ISSN: 2633-4828 Vol. 5 No.4, December, 2023

International Journal of Applied Engineering & Technology

HYBRID OPTIMIZATION ALGORITHM WITH RANDOM FOREST CLASSIFIER FOR IOT TRAFFIC CLASSIFICATION

[1]Abhishek Kumar Dubey and [2]Vinod Kumar Singh

[1]M.Tech Scholar, Department of Electronics & Communication Engineering, SR Group of Institutions, Jhansi, India

[2]Professor and Dean, Research & Development, SR Group of Institutions, Jhansi, (U.P.) India

information.dubey.abhishek@gmail.com, singhvinod34@gmail.com

ABSTRACT

The IOT network traffic classification is a scheme for analysing traffic in Internet of Things. This scheme is executed for classifying diverse activities in network. The traffic is classified in diverse stages in which the data is utilized into input, pre-processed, features are extracted, traffic is classified and performance is analyzed. Over past few decades, several ML techniques have deployed, however they are not effective concerning accuracy. These techniques have not potential for extracting features from the dataset. This work suggests an approach, which is efficient for extract features from the dataset at superior accuracy while classifying traffic. It is a hybrid algorithm in which GA is integrated with particle swarm optimization algorithm. The suggested algorithm is capable of extracting features and classifying them on the basis of RF technique. Python is executed for simulating the suggested algorithm concerning accuracy, precision, recall.

Keywords: IOT, PSO, Genetic, Random Forest

1. INTRODUCTION

Internet of Things (IoT) is a network in which embedded systems are exploited to communicate with either wireless or wired schemes. The network of physical items whose embedding is done with electronics and internet to allow them for gathering, occasionally processing, and exchanging data is another definition of IoT. These devices have limited computing, storage, and communication abilities. The term "things" in IoT is used to define items from daily basis, ranging from smart home devices, including tubes, connecters, fridges, stoves, and ACs to recent appliances such as RFID, heartbeat detection method, etc. [1]. It offers an extensive variety of applications and services, from vital structure to farming, army, appliances, and medical field. According to safety professionals, a great threat is occurred to diverse gadgets having association with Internet due to ignorance of providers for utilizing security mechanism in these gadgets. The safety and privacy threats are taken place, setting the standard for customer anxieties regarding securing data online, becomes a major task. The reason is, such gadgets can both monitor and collect personal data from users. Users are cautious of storing too much private information in clouds for decent cause in the wake of the apparently endless stream of revelations about significant data breaches.

1.1 Security Features in IoT

Depending on its function, an IoT system may have different security requirements. An IoT system should perform the specified tasks, defend against attacks, and continue to function normally in the event of an attack. This system has to work securely and robustly to keep the data secret, reliable, available, and authentic. The goal of secrecy is to prevent unauthorized individuals from accessing information [2]. An IoT system includes numerous components, and each one is capable of transmitting data with another motes. Reliability refers to making sure the authenticity of information (the data source is legitimate, and the data hasn't been changed.). Data integrity guarantees service quality as well as data security and privacy.

Maintaining the integrity of the data is crucial, especially for outsourced data or data from third parties, as cloud computing becomes more widely available and data quantities continue to rise. Accessibility refers to having data or services available when and when users need them. The main objective of IoT security is to guarantee each IoT application's operation and availability. Authentication is the process of confirming an ID of an individual and to

prevent from unapproved access and improper usage of any secret data. The extensive usage of Internet of Things leads to create complexity to implement an authentication system in this scenario [3]. Several applicants, like tools, users, SPs, are comprised in this system.

1.2 Security Attacks in IoT Layers

In particular, its applications and expertise are not advanced properly. To develop IoT and its deployment is much complex due to numerous problems, like technical restraints, inadequate criteria, and lower safety and secrecy features. New IoT technologies and services must be widely accepted in order to become commonplace. This mostly depends on the state of data security and privacy protection, which are complex challenges for IoT due to its deployment, portability, and intricacy.

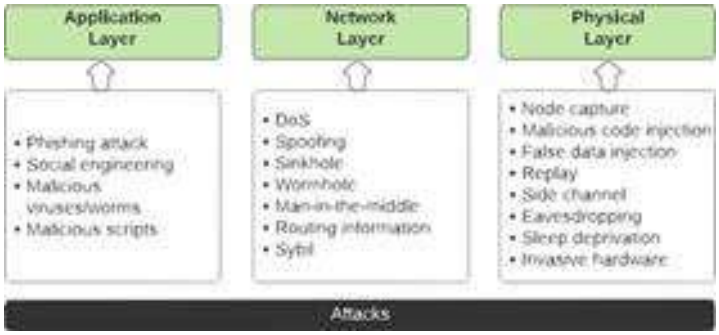

Fig. 1. Main attacks in each layer Security challenges associated with each IoT layer are shown in Fig. 1.Physical Layer and Security Issues

Its primary functions are to provide device identification and provide service discovery. The bottom layer of the architecture is essentially the layer responsible for the networked devices [4]. Physical layer assaults are occurred on the hardware components of mechanism and the physical presence of malevolent is essential near or within this mechanism during their activation. Node capture, malicious node injection, fake data injection, replay, side channel, eavesdropping, sleep deprivation, and intrusive hardware are the main attacks in this layer.

I. Network Layer and Security Issues

Communication between different systems and devices within the Internet of Things (IoT) ecosystem is made possible in largely by this layer. The assailant is focused on routing to activate assault while analysing data and traffic, deceiving, and MITM assaults. Moreover, sybil assaults are also feasible, in which phoney and sybil IDs result in dodging network. One of the key services offered at the network layer is routing, which requires current routing data in the routing tables [5]. The normal operation is compromised, and secret data may potentially leak in case of spoofing and modification of routing information via assailants. The QoS is also compromised by changes in forwarding nodes. The attacker can transmit their own data rather than legal data when it is forged and cloned.

II. Application Layer and Security Issues

The service-oriented nature of the application layer ensures that all connected devices receive the same level of service. It can provide storage capabilities for the gathered data by storing it in a database. Also, this layer assists in associating expedients to others through the use of various applications that meet the requirements of the users. IoT apps are particularly profitable targets for attackers since application-level attacks are comparatively easy to initiate. BO, malware, DoS, MITM intrusions are easily launched in it.

1.3 Machine Learning in IoT Security

ML is a branch of AI that combines a variety of methods and schemes to give computers and other smart devices intelligence. The network security landscape has adopted Machine Learning approaches such as SL, USL, and RL to a large extent [6]. It is used to precisely identify and specify the specific security rules that must be applied in

the data plane. Defining rules for limiting access or labelling network traffic are two ways to fine-tune the various security protocol parameters in to prevent a certain sort of attack. In fact, many ML methods can defend against a range of IoT attacks. Supervised learning algorithms are used in Internet of Things to detect range, estimate channel, secure data and localizing issues. They operate on labelled data. The present datasets are predicted and modelled after classifying data for which, supervised machine learning is beneficial. Constant mathematical values can be predicted using regression. Support Vector Machine, NB, RF, and DT are extensive mechanisms. Nonetheless, LR and KNN are utilized in second category. Such methods emphasize on data, used to train system, for predicting novel statement.

The subsequent category of ML aims to deploy unlabelled data and utilizes the input data heuristically [7]. Such approaches are useful to balance load, cluster cell, detect attack and irregularity. Data is clustered dependent on some underlying differences and connections are accomplished when the data is clustered, that comes under the area of USL. The USL nature of clustering avoids the responses whether they are precise or not. The data is visualized for computing the accuracy in outcomes. It is possible to pre-label the clusters in datasets in case of probability of classifying responses as right and wrong answer, and in case of supremacy of classifiers.

Reinforcement strategies strengthen the reward and action relationship of agent with scenario on the basis of diverse phases. Such an association is utilized to deal with diverse problems in Internet of Things. The state transition function must be known to agent, but there is no need for a large training data set. Almost all the applications benefit from USL methods at slight initial ecological information. For instance, IoT network zero-day attacks typically start without much or no information. As a result, the class of machine learning (ML) called USL has robustness in internet of things to thwart such assaults.

The initial category of Machine Learning such as SVM, RF, DT, and NB also acts significantly in securing the network. For example, the first algorithm is capable of modelling non-linear decision boundaries [8], but their use becomes challenging with huge data sets. The second algorithm has more efficacy as compared to first one. These techniques are simpler to use and can adapt to any size of the dataset at hand. It predicts more quickly and with a higher degree of accuracy. However, different from other algorithms, training of this approach is longer. Naïve Bayes is appropriate for concerns related to spam detection and text classification. The KNN and LR models have higher memory usage and struggle with data of higher dimensionality. KM and HC are two well-liked models when it comes to unsupervised ML techniques. The most widely used clustering technique is K-means clustering, for clustering the data according to geometric partings among data points. The centroids are considered to integrate clusters for generating the globular arrangements having similar size [9]. The data is clustered after determining various clusters which cause infeasibility or inefficiency. Moreover, clusters are not generated precisely in case of absence of spherical clusters.

The last kind of methods are simple concerning computation, though their time usage is higher for attaining a stable position. The primary difficulties in using these methods in the dynamic environments of Internet of Things are their late convergent rate, and the state transition function is much difficult to understand. The ideal course of action is discovered by trial and error and after numerous changes. Deep learning also makes use of robust function approximation, estimate, and learning capabilities, which results in more effective solutions in a number of IoT problem areas, such as safety and protection [10]. Owing to resource limitations, its devices might incapable of hosting or operating sophisticated models to execute any kind of job, including communication, analysis, and predicting data. Thus, DL-based methods are proved superior in contrast to existing approaches due to least latency and difficulty. The ability of DNNs to identify and locate low dimensional illustrations from higher dimensionality data patterns is another advantage. In heterogeneous scenario, DRL and its derivatives are used for authentication and DDoS detection. Sone of these methods are DDPG, PER, ANSQL and DNDQN which are effective to secure network.

1.4 Metaheuristic Algorithms for Feature Selection

Feature Selection is a crucial task in optimizing the feature space, which is an n-dimensional space that represents each sample as a point. It involves identifying the most relevant and significant features while discarding irrelevant or unnecessary ones, particularly when dealing with large feature spaces. This process is considered challenging in machine learning, and researchers have recently turned to metaheuristic algorithms to address it. These algorithms are intelligent optimization techniques that combine random and local search algorithms, allowing for global search abilities without requiring domain knowledge or advanced assumptions about the search space. Metaheuristic algorithms are based on computational intelligence mechanisms that can solve complex optimization problems and produce best or reasonable solutions. Several metaheuristic algorithms are commonly used for feature selection which are described as follow:

I. ***Particle Swarm Optimization***: This algorithm is developed on the basis of the social behaviour of some creatures. The major focus of this algorithm is on tackling the optimization issues in continuous search space. Its standard algorithm employs various particles which lead to establish a population. In this, every particle, responsible for maintaining a location and a velocity in a n- dimensional search space, is considered for illustrating a potential solution of the optimized issue [11]. The evolutionary procedure is consisted of two examples: Gbest and Pbest which are utilized to adjust the velocity and location of every particle. This procedure is executed for discovering the global optimization solution which Gbest has shown. Each iteration assists in updating the novel velocity and position of every particle as:

$$v_{id}^{t+1} = w \cdot v_{id}^{t} + c_1 r_1 \cdot \left(Pbest_{id}^{t} - x_{id}^{t}\right) + c_2 r_2 \cdot \left(Gbest_{d}^{t} - x_{id}^{t}\right) \quad (1)$$

$$x_{id}^{t+1} = x_{id}^{t} + v_{id}^{t+1} \quad (2)$$

In this, t denotes the generation number, d is used to illustrate the dimension of a vector, the velocity of the ith particle in the tth generation is represented with v_{id}^{t} and position with x_{id}^{t}. All the particles lead to discover the finest solution within the tth generation which is shown via $Gbest_{d}^{t}$ and $Pbest_{id}^{t}$ illustrates the finest solution in ith particle; an inertia weight is denoted with w for verifying the amount of previously preserved velocity; c_1 and c_2 are utilized to represent 2 positive constants. These constants are also known as acceleration coefficients. r_1 and r_2 denote 2 random numbers whose generation is generated within [0,1].

II. ***Ant Search***: The colonies behaviour of ants has great influence on developing this algorithm. This algorithm employs a colony of cooperating agents to resolve the optimization issue. In this algorithm, every ant has potential for following a prior ant that results in releasing a substance. Hence, the entire ant colony is capable of self-organizing itself. A positive feedback loop is responsible for the emerging collective behavior. It is a probability of selecting a certain route by an ant, which is maximized when the similar route is selected by numerous ants. The reason is the constant reinforcement of the track with novel pheromone. This algorithm is useful for addressing the issues related to select the features. For this, the probability p_{ij}^{k} of arrival of the kth ant to feature j starting from the attribute i, is computed as:

$$p_{ij}^{k} = \begin{cases} \dfrac{[\tau_{ij}]^{\alpha} \cdot [\eta_{ij}]^{\beta}}{\sum_{k \epsilon \mathcal{U}_k} \tau_{ik}^{\alpha} [\eta_{ik}]^{\beta}} & if\ j \epsilon \mathcal{U}_k \\ 0 & otherwise \end{cases} \quad (3)$$

In this, $\mathcal{U}_k$ defines the set of feasible features, τ_{ij} is used to illustrate the amount of pheromone across the ij route, the heuristic information denoted with η_{ij} which is related to the selected attribute j; α and β are used to represent the metrics. The first one is utilized in order to control the pheromone trials and latter one for heuristic information [12].

III. ***Cuckoo Search***: This algorithm is planned on the basis of brood parasitism approach which characterizes some cuckoo's species. Generally, such species aims to lay their eggs in the nests of other birds. The host birds are capable of engaging a conflict due to their ability of recognizing the alien eggs. Some specific cuckoo species are grown in such a manner that females can imitate the colour and size of eggs of some host species. It results in cheating the hosts, and the probability of cuckoos reproductivity is developed. Let an egg in a nest be a solution, 3 rules are considered in Cuckoo Search (CS) approach in which every cuckoo focuses on laying one egg at time, in a random nest; the bests nests are considered as the candidate for carrying over next generations; a limited number of nests are utilized and, as a host bird is employed for determining the alien (cuckoo) eggs of a probability p_d, and escape from it. This algorithm emphasizes on deploying novel solutions rather than others in the nest. The given equation is executed for attaining novel solutions for cuckoo i for random walk as:

$$x_i^{t+1} = x_i^t + \alpha \otimes \mathcal{L}(\lambda) \quad (4)$$

In this, $\alpha > 0$ illustrates a scale factor, the product $\otimes$ is employed for entry-wise multiplications, and the Lévy distribution with $(1 < \lambda \leq 3)$ is denoted with $\mathcal{L}$.

IV. ***Ant Lion Optimization (ALO)***: This approach is developed according to the stalking actions of ant lions [43]. This algorithm is implemented for simulating the connections amid ant lions and the prey in the trap. For searching the food, ants are moved at random. Ant lions make those ants their prey with their traps during wandering of ants. The modelling of random movement of ants is done as:

$$X(t) = \sum_{i=1}^{t} 2r(t_i) - 1 \quad (5)$$

In this, the number of random walk steps is denoted with t, and $r(t)$ represents a random value within 0 and 1

2. LITERATURE REVIEW

M. Bagaa, et.al (2020) suggested an innovative ML-enabled security mechanism for tackling security aspects about IoT domain, in automatic way [13]. In this Artificial Intelligence (AI) model, the monitoring agent was integrated with AI-based reaction agent in which ML-methods were utilized and tested. The distributed Data Mining (DM) system and Neural Network (NN) algorithm had implemented to mitigate the attacks. The results indicated that the suggested model was effective. Particularly, DM algorithm was utilized to distribute the assaults due to which the attacks were detected efficiently at lower cost. Furthermore, one-class SVM was implemented to compute the suggested model on real smart building case. This model offered an accuracy of 99.71% and it was proved feasible.

H. Lee, et.al (2023) projected an approach in which features of lower-dimensionality and fixed- length were considered based on opcode category information on Machine Learning (ML) algorithms [14]. This approach focused on transforming the binary Internet of Things (IoT) dataset into opcode for generating the features. A sequence and the entropy values of opcode kinds were employed for generating the attributes. A two-dimensional (2D) image was utilized for visualizing these features so that the patterns were observed. Different ML methods called 5- NN, SVM, DT, RF and MLP were utilized in computing the projected approach concerning Accuracy, Precision, Recall, F1-score, and MCC. The projected approach yielded 98% accuracy. According to the outcomes, the effectiveness and strength of the projected approach is useful to recognize different kinds of IoT malware.

W. Ma, et.al (2021) recommended a ML-based trust evaluation method [15]. A deep learning (DL) algorithm was exploited for aggregating the trust possessions of network such as QoS, and a behavioral method was developed for a given Internet of Things device considering the time- reliant attributes. The resemblance amid real and network behaviors which this method had predicted, was computed for evaluating the trust values. These values helped in illustrating the trust rank of an IoT node, and employed to make the decision. In the end, experiments were conducted, and the investigated approach was worked more robustly than other approaches.

T. Gaber, et.al (2022) recommended an intrusion detection system (IDS) for detecting the injection assaults in Internet of Things (IoT) applications [16]. This system deployed two methods: constant removal (CR) and recursive feature elimination (RFE), and various Machine Learning (ML) algorithms called SVM, RF, and DT were utilized to detect the attacks. An analysis was conducted on the recommended system using AWID dataset. The experimental results exhibited that the recommended system performed effectively with DT and attained 99% accuracy for detecting the injection assaults. Furthermore, the supremacy of this system was proved over the traditional methods.

D. Mishra, et.al (2022) developed an LGBM method for detecting the intrusive doings in IoT [17]. This method was useful to recognize the malicious activities in IoT network. The ideal set of hyper-parameters of this method was discovered using an effective evolutionary optimization technique. Thus, a GA with KWTS and UCO had adopted to examine the hyper-parameter search space. At last, the simulation was conducted on the developed method with respect to generalized potential and efficacy. In results, the developed approach outperformed the traditional methods, and worked robustly against the intrusions in an IoT environment.

R. Banavathu, et.al (2023) introduced a NAIBSM for securing the way to store data effectively in Internet of Things-based SCSs [18]. It was flexible for capturing every user authentication so that diverse assaults such as DDOS were detected. For this, Artificial Intelligence (AI) technique called Leverage Bat algorithm (LBA) for discovering the complicated features. Blockchain was developed to establish a communication securely and make the stored data reliable on the terminal of IoT. The hash values were obtained for making the data reliable. The data was arranged, stored and classified for every user using a weight-based data storage process. The experimental results revealed that the introduced model was secure, and provided lower communicating overhead and higher accuracy.

E. Gelenbe, et.al (2022) presented a new online CDIS which identified IoT devices and IP addresses affected with a Botnet attack [19]. This system focused on selecting the precise metric after extracting them from traffic. An AADRNN technique was executed to train this system online based on traffic parameters. An AAL was assisted in training this approach for estimating the traffic as authentic and attacked. The experimental results on Mirai data confirmed that the presented system attained an accuracy up to 97%, at lower training and execution time. Furthermore, this system was assisted in acquiring significant information for preventing the spread of Botnet assaults in IoT networks in which several devices and IP addresses were utilized.

H. Kim, et.al (2021) established an approach called Panop that was an Artificial Neural Network (ANN)-based network intrusion detection system (NIDS) for a distributed network system for detecting the malevolent packets, host-oriented assaults [20]. This method was relied on learning the network and device behaviors about transmitting the packet in Internet of Things (IoT) network. This approach was adaptable on low-end gateway operating in IoT. The results of experimentation depicted that the established approach was effective, practical and robust to detect real attacks. Raspberry Pi was executed to compute this approach. The results reported that the established approach offered a superior accuracy in comparison with the traditional methods.

M. S. A. Muthanna, et.al (2022) designed an intelligent cuLSTMGRU algorithm was implemented to efficiently detecting the attacks in Internet of Things (IoT) [21]. The IoT-based dataset was executed to quantify the designed system in terms of diverse parameters. This system yielded an accuracy of 99.23% for detecting attacks and lower false-positive rate (FPR). The designed system was simulated with two techniques. The results validated the superiority of the designed system over others with respect to speed efficacy, accuracy, and precision.

A. K. Dey, et.al (2023) integrated STF approaches (e.g., CS, PCC, and MI) with a NSGA-II- based metaheuristic solution to develop a hybrid feature selection approach []22. Filter-based techniques were used to rank the features, guiding population initialization in NSGA-II for quicker convergence in the direction of a solution. The performance of the new approach was assessed by applying it to the ToN-IoT dataset. The accuracy and amount of attributes were compared to those obtained by other cutting-edge schemes. In analysis, the new approach outperformed other techniques, with only 13 optimized features selected and a highest accuracy of 99.48%.

M. S. Hossain et.al (2022) presented a novel approach for detecting Android ransomware using traffic analysis to overcome previous challenges [23]. The approach utilized particle swarm optimization (PSO) to identify relevant traffic features. The data traffic was classified using decision tree and random forest classifiers according to the selected traffic features, allowing the researchers to detect ransomware cyber-attacks in two different scenarios. The presented PSO- based feature selection approach substantially improved detection accuracy. According to the study, the random forest classifier performed better in detecting ransomware, while the decision tree classifier was found to be more effective in detecting specific types of ransomwares. The new approach achieved accuracy improvements of 2.26% in the first and 3.7% in the second scenario, and reduced the number of features used by 56.01% to 91.95%. The optimization process reached an optimal value in approximately ten iterations, demonstrating fast convergence.

M. Stankovic, et.al (2022) projected an ensemble of ABC metaheuristics for maximizing the accuracy of the extreme learning machine classifier for feature selection [24]. The proposed framework was tested using two well-known network security datasets named UNSW-NB15 and CICIDS-2017 to demonstrate its performance improvement. The study evaluated the performance of suggested method in simulation findings against traditional methods in comparable situations for the same purpose.

M. Tubishat, et.al (2020) introduced two significant improvements to the original BOA, including LSAM model to overcome local optima and for augmenting the solutions range [25]. They then employed 20 UCI data sets for computing the Dynamic Butterfly Optimization Algorithm (DBOA) for feature selection problems. This approach was evaluated against its comparative algorithms based on various performance metrics such as accuracy, fitness values, and convergent rate. The experimentation indicated that DBOA exhibited better performance than the relative algorithms across most of the performance indicators considered.

R. Al-Wajih, et.al (2021) introduced a new method called HBGWOHHO, which was a hybrid of GWO and HHO algorithms for feature selection. They used a STF to convert the continuous search space into a binary one, as required for selecting attributes [26]. Researchers further evaluated the quality of selected features using a wrapper-based k-Nearest neighbour. They assessed the effectiveness of their approach on 18 standard UCI benchmark datasets and compared it with other metaheuristic solutions. The findings demonstrated that HBGWOHHO achieved better performance in light of accuracy, dimensionality, and computing time in contrast to both the BGWO algorithm and other feature selection algorithms such as BPSO and BGA.

3. RESEARCH METHODOLOGY

The algorithms for classifying IoT traffic will assist in determining the kind of traffic flowing through the network. There are several phases involved in the IoT traffic classification models, including pre-processing the data collection, extracting features, classifying the data, and analyzing efficacy. Many strategies have been put out in recent years for the effective classification of network traffic. We must take into consideration the numerous shortcomings of the current schemes when conducting our investigation. Because of the vastness of the KDD dataset, current approaches are unable to establish a relationship between each property and the target set. This research project will provide a novel approach that may retrieve dataset features for effective categorization. The efficacy of the IoT traffic classification will be enhanced by the hybrid models for classification that will be suggested. The following technique is a description of the research project's motivation, which is to increase accuracy: -

1. Data set input and Pre-processing: - Data collected from the official source known as KDD is used as input in the first step, which is the dataset input. 42 attributes in the NSL-KDD dataset used in this investigation are compromised. In order to improve the KDD'99 datasets and get rid of the biased classification outcomes, duplicate occurrences are removed. Just twenty percent of the training set is used. There are, however, several versions of the data set available. The format used to express this data is KDDTrain+_20Percent.

2. Feature Extraction: - The crucial step in establishing the relationship between the attribute set and the target set is extraction of features. Combining the PSO and genetic algorithms results in the hybrid optimization algorithm. The hybrid form of the PSO and Genetic algorithms is represented by the suggested flowchart. This algorithm is helpful in choosing the optimization characteristics and encoding a workable solution to a problem in a person. In actuality, each person is seen as an entity that supports chromosomal traits. Together, a number of people form a population. Prior to using a genetic algorithm (GA), the main goal is to randomly create an assortment of chromosomes and surround it with problem-related variables. The subsequent stage focuses on analyzing the generated chromosomal data. The chromosomes are important for creating more chromosomes since they may clearly illustrate the best way to approach the problem. The population in this algorithm refers to the main collection of alternate solutions that are accessible. To perform coding for an answer to address the problem, each member of the population is illustrated using a chromosome. The formula for decoding is as follows:

$$X = X_{min} + \frac{X_{max} - X_{min}}{2^{Nx} - 1} \sum_{n=0}^{N^x - 1} b_n^X 2^n \qquad (6)$$

which describe the binary representations of X's as b_0^X ,.....,$b_{N^x - 1}^X$. The generations, or different iterations, are used to create the chromosomes. A variety of fitness indicators are used in each generation to assess the chromosomes' fitness value. A relation exists between each particle i and two vectors: the velocity vector $\boldsymbol{V}_i = \left[v_{i,1}, v_{i,2},...,v_{i,n}\right]$ and the position vector $\boldsymbol{X}_i = \left[x_{i,1}, x_{i,2},...,x_{i,n}\right]$. To carry out their search procedure, the positions of novel ideas, $x_{i,d}$, are changed at a steady rate. This approach aims to remind each particle of its past location, denoted as $\boldsymbol{pbest}_i$,, and defines $\boldsymbol{gbest}$ as the current global optimal position that the particle swarm as a whole has found. The sites were found, which resulted in updating each particle's position and velocity in accordance with the provided equations as:

$$v_{i,d}(t+1) = \omega \cdot v_{i,d}(t) + c_1 \cdot rand_1 \cdot \left(\boldsymbol{pbest}_{i,d} - x_{i,d}(t)\right) + c_2 \cdot rand_2 \cdot \left(\boldsymbol{gbest}_d - x_{i,d}(t)\right) \qquad (7)$$

$$x_{i,d}(t+1) = x_{i,d}(t) + v_{i,d}(t+1) \qquad (8)$$

This uses t to show the t-th iteration, d to show the particle's d-th dimension, ω to indicate the inertia weight, c_1 and c_2 to show the acceleration constants, and $rand_1$ and $rand_2$ to specify the random numbers, and their spread is carried out at random within the range [0, 1]. The algorithm's effectiveness is improved by reducing the inertia weight ω. The definition of this weight is:

$$\omega = \omega_{max} - (\omega_{max} - \omega_{min}) \cdot \frac{t}{t_{max}} \qquad (9)$$

This uses ω_{max} to indicate the maximum weight, ω_{min} to indicate the minimum weight, t to indicate the number of the current iteration, and t_{max} to specify the number of the maximum iteration. To combine two chromosomes from the present generation to create the child, a crossover operator or a mutation operator (MO) is used. An inventive generation keeps the population size constant. Based on fitness standards, some parents and kids are chosen, while others are disqualified for giving birth to this generation. Fitter chromosomes can come in a variety of forms.

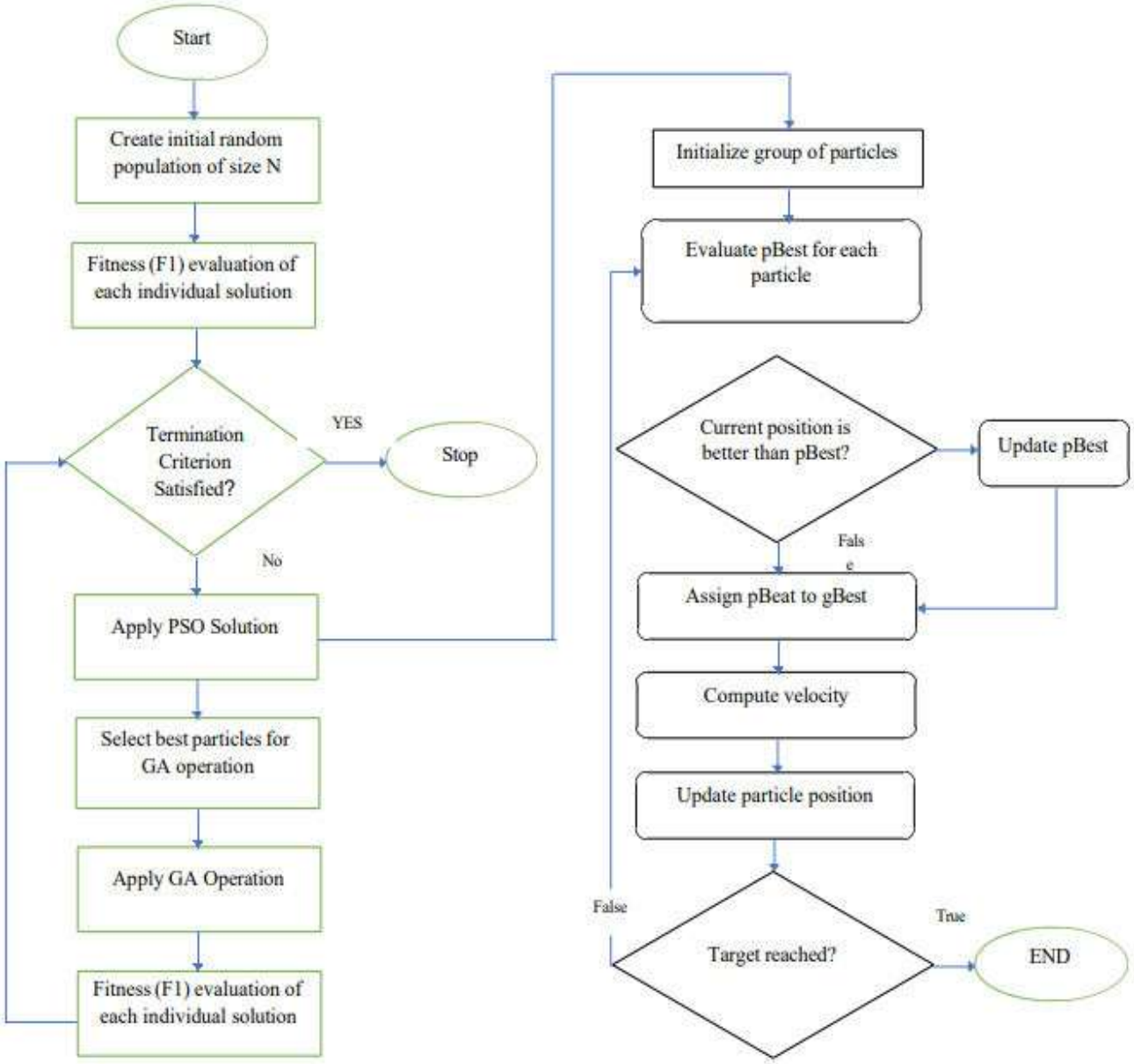

Fig 2: Proposed Hybrid Optimization Algorithm

3. Classification: - The classifier used for the classification is the random forest. The optimization procedure is removed from the random forest model's input for classification. When multiple small, weak DTs grow simultaneously in Random Forest, the trees are averaged or receive the majority vote to form only one, powerful learner. The RFs are often investigated as the most accurate learning algorithms available for training. In formal terms, a predictor made of a series of randomized base regression trees is called an RF. The uniformly dispersed outputs of an arbitrarily produced variable are represented by the variables $\left\{r_n\left(x, \ominus_m, D_n\right), m \geq 1\right\}$, where $\ominus_1, \ominus_2$ ..... These RT integrations are carried out in order to provide the aggregated

regression estimation.

$$\bar{r}_n(X, D_n) = \mathbb{E}_{\Theta}[r_n(X, \Theta, D_n)], \quad (10)$$

Here, $\mathbb{E}_{\Theta}$ indicates what is anticipated as a function of the random parameter, subject to X and the data set D_n. To make the sample a bit simpler, the estimations' dependence would be removed, and it would be expressed as $\bar{r}_n(X)$ instead of $\bar{r}_n(X, D_n)$.. Monte Carlo was utilized to derive the above expectation when the M RTs are produced and the average of the participant's outcomes is obtained. In the process of building individual trees, when the split position and split coordinate are chosen, the effectiveness of the succeeding cuts is evaluated using the randomization variable . The variable is inferred as the independent of X and the training sample D_n.

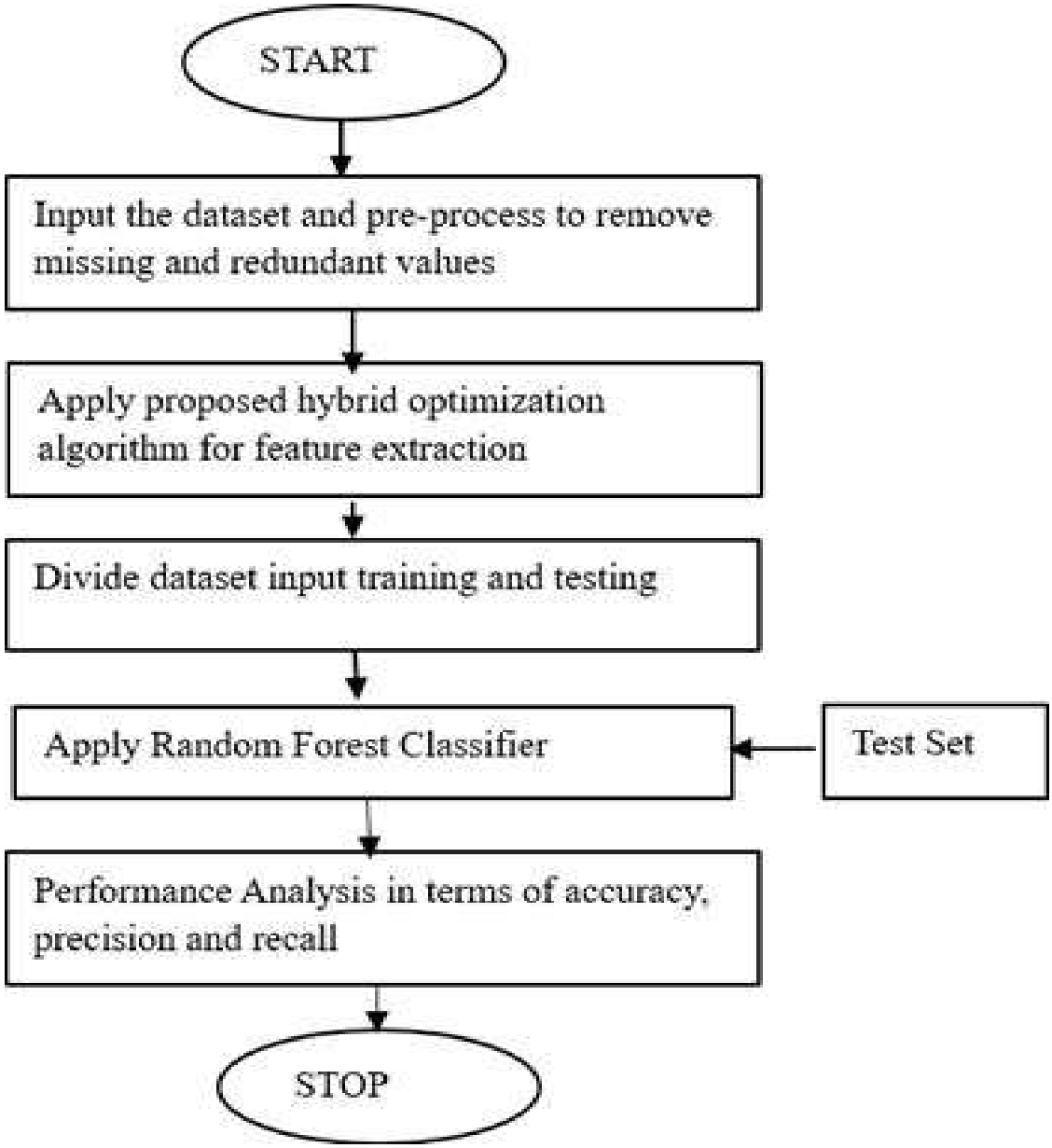

Fig 3: Proposed Model

3.1. Proposed Hybrid Optimization Algorithm

Algorithm 1: The structure of genetic algorithm

1: Set the generation counter $t := 0$
2: Generate an initial population p^0 randomly.
3: Evaluate the fitness function of all individuals in p^0.
4: **repeat**
5: Set $t = t + 1$. {Generation counter increasing}.
6: Select an intermediate population p^t from p^{t-1}. {Selection operator}.
7: Associate a random number r from (0,1) with each row in p^t.
8: **if** $r < p_c$ **then**
9: Apply crossover operator to all selected pairs of p^t. {Crossover operator}.
10: Update p^t.
11: **end if**
12: Associate a random number r_1 from (0,1) with each gene in each individual in p^t.
13: **if** $r_1 < p_m$ **then**
14: Mutate the gene by generating a new random value for the selected gene with its domain. {Mutation operator}.
15: Update p^t.
16: **end if**
17: Evaluate the fitness function of all individuals in p^t.
18: **until** Termination criteria are satisfied.

Procedure 1 $(Crossover\ (p^1, p^2))$.

1. Randomly choose $\lambda \epsilon (0,1)$.
2. Two offspring $c^1 = (c_1^1, \dots, c_D^1)$ and $c^2 = (c_1^2, \dots, c_D^2)$ are generated from parents $p^1 = (p_1^1, \dots, p_D^1)$ and $p^2 = (p_1^2, \dots, p_D^2)$ where
$c_i^1 = \lambda p_i^1 + (1-\lambda) p_i^2$,
$c_i^2 = \lambda p_i^2 + (1-\lambda) p_i^1$,
$i = 1, \dots, D$.
3. Return.

Algorithm 2. Proposed algorithm.

1: Set the initial values of the population size P, acceleration constant c_1 and c_2, crossover probability P_c, mutation probability P_m, partition number $Part_{no}$, number of variables in each partition v, number of solutions in each partition η and the maximum number of iterations Max_{itr}.
2: Set $t := 0$. {Counter initialization}.
3: **for** $(i = 1 : i \leq P)$ **do**
4: Generate an initial population$X_i(\vec{t})$ randomly.
5: Evaluate the fitness function of each search agent (solution) $f(\overrightarrow{X_i})$
6: **end for**
7: **repeat**
8: Apply the standard particle swarm optimization (PSO) algorithm as shown in Algorithm 1 on the whole population $X(\vec{t})$.
9: Apply the selection operator of the GA on the whole population $X(\vec{t})$.
10: Partition the population $X(\vec{t})$ into $Part_{no}$sub-partitions, where each sub-partition $X'(\vec{t})$size is $v \times \eta$.
11: **for** $(i = 1 : i \leq Part_{no})$ **do**
12: Apply the arithmetical crossover as shown in Procedure 1 on each sub-partition $X'(\vec{t})$
13: **end for**
14: Apply the GA mutation operator on the whole population $X(\vec{t})$.
15: Update the solutions in the population $X(\vec{t})$.
16: Set $t = t + 1$. {Iteration counter is increasing}.
17: until $(t \geq Max_{itr})$. {Termination criteria are satisfied}.
18: Produce the best solution.

4. RESULT AND DISCUSSION

The categorization of network traffic forms the core of this research project. The methodology for categorizing network traffic is implemented in several stages, involving the pre-processing, feature extraction, and categorization of the data. KDD is the dataset that is used to test the model. The 42 attributes and target set in the KDD dataset comprise several kinds of distinct assaults. A number of criteria, including recall, accuracy, and precision, are taken into account while assessing the newly presented method. The classification scheme used for Internet of Things network traffic. To classify IoT network traffic, the SVM, KNN, Random Forest, and Logistic Regression are used. Table 1 describes the categorization algorithm's findings.

Table 1: Classification Algorithms

Model	Accuracy	Precision	Recall
SVM Classifier	75.74 Percent	81 Percent	76 Percent
Logistic Regression	72.67 Percent	80 Percent	77 Percent
KNN Classifier	70 Percent	72 Percent	76 Percent
Random Forest Classifier	75.78 Percent	76.89 Percent	75.90 Percent

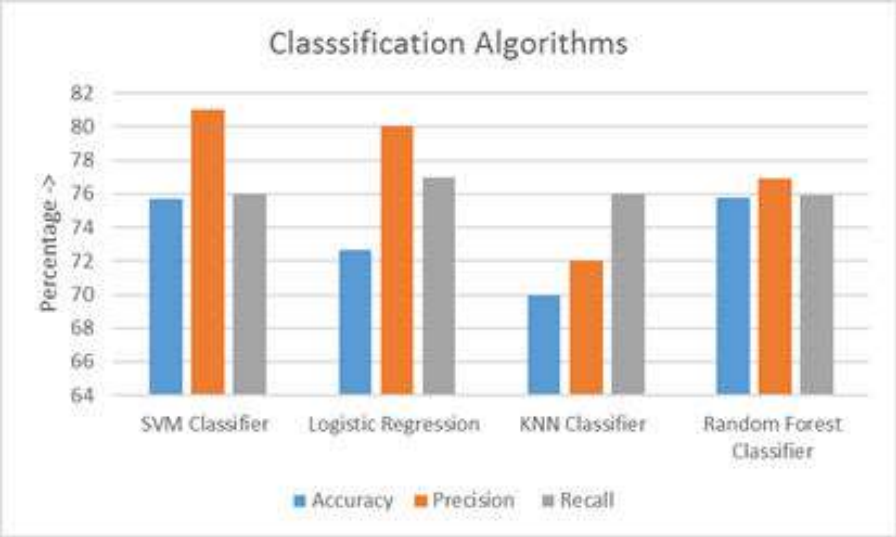

Fig 4: Classification Algorithms Results

Figure 4 illustrates the application of several classification techniques, including SVM, Random Forest, KNN, and Logistic Regression, for classifying IoT data. The random forest algorithm achieves the highest accuracy at 75.78 percent;

SVM has the highest precision value at 81 percent; and logistic regression has the highest recall at 77 percent.

Table 2: Optimization Algorithm Results

Models	Accuracy	Precision	Recall
Gray Wolf+ Random Forest	73.88 Percent	81 Percent	74 Percent
BAT +Random Forest	76.64 Percent	82 Percent	77 Percent
Firefly+ Random Forest	77.36 Percent	77 Percent	74 Percent
PSO+ Genetic+Random Forest	99.76 Percent	99 Percent	99 Percent

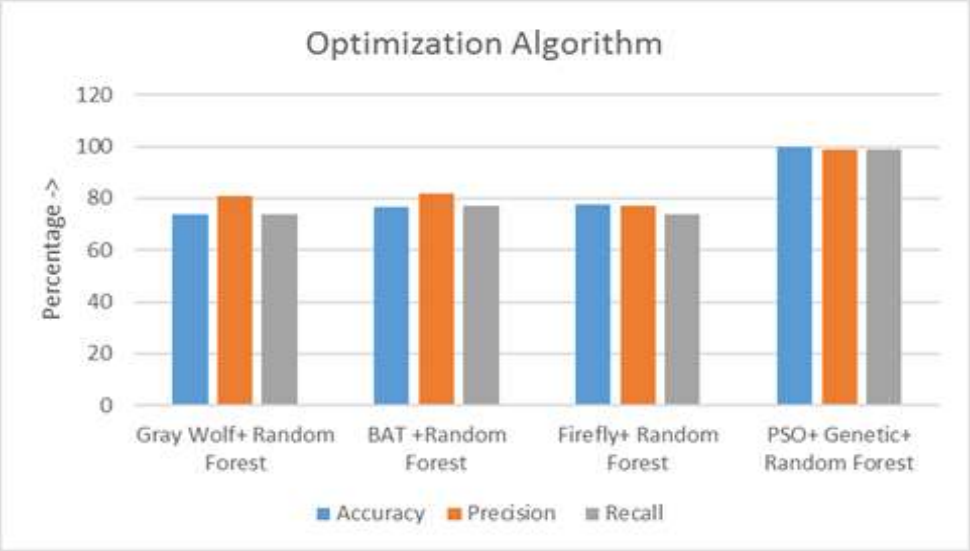

Fig 5: Optimization Outcomes Using Random Forest

The outcomes of several optimization algorithms using a random forest classifier are displayed with regard to of accuracy, precision, and recall in Figure 5. For the purpose of classifying IOT network traffic, the Gray Wolf, BAT, Firefly, and suggested hybrid optimization method are put into practice. In comparison to the firefly, BAT, and gray wolf algorithms, the suggested hybrid optimization method achieves the highest levels of accuracy, precision, and recall.

5. CONCLUSION

This article concludes that there is a high level of vulnerability of the Internet of Things network to security assaults. A variety of assaults, including DOS, reply, version number, and others, are conceivable on the network. The most sophisticated algorithms available for categorizing network threats are machine learning algorithms. Due to flaws in the feature extraction process, algorithmic methods for machine learning that have been suggested in previous years have not been able to attain good accuracy. For feature extraction, a hybrid optimization technique—a PSO and genetic algorithm combination—is suggested. For the purpose of feature extraction, several more optimization techniques, such as Firefly, BAT, and Gray Wolf, are also used. The suggested approach is put into practice using Python, and the outcomes are contrasted with those of machine learning algorithms such as SVM, BAT, and firefly. The suggested model classifies IOT network traffic with 99 percent accuracy, precision, and recall. The suggested approach yields 30 to 35 percent better results than other optimization techniques and machine learning algorithms. Models created using deep learning may be used in the future to classify IoT traffic.

REFERENCES

[1] J. R. Elias, R. Chard, J. A. Libera, I. Foster and S. Chaudhuri, "The Manufacturing Data and Machine Learning Platform: Enabling Real-time Monitoring and Control of Scientific Experiments via IoT," 2020 IEEE 6th World Forum on Internet of Things (WF-IoT), New Orleans, LA, USA, 2020, pp. 1-2

[2] S. S. Swarna Sugi and S. R. Ratna, "Investigation of Machine Learning Techniques in Intrusion Detection System for IoT Network," 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), Thoothukudi, India, 2020, pp. 1164-1167

[3] M. Afroz, N. Hasan and M. I. A. Hossain, "IoT Based Two Way Safety Enabled Intelligent Stove with Age Verification Using Machine Learning," 2021 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2021, pp. 1-5

[4] N. Karmous, M. O. -E. Aoueileyine, M. Abdelkader and N. Youssef, "IoT Real-Time Attacks Classification Framework Using Machine Learning," 2022 IEEE Ninth International Conference on Communications and Networking (ComNet), Hammamet, Tunisia, 2022, pp. 1-5

[5] H. Pandey and S. Prabha, "Smart Health Monitoring System using IOT and Machine Learning Techniques," 2020 Sixth International Conference on Bio Signals, Images, and Instrumentation (ICBSII), Chennai, India, 2020, pp. 1-4

[6] S. M. Shahriar, H. I. Peyal, M. Nahiduzzaman and M. A. H. Pramanik, "An IoT-Based Real- Time Intelligent Irrigation System using Machine Learning," 2021 13th International Conference on Information & Communication Technology and System (ICTS), Surabaya, Indonesia, 2021, pp. 277-281

[7] U. Arul, A. A. Prasath, S. Mishra and J. Shirisha, "IoT and Machine Learning Technology based Smart Shopping System," 2022 International Conference on Power, Energy, Control and Transmission Systems (ICPECTS), Chennai, India, 2022, pp. 1-3

[8] V. T. Hayashi et al., "Improving IoT Module Testability with Test-Driven Development and Machine Learning," 2021 8th International Conference on Future Internet of Things and Cloud (FiCloud), Rome, Italy, 2021, pp. 406-412

[9] S. Kavitha, V. R. Karumanchi, T. S. Rajeswari, V. C. Jadala, S. H. Raju and M. Kavitha, "Machine Learning based Authentication of loT Devices in Traffic Prediction for ITS," 2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC), Salem, India, 2022, pp. 1530-1534

[10] Z. Liu, N. Thapa, A. Shaver, K. Roy, X. Yuan and S. Khorsandroo, "Anomaly Detection on IoT Network Intrusion Using Machine Learning," 2020 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD), Durban, South Africa, 2020, pp. 1-5

[11] M. Shen, Y. Liu, L. Zhu, K. Xu, X. Du, and N. Guizani, “Optimizing Feature Selection for Efficient Encrypted Traffic Classification: A Systematic Approach,” IEEE Network, vol. 34, no. 4, pp. 20–27, Jul. 2020,

[12] S. S. Kareem, R. R. Mostafa, F. A. Hashim, and H. M. El-Bakry, “An Effective Feature Selection Model Using Hybrid Metaheuristic Algorithms for IoT Intrusion Detection,” Sensors, vol. 22, no. 4, p. 1396, Feb. 2022

[13] M. Bagaa, T. Taleb, J. B. Bernabe and A. Skarmeta, "A Machine Learning Security Framework for Iot Systems," in IEEE Access, vol. 8, pp. 114066-114077, 2020

[14] H. Lee, S. Kim, D. Baek, D. Kim and D. Hwang, "Robust IoT Malware Detection and Classification Using Opcode Category Features on Machine Learning," in IEEE Access, vol. 11, pp. 18855-18867, 2023

[15] W. Ma, X. Wang, M. Hu and Q. Zhou, "Machine Learning Empowered Trust Evaluation Method for IoT Devices," in IEEE Access, vol. 9, pp. 65066-65077, 2021

[16] T. Gaber, A. El-Ghamry and A. E. Hassanien, “Injection attack detection using machine learning for smart IoT applications”, Physical Communication, vol. 173, no. 4, pp. 5363-5365, 16 March 2022

[17] D. Mishra, B. Naik and S. Vimal, “Light gradient boosting machine with optimized hyperparameters for identification of malicious access in IoT network”, Digital Communications and Networks, vol. 9, no. 1, pp. 125-137, 12 October 2022

[18] R. Banavathu and S. Meruva, “Efficient secure data storage based on novel blockchain model over IoT-based smart computing systems”, Measurement: Sensors, 10 March 2023

[19] E. Gelenbe and M. Nakıp, "Traffic Based Sequential Learning During Botnet Attacks to Identify Compromised IoT Devices," in IEEE Access, vol. 10, pp. 126536-126549, 2022

[20] H. Kim et al., "Panop: Mimicry-Resistant ANN-Based Distributed NIDS for IoT Networks," in IEEE Access, vol. 9, pp. 111853-111864, 2021

[21] M. S. A. Muthanna, R. Alkanhel, A. Muthanna, A. Rafiq and W. A. M. Abdullah, "Towards SDN-Enabled, Intelligent Intrusion Detection System for Internet of Things (IoT)," in IEEE Access, vol. 10, pp. 22756-22768, 2022

[22] A. K. Dey, G. P. Gupta and S. P. Sahu, "Hybrid Meta-Heuristic based Feature Selection Mechanism for Cyber-Attack Detection in IoT-enabled Networks", Procedia Computer Science, vol. 218, pp. 318-327, 31 January 2023

[23] M. S. Hossain et al., "Android Ransomware Detection From Traffic Analysis Using Metaheuristic Feature Selection," in IEEE Access, vol. 10, pp. 128754-128763, 2022

[24] M. Stankovic, M. Antonijevic, N. Bacanin, M. Zivkovic, M. Tanaskovic and D. Jovanovic, "Feature Selection by Hybrid Artificial Bee Colony Algorithm for Intrusion Detection," 2022 International Conference on Edge Computing and Applications (ICECAA), Tamilnadu, India, 2022, pp. 500-505

[25] M. Tubishat, M. Alswaitti, S. Mirjalili, M. A. Al-Garadi, M. T. Alrashdan and T. A. Rana, "Dynamic Butterfly Optimization Algorithm for Feature Selection," in IEEE Access, vol. 8, pp. 194303-194314, 2020

[26] R. Al-Wajih, S. J. Abdulkadir, N. Aziz, Q. Al-Tashi and N. Talpur, "Hybrid Binary Grey Wolf With Harris Hawks Optimizer for Feature Selection," in IEEE Access, vol. 9, pp. 31662- 31677, 2021