajjar et al., [7] (2021). For the tools of adversarial example generation, the particle swarm optimization and genetic algorithm evolutionary computing, generative adversarial deep learning network and Monte Carlo methods has been used. The NSL-KDD, UNSW-NB15 benchmark datasets were used for performance analysis of the algorithms in evading a NIDS.

Mahdis Saharkhizan et al., [13] (2020) proposed the advanced deep learning approach for cyber-attack detection against IoT systems. Their method was integrated with set of Long-Short-Term-Memory (LSTM) modules into ensemble detectors. At the last, decision tree classifier was used to merge all the models in-order to obtain the aggregated output at the final stage of the classifier.

Junaid Shabbir Abbasi et al.,[11] (2021) proposed a Deep Learning-based Feature Extraction (DLFE) and Optimization of Pattern Matching (OPM) methods for Network Intrusion Detection and Prevention systems for pattern matching engine optimization in intrusion detection. Here, the attack detection was done by using support vector machine (SVM) with optimal separating hyperplane was created for classification. The performance of the proposed method was evaluated by using time, throughput and memory utilization.

Ly Vu et al., [12](2020) proposed a novel deep transfer learning (DTL) for learning non-labelled data which was collected from multiple IoT devices. The proposed DTL model was based on two AutoEncoders (AE). The attacks of incoming samples were detected by auto encoder in target dataset. The Area Under the Curve (AUC) was used as the performance metric to evaluate the model.

Imtiaz Ullah et al., [10] (2021) proposed an anomaly-based intrusion detection model for IoT network. In this paper, the multiclass classification model was built using convolutional neural network with different dimensions such as 1D, 2D, 3D. The proposed CNN model was validated against various datasets includes BoT-IoT, IoT Network Intrusion, MQTT-IoT-IDS2020, and IoT-23 intrusion detection datasets. The performance metrics used to evaluate the proposed model was accuracy, precision, recall and f1-score.

Segun I. Popoola et al., [21] (2021) proposed an efficient deep learning method for botnet attack detection in memory-constrained IoT devices. Here, deep Bidirectional Long Short-Term Memory (BLSTM) was used where the long-term inter-related changes were analysed in the low-dimensional feature set which was produced by LAE. The proposed hybrid DL model was evaluated against BoT-IoT dataset.

The aforementioned related worksfocussed on attack detection in IoT network with various types of publicly available datasets, self-built dataset, real-time sensor data by using machine learning and deep learning techniques. Due to complexity, deep learning model is expensive to train the data when compared with machine learning model. The traditional machine learning also predicts the data samples using the single ML classifier. The single ML may produce the biased outcomes if the data is not uniformly distributed. The bio-inspired feature selection algorithms are used to find the optimal solution for a complex optimization problem [41]. Therefore,in this proposed work, the attacks in the IoT network are detected byusing multiple ensembled machine learning models using prominentbio-inspired feature selection using PSO and ISSA bio-inspired algorithm.

III. PROPOSED SYSTEM

The detailed description of the proposed model is depicted in Fig. 1. In this work, three IoT datasets namely Distributed Smart Space Orientation System (DS2OS), IoT23, BoT-IoT datasets are used as input datasets. In the pre-processing stage, the datasets are pre-processed by using various pre-processing techniques such as Normalization, Ordinal Encoding, etc., The data pre-processing techniques improves the quality of input data for efficient classification with different machine learning algorithms. After pre-processing, the most significant features are selected from the entire features in the dataset by using bio-inspired feature selection algorithms (PSO, ISSA). The input data with the selected features are then given as the input to the Machine Learning and the Ensemble Learning model classifiers for accurate classification.

For the testing phase, the unseen data is used to evaluate the model. The modules in the proposed work are (i) Data pre-processing (ii) Target Class data Balancing using SMOTE. (iii) Feature selection using Particle Swarm Optimization (PSO) and Improved Salp Swarm Optimization algorithms (ISSA). (iv) Classification using Machine Learning models and the Ensemble Learning model. (v) Performance Evaluation

A. Data Pre-processing

The pre-processing is a method to convert the data of any format into a standard format appropriate for model construction and classification which will increase the efficiency and performance of the system.

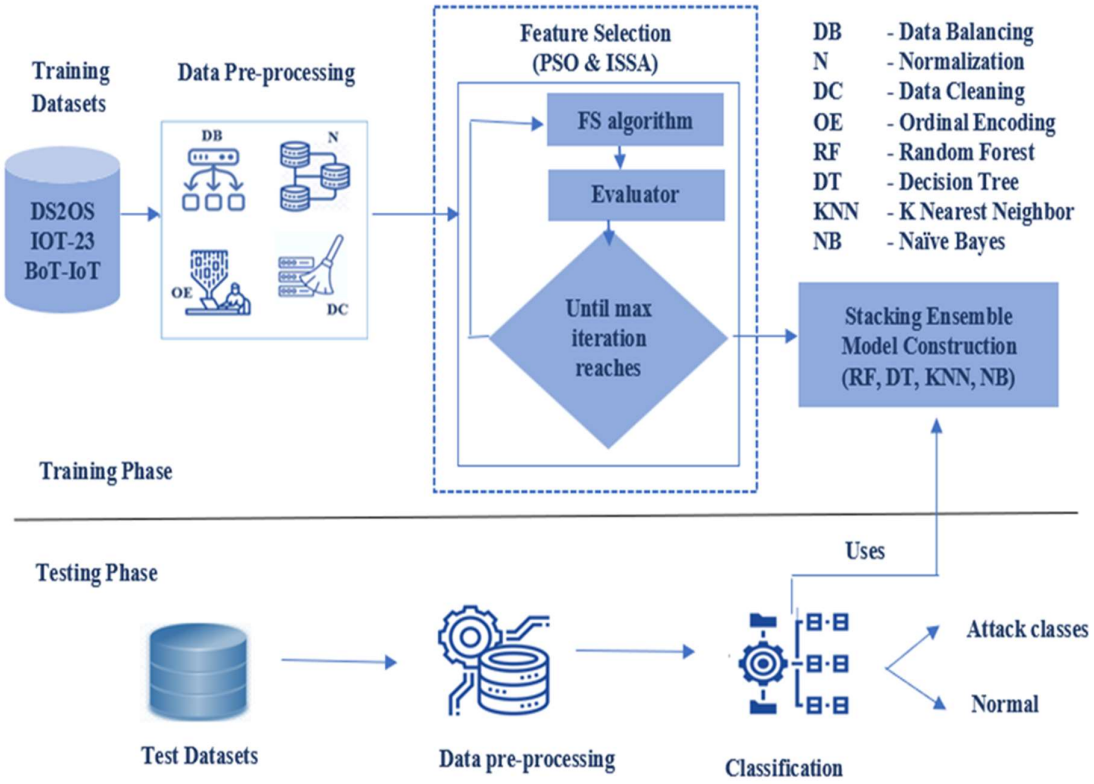


Fig.1. System Diagram

The pre-processing approaches used in this paper include:

Data Cleaning

The null values, infinite values, nan values and the empty features are removed or assign any numerical value (‘0’) for specific features or specific columns from the datasets.

Normalization

It is a scaling technique that transforms the data into other formats by applying shifting or rescaling operation in-order to bring the input data within the range of 0 and 1. It is also known as Min-Max Scaling.

Ordinal Encoding

The machine learning model requires all the features to be numerical to fit the model. So, the categorical feature values will be converted into numerical format by using the encoding techniques.

B. Class Balancing Approach

The imbalanced dataset contains classes of skewed proportions i.e., the target class with majority and minority classes. For example, if a dataset contains 100 instances for attack class and 20 instances for normal class then it is imbalanced dataset which makes the classification result biased i.e., for any kind of new data, the model will always predict attack as output. In-order to improve the model performance and obtain higher accuracy the class should be balanced. In this, the imbalanced dataset is handled by using SMOTE approach.

SMOTE [33] is an oversampling technique that generates the synthetic samples for minority class by computing k-nearest neighbors for this point and the new samples are added between

the chosen point and its neighbors. The over fitting problem caused by randomly sampling technique can also be overcome by using SMOTE approach.

C. Feature Selection

The purpose of feature selection is to identify the best set of features to build the model. It is the optimization techniques in which the best features can be used instead of the entire features in the dataset. It will reduce the model training time. In this proposed work, the Particle Swarm Optimization (PSO) and Improved Salp Swarm Optimization Algorithms (ISSA) are used to select the prominent features from the dataset. Both the algorithms are worked with different evaluator algorithms such as KNN, NB, DT and RF using wrapper mode. In each iteration, the algorithm will be applied to the training dataset to identify the feature set for training the Machine Learning models and the Ensemble Learning model

Particle Swarm Optimization (PSO) Algorithm

PSO is one of the simplest bio-inspired algorithms. It is a stochastic optimization technique inspired by the movement and intelligence of swarm. It uses a number of particles that represents a swarm moving around in search space to find the best optimal solution. It optimizes a problem by improves the candidate solution iteratively. It has the population of particle solution and moving those around in the search space based on the particle's position and velocity. Each particle movement will be considered as their own best position in their search space is known as personal best or pbest. Another best value is global best value which is obtained as yet by any particle nearest to that particle. The Algorithm 1 shows the steps involved in the Particle swarm optimization algorithm. Here, p_best represents the best position of specific particle and the g_best represents the best of all the particle i.e., the global best optimal solution obtained so far.

The detailed labels are given in nomenclature TABLE I.

TABLE I
NOMENCLATURE

P	Set of all particles or population size (P ∈ 1,2,3, ...)
max_iterations	Maximum number of Iterations
W	Inertia coefficient
c1	Cognitive coefficient
c2	Social coefficient
K	Dimension - which represents number of features
x_{ik}	$x \in 1,2,3, \ldots k$
v_{ik}	$v \in 1,2,3, \ldots k$
j	Iterating variable
p_best	Personal best value in the search space
g_best	Global best value in the search space

Algorithm 1

Input: fitness function, population size/number of particles P, max_iterations, w, c1, c2
Output: best features
1: for each particle $i \in P$
2: for each dimension k
3: initialize position x_{ik} randomly
4: initialize velocity v_{ik} randomly

5: Iteration $j = 1$
6: while max_iterations
7: for each particle i
8: calculate fitness function value
9: if fitness value $> p\_best_{ik}$
10:set current fitness value as p_best_{ik}
11:choose the particle with best g_best_k
12:for each particle i
13:for each dimension k
14: calculate velocity by
15: $v_{ik}(j+1) = w * v_{ik}(j) + c_1 r_1(p_{ik} - x_{ik}) + c_2 r_2(p_{gk} - x_{ik})$
16: update particle position by $x_{ik}(j+1) = x_{ik}(j) + v_{ik}(j+1)$
17: $j = j + 1$

Improved Salp Swarm Optimization Algorithm (ISSA)

Salp is a barrel-shaped planktic tunicate which moves by contracting and pumping water through its gelatinous body. The most interesting behavior of salp is when it forms by align next to each other as stringy colony. In S. Mirjaliliet al., (2017) [22]work, they model the swarming behavior of salp mathematically by dividing the salp population into a leader and followers. The first salp was designated in the direction of motion of the salp chain as the leader salp and the rest of the salp as follower salp.

The position of salp was defined in n-dimensional space where n is the number of variable or features in the datasets. The position of all the salps were stored in the matrix z and the food source are given in the search space F. The position of leader was updated by using the formula in equation (1):

$$z_j^1 = \begin{cases} F_d + c1\big((ub_d - lb_d)c2 + lb_d\big)\, c3 \geq 0.5 \\ F_d - c1\big((ub_d - lb_d)c2 + lb_d\big)\, c3 < 0.5 \end{cases} \quad (1)$$

In this, z_d^1represents the position of the leader salp and F_d denotes the position of the food source in d^{th} dimension, ub_d and lb_d represents the upper and lower bounds in jth dimension, $c1, c2$ and $c3$ are the random variable to control the exploration and exploitation. The coefficients are improved by using equation 3 formula

$$c1 = 2e^{-\left(\frac{4l}{L}\right)^2} \quad (2)$$

The position of the follower salps are updated based on the Newton's law of motion given in equation 3.

$$z_d^m = \frac{1}{2}(z_d^m + z_d^{m-1})\ m \geq 2 \quad (3)$$

The leader salp update its position based on the direction of food source and the followers salps update their position based on each other. For feature selection optimization problem, the food source is replaced by global optimal solution based on the optimal solution obtained till now. The salp swarm problem was improved in Mohammad Tubishat et al.,(2020) [15] work. The SSA was improved by using Opposition Based Learning (OBL) approach to improve the population diversity of SSA at the initialization phase and Local Search Algorithm (LSA) was used with SSA for the exploitation improvement and avoid it from stuck into local optima. The algorithm 2 describes the steps involved in the Improved Salp swarm optimization algorithm. It returns the best set of features obtained until the maximum iteration reaches. The further notations are given in the nomenclature TABLEII:

TABLE II
NOMENCLATURE

Z	Number of salps or population size ($z_i \in 1,2,3,\dots n$)
L	Maximum number of Iterations
l	Current iteration
W	Inertia coefficient
c1	Controlling parameter
c2, c3	Random values between [0,1]
J	Dimension - which represents number of features
T	Iterating variable
lb	lower bound value of jth dimension
ub	upper bound value of jth dimension

Algorithm 2

Input: fitness function, population size/number of salps, max_iterations
Output: best fitness solution
1: Initialize the salps positions z_i
2: Select the n fittest salps, which represent the initial SSA population
3: while (t <max_iterations)
4: determine the fitness value of each salp
5:*F*eat = best_salp
6: Update the value of c_1 parameter by $c1 = 2e^{-\left(\frac{4l}{L}\right)^2}$
7:for every salp (z_i)
8:if (i ==1)
9:Update leader position by

$$z_j^1 = \begin{cases} F_d + c1\left((ub_d - lb_d)c2 \ + \ lb_d\right) c3 \geq 0.5 \\ F_d - c1\left((ub_d - lb_d)c2 \ + \ lb_d\right) c3 < 0.5 \end{cases}$$

10:else
11: Update follower position by

$$z_j^i = \frac{1}{2}\left(z_j^i + \ z_j^{i-1}\right) i \geq 2$$

12: end if
13:end for
14: reposition the salp based on lower and upper bounds of problem variables
15: Apply LSA on *F* to find if there is a better solution (if better solution found then update *F*; otherwise *F* left unchanged)
16:$t = t + 1$
17: end while

D. Classification using Machine Learning and Ensemble Learning approaches

In machine learning, the mathematical models can be building using various machine learning algorithms and the predictions can be made based on the historical information. It is useful in various real-time applications such as image recognition, speech recognition, email filtering, security systems and many more.

The machine learning techniques has been classified as supervised, unsupervised and reinforcement learning algorithm. In this paper, the supervised machine learning models such as Random Forest (RF), Decision Tree (DT), K-Nearest Neighbor (KNN), Naïve Bayes (NB)

are used to classify attacks in the IoT Network. For Ensemble Learning, stacking ensemble learner is usedfor final model construction and prediction.

In stacking ensemble classifier, the standard of model prediction is improved by combining all the week classifiers. It takes several regression or classification models as base classifier and their outcomes will be used as an input to the final meta classifier/regressor. In this proposed work, the multi-layer stacking model was built with four machine learning algorithms. The base layer is constructed with the KNN and NB algorithms in the first layer and the RF and DT algorithms in the second layer. The final stacking layer is built with logistic regression as the meta classifier. The Fig. 2 depicts the proposed multi-layer stacking classifier.

IV. EXPERIMENTAL RESULT

A. Dataset Description

In this work, DS2OS, BoT-IoT and IoT-23 datasets are used to evaluate the model as listed in Table 1 – Table 3.

1) DS2OS dataset

The DS2OS dataset is publicly available open source dataset collected from Kaggle [4]. DS2OS dataset contains traces collected in the Distributed Smart Space Orchestration System (DS2OS) IoT environment. These data are different from other network traces because it is collected from application layer. They captured the data using four simulated IoT sites with different type of service such as light controller, thermometer, movement sensors, washing machines, batteries, thermostats, smart doors and smart phones. This dataset contains 357952 records and 13 features (12 input feature + 1 attack classes). The distribution of attack classes in the dataset is given in the Table III.

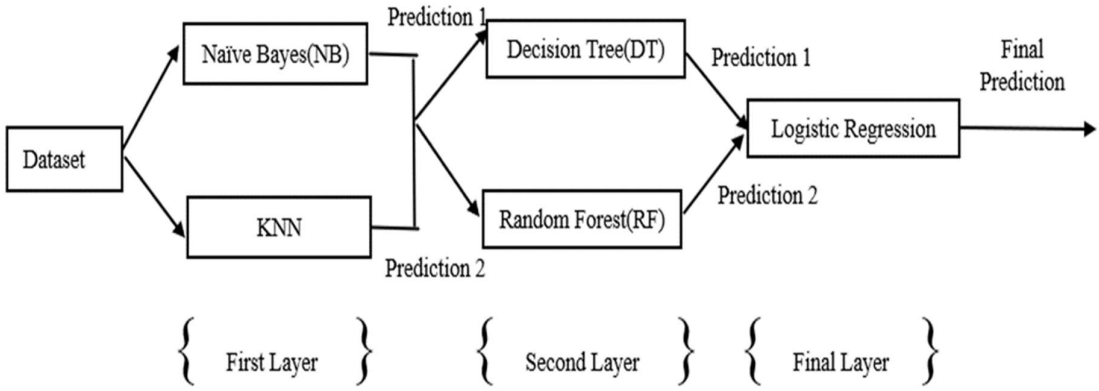


Fig. 2. Stacking classifier with 2 layers

TABLE III
DISTRIBUTION OF ATTACK CLASSES IN DS2OS DATASET

Types of Attack	No. of instances
Normal	347934
DoS	5780
Scan	1547
Malicious control	889
Malicious operation	805
Spying	532

Data Probing	342
Wrong Setup	122

2)BoT-IoT Dataset

The BoT-IoT dataset was created by designing a realistic network environment in the Cyber Range Lab of the centre of UNSW Canberra Cyber [17]. The environment contains a combination of both normal and botnet traffic. The dataset's source files are available in different formats, including the original pcap files, the generated argus files and csv files. The captured pcap files are 69.3 GB in size, with more than 72M records. The dataset includes DDoS, DoS, OS and Service Scan, Keylogging and Data exfiltration attacks, with the DDoS and DoS attacks further organized, based on the protocol used. Here, they extracted the 5% of the original dataset via the use of select MySQL queries. In this work, the 100000 instances from each attack classes are extracted from the dataset if the classes are available. Some classes contain a smaller number of instances only. The below TABLE IV show the attack classes in the 5% of entire dataset.

TABLE IV
ATTACK CATEGORY AND NUMBER OF INSTANCES IN BOT-IOT DATASET

Category	Subcategory	No. of instances
Normal	Normal	477
DDoS/DoS	TCP UDP HTTP	1593180 1981230 2474
Reconnaissance	OS Fingerprinting Service Scanning	17914 73168
Information Theft	Keylogging Data Exfiltration	73 6

3)IoT-23 Dataset

IoT-23 dataset incorporates network traffic from Internet of Things (IoT) devices. It contains captures from 20 malware devices and 3 normal devices. This IoT network traffic was captured in the Stratosphere Laboratory, AIC group, FEL, CTU University, Czech Republic [20]. In this, the original .pcap file is analysed manually. The suspicious flows are detected and labels are assigned. The Zeek conn.log file was obtained by running the Zeek network analyser using the original pcap file and stored it in conn.log.labeled file. The malware samples used to infect devices are Mirai, Torii, Trojan, Gagfyt, Hakai, Okiru, IRCBot and other samples also. The dataset contains more than 760 million packets and 325 million labelled flows of more than 500 hours of traffic. The captures were taken during 2018 and 2019. The dataset contains 21 features (20 input traffic + 1 target class(attack or normal)).

The TABLE V describes the attack class distribution and number of instances in detail. In this work, the label with less than 50 instances were eliminated and 100000 instances were extracted from the labels whose instances were more than 1 million for ease of use.

TABLE V
ATTACK CLASS DISTRIBUTION AND NUMBER OF INSTANCES IN IOT-23 DATASET

Class Label	No. of instances
PartOfAHorizontalPortScan	213852929

Okiru	60990711
Benign	30858735
DDoS	19538713
C&C	22883
Attack	9398
C&C-HeartBeat	34507
C&C-FileDownload	53
C&C-Torii	30
FileDownload	18
C&C-HeartBeat-FileDownload	11
C&C-Mirai	2

B.Result Analysis

In this work, the proposed model is evaluated against three datasets. The experimental result has been analysed and performance of the model was identified with the various performance metrics such as accuracy, precision, recall, F1-score and FPR. For each dataset, the performance is evaluated with various categories as follows:

without feature selection

using feature selection algorithms (PSO & ISSA)

using Ensemble Learning

1)Result Analysis for DS2OS Dataset

The DS2OS dataset was evaluated and the corresponding experimental results with all the features, with sets of selected features based on feature selection algorithms used in this work and the Ensemble model result is shown in this section. The result with all the features is given in the Fig. 3.

The significant features are selected from the dataset for more efficient and accurate classification by using Particle Swarm Optimization (PSO) and Improved Salp Swarm Optimization Algorithms (ISSA). The evaluator in each feature selection algorithm has been changed with different machine learning algorithms such as RF, DT, NB, KNN i.e., if the evaluator algorithm is Random Forest (RF), the features selected based on RF algorithm is used to train and test the all-other models. Similarly, for all other algorithms as evaluator at feature selection algorithm.

The number of features selected using both PSO and ISSA with different evaluator algorithm was given in the TABLEVI. In this table, the features are selected automatically by feature selection algorithms.

TABLE VI

NO. OF FEATURES SELECTED BY BOTH PSO AND ISSA ALGORITHM

FS Algorithm	Evaluator	Selected Features
PSO	PSO + RF	1, 3, 4, 5, 8, 9, 10
	PSO + DT	1, 2, 5, 7, 8
	PSO + NB	2, 3, 4, 5, 7, 8
	PSO + KNN	0, 5, 8, 9, 10
ISSA	ISSA + RF	1, 2, 3, 5, 7, 8
	ISSA + DT	1, 5, 8, 9, 10
	ISSA + NB	0, 1, 2, 3, 5, 8
	ISSA + KNN	1, 3, 5, 7, 8

The result obtained for all the machine learning classifiers used in the proposed work with the features selected by PSO is shown in the TABLE VII and ISSA is shown in Table VIII.
From TABLEVII, it is inferred that the NB algorithm's performance is lower than the other ML algorithms, but its performance increased when the evaluator for PSO algorithm is used as Random Forest algorithm.
From TABLE VIII, it is observed that for DS2OS dataset, the models trained with the features selected by PSO performs better than the models trained with the features selected by ISSA algorithm.
The validation performance of ensemble model by using both feature selection algorithm is given in the Fig. 4 and the detection performance of ensemble model by using both feature selection algorithm is given in the Fig. 5 and the evaluation metrics obtained for each attack in DS2OS dataset is given in the Fig. 6.

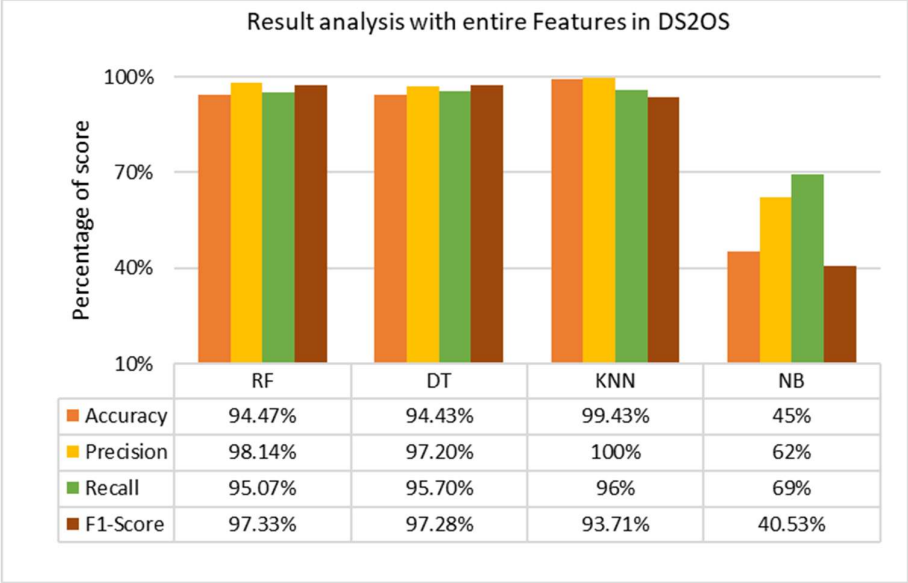


Fig. 3.Result obtained for DS2OS dataset with all the features

TABLE VII
RESULT OBTAINED FOR DS2OS DATASET WITH THE SELECTED FEATURES BY PSO

PSO + Evaluator \ Performance metrics	ML Classifiers	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
PSO + RF	RF	99.26	96.32	99.13	97.80
	DT	99.26	96.32	99.13	97.80
	KNN	98.32	97.75	95.70	97.31
	NB	70.23	45.32	69.58	72.59
PSO + DT	RF	97.65	84.32	99.54	85.52
	DT	96.44	83.14	99.54	85.78
	KNN	93.44	97.58	91.53	95.09
	NB	59.52	40.51	64.66	42.38
PSO + NB	RF	96.64	83.20	99.42	86.81
	DT	96.44	83.20	95.42	85.81

PSO + Evaluator \ Performance metrics	ML Classifiers	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
PSO + KNN	KNN	93.64	95.78	91.93	95.90
	NB	44.50	40.03	65.18	36.51
	RF	99.27	97.23	99.07	97.65
	DT	99.27	96.11	99.07	97.65
	KNN	93.85	92.25	95.69	97.12
	NB	45.13	44.18	58.77	38.91

TABLE VIII
RESULT OBTAINED FOR DS2OS DATASET WITH THE SELECTED FEATURES BY ISSA

ISSA + Evaluator \ Performance metrics	ML Classifiers	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
ISSA + RF	RF	96.44	83.21	99.54	85.82
	DT	96.44	83.20	99.54	85.81
	KNN	99.36	99.75	91.93	95.09
	NB	55.56	42.29	68.36	45.71
ISSA + DT	RF	96.65	83.19	99.55	85.77
	DT	96.38	83.19	99.53	85.78
	KNN	99.36	99.75	91.93	95.05
	NB	60.52	40.05	64.66	45.44
ISSA + NB	RF	95.76	73.06	99.45	77.37
	DT	95.76	73.06	99.45	77.36
	KNN	99.20	98.71	86.87	91.57
	NB	45.08	43.48	67.58	40.36
ISSA + KNN	RF	99.27	97.23	99.07	97.65
	DT	96.44	83.20	99.54	85.81
	KNN	99.35	99.23	91.93	94.88
	NB	49.83	40.46	66.22	37.65

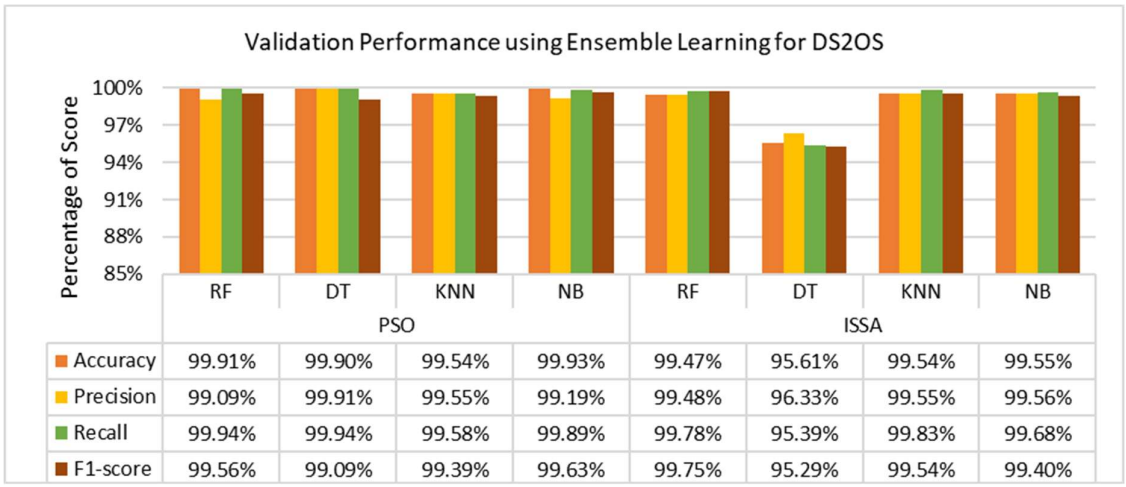
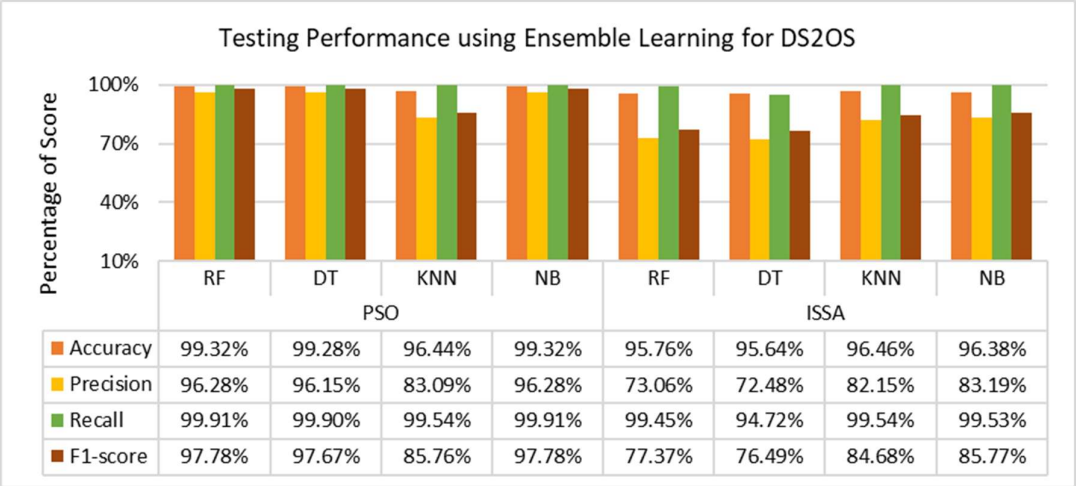
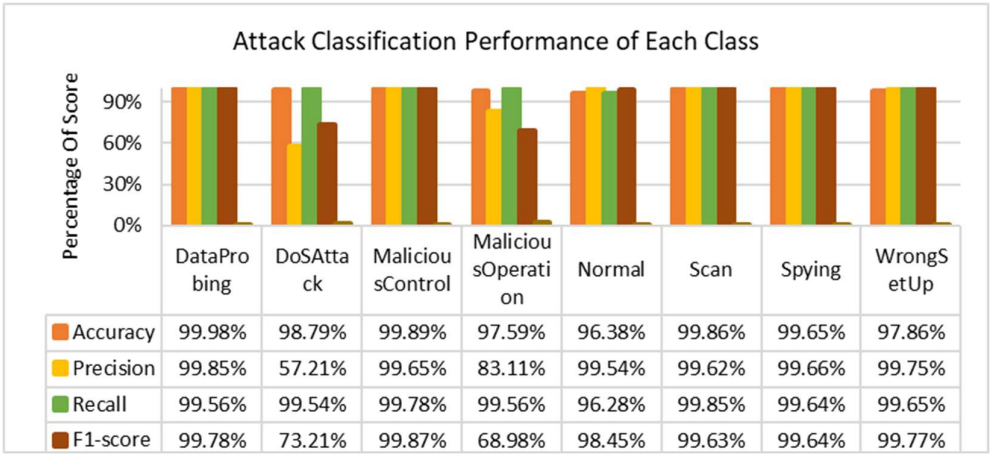


Fig. 4.Validation performance of DS2OS dataset using Ensemble Learning



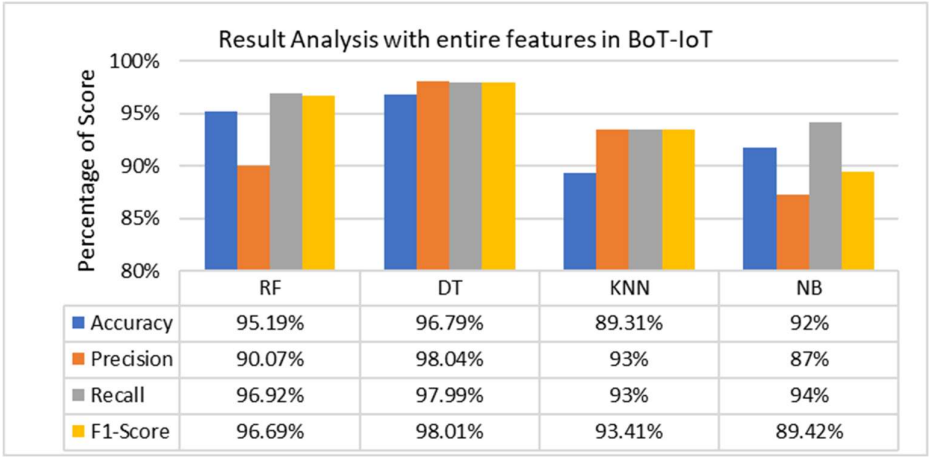
	PSO				ISSA			
Accuracy	99.32%	99.28%	96.44%	99.32%	95.76%	95.64%	96.46%	96.38%
Precision	96.28%	96.15%	83.09%	96.28%	73.06%	72.48%	82.15%	83.19%
Recall	99.91%	99.90%	99.54%	99.91%	99.45%	94.72%	99.54%	99.53%
F1-score	97.78%	97.67%	85.76%	97.78%	77.37%	76.49%	84.68%	85.77%

Fig. 5.Detection performance of DS2OS dataset using Ensemble Learning



	DataProbing	DoSAttack	MaliciousControl	MaliciousOperation	Normal	Scan	Spying	WrongSetUp
Accuracy	99.98%	98.79%	99.89%	97.59%	96.38%	99.86%	99.65%	97.86%
Precision	99.85%	57.21%	99.65%	83.11%	99.54%	99.62%	99.66%	99.75%
Recall	99.56%	99.54%	99.78%	99.56%	96.28%	99.85%	99.64%	99.65%
F1-score	99.78%	73.21%	99.87%	68.98%	98.45%	99.63%	99.64%	99.77%

Fig. 6. Performance measure for each attack in DS2OS dataset



	RF	DT	KNN	NB
Accuracy	95.19%	96.79%	89.31%	92%
Precision	90.07%	98.04%	93%	87%
Recall	96.92%	97.99%	93%	94%
F1-Score	96.69%	98.01%	93.41%	89.42%

Fig. 7.Result Obtained for BoT-IoT dataset with all the features

From Fig. 4, 5 and 6, it is inferred that the performance of the models increased when Ensemble Learning is used to train, validate and test the dataset. When compared to the feature selection algorithm with model training, the accuracy, precision, recall, f-score are increased for the Ensemble model and the FPR rate is also low for all the attacks.

2)Result Analysis for BoT-IoT Dataset

The BoT-IoT dataset was evaluated and the corresponding experimental results with all the features, with sets of selected features based on feature selection algorithms used in this work and the Ensemble model result is shown in this section. The result with all the features is given in the Fig. 7.

The number of features selected using both PSO and ISSA with different evaluator algorithm was given in the Table IX. In this, the list features are selected automatically by the algorithms.

TABLE IX
NO. OF FEATURES SELECTED BY BOTH PSO AND ISSA ALGORITHM

FS Algorithm	Evaluator	Selected Features
PSO	PSO + RF	0, 1, 5, 9, 11, 13, 19, 21, 25, 27
	PSO + DT	2, 5, 6, 7, 8, 11, 14, 15, 18, 26, 27
	PSO + NB	6, 10, 14, 15, 20, 23, 26, 27
	PSO + KNN	2, 3, 5, 6, 10, 14, 15, 18, 27
ISSA	ISSA + RF	2, 6, 14, 17, 24, 26, 27
	ISSA + DT	2, 4, 6, 8, 16, 19, 23, 27
	ISSA + NB	3, 6, 15, 16, 18, 21, 27
	ISSA + KNN	0, 2, 3, 5, 6, 7, 11, 12, 15, 17, 26

The result obtained for all the machine learning classifiers used in the proposed work for BoT-IoT dataset with the features selected by PSO is shown in the TABLE X and ISSA is shown in TABLE XI.

From TABLE X, it is inferred that the NB algorithm's performance is lower than the other ML algorithms, but its performance increased when the evaluator for PSO algorithm is used as KNN algorithm. In this, the performance of KNN algorithm also minimum when compared with the performance of DS2OS dataset with PSO.

From TABLE XI, it is observed that for BoT-IoT dataset, the models trained with the features selected by ISSA performs better than the models trained with the features selected by PSO algorithm. The performance all the algorithm increased, particularly, when KNN is used as evaluator algorithm performance of all the classifiers increases.

The validation performance of ensemble model by using both feature selection algorithm is given in the Fig. 8 and the detection performance of ensemble model by using both feature selection algorithm is given in the Fig. 9 and the evaluation metrics obtained for each attack in BoT-IoT dataset is given in the Fig. 10

From Fig. 8, 9 and 10, it is inferred that the performance of the models increased when Ensemble Learning is used to train, validate and test the dataset. When compared to the feature selection algorithm with model training, the accuracy, precision, recall, f-score are increased for the Ensemble model and the FPR rate is also low for all the attacks.

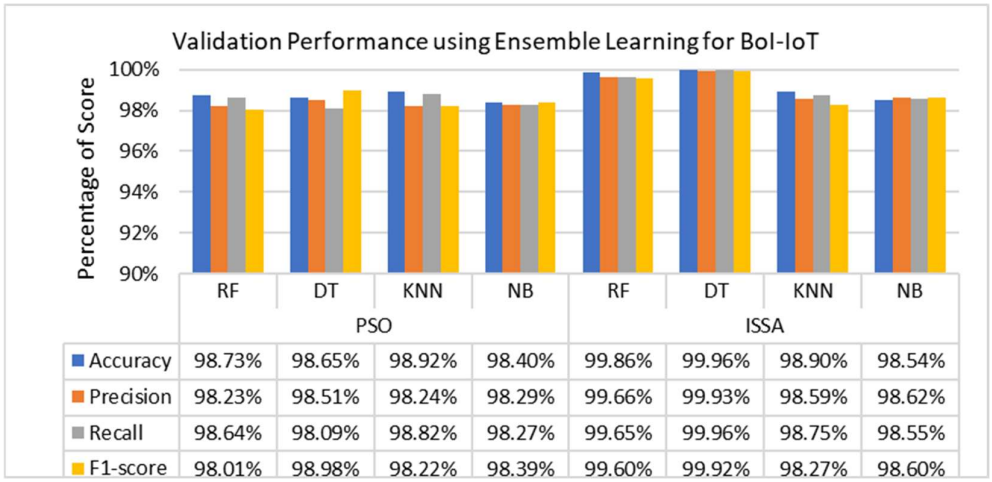
TABLE X
RESULT OBTAINED FOR BOT-IOT DATASET WITH THE SELECTED FEATURES BY PSO

PSO + Evaluator \ Performance metrics	ML Classifiers	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
PSO + RF	RF	97.08	98.04	98.95	98.66
	DT	97.63	97.52	98.77	98.58
	KNN	65.77	51.97	64.73	55.01
	NB	42.81	49.86	52.82	40.03
PSO + DT	RF	98.43	96.02	96.02	96.26
	DT	98.36	90.04	98.98	98.99
	KNN	88.64	76.09	87.51	78.81
	NB	55.79	61.10	61.57	52.91
PSO + NB	RF	96.75	98.65	98.77	97.25
	DT	95.16	97.01	96.99	96.99
	KNN	65.40	48.41	61.68	47.37
	NB	67.61	59.80	66.68	56.72
PSO + KNN	RF	96.24	90.02	98.94	94.16
	DT	97.39	95.45	98.56	96.99
	KNN	92.74	92.94	98.89	95.56
	NB	69.01	70.77	78.32	67.25

TABLE XI

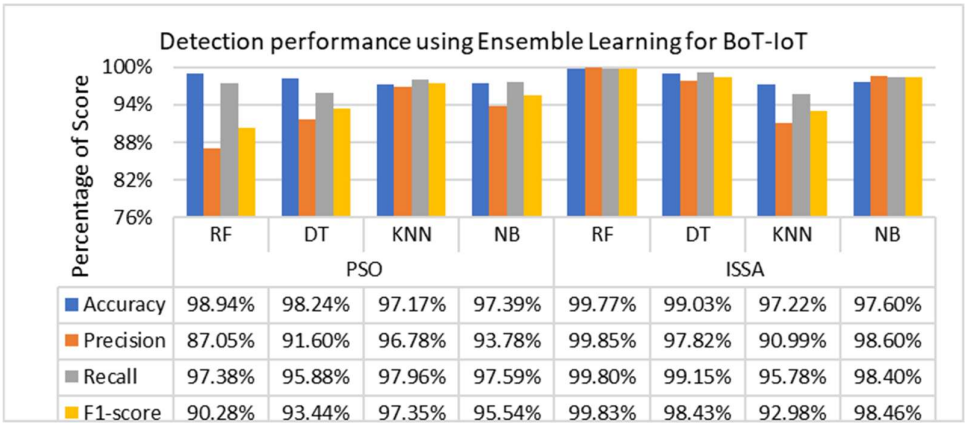
RESULT OBTAINED FOR BOT-IOT DATASET WITH THE SELECTED FEATURES BY ISSA

ISSA + Evaluator \ Performance metrics	ML Classifiers	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
ISSA + RF	RF	99.89	98.55	99.25	99.54
	DT	99.09	98.44	99.25	99.64
	KNN	95.29	97.02	96.94	96.97
	NB	76.21	86.97	85.47	82.07
ISSA + DT	RF	99.97	99.57	98.15	96.52
	DT	99.75	97.50	97.73	97.62
	KNN	88.44	80.08	91.05	84.20
	NB	66.95	59.74	61.80	51.25
ISSA + NB	RF	99.83	99.04	99.89	99.95
	DT	98.36	99.04	98.99	99.05
	KNN	68.81	52.96	64.48	53.86
	NB	61.62	61.16	49.30	49.82
ISSA + KNN	RF	96.24	90.02	98.94	94.16
	DT	97.39	95.45	98.56	96.99
	KNN	92.74	92.94	98.89	95.56
	NB	69.01	70.77	78.32	67.25



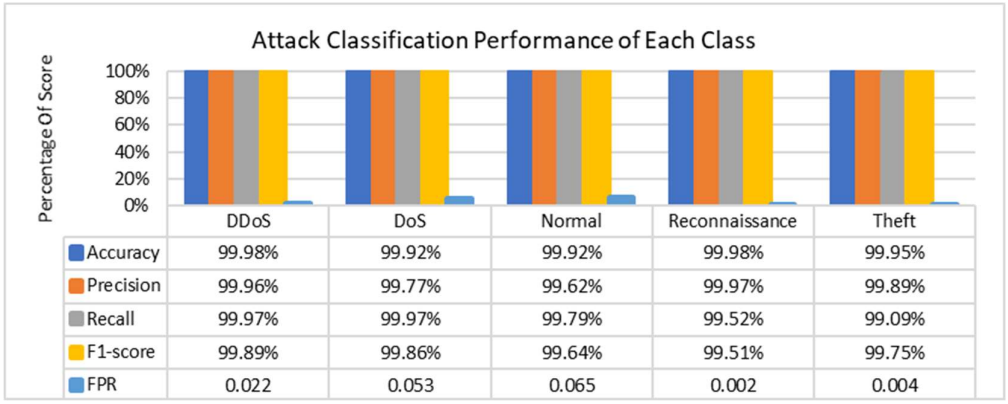
	PSO				ISSA			
Accuracy	98.73%	98.65%	98.92%	98.40%	99.86%	99.96%	98.90%	98.54%
Precision	98.23%	98.51%	98.24%	98.29%	99.66%	99.93%	98.59%	98.62%
Recall	98.64%	98.09%	98.82%	98.27%	99.65%	99.96%	98.75%	98.55%
F1-score	98.01%	98.98%	98.22%	98.39%	99.60%	99.92%	98.27%	98.60%

Fig. 8.Validation performance of BoT-IoT dataset using Ensemble Learning



	PSO				ISSA			
Accuracy	98.94%	98.24%	97.17%	97.39%	99.77%	99.03%	97.22%	97.60%
Precision	87.05%	91.60%	96.78%	93.78%	99.85%	97.82%	90.99%	98.60%
Recall	97.38%	95.88%	97.96%	97.59%	99.80%	99.15%	95.78%	98.40%
F1-score	90.28%	93.44%	97.35%	95.54%	99.83%	98.43%	92.98%	98.46%

Fig. 9.Detection performance of BoT-IoT dataset using Ensemble Learning



	DDoS	DoS	Normal	Reconnaissance	Theft
Accuracy	99.98%	99.92%	99.92%	99.98%	99.95%
Precision	99.96%	99.77%	99.62%	99.97%	99.89%
Recall	99.97%	99.97%	99.79%	99.52%	99.09%
F1-score	99.89%	99.86%	99.64%	99.51%	99.75%
FPR	0.022	0.053	0.065	0.002	0.004

Fig. 10.Performance measure for each attack in BoT-IoT dataset

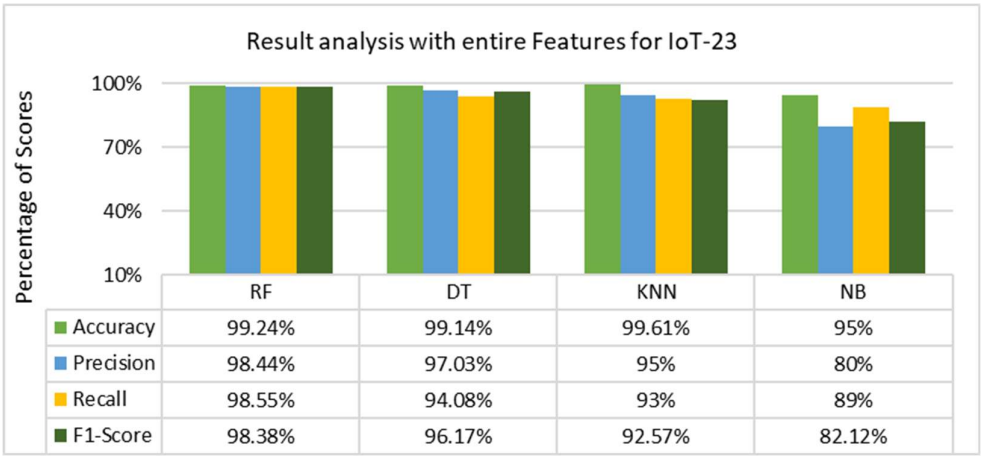


Fig. 11.Result Obtained for IoT-23 dataset with all the features

Result Analysis for IoT-23 Dataset

The IoT-23 dataset was evaluated and the corresponding experimental results with all the features, with sets of selected features based on feature selection algorithms used in this work and the Ensemble model result is shown in this section. The result with all the features is given in the Fig.11.

The number of features selected using both PSO and ISSA with different evaluator algorithm was given in the Table 10. In this, the list features are selected automatically by the feature selection algorithms.

The result obtained for all the machine learning classifiers used for IoT-23 dataset with the features selected by PSO is shown in the TABLE XIII and ISSA is shown in TABLE XIV.

From TABLE XIII, it is inferred that the performance of all the classifiers is higher than the performance obtained from DS2OS and BoT-IoT datasets. In this, performance of NB classifier also increased for this dataset. Among the other evaluator, the PSO+RF performance is higher than other evaluator performance.From TABLE XIV, it is inferred that the performance of all the classifiers is higher than the performance obtained from other datasets with ISSA algorithm. In this, performance of NB classifier also increased for this dataset. Among the other evaluator, the ISSA+KNN's performance is higher than other evaluator performance. For the IoT-23 dataset, ISSA performs better than PSO.

TABLE XII
NO. OF FEATURES SELECTED BY BOTH PSO AND ISSA ALGORITHM

FS Algorithm	Evaluator	Selected Features
PSO	PSO + RF	0, 3, 4, 5, 7, 11, 17
	PSO + DT	2, 3, 4, 5, 8, 14, 16
	PSO + NB	0, 3, 4, 6, 7, 9, 16
	PSO + KNN	0, 3, 4, 8, 10, 14
ISSA	ISSA + RF	3, 4, 8, 9, 11, 19
	ISSA + DT	3, 4, 8, 14, 16, 19
	ISSA + NB	1, 3, 4, 8, 9, 11, 12, 15
	ISSA + KNN	3, 4, 8, 11, 13, 15

The validation performance of ensemble model by using both feature selection algorithm is given in the Fig. 12 and the detection performance of ensemble model by using both feature

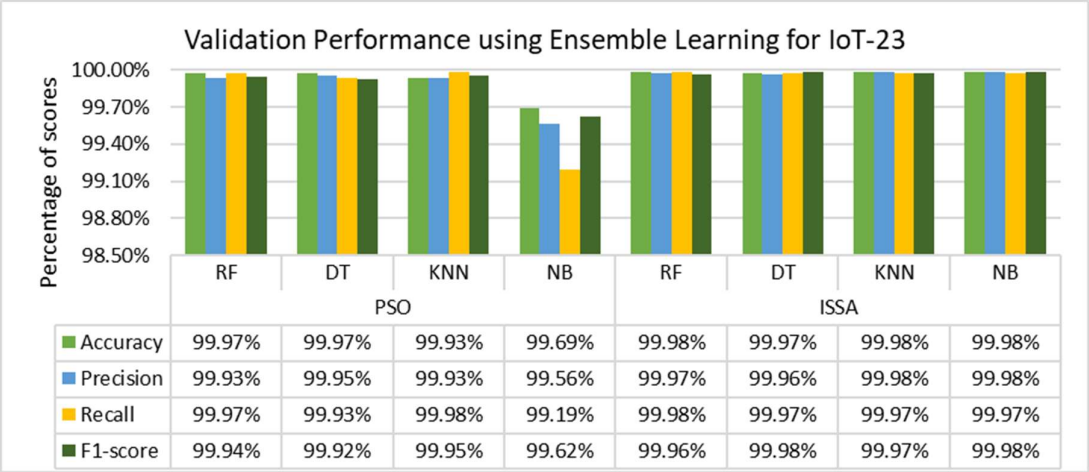
selection algorithm is given in the Fig. 13 and the evaluation metrics obtained for each attack in IoT-23 dataset is given in the Fig. 14.

TABLE XIII
RESULT OBTAINED FOR IOT-23 DATASET WITH THE SELECTED FEATURES BY PSO

PSO + Evaluator \ Performance metrics	ML classifiers	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
PSO + RF	RF	99.97	99.16	94.18	95.84
	DT	99.73	98.91	91.87	94.42
	KNN	99.81	94.46	94.08	92.74
	NB	91.46	82.42	84.06	80.89
PSO + DT	RF	99.98	99.94	98.95	99.47
	DT	99.86	98.52	97.97	98.08
	KNN	99.88	96.18	93.23	92.80
	NB	94.81	79.11	87.18	80.13
PSO + NB	RF	99.92	99.89	99.45	98.72
	DT	99.95	99.18	99.12	99.15
	KNN	99.83	96.05	99.86	97.50
	NB	93.68	84.03	88.82	82.36
PSO + KNN	RF	99.98	99.58	97.46	98.53
	DT	99.84	99.49	97.46	98.52
	KNN	99.85	94.20	99.87	96.42
	NB	93.09	79.02	88.34	80.86

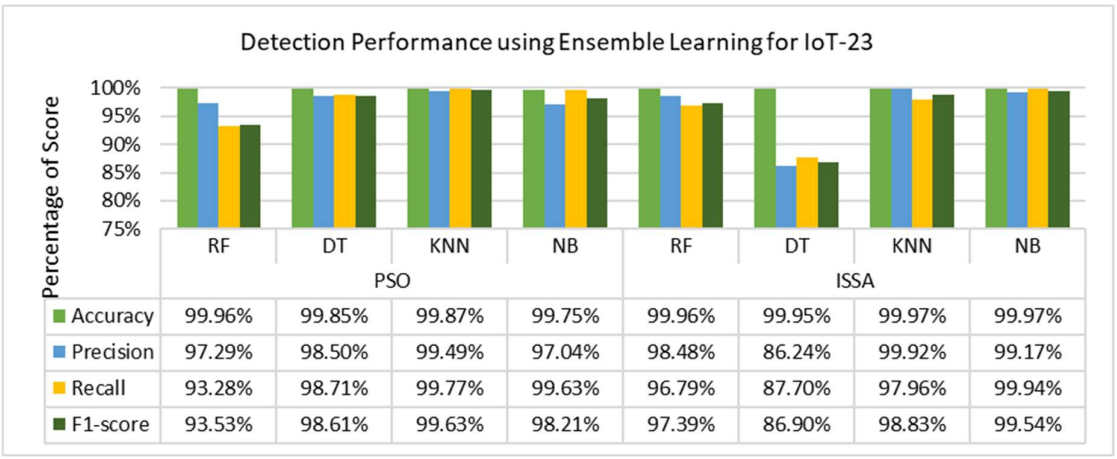
TABLE XIV
RESULT OBTAINED FOR IOT-23 DATASET WITH THE SELECTED FEATURES BY ISSA

ISSA + Evaluator \ Performance metrics	ML classifiers	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
ISSA + RF	RF	99.76	99.88	98.02	99.39
	DT	99.79	99.02	99.06	99.46
	KNN	99.78	90.15	98.93	92.52
	NB	93.58	76.28	86.28	75.24
ISSA + DT	RF	99.82	98.43	99.15	98.78
	DT	99.03	97.82	99.15	98.43
	KNN	99.78	89.15	98.93	92.52
	NB	93.58	76.28	86.28	75.24
ISSA + NB	RF	99.98	99.31	97.46	98.52
	DT	99.97	99.42	96.23	97.64
	KNN	99.84	94.23	99.86	95.98
	NB	94.09	80.44	92.47	82.95
ISSA + KNN	RF	99.74	99.15	96.45	97.97
	DT	99.75	99.14	96.45	97.97
	KNN	99.88	97.86	96.41	96.82
	NB	95.41	85.14	89.60	83.93



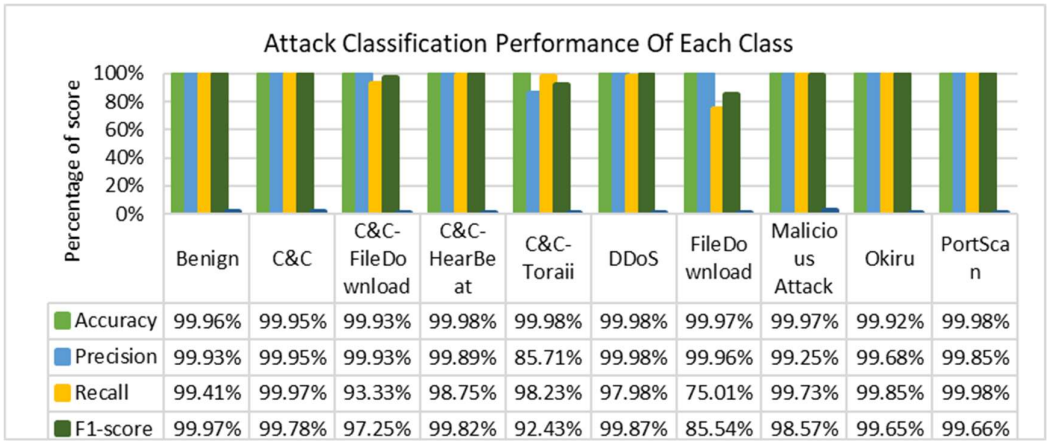
	PSO				ISSA			
Accuracy	99.97%	99.97%	99.93%	99.69%	99.98%	99.97%	99.98%	99.98%
Precision	99.93%	99.95%	99.93%	99.56%	99.97%	99.96%	99.98%	99.98%
Recall	99.97%	99.93%	99.98%	99.19%	99.98%	99.97%	99.97%	99.97%
F1-score	99.94%	99.92%	99.95%	99.62%	99.96%	99.98%	99.97%	99.98%

Fig. 12.Validation performance of IoT-23 dataset using Ensemble Learning



	PSO				ISSA			
Accuracy	99.96%	99.85%	99.87%	99.75%	99.96%	99.95%	99.97%	99.97%
Precision	97.29%	98.50%	99.49%	97.04%	98.48%	86.24%	99.92%	99.17%
Recall	93.28%	98.71%	99.77%	99.63%	96.79%	87.70%	97.96%	99.94%
F1-score	93.53%	98.61%	99.63%	98.21%	97.39%	86.90%	98.83%	99.54%

Fig. 13.Detection performance of IoT-23 dataset using Ensemble Learning



	Benign	C&C	C&C-FileDownload	C&C-HearBeat	C&C-Toraii	DDoS	FileDownload	Malicious Attack	Okiru	PortScan
Accuracy	99.96%	99.95%	99.93%	99.98%	99.98%	99.98%	99.97%	99.97%	99.92%	99.98%
Precision	99.93%	99.95%	99.93%	99.89%	85.71%	99.98%	99.96%	99.25%	99.68%	99.85%
Recall	99.41%	99.97%	93.33%	98.75%	98.23%	97.98%	75.01%	99.73%	99.85%	99.98%
F1-score	99.97%	99.78%	97.25%	99.82%	92.43%	99.87%	85.54%	98.57%	99.65%	99.66%

Fig.14.Performance measure for each attack in IoT-23 dataset

From Fig. 12, 13 and 14, it is inferred that the performance of the classifiers increased when Ensemble Learning is used to train, validate and test the dataset. When compared to the feature selection algorithm with model training, the accuracy, precision, recall, f-score are increased

for the Ensemble model validation as well as detection and the FPR rate is also low for all the attacks. Especially, the classifier's performance obtained for IoT-23 is higher when compared to the performance obtained for other datasets. The comparison of proposed work with the existing work is given in the TABLE XV. When comparing the proposed work with existing work, the proposed model works better than other previous work. In the proposed work, the ensemble with ISSA model obtained higher

From the experimental result, it is observed that the performance of classifiers which is trained using IoT-23 dataset was better than other datasets. For BoT-IoT and Iot-23 dataset, the performance of classifiers increases from without feature outcome to with feature selection outcome and in these feature selection algorithms the result obtainedfrom ISSA is better than PSO. For DS2OS dataset, the classifiers performance with PSO was higher than ISSA based on the features selected in PSO algorithm. When the feature selection approach is compared with Ensemble Learning approach, the Ensemble classifier outperforms for all the three datasets. When the Ensemble classifier is compared with three datasets, the ensemble classifier's performance of IoT-23 is greater than the other two datasets.

V. CONCLUSION

In this work, the attack in IoT network has been detected by using three IoT-Datasets namely DS2OS, BoT-IoT and IoT-23 datasets. The imbalanced classes in the dataset are handled by using SMOTE which also create greater impact in performance of the model in terms of accuracy, precision,recall and f1-score. The ML models and the Ensemble model

TABLE XV
COMPARISON OF PROPOSED WORK WITH EXISTING WORK

References	Dataset Sources	Methods	Results			
			Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Bedi et al. [28]	DS2OS	ANN, DT, RF, LR, SVM	99.40	99.00	99.00	99.00
I. Ullah et al. [24]	BoT-IoT	CNN1D, CNN2D, CNN3D	99.90	99.96	99.91	98.10
N. Abdalgawad et al. [16]	IoT-23	AAE, BiGAN	97.00	1.00	95.00	94.00
Sahu et al. [1]	IoT-23	CNN-LSTM	96.23	95.94	97.06	96.50
Shivanjali Khare et al. [29]	DS2OS	AdaBoost	99.00	99.00	99.00	99.00
Selvakumar. B, et al. [34]	BoT-IoT	CNN LeNet-5	95.86 96.20	-	-	-
Proposed Work	**DS2OS BoT-IoT IoT-23**	Without FS	94.47 95.19 99.24	98.14 90.07 98.44	95.07 96.92 98.55	97.33 96.69 98.38
		With FS (PSO)	99.26 97.63 99.98	96.32 97.52 99.94	99.13 98.77 98.95	97.80 98.58 99.47
		With FS (ISSA)	99.27 99.97 99.76	97.23 99.57 99.88	99.07 98.15 98.02	97.65 96.52 99.39

References	Dataset Sources	Methods	Results			
			Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
		Ensemble with PSO	99.91 98.76 99.97	99.09 98.23 99.93	99.94 98.64 99.97	99.56 98.01 99.94
		Ensemble with ISSA	**99.47** **99.97** **99.98**	**99.48** **99.96** **99.97**	**99.78** **99.93** **99.98**	**99.75** **99.92** **99.98**

is trained, validated and tested efficiently with the prominent features selected by using PSO and ISSA feature selection algorithms. From the results obtained, it is inferred that the Ensemble Model's accuracy is greater than the accuracy obtained for Feature selection model. When compared to the dataset, the IoT-23 dataset performs higher than the
performance of other datasets used in the proposed work with more than 99% accuracy, precision, recall, f1-score and less than 0.030% False positive Rate for all the attacks in all datasets.

ACKNOWLEDGEMENT

This work is supported by DST ASEAN-India Collaborative Research Project scheme with sanction order no: DST/CRD/2020/000352 titled"A Resilient Software-defined IoT Architecture for Net House Monitoring and Management for g+rowing seasonal crops, fruits and flowers throughout the year in Tamilnadu, a southern state of India". The authors express their sincere gratitude to the in-charge of the scheme Mr. R.K. Sharma, Senior Director Scientist-F, Department of Science & Technology (Govt of India), New Delhi for encouragement and guidance.

REFERENCES

Amiya Kumar Sahu, Suraj Sharma, M. Tanveer, Rohit Raja, "Internet of Things attack detection using hybrid Deep Learning Model", Computer Communications, Elsevier, Volume 176, Pages 146-154, 2021.

Asmaa A. Elsaeidy, Abbas Jamalipour and Kumudu S. Munasinghe, "A Hybrid Deep Learning Approach for Replay and DDoS Attack Detection in a Smart City", IEEE Access, Volume: 9, Pages: 154864 – 154875, 2021.

Ah. E. Hegazy, M.A. Makhlouf, Gh. S. El-Tawel, "Improved salp swarm algorithm for feature selection", Journal of King Saud University - Computer and Information Sciences, Volume 32, Issue 3, Pages 335-344, 2020

Aubet, FX and Pahl, MO, "DS2OS traffic traces", 2018, https://www.kaggle.com/datasets/francoisxa/ds2ostraffictraces.

Ban Mohammed Khammas, "Ransomware Detection using Random Forest Technique", Elsevier - ICT Express, Volume 6, Issue 4, Pages 325-33, 2020.

Eirini Anthi, Lowri Williams, Amir Javed, Pete Burnap, "Hardening machine learning denial of service (DoS) defences against adversarial attacks in IoT smart home networks", Elsevier – computer & security, volume 108, 2021.

Elie Alhajjar, Paul Maxwell, Nathaniel Bastian, "Adversarial machine learning in Network Intrusion Detection Systems", Elsevier - Expert Systems with Applications, Volume 186, https://doi.org/10.1016/j.eswa.2021.115782, 2021.

Georgios Tertytchny, Nicolas Nicolaou, Maria K. Michael, "Classifying network abnormalities into faults and attacks in IoT-based cyber physical systems using machine

learning", Elsevier - Microprocessors and Microsystems, doi: https://doi.org/10.1016/j.micpro.2020.103121, 2020.

H.T. Manjula, Neha Mangla, "An approach to on-stream DDoS blitz detection using machine learning algorithms", Elsevier - Materials Today: Proceedings, https://doi.org/10.1016/j.matpr.2021.07.280, 2021

Imtiaz Ullah and Qusay H. Mahmoud, "Design and Development of a Deep Learning-Based Model for Anomaly Detection in IoT Networks", IEEE Access, volume 9, Pages 103906 – 103926, 2021.

Junaid Shabbir Abbasi, Faisal Bashir, Kashif Naseer Qureshi, Muhammad Najam ul Islam, Gwanggil Jeon, "Deep learning-based feature extraction and optimizing pattern matching for intrusion detection using finite state machine", Elsevier - Computers and Electrical Engineering, Volume 92, 2021.

Ly Vu, Quang Uy Nguyen, Diep N. Nguyen, Dinh Thai Hoang and Eryk Dutkiewicz, "Deep Transfer Learning for IoT Attack Detection", IEEE Access, volume 8, pages 107335 – 107344, 2020

Mahdis Saharkhizan, Amin Azmoodeh, Ali Dehghantanha, Kim-Kwang Raymond Choo, Reza M. Parizi, "An Ensemble of Deep Recurrent Neural Networks for Detecting IoT Cyber Attacks Using Network Traffic", IEEE Internet Of Things Journal, volume 7, Issue 9, Pages 8852-8859, 2020.

Malathi. C, Naga Padmaja. I, "Identification of cyber-attacks using machine learning in smart IoT networks", Materials Today: Proceedings, Elsevier, 2020.

Mohammad Tubishat, Norisma Idris, Liyana Shuib, Mohammad A.M. Abushariah, SeyedaliMirjalili, "Improved Salp Swarm Algorithm based on opposition based learning and novel local search algorithm for feature selection", Expert Systems with Applications, Volume 145, 2020

N. Abdalgawad, A. Sajun, Y. Kaddoura, I. A. Zualkernan and F. Aloul, "Generative Deep Learning to detect Cyberattacks for the IoT-23 Dataset", IEEE Access, Volume 10, Pages: 6430-6441, 2021.

Nour Moustafa, October 16, 2019, "The Bot-IoT dataset", IEEE Dataport, doi: https://dx.doi.org/10.21227/r7v2-x988.

Ömer KASIM, "An Efficient and Robust Deep Learning based Network Anomaly Detection against Distributed Denial of Service Attacks", Computer Networks, Elsevier, 2020.

Reddy SaiSindhuTheja, Gopal K. Shyam, "A machine learning based attack detection and mitigation using a secure SaaS framework", Journal of King Saud University – Computer and Information Sciences, https://doi.org/10.1016/j.jksuci.2020.10.005, 2020.

Sebastian Garcia, Agustin Parmisano, & Maria Jose Erquiaga. (2020). IoT-23: A labeled dataset with malicious and benign IoT network traffic (Version 1.0.0) [Data set]. Zenodo. http://doi.org/10.5281/zenodo.4743746

Segun I. Popoola, Bamidele Adebisi, Mohammad Hammoudeh, Guan Gui, Haris Gacanin, "Hybrid Deep Learning for Botnet Attack Detection in the Internet of Things Networks", IEEE Internet of Things Journal, Volume: 8, Issue: 6, Pages: 4944 – 4956, 2021.

SeyedaliMirjalili, Amir H. Gandomi, Seyedeh Zahra Mirjalili, Shahrzad Saremi, Hossam Faris, Seyed Mohammad Mirjalili, "Salp Swarm Algorithm: A bio-inspired optimizer

for engineering design problems", Advances in Engineering Software, Volume 114, Pages 163-191, 2017.

Swathi Sambangi and Lakshmeeswari Gondi, "A Machine Learning Approach for DDoS (Distributed Denial of Service) Attack Detection Using Multiple Linear Regression", 14th International Conference on Interdisciplinarity in Engineering—INTER-ENG, MPDI Proceedings, 2020.

Ullah, I., Ullah, A., Sajjad, M., "Towards a Hybrid Deep Learning Model for Anomalous Activities Detection in Internet of Things Networks", MDPI – IoT, volume 2, Issue 3, Pages: 428-448, 2021.

Zhida Li, Ana Laura Gonzalez Rios, and Ljiljana Trajkovic, "Machine Learning for Detecting Anomalies and Intrusions in Communication Networks", IEEE Journal on Selected Areas in Communications, Volume: 39, Issue: 7, Pages 2254 – 2264, 2021.

Zhihong Tian, Chaochao Luo, Jing Qiu1, Xiaojiang Du, Mohsen Guizani, "A Distributed Deep Learning System for Web Attack Detection on Edge Devices", IEEE Transactions on Industrial Informatics, Volume: 16, Issue: 3, Pages: 1963 – 1971, 2020.

Abdullahi, M.; Baashar, Y.; Alhussian, H.; Alwadain, A.; Aziz, N.; Capretz, L.F.; Abdulkadir, S.J. Detecting Cybersecurity Attacks in Internet of Things Using Artificial Intelligence Methods: A Systematic Literature Review. Electronics 2022, 11, 198. https://doi.org/10.3390/electronics11020198

Pradeep Bedi, Shivlal Mewada, RasmbabuArjunaraoVatti, Chaitanya Singh, Kanwalvir Singh Dhindsa, Muruganantham Ponnusamy, Ranjana Sikarwar, "Detection of attacks in IoT sensors networks using machine learning algorithm", Microprocessors and Microsystems, Volume 82, https://doi.org/10.1016/j.micpro.2020.103814, 2021.

Shivanjali Khare; Michael Totaro, "Ensemble Learning for Detecting Attacks and Anomalies in IoT Smart Home", IEEE - 2020 3rd International Conference on Data Intelligence and Security (ICDIS), 2020, https://doi.org/10.1109/ICDIS50059.2020.00014.

UshamSanjota Chanu, Khundrakpam Johnson Singh, YambemJina Chanu, "An ensemble method for feature selection and an integrated approach for mitigation of distributed denial of service attacks", Concurrency and Computation: Practice and Experience, Volume 34, Issue 13, https://doi.org/10.1002/cpe.6919, 2022.

Reneilson Santos, Danilo Souza, Walter Santo, Admilson Ribeiro, Edward Moreno, "Machine learning algorithms to detect DDoS attacks in SDN", Concurrency and Computation: Practice and Experience, Volume 32, Issue 16, https://doi.org/10.1002/cpe.5402, 2019.

Seong Il Bae, Gyu Bin Lee, Eul Gyu Im, "Ransomware detection using machine learning algorithms", Concurrency and Computation: Practice and Experience, Volume 32, Issue 18, https://doi.org/10.1002/cpe.5422, 2019.

Selvakumar, B., Sridhar Raj, S., Vijay Gokul, S., Lakshmanan, B, "Deep Learning Framework for Anomaly Detection in Iot Enabled Systems", In: Makkar, A., Kumar, N. (eds) Deep Learning for Security and Privacy Preservation in IoT. Signals and Communication Technology. Springer, Singapore. https://doi.org/10.1007/978-981-16-6186-0_5

Selvakumar B, Muneeswaran K, "Firefly algorithm based feature selection for network intrusion detection", Computers & Security, Volume 81, Pages 148-155, 2019.

Web Reference: https://www.scmagazine.com/news/architecture/torii-malware-could-be-gateway-to-more-sophisticated-iot-botnet-attacks

Web Reference: https://cyware.com/news/the-mirai-mania-a-brief-look-into-the-notorious-mirai-botnet-and-its-variants-37c443f8

Kelton A.P. da Costa, João P. Papa, Celso O. Lisboa, Roberto Munoz, Victor Hugo C. de Albuquerque, "Internet of Things: A survey on machine learning-based intrusion detection approaches", Computer Networks, Volume 151, Pages 147-157, 2019.

Kumar, R.; Subbiah, G, "Zero-Day Malware Detection and Effective Malware Analysis Using Shapley Ensemble Boosting and Bagging Approach". Sensors, Volume 22, Issue 7. https://doi.org/10.3390/s22072798, 2022.

Yeduri Sreenivasa Reddy, Ankit Dubey, Abhinav Kumar, and Trilochan Panigrahi, "A Successive Interference Cancellation Based Random Access Channel Mechanism for Machine-to-Machine Communications in Cellular Internet-of-Things", IEEE Access, volume 9, Pages 8367-8380, doi: 10.1109/ACCESS.2021.3049439, 2021.

Shubham Gupta, Balu L. Parne, Narendra S. Chaudhari, "ISAG: IoT-enabled and Secrecy Aware Group-based handover scheme for e-health services in M2M communication network", Future Generation Computer Systems, Volume 125, Pages 168-187, 2021.

Ashraf Darwish, "Bio-Inspired Computing: Algorithm Revie, Deep Analysis And The Scope Of Applications", Future Computing And Informatics Journal, Volume 3, Issue 2, Pages 231-246, 2018.

Danthala, S. W. E. T. H. A., Et Al. "Robotic Manipulator Control By Using Machine Learning Algorithms: A Review." International Journal Of Mechanical And Production Engineering Research And Development 8.5 (2018): 305-310.

Durgabai, R. P. L., And P. Bhargavi. "Pest Management Using Machine Learning Algorithms: A Review." International Journal Of Computer Science Engineering And Information Technology Research (IJCSEITR) 8.1 (2018): 13-22.

Nawalagatti, Amitvikram, And R. Kolhe Prakash. "A Comprehensive Review On Artificial Intelligence Based Machine Learning Techniques For Designing Interactive Characters." International Journal Of Mathematics And Computer Applications Research (IJMCAR) 8.3 (2018): 1-10.

Mathur, Geetika, Harshit Sharma, And Rishabh Pandey. "A Study On Self-Driving Car An Application Of Iot." International Journal Of Computer Networking, Wireless And Mobile Communications (IJCNWMC) 9 (2019): 25-34.

Manivannan, T., And P. Radhakrishnan. "Preventive Model On Quality Of Service In IOT Applications." Int. J. Mech. Prod. Eng. Res. Dev 10.3 (2020): 1247-1264.