I. J. Computer Network and Information Security, **2025, 2, 72-87**
Published Online on April 8, 2025 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijcnis.2025.02.05

An Efficient IoT Based Intrusion Detection System Using Optimization Kernel Extreme Learning Machine

Laiby Thomas*
Computer Science and Information Science, Srinivas University, Mangaluru, Karnataka -575001, India
E-mail: laibymary@gmail.com
ORCID iD: https://orcid.org/0000-0002-2608-3866
*Corresponding author

Anoop B. K.
Artificial Intelligence & Machine Learning, Srinivas Institute of Technology, Mangaluru, Karnataka -574143, India
E-mail: dranoopbk@sitmng.ac.in
ORCID iD: https://orcid.org/0000-0003-4288-5065

Received: 15 September 2023; Revised: 24 February 2024; Accepted: 28 June 2024; Published: 08 April 2025

Abstract: The Internet of Things (IoT) is an ever-expanding network that links all objects to the web so that they can communicate with one another using standardized protocols. Recently, IoT networks have been extensively used in advanced applications like smart factories, smart homes, smart grids, smart cities, etc. They can be used in conjunction with artificial intelligence (AI) and machine learning to facilitate a data collection procedure that is both simplified and more dynamic. Along with the services provided by IoT applications, various security issues are also raised. The accessing of IoT devices is mainly through an untrusted network like the Internet which makes them unprotected against a wide range of malicious attacks. The detection performance of current IDSs is hindered by issues including false alarms, low detection rate, an unbalanced dataset, and slow response time. This study proposes a new intrusion detection system (IDS) for the IoT that utilizes the chaotic improved Black Widow Optimization Kernel Extreme Learning Machine (CIBWO-KELM) algorithm to address these problems. Initially, the pre-processing of the dataset is carried out using min-max normalization, changing string values to numerical values and changing IP address to numerical values. The selection of the highest performing feature set is achieved through the information gain method (IGM), and finally, the intrusion detection is performed by the CIBWO-KELM algorithm. Python is the tool utilized for testing, while the BoT-IoT dataset is used for simulation analysis. The suggested model achieves an accuracy level of 99.7% when applied to the BoT-IoT dataset. In addition, the results of the studies demonstrate that the proposed model outperforms other current techniques.

Index Terms: Internet of Things, Machine Learning, Intrusion Detection System, Chaotic Improved Black Widow Optimization, Kernel Extreme Learning Machine.

1. Introduction

For instance, the Internet of Things (IoT) imagines a society where every aspect of daily life is interconnected. This indicates a significant change in the environment that interacts and creates new ways of dealing with the constantly available stream of information [1]. The IoT goal is not without significant obstacles, though. Small, battery-powered gadgets will rule IoT surroundings. For instance, sensors are devices that can carry out autonomous monitoring activities but will avoid carrying out computationally demanding jobs because these processes will soon drain their battery [2]. However, the massive connectivity of the IoT has a drawback, even if the right to privacy is essentially threatened. In the modern world, such a privacy concern is growing more significant [3]. However, introducing IoT technologies is more likely to worsen the issue than solve it, as most of the time, information sharing across IoT nodes happens silently in the background, going unobserved by users [4].

Several service sectors, including cattle tracking systems, home automation, medical monitoring devices, traffic monitoring systems, and many more, are increasingly relying on IoT devices. More security concerns are being posed to the environment and its resources [5,6]. Firewalls and other conventional security measures fall short of stopping modern attackers. Many security threat prevention systems are already available, but the bulk of them cannot offer protection

from fresh attacks [7,8]. Regrettably, IoT devices are still open to outside threats in the modern world [9]. Although it is common practice to identify attacks based on their recognized signatures, identifying the signature still necessitates the expert study of the attack [10,11].

Intrusion Detection Systems (IDS) should be able to spot both known and undiscovered attacks and let sensor hubs know about them to protect the Internet of Things (IoT) from many security risks [12,13]. IDS can detect suspicious or unusual behavior when there is an interruption and issue a warning [14]. IDSs for WSNs are more challenging than other frameworks since sensor hubs are frequently made of cheap, tiny, and insufficient hardware [15,16]. An actual dataset with regular profiles and attacks that recognize an attacker's signature is not present in the WSN [17,18]. Researchers employ machine learning within an IoT gateway to help protect the system framework [19,20]. This helps them deal with the problems that come with protecting IoT devices. Machine learning (ML), which is a crucial part of artificial intelligence (AI), can be used to give computers additional intelligence in terms of context, patterns, and relationships.

1.1. Motivation

Internet technology has expanded its applications in numerous sectors of our lives over the past few decades, including online auctions, online applications, banking operations, social networking, electronic commerce applications, and so on. Hackers have compromised many electronic networks due to inadequate computer system security, primarily through denial-of-service or distributed denial-of-service attacks. IDS can employ a wide variety of algorithms for attack detection. Because they particularly address similar issues, including data collection for analysis or numerical exploration processes, many of these techniques are derived from data mining techniques. Almost all well-known data mining techniques have been applied for abnormality detection in IDS, including fuzzy logic and evolutionary algorithms. The abuse and abnormality discovery procedures were combined to create the Fusion method, which was then used with the NIDS. Due to their durability, the IDS has gained popularity during the last few years. IDSs are designed to find intruders in a space. A host seeking to get unauthorized access to some other nodes can be considered an intrusion in an IoT environment. The field of IDS has a research gap. For IDS in IoT, many ML approaches are employed. Moreover, it does not adequately address problems with complexity. These methods do not offer excellent precision and are also restricted to certain strikes. These solutions can be used in the real world system by reducing the complexity difficulties with IDS.

1.2. Objectives

The key objectives of the research work are listed below.

- To design a hybrid optimized ML technique known as KELM to detect the attacks in the IoT by selecting important features using IGM.
- To perform pre-processing using various techniques such as min-max normalization, changing strings to numerical, and changing IP address to the numerical value.
- To optimize the detection model using CIBWO to achieve the best detection accuracy rate compared to other methods.
- To test the effectiveness of the proposed method by training the ML model with an effective botIoT dataset and to examine the performance using various performance metrics.

The rest of the sections in the research work are as follows: Section 2 provides the related work, and section 3 briefly explains the core of the research work, which is the proposed intrusion detection methodology in IoT. Section 4 contains the result and discussion, and in section 5 conclusion is included for the overall research work.

2. Related Works

An IDS for wormhole risks in the RPL-based IoT was suggested by SD Bhosale et al. [21]. When it comes to attacks against the 6LoWPAN adaption layer of the RPL network, wormhole attacks are among the worst of the worst. This sort of attack involves a group of malicious nodes faking a direct connection between two targets by establishing a "tunnel" between the nodes. The attack and attacker node were tracked down thanks to the received signal strength indicator (RSSI). Deka et al. [22] developed a parallel cumulative ranker technique to rank the attributes of a dataset to identify network traffic cost-effectively. Active learning was also mentioned when discussing the training of an SVM binary classifier for the unsupervised identification of DDoS attack traffic by an expert module. With the suggested strategy, large improvements in network traffic categorization accuracy can be made through the selection of small batches of training samples from a dataset. Venkatraman and B Surendaran [23] suggested timed automata controller-based hybrid adaptive IDS. The proposed Hybrid IDS utilized its additional knowledge of prevalent multimedia file types to inspect multimedia data-containing packets fully. To identify the infiltration in IoT networks, a crowdsourced online library for developing harmful pattern sets based on signatures, and a self-tuning timed automaton were built.

Patil et al. [24] designed a virtual environment monitoring system to prevent intrusions. Security is a major concern in the latest innovations of computer systems based on virtualization. In order to deliver secure internet computing services, the underlying virtual environment must be protected against the system- and network-level threats. It scanned freshly linked virtual machines for system and network model vulnerabilities. Diro et al. [25] developed a new deep

learning-based cyber security solution to aid in the identification of dangers in the social Internet of things (DL). DL's high-level feature extraction may make cyber-attack detection impervious to minute modifications or novel attacks. The distributed attack detection system's performance is compared to that of a centralized system, and the deep model's performance is compared to a traditional machine learning approach. Bany Salameh et al. [26] created a probabilistic-based channel assignment technique by studying reactive and proactive jamming attacks. This method reduced invalid CR packet emissions in a certain time frame. The proposed method used statistical data, jamming attacks, and fading circumstances from licensed primary users to supply CR IoT devices with the best secure channels and lowest invalidity ratios.

The Mobile Code-driven Trust Mechanism (MCTM) was designed by N Tariq et al. [27] as a software-defined-network (SDN) based energy-efficient solution to the problem by evaluating SN trust based on their forwarding behaviors. Findings from the research indicate that SNs can benefit greatly from state-of-the-art Software-Defined Network (SDN)-based energy-efficient management. Hasan et al. [28] claim that by comparing the efficacies of several machine learning models, it can accurately detect assaults and irregularities on IoT devices. Machine learning (ML) algorithms like decision trees, random forests, support vector machines, logistic regression, and artificial neural networks have been used here (ANN). Attacks and irregularities include denial of service, surveillance, malicious control, malicious operation, data type probing, scan, and faulty configuration.

Since WSNs require authenticity and reliability, C. Lyu et al. [29] suggested selective authentication-based geographic opportunistic routing (SelGOR) to counteract DoS attacks. In order to hasten the isolation of attackers, a distributed cooperative verification mechanism was also developed. Due to opportunistic routing, this technique also prevents SelGOR from avoiding repeated signature verification and duplication of data delivery. A safe fusion estimation was created by B Chen et al. [30] for bandwidth-constrained cyber-physical systems subject to replay assaults. Because Multisensor Information Fusion Estimation (MIFE) may boost estimate accuracy, dependability, and robustness against 8 attacks, it offers a compelling option to research secure 6 estimation problems. A new mathematical model, including a compensatory strategy, was published to characterize replay dangers and bandwidth constraints.

A novel approach to feature clustering utilizing Flow, Transmission Control Protocol (TCP) and Message Queuing Telemetry Transport (MQTT) features within the UNSW-NB15 dataset was presented by Ahmad et al. [31]. This methodology addressed the challenges like over-fitting, data-set imbalance and the curse of dimensionality. Employing supervised ML algorithms like Random Forest (RF), Support Vector Machine, and Artificial Neural Networks, the results were computed. Utilizing RF, the accuracy rates of 98.67% and 97.37% in binary and multi-class classification was attained respectively. In the context of cluster-based techniques, this approach yields classification accuracies of 96.96% and 91.4% for binary and multi-class. Table 1 presents the comparison of existing techniques with its drawbacks.

Table 1. Comparison of existing techniques

Reference	Method	Advantages	Disadvantages
[21]	IDS for wormhole attack	RSSI is high	When networks are many, the detection rate is low.
[22]	ranked dataset properties using a parallel cumulative ranker for network traffic classification	improves classification accuracy while requiring fewer training samples	The power system attack dataset is not compatible with it.
[23]	Timed automata controllers power adaptive hybrid IDS	It significantly contributes to the sustainability of the Internet and our intelligent society	Detection accuracy can be improved
[24]	created a mechanism for monitoring the virtual environment to stop incursions	From a system and network security aspect, it is suitable for the virtual environment	The detection rate can be improved
[25]	This work aims to present a new deep learning technique to cyber security that can detect threats in social IoT environments	Cyberattacks can be detected more effectively using distributed attack detection than centralized techniques.	The performance can be improved
[26]	New channel assignment algorithm based on probability theory	Delay limits are used to minimize the proportion of invalid CR packet transfers.	Different channels' CR link quality is ignored.
[27]	Energy-efficient, software-defined-network-based Mobile Code-driven Trust Mechanism (MCTM)	The identification and removing compromised SNs from SN-enabled Internet of Things applications	Routing attacks like a wormhole, Sybil, and a sinkhole for resource-constrained SNs are not considered
[28]	Several machine learning methods have been tested to accurately anticipate IoT attacks and abnormalities	High system performance	Things like Big Data and other related concerns are ignored.
[29]	Selective authentication-based geographic opportunistic routing (SelGOR)	Eliminate redundant signature verification and repeated data transfer caused by opportunistic routing	The DoS attackers' behavior model is not taken into account
[30]	Examine the network assaults that IoT gateways are susceptible to.	High performance	It does not detect a broad range of attacks
[31]	Random Forest (RF), Support Vector Machine, and Artificial Neural Networks	Disables over-fitting and dataset imbalance issues.	Higher training time is required.

Problem Statement

An intrusion detection system can find a breach at any stage, before, during, or after it occurs. Due to anomalies, the IDS is notorious for triggering false alarms. Efforts are being made to lower the current false-positive rate. Examining the data analysis procedure of intrusion detection can illuminate a problem with accurate data classification. This means that, from the perspective of anomaly-based IDS, we may drastically reduce the number of false positives by identifying and extracting the features that distinguish normal data from anomalous data. Most machine learning and data mining-based intrusion detection technologies rely on tried-and-true methods and equipment. However, the analysis of more data takes longer, which delays the discovery of assaults. If an IDS can sound the warning as soon as possible to lessen the damage a continuous attack can do, it will be more beneficial. IDS needs to be as rapid as it can be when working online. This is possible if the amount of data to be evaluated can be reduced while retaining the data. Current advancements in AI have prompted scientists to integrate distributed IDSs with other machine learning techniques. The complexity of IDSs presents a number of challenges for typical machine learning algorithms. The promise of IDSs can be realized in the real world by advancing technology to fix these faults.

3. Proposed Methodology

The IoT needs to be protected against various security risks, so an IDS is expected to find known and undetected attacks and alert sensor hubs. IDS detects suspicious or unusual activity during disruption and issues an alert. Within an IoT gateway, ML is used to solve the difficulties of protecting IoT devices while securing the system fabric. The collected attributes are examined for both normal and abnormal behavior to select the most significant traits. Initially, the dataset is pre-processed using three techniques, e.g. Min-Max normalization, changing string values to numeric values, and changing IP address to numeric values. The information acquisition approach is then used to select features (IGM). The method executes quickly and extracts the best performing feature set. A new detection algorithm called CIBWO-KELM is proposed. In order to improve the parameters of the Kernel Extreme Learning Machine, an improved Black Widow Optimization (IBWO) is presented in this chaotic study (KELM). Fig.1 shows the overall workflow of the proposed methodology.

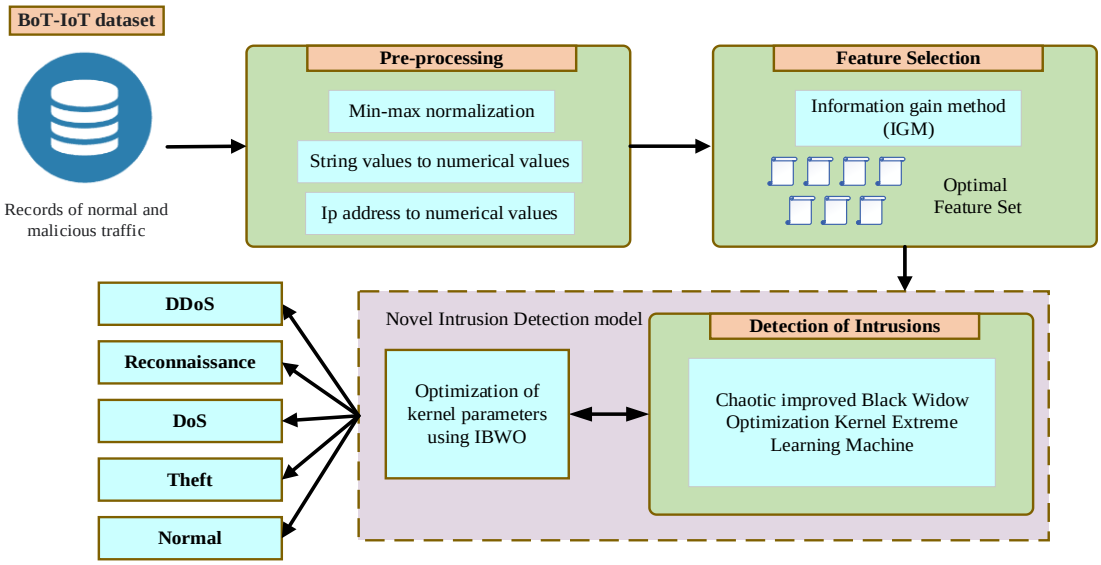

Fig.1. Overall structure of the proposed model

As IoT devices function within an interconnected and interdependent system, a continuous invasion of new threats arises. Furthermore, as the IoT devices typically operate in unattended environments, there exists a vulnerability wherein intruders can maliciously gain access to these devices. The use of eavesdropping techniques poses a risk of unauthorized access to private information transmitted over wireless networks that connect IoT devices. These security concerns are the limitation faced by IoT devices in incorporating advanced security features due to the restricted energy and processing capabilities. Consequently, there is a persistent need to implement an effective intrusion detection network to safeguard organizations utilizing IoT from potential cyber threats. In this research, the IGM approach helps in capturing highest information value and reduce the dimensionality issues. The novel CIBWO-KELM method is employed for promoting the generalization performance of features thereby reducing the overfitting and convergence issue by optimal tuning of kernel parameters.

3.1. Pre-processing

Data pre-processing is defined as the function of converting the raw data into meaningful form. Initially, the process

is achieved because of various issues in real-world data. Some of the issues are inconsistent, unstructured, redundant, and noisy. These issues can be eradicated in this particular stage. Also, the pre-processing techniques can minimize the size of the data, normalize the data, identify relationships between the data, remove outliers, and enhance the efficiency of attack detection. In this proposed work, normalization is performed. It is said to be a data transformation strategy used to convert a wide range of numerical values in a dataset to a common scale without making any alterations in the range of the values. Every attribute will attain equal weight after performing normalization. This technique speeds up model training for neural network and ML classification algorithms. Z-score, min-max, and decimal scaling are normalizing approaches. Min-max normalization normalizes the dataset. This method conserves all data relationships and does not find any potential consequences. Min-max is mathematically represented as,

$$M_m = \frac{[v - min(a) * [new\_max(a) - new\_min(a)]]}{[max(a) - min(a)]} + new\_min(a) \quad (1)$$

Here, the original value of the attribute $a$is denoted as v, the new value gained after normalization is denoted asM_m, the max and min value of the attribute is denoted as $max(a)\ and\ min(a)$. In this technique, it maps a value v of a to M_m in the range $[new\_max(a) - new\_min(a)]$. For example, using this technique, a value is mapped to the attribute between the range [0, 1]; therefore, $new\_max(a) = 1$ and $new\_min(a) = 0$. The simplified form of equation (1) is represented as,

$$M_m = \frac{v - min(a)}{max(a) - min(a)} \quad (2)$$

Furthermore, the dataset is composed of IP addresses and strings that should be converted into numerical values. Therefore, the techniques such as the conversion of IP addresses and strings to numerical values are effectively performed.

3.2. Feature Selection using IGM

In data mining, the ML classification algorithms employ a feature selection strategy in ID datasets. This is a practical method for improving IDS's functionality. The two main categories of feature selection strategies are filter and wrapper methods. In filter-based methods, they use any of the filter models to select the best optimal features based on the method known as rank search. The top features are chosen, while the lower-ranked features are removed from the feature set. This type of technique is computationally simple, more applicable on large datasets, and fast scalable. The wrapper-based strategy uses a learning algorithm to test the quality of the selected feature set. Many feature subsets are validated based on the performance of the classification, and the best feature set that attains this quality is selected as the optimal feature set. In the proposed work, ranking based technique is used known as IGM. This technique provides a classification ranking for the entire attributes, also known as features related to the class. Then a threshold value is allocated to choose various features based on the order obtained. Therefore, the feature that matches the class is considered important, and others are eliminated. GM, or mutual information maximization (MIM), is a method for reducing or eliminating correlated features in a dataset by focusing on the relationship between features and classes. It is widely used for feature selection because of its computational efficiency and simplicity.

A. Information Theory Concept

Mutual, conditional, and joint entropies constitute information theory. A random variable's entropy measures uncertainty and information needed to comprehend random variable. Consider a discrete random variable P with Udifferent number of values as$P = \{p_1, p_2, \ldots p_U\}$, another discrete random variable $Q$with Vdifferent number of values as$Q = \{q_1, q_2, \ldots q_V\}$, and Pas entropy is represented as$E(P)$:

$$E(P) = -\sum_{x=1}^{U} g(p_x)\, log\big(g(p_x)\big) \quad (3)$$

Here, $g(p_x)$is represented as,

$$g(p_x) = \frac{Number\ of\ ins\,tan\,ts\ with\ value\ p_x}{Total\ number\ of\ ins\,tan\,ts\ of P} \quad (4)$$

The definition of entropy for continuous random variables is mathematically represented as,

$$E(P) = -\int g(p)\, log\ g\ (p) dp \quad (5)$$

The quantity of uncertainty still existing in the variable Qafter its variable introduction $P$is denoted in eqn (6) by its conditional entropy Q

$$E(Q|P) = -\sum_{p_x \in P} \sum_{q_y \in Q} g\big(q_y, p_x\big)\, log\, \Big(g\big(q_y|p_x\big)\Big) \quad (6)$$

Whenever two variables are involved, the resultant uncertainty is mentioned by the joint entropy of $P$and$Q$, which is represented as $E(Q,P)$

$$E(Q,P) = -\sum_{p_x \in P} \sum_{q_y \in Q} g(q_y, p_x)\, log\left(g(q_y, p_x)\right) \tag{7}$$

Here, the joint probability of $q_y$and $p_x$is denoted as$g(q_y, p_x)$.

The mathematical link between entropy, conditional entropy, and joint entropy is given as,

$$E(Q,P) = E(P,Q) = \begin{cases} E(P) + E(Q|P) \\ E(Q) + E(P|Q) \end{cases} \tag{8}$$

Mutual information (MI) is another notion in information theory that is defined as the quantity of information released by the newly formed variable and the information issued by two variables. The MI between the variables $Q$and$P$is represented as $M(Q;P)$, and it is mathematically represented as,

$$M(Q;P) = \begin{cases} E(Q) - E(Q|P) \\ E(P) - E(P|Q) \\ E(Q) + E(P) - E(Q,P) \end{cases} \tag{9}$$

$$M(Q;P) = \sum_{p \in P} \sum_{q \in Q} g(q,p)\, log \frac{g(q,p)}{g(q)g(p)} \tag{10}$$

The MI for continuous random variables is represented as,

$$\text{To exactly identify } M(P;Q) = \int g(p,q)\, log \frac{g(p,q)}{g(p)g(q)} dpdq \tag{11}$$

the probability density functions $\left(g(p), g(q), g(p,q)\right)$ is too difficult. Hence, the continuous variables are initially discretized, and then the MI and the entropy are calculated with discrete descriptions. Conditional MI $M\left(Q; P_x|P_y\right)$ determine the new discriminative information issued by $P_x$when $P_y$is chosen, which is represented as,

$$M\left(Q; P_x|P_y\right) = \begin{cases} E\left(Q|P_y\right) - E\left(Q|P_x, P_y\right) \\ E\left(P_x|P_y\right) - E\left(P_x|Q, P_y\right) \\ M\left(P_x; Q|P_y\right) \end{cases} \tag{12}$$

Joint mutual information (JMI) identifies the information issued by $Q$and the joint variables$P_x, \ldots, P_n$, which is represented as,

$$M(Q; P_x, \ldots, P_n) = E(Q) - -E(Q|P_x, \ldots, P_n) \tag{13}$$

The quantity of information issued among three variables is said to be interaction information, which is represented as,

$$M\left(Q; P_x; P_y\right) = M(Q; P_x) + M\left(Q; P_y\right) - M\left(Q; P_x, P_y\right) \tag{14}$$

The value of $M\left(Q; P_x; P_y\right)$can be negative, zero, or positive, and it is known as the shared discriminative information of two features.

B. Proposed IGM

To attain the best optimal feature subset$F_{optimal} = max\, M\,(L;F)$, where output class is denoted as$L$and Fdenotes the feature subset. The proposed feature selection is based on MI, and the features are validated using this MI. According to equation (9), $M(L;F) = E(L) - E(L|F)$, where $E(L|F)$and$E(L)$represents the conditional entropy and entropy. $E(L)$Represents the uncertainty of the variable L , and the remaining uncertainty after the introduction $F$is denoted as$E(L|F)$. For a particular problem to get maximum$M(L;F)$, $E(L)$is considered as a constant. It just requires choosing the feature subset Fthat can make $E(L|F)$minimum. Consider Kbe the entire feature subset of the data, and each feature $k_x \in K$can be regarded as a division of Lthat classifies Linto various segments and$E(L|k_x)$is the average uncertainty of each part. Lcan be divided into more or equal segments than any single feature when $k_x$and$k_y$are introduced. Hence, the new partition of $L$are $\{k_x, k_y\}$and Lcan be classified into various segments based on two features $k_x$andk_y. Hence, $E\left(L|k_x, k_y\right)$is considered as the average uncertainty of each part.

Table 2. Example data of XOR

k_1	k_2	k_3	k_4	$L = k_2 \oplus k_3$
1	0	0	1	0
0	0	0	0	0
0	1	1	0	0
1	1	0	1	1
0	1	1	0	0
1	0	1	1	1

Assume that $F = \{k_1, k_2, .., k_{n-1}\}$is chosen, and$\hat{f} = F \cup k_x$, where the next validated feature is denoted ask_x. Instead of measuring MI between feature $k_x$and each feature in the chosen feature set $F$to attain redundant information, $F$is considered as the new division as $F_n$and calculates JMI of $\{F_n, k_x\}$directly with$L$output class and is mathematically represented as,

$$H_{IGM}(k_x) = argmax\ M\ (L; F_n, k_x) \tag{15}$$

Here$k_x \in K - F$. For example, consider the features k_1 and $k_2$have been elected and try to identify the features $k_3$and $k_4$that is to be elected next. The proposed technique will take the entire $\{k_1, k_2\}$and become the equivalent division which is represented asF_n. Then compute $M(L; F_n, k_3)$and $M(L; F_n, k_4)$and select the one which is bigger in value and k_3 is selected. Then, $\{F_n, k_3\}$is used as the new equivalent division for the next epoch. To attain equivalent divisionF_n, a unique label must be provided for the combination. For instance, the features such as $k_1$and $k_2$are selected from Table 2 then to calculate the features such as k_3 and k_4, to form an equivalent division for $\{k_1, k_2\}$. The combinations attained from $\{k_1, k_2\}$are $\{(0,0), (1,0), (1,1), (0,1)\}$and unique labels are given for each combination as 1 for (0, 0), 2 for (1, 0), 3 for (1, 1), and 4 for (0, 1). From these unique labels, the new equivalent division can be obtained as $F_n = [1,2,3,3,2,1,4,4]$. In this way, an optimal feature subset is selected to detect the type of attack. The pseudocode for IGM is shown in Table 3.

Table 3. Pseudocode for IGM

Input: Pre-processed dataset Swill entirely feature set $K = \{k_1, k_2, \ldots, k_M\}$, output class label L, and user-denoted feature number threshold T
Output: Optimal feature subset K

1. $F \leftarrow \varphi$;
2. $t \leftarrow 0$;
3. Initialize $F_n$to keep the equivalent division of the chosen features;
4. Initialize M_I_a[]array to keep the MI between the class and features;
5. **for** $i = 1$to$M$do
6. Measure the MI $M(L; k_x)$
7. Keep $M(L; k_x)$ in M_I_a[]
8. **end for**
9. Choose the feature $k_x$that has maximum value in M_I_a[]
10. $F = F \cup k_x$;
11. $K = K - k_x$;
12. $t = t + 1$;
13. $F_n = k_x$;
14. **while** $t < T$
15. Initialize CMI_a[]array to keep JMI
16. **for** each feature $k_x \in K$do
17. Measure JMI $M(L; F_n, k_i)$
18. Keep$M(L; F_n, k_i)$in CMI_a[]
19. **end for**
20. elect the feature $k_x$with maximum value in CMI_a[];
21. $F = F \cup k_x$;
22. $K = K - k_x$;
23. $t = t + 1$;
24. update $F_n$based on the previous $F_n$andk_x
25. **end while**
26. **output** K

3.3. Attack Detection Using CIBWO-KELM

A KELM-based ML model is used to detect the nature of attacks in the IoT. It is an integration of ELM and a kernel function. The kernel function used in the proposed work is named RBF. A brief description of ELM and KELM is given below.

A. Extreme Learning Machine (ELM)

One-hidden-layer feed forward neural networks are efficient, and this one is no exception. It is simple to obtain local optimal solutions while training a typical neural network, and a broad number of parameters is required. But in ELM, the total number of hidden nodes is set from scratch without adjusting the hidden layer's bias or the input layer's weight. The model's path to the best possible global response becomes more direct. Hence, the ELM model can converge faster and more effectively in terms of learning performance.

For a training dataset $D_0 = \{(u_n, t_n), n = 1, \ldots\ldots, M\}$, where the input feature vector is denoted as $u_n = [u_{n1}, \ldots u_{nm}] \in \Re^m$, and the target vector is denoted as $t_n = [t_{n1}, \ldots . t_{ng}] \in \Re^g$. The main aim is to attain an optimal model to continue with other testing tasks. The output vector is denoted as $c_n = [c_{n1}, \ldots, c_{ng}] \in \Re^g$ achieved through the network ELM. The ELM network is mathematically represented as,

$$c_n = \sum_{m=1}^{s} \alpha_m h_m(u_m) = \sum_{m=1}^{s} \alpha_m h(\beta_m \cdot u_n + k_m), n = 1, \ldots\ldots M \tag{16}$$

Here, β_m is represented as the weight vector between the input and hidden layers, the weight vector between the output and hidden layers is denoted as α_m, the m^{th} hidden node's bias is denoted as k_m, and the hidden layer's activation function is denoted as $h(\cdot)$. The parameters of the node β_m and k_m are assigned randomly, and only the quantity of hidden nodes s must be identified in the model. The error is assigned as zero in between the target t and output c, then the mathematical expression is formulated as follows,

$$\sum_{n=1}^{X} \| t_n - c_n \| = 0 \tag{17}$$

On combining equations (16) and (17), the parameters α_m, β_m and k_m are obtained. The equations can be formulated as,

$$\sum_{m=1}^{s} \alpha_m h(\beta_m \cdot u_n + k_m) = t_n, n = 1, \ldots . M \tag{18}$$

The matrix form of equation (18) can be represented as,

$$\underbrace{\begin{bmatrix} h(\beta_1 \cdot u_1 + k_1) \ldots . h(\beta_s \cdot u_1 + k_s) \\ \vdots \vdots \vdots \\ h(\beta_1 \cdot u_M + k_1) \ldots \ldots h(\beta_s \cdot u_M + k_s) \end{bmatrix}_{M \times s}}_{G_{M \times s} = [g(u_1), \ldots, g(u_M)]^A} \cdot \underbrace{\begin{bmatrix} \alpha_1^A \\ \vdots \\ \beta_s^A \end{bmatrix}}_{\alpha_{s \times n}\ s \times n} = \underbrace{\begin{bmatrix} t_1^A \\ \vdots \\ t_M^A \end{bmatrix}}_{A_{M \times n} M \times n} \tag{19}$$

It can be written as:

$$G\alpha = A \tag{20}$$

Here, the output matrix is denoted as A, α denotes the weight matrix of the hidden layer and G represents the output matrix of the hidden layer. The hidden layer's weight matrix is mathematically represented as,

$$\alpha = G^+ A \tag{21}$$

The Moore-Penrose generalized inverse of the matrix G is denoted as G^+, and it is mathematically represented as,

$$G^+ = G^A (G G^A)^{-1} \tag{22}$$

B. Proposed KELM

The accuracy of detection by the traditional model becomes low when unknown testing datasets are provided. Hence, a kernel parameter K/L was applied to enhance the generalization capacity of the model. This combination is termed as KELM and its output function [32] is mathematically represented as follows.

$$O(u) = g(u)\alpha \; which \; implies \; O(u) = g(u) G^A \left(\frac{K}{L} + G G^A \right)^{-1} A \tag{23}$$

From the above equation, $\alpha = G^A (K/L + G G^A)^{-1} A$ that are derived using equations (21) and (22). Here, the identity

matrix is denoted as K, and the regularization parameter is denoted as$L$. The kernel function in the KELM model is mathematically represented as,

$$\Psi_{KELM} = GG^{A}, \Psi_{KELM_{m,n}} = g(u_m)g(u_n) = S(u_m, u_n) \tag{24}$$

The KELM model function is mathematically formulated as,

$$O(u) = \begin{bmatrix} S(u,u_1) \\ : \\ S(u,u_M) \end{bmatrix} \left(\frac{K}{L} + \Psi_{KELM}\right)^{-1} A \tag{25}$$

Choosing the best kernel function can enhance the KELM model. There are various kernel functions. Among all the kernel functions, the RBF kernel function performs well with the KELM to detect attacks in IoT.

RBF kernel function: The suggested KELM model has a high generalization and great learning capacity due to the use of the ELM and the conventional local kernel function. To mathematically describe the RBF kernel function mathematically as

$$R_{bf}(u, u_m) = exp\left(-\frac{\|u-u_m\|^2}{2\omega^2}\right) = exp\left(-\frac{\|u-u_m\|^2}{b}\right) \tag{26}$$

Here, the exponent parameter of the selected kernel function is expressed as$b = 2\omega^2$.

The important parameters of KELM that are optimized are the kernel parameters $\varphi$and regularization parameters$L$

3.4. Kernel Parameter Optimization using CIBWO

The black widow is a medium-sized spider belonging to the Orygiidae family, mainly available along the Mediterranean Sea in European countries. The female spider spends their entire time in their webs feeding, hatching offspring, and mating. A female spider releases a pheromone in the web that attracts the male spider for mating. The female spider will utilize the male spider in the post-mating and perform hatching by transferring the eggs into the sock. After the hatching process, sibling cannibalism is engaged by the offspring. In this stage, the strong widow will eat the weak, and they even utilize their mother when they are on the web for a short period.

A. Mathematical Expression of BWO

The traditional BWO algorithm undergoes four stages: population initialization, cannibalism, procreation, and mutation. The four stages of the BWO algorithm are described below.

a. Initialization

The spider population is composed of Mwidows $R_1, R_2, \ldots, R_M$is denoted as$P_{M\times d} = [R_1, R_2, \ldots, R_M]$. Where, the dimension of the optimization problem is denoted as d. The k^{th} widow in the population is denoted as $R_k = [r_{k,1}, r_{k,2}, .., r_{k,d}](1 \le k \le M)$. Hence, the initial population is mathematically represented as,

$$\text{Here, } g\text{and}f\text{represent}r_{k,b} = f_b + rand(0,1) \cdot (g_b f_b), 1 \le b \le d \tag{27}$$

s the lower and upper bounds of the variables in the optimization design.

b. Procreate

The production of a new generation is obtained only by the mating behavior of widows. When the mating process starts, a group of spiders, as father and mother, are chosen randomly from the population depending on the procreating rate. The production of offspring is mathematically represented as,

$$\begin{cases} F_x = \beta M_x + (1-\beta)M_y \\ F_y = \beta M_y + (1-\beta)M_x \end{cases} \tag{28}$$

The mother spiders are denoted as $M_x$andM_y, father spiders are denoted as $F_x$andF_y, and $\beta$is the random number.

c. Cannibalism

There are three types of cannibalism: sibling cannibalism, sexual cannibalism, and cannibalism between mother and offspring. In this stage, the weak spiders are eliminated, and the excellent spiders are conserved.

d. Mutation

Using the mutation rate, the quantity of population to mutate is identified. From a chosen individual $R_x(1 \le x \le M)$,

two elements are randomly selected and then exchanged from the array$r_{x,n}$and$r_{x,m}(1 \leq n, m \leq d)$.

B. Mathematical expression of CIBWO

The population of the widow spiders are initialized using the chaos map, which is termed CIBWO. The logistic map is mathematically represented as,

$$R_{m+1} = \delta \times R_m \times (-R_m)\delta \in [0,4], R_m \in (0,1) \tag{29}$$

Here, the degree of chaos is controlled using the parameterδ, and when the parameter possesses a value of 4, it is in a completely chaotic state. Chaotic sequences are widely used for various practical problems, and these sequences effectively optimize the initialization process of an optimization algorithm. The detection accuracy is said to be the fitness function and is mathematically represented as,

$$F_f = Max(avg_{AC}) \tag{30}$$

Here, avg_{AC}denotes the average test detection accuracy which is given as $\sum_{j=1}^{f} testAC_j \,/f$ achieved by the classifier known as KELM. Here, the number of iterations $f$is set to be 100 and the iteration which yields maximum testing accuracy is replaced as the best outcome. The iteration process promoting better detection performance on optimizing the kernel parameters in case of test accuracy performance outcome is considered. The pseudocode for the proposed CIBWO is shown in Table 4.

Table 4. Pseudocode for CIBWO

Start Input: training set, interval of *L*and φ Output: Optimal value of *L*and φ
1. Initialize size of population p, sub-population attained procreating, cannibalism, and mutation P_{cm}
2. **for** $n = 1, \ldots a$**do**
3. **for** $m = 1, \ldots, b$**do**
4. Compute the fitness value $F_f$using equation (30)
5. Determine the worst and best solutions
6. Update the solutions based on equation (29)
7. Obtain a new fitness value F_f'
8. **if** $F_f' < F_f$ **then**
9. Consider the new solution $F_f = F_f'$
10. **else**
11. eliminate the new solution
12. **end if**
13. **if** the termination criteria is reached then
14. Update the result in the m^{th}population
15. **else**
16. return to step (4)
17. end if
18. end for
19. Determine the best optimal solutions F_{f_m}among p_{cm}sub-populations
20. **end for**
21. Identify the best optimal solutions $F_{f_n}$from $p$population and particular *L*and φ
22. **return** best *L*and best φ

4. Results and Discussion

The efficiency and superiority of the suggested model's performance over the state-of-the-art methodologies are demonstrated through a number of different evaluation methods. Here, the simulation analysis and the derived results are discussed. Classification and Regression Trees (CART), Naive Bayes (NB), random forest (RF), and Support vector machine (SVM) are all applied to the same dataset, and their performances are compared in this analysis.

4.1. Simulation Scenario

Python is used to implement the work, and simulation results are obtained. The BoT-IoT dataset is used to show that the suggested model works. When botnet attacks on IoT networks arise, the Bot-IoT is the best publicly available dataset

to study the resulting traffic. The cyber range lab at UNSW Canberra established an authentic network setting to generate the dataset. In the network environment, both regular and botnet traffic coexist. A total of over 72 million records of benign and malicious traffic have been recorded for this dataset. The dataset includes information on several types of attacks, including denial of service (DoS), data exfiltration, key logging, operating system (OS) and service scan, and distributed denial of service (DDoS). There were 4,55,786 instances of DoS (0) in the IoT Botnet training set and 71,971 instances of DoS (0) in the IoT Botnet testing set. The BoT-IoT dataset contains 74 .csv files, each consisting of 1 million records combining botnet and normal traffic.

4.2. Performance Metrics

Many indicators are used to assess the suggested model's effectiveness. Accuracy, kappa coefficient, recall, false positive rate (FPR), and precision are among the parameters considered. Here are the corresponding mathematical expressions:

Accuracy: It is defined as the fraction of correctly classified positive or negative cases. The following is a formulation of accuracy:

$$\text{Accuracy} = \frac{\text{TP}+\text{TN}}{X} \tag{31}$$

In this case, the total number of input samples is denoted as$X$. True positive (TP) indicates cases in which the model correctly identified as intrusions, while True negative (TN) indicates instances that correctly identified as normal.

Precision: Precision is the model's ability to correctly identify the target instances. It shows the chances of how much the positively labelled sample remains accurate. Accuracy can be expressed mathematically in the following way:

$$\text{Precision} = \frac{\text{TP}}{\text{TP}+\text{FP}} \tag{32}$$

Where False positive (FP) denotes the non-intrusion instances that are wrongly labelled as intrusions.

Recall: It is the proportion of correct positive detections to the actual abnormal samples. The mathematical formulation can be given as:

$$\text{recall} = \frac{\text{TP}}{\text{TP}+\text{FN}} \tag{33}$$

False negative (FN) denotes the intrusion instances wrongly labelled as non-intrusions.

F1-score: The harmonic mean of precision and recall is denoted as F1-score. The mathematical formulation of the F1-score can be given as follows:

$$\text{F1-score} = \frac{2(\text{recall}*\text{precision})}{\text{recall}+\text{precision}} \tag{34}$$

False positive rate (FPR): The percentage of incorrect classification of normal samples as intrusions by the model are termed FPR. The mathematical formulation can be given as

$$\text{FPR} = \frac{\text{FP}}{X_{\text{Normal}}} \tag{35}$$

Where X_{Normal} denotes the overall actual normal samples in$X$.

Kappa coefficient: It demonstrates the model's superiority over other models as a statistical metric. The mathematical expression is as follows:

$$K = \frac{F_o - F_a}{1 - F_a} \tag{36}$$

Where $F_o$denotes the proportion of observed agreement among existing classes of behavior and predicted class. F_a denotes the proportion of agreement expected by chance.

4.3. Performance Analysis

The suggested model's efficacy is measured against BoT-IoT dataset records of both benign and malicious traffic.

Fig.2 compares the proposed CIBWO-KELM model's accuracy to existing models. As can be seen in the figure, the proposed model achieved excellent classification accuracy. When unknown testing datasets are provided, the accuracy of detection by the traditional models becomes low. In order to improve the model's ability to generalize, the KELM classification model is presented and then optimize the KELM model's kernel parameters with CIBWO to ensure the highest possible detection accuracy. As a result of the model's ability to detect the incursions, it was found to have greater accuracy than existing models. An impressive 99.7% precision is achieved by the proposed model. Of the compared models, CART provided the highest accuracy at 92.2%, whereas Naive Bayes produced an extremely low accuracy at 87.1%.

Fig.3 displays a comparison of accuracy of the suggested model with alternative methods. The proposed model outperformed the competition in terms of precision, as seen in the figure. Since the existing models obtained very few positive predictions as true intrusions, the precision remained low. The proposed CIBWO-KELM model obtained a precision of 99.79%, and the Naïve Bayes obtained the least precision of 85.71%.

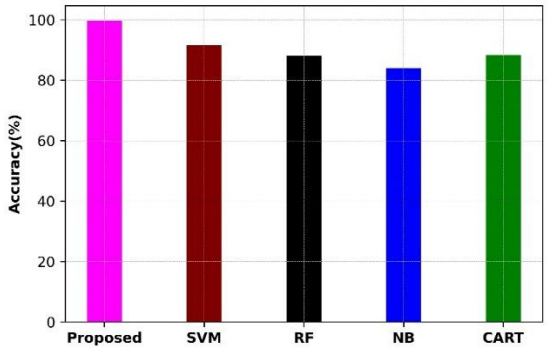

Fig.2. Accuracy comparison

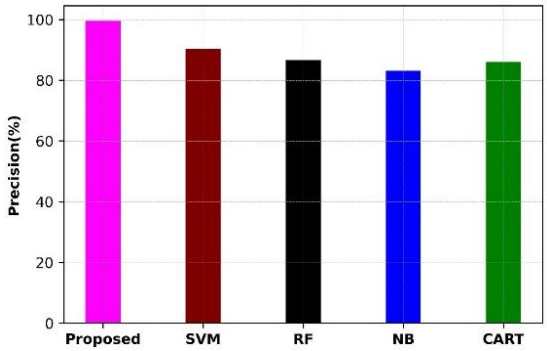

Fig.3. Precision comparison

Fig.4 shows a comparison of the proposed model's recall. The proposed CIBWO-KELM model attained a very high precision than the traditional techniques. The compared models obtained a low recall score due to a high number of false negatives (i.e.) a large number of intrusions are wrongly predicted as non-intrusions. The proposed model obtained an overall recall score of 99.75%, and the Naïve Bayes obtained the least recall score of 86.83%.

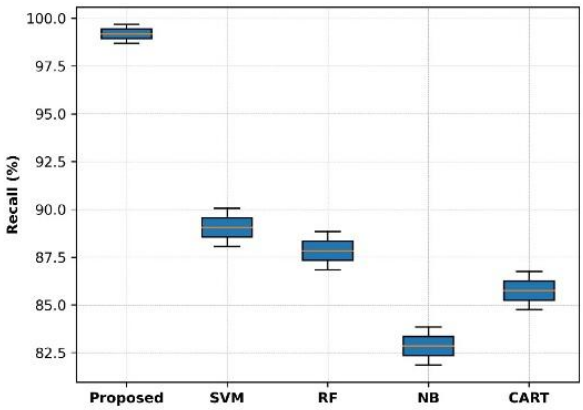

Fig.4. Comparison of recall with existing classifiers

The F1-score comparison of the suggested model is displayed in Fig.5. The proposed CIBWO-KELM model attained a very high F1 score compared to existing classifiers. The comparing models' low precision or recall value correspondingly results in a low F1-score. The proposed model obtained an overall F1-score of 99.77%, and the Naïve Bayes model obtained the least F1-score of 86.25%.

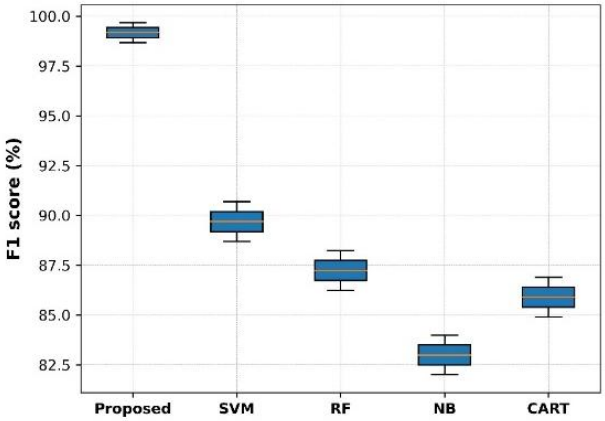

Fig.5. F1-score comparison

Fig.6 displays the results of comparing the proposed and existing classifiers' kappa coefficients. In terms of kappa, the proposed CIBWO-KELM model performed better than the other models. The proposed model attained an overall kappa coefficient of 99.70%, and the Naïve Bayes model obtained the least Kappa value of 82.71%.

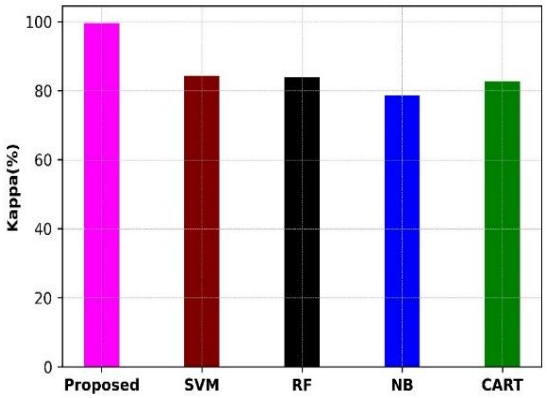

Fig.6. Kappa coefficient

Table 5 compares the proposed model's performance with some existing models, including SVM, RF, NB, and CART. Compared to conventional approaches, the new model achieved superior performance matrix values.

Table 5. Performance comparison using the Bot-IoT dataset

Methods	Accuracy	Precision	Recall	F1 score	kappa
CIBWO-KELM	99.7%	99.7%	99.7%	99.7%	99.7%
SVM	90.8%	88.8%	88.8%	88.8%	83.04%
RF	91.1%	89.4%	90.4%	89.9%	87.5%
NB	87.1%	85.7%	86.8%	86.2%	82.7%
CART	92.2%	92.3%	91.3%	91.8%	89.6%

Confusion matrix analysis: As shown in Fig.7, the suggested model's confusion matrix on the Bot-IoT dataset is rather low. In the total of 1449 DDoS labels, 1443 samples are correctly predicted as DDoS, and the remaining 6 labels are wrongly classified as DoS and Theft. But two of DoS's 2045 labels were properly categorized as regular labels. From 543 normal labels, the proposed model classifies 542 labels correctly, and the remaining 1 is incorrectly classified as DoS. The simulation results demonstrate that the proposed strategy produces greater classification accuracy.

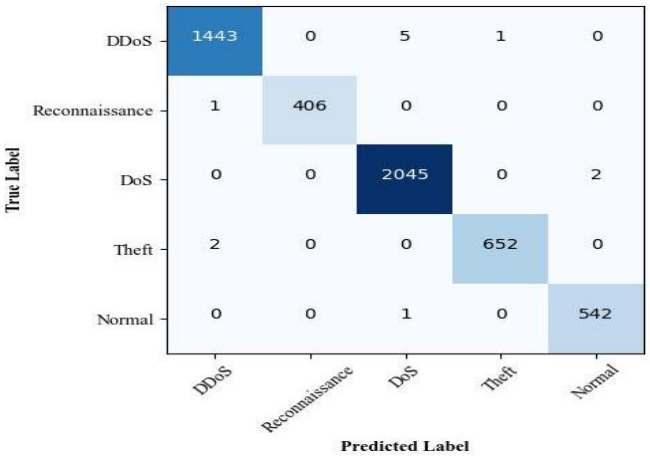

Fig.7. Confusion matrix of proposed CIBWO-KELM model using Bot-IoT dataset

The ROC curve compares the proportion of correct classifications to those that are incorrect at different cutoffs. By decreasing the positive classification threshold, both the count of false and true positives are increased. For intrusion detection on the Bot-IoT dataset, the suggested model achieves an AUC of 99.7%. The true positive rate of the trained CIBWO-KELM model is illustrated in Fig.8.

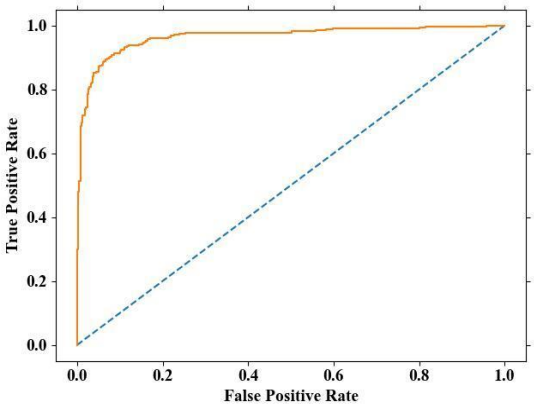

Fig.8. True positive rate VS False positive rate

4.4. Discussion

The utilization of IGM in intrusion detection assists in investigating the most pertinent features by assessing the contribution to the overall information content. This ensures that the chosen features are better composed to capture significant patterns pertaining to security threats. By concentrating on features that offer the highest information value, the IGM aids in diminishing the dimensionality of the feature space. This proves indispensable for intrusion detection systems, restructuring the analysis, advancing processing, and developing more efficient model training. The outcome of this process is an efficient set of informative features, leading to quicker model training and testing. This holds particular significance for real-time intrusion detection systems where a quick response to security threats is of greatest importance. The KELM approach proves highly suitable for intrusion detection due to its adept handling of non-linear relationships within the data. This is particularly critical, given that malicious activities in cybersecurity often manifest intricate and non-linear patterns and the linear models may fail to capture. In the context of intrusion detection, KELM exhibits a diminished risk of overfitting compared to certain other sophisticated models. This characteristic is advantageous as it ensures the model's ability to generalize effectively to unseen data and adapt to emerging types of cyber threats. By optimizing the kernel parameters using CIBWO, the KELM proficiently maps input features into a higher-dimensional space. This strategic transformation enhances the visibility of intrusion-related patterns, thereby improving the model's efficacy in capturing intricate relationships within the data. Optimization guarantees that the chosen parameters yield a model with robust generalization to unseen data, mitigating the chances of overfitting or underfitting. The process of finely adjusting kernel parameters through CIBWO contributes to heightened accuracy in predicting outcomes. It aids in the identification of kernel parameters that enhance the model's ability to generalize effectively. CIBWO is designed to efficiently seek optimal parameter values, expediting the convergence process during model training and thereby reducing the overall training time. On considering the efficiency of proposed models, effective performance outcomes are attained through the implementation of novel intrusion detection model.

5. Conclusions

In case of network security, IDS plays a crucial role and the security of smart environment applications is exposed by the significant vulnerabilities prevalent in heterogeneous networks such as IoT. This study introduces an IDS based on machine learning, designed to safeguard IoT devices from malicious intrusions. The process involves three key steps: pre-processing, feature selection, and intrusion classification. During the pre-processing phase, various tasks are executed to enhance attack detection efficiency. These include minimizing data size, normalizing data, identifying interrelations, erasing outliers, and improving overall data quality by eliminating inconsistent, redundant, unstructured, and noisy aspects from real-world data. Feature selection is carried out using IGM, which involves categorizing and ranking a class's characteristics. The detection of intrusion types in IoT relies on a machine learning model based on KELM that enhances the generalization and feature learning performance. The model's parameters are optimized using a meta-heuristic approach known as CIBWO algorithm that helps to diminish the overfitting issues and enhances the detection accuracy. The proposed model demonstrates impressive performance metrics, achieving an accuracy of 99.76%, precision of 99.79%, recall of 99.75%, F1-score of 99.77%, and kappa of 99.70%. In future, there is a potential for developing real-time anomaly prediction systems to prevent attacks before they occur. Additionally, further exploration is needed in IDS for smart vehicles. Also, the efficiency of the proposed model will be evaluated against various benchmark datasets.

References

[1] M.A. Umer, K.N. Junejo, M.T. Jilani and A.P. Mathur, "Machine learning for intrusion detection in industrial control systems: Applications, challenges, and recommendations." International Journal of Critical Infrastructure Protection vol. 38, pp. 100516, 2022.

[2] A. Chowdhury, G. Karmakar, J. Kamruzzaman, “The co-evolution of cloud and IoT applications: Recent and future trends,” In Handbook of Research on the IoT, Cloud Computing, and Wireless Network Optimization. IGI Global pp. 213-234, 2019.

[3] A. Raghuvanshi, U.K. Singh, G.S. Sajja, H. Pallathadka, E. Asenso, M. Kamal, A. Singh and K. Phasinam, "Intrusion detection using machine learning for risk mitigation in IoT-enabled smart irrigation in smart farming." Journal of Food Quality vol. 2022, pp. 1-8, 2022.

[4] A. Khraisat, I. Gondal, P. Vamplew, J. Kamruzzaman, “Survey of intrusion detection systems: techniques, datasets and challenges,” Cybersecurity. vol. 2, no.1, pp. 1-22, 2019.

[5] K. Albulayhi, Q.A. Al-Haija, S.A. Alsuhibany, A.A. Jillepalli, M. Ashrafuzzaman, and F.T. Sheldon, "IoT intrusion detection using machine learning with a novel high performing feature selection method." Applied Sciences vol. 12, no. 10, pp. 5015, 2022.

[6] B.B. Zarpelão, R.S. Miani, C.T. Kawakani, S.C. de Alvarenga, “A survey of intrusion detection in Internet of Things,” Journal of Network and Computer Applications. vol. 84, pp. 25-37, 2017.

[7] I.D. Mienye, Y. Sun, Z. Wang, “Prediction performance of improved decision tree-based algorithms: a review,” Procedia Manufacturing. vol. 35, pp. 698-703, 2019.

[8] Y. Lu, L. Da Xu, “Internet of Things (IoT) cybersecurity research: A review of current research topics,” IEEE Internet of Things Journal. vol. 6, no. 2, pp. 2103-2115, 2018.

[9] N. Moustafa, B. Turnbull, K.K.R. Choo, “An ensemble intrusion detection technique based on proposed statistical flow features for protecting network traffic of Internet of things,” IEEE Internet of Things Journal. vol. 6, no. 3, pp. 4815-4830, 2018.

[10] M. Asif, S. Abbas, M.A. Khan, A. Fatima, M.A. Khan, and S.-W. Lee, "MapReduce based intelligent model for intrusion detection using machine learning technique." Journal of King Saud University-Computer and Information Sciences 2021.

[11] W.L. Al-Yaseen, Z.A. Othman, M.Z.A. Nazri, “Multi-level hybrid support vector machine and extreme learning machine based on modified K-means for intrusion detection system,” Expert Systems with Applications. vol. 67, pp. 296-303, 2017.

[12] K.-A. Tait, J.S. Khan, F. Alqahtani, A.A. Shah, F.A. Khan, M. Ur Rehman, W. Boulila, and J. Ahmad, "Intrusion detection using machine learning techniques: an experimental comparison." In 2021 International Congress of Advanced Technology and Engineering (ICOTEN), IEEE, pp. 1-10, 2021.

[13] G.S. Sajja, M. Mustafa, R. Ponnusamy and S. Abdufattokhov, "Machine learning algorithms in intrusion detection and classification." Annals of the Romanian Society for Cell Biology vol. 25, no. 6, pp. 12211-12219, 2021.

[14] H. Bostani, M. Sheikhan, “Hybrid of anomaly-based and specification-based IDS for Internet of Things using unsupervised OPF based on MapReduce approach,” Computer Communications. vol. 98, pp. 52-71, 2017.

[15] L. Santos, C. Rabadao, R. Gonçalves, “Intrusion detection systems in Internet of Things: A literature review,” In 2018 13th Iberian Conference on Information Systems and Technologies (CISTI), pp. 1-7, 2018. IEEE.

[16] N. Islam, F. Farhin, I. Sultana, M.S. Kaiser, M.S. Rahman, M. Mahmud, A.S.M. SanwarHosen, and G.H. Cho, "Towards Machine Learning Based Intrusion Detection in IoT Networks." Computers, Materials & Continua vol. 69, no. 2, 2021.

[17] C. Gomez, A. Arcia-Moret, J. Crowcroft, “TCP in the Internet of Things: from ostracism to prominence,” IEEE Internet Computing. vol. 22, no. 1, pp. 29-41, 2018.

[18] D. Midi, A. Rullo, A. Mudgerikar, E. Bertino, “Kalis—A system for knowledge-driven adaptable intrusion detection for the Internet of Things,” In 2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS). pp. 656-666, 2017. IEEE.

[19] Q. Shafi, A. Basit, S. Qaisar, A. Koay, I. Welch, “Fog-assisted SDN controlled framework for enduring anomaly detection in an IoT network,” IEEE Access, vol. 6, pp. 73713-73723, 2018.

[20] S. Prabavathy, K. Sundarakantham, S.M. Shalinie, “Design of cognitive fog computing for intrusion detection in Internet of Things,” Journal of Communications and Network. vol. 20, no. 3, pp. 291-298, 2018.

[21] S. Deshmukh-Bhosale, S.S. Sonavane, “A real-time intrusion detection system for wormhole attack in the RPL based Internet of Things,” Procedia Manufacturing. vol. 32, pp. 840-847, 2019.

[22] R.K. Deka, D.K. Bhattacharyya, J.K. Kalita, “Active learning to detect DDoS attack using ranked features,” Computer Communications. vol. 145, pp. 203-222, 2019.
[23] S. Venkatraman, B. Surendiran, “Adaptive hybrid intrusion detection system for crowd sourced multimedia Internet of things systems,” Multimedia Tools and Applications. vol. 79, no. 5, pp. 3993-4010, 2020.
[24] R. Patil, C. Modi, “Designing a Virtual Environment Monitoring System to Prevent Intrusions in Future Internet of Things,” In Recent Findings in Intelligent Computing Techniques. pp. 345-351, 2019. Springer, Singapore.
[25] A.A. Diro, N. Chilamkurti, “Distributed attack detection scheme using deep learning approach for Internet of Things,” Future Generation Computer Systems, vol. 82, pp. 761-768, 2018.
[26] H.A.B. Salameh, S. Almajali, M. Ayyash, H. Elgala, “Spectrum assignment in cognitive radio networks for internet-of-things delay-sensitive applications under jamming attacks,” IEEE Internet of Things Journal. vol. 5, no. 3, pp. 1904-1913, 2018.
[27] N. Tariq, M. Asim, Z. Maamar, M.Z. Farooqi, N. Faci, T. Baker, “A Mobile Code-driven Trust Mechanism for detecting internal attacks in sensor node-powered IoT,” Journal of Parallel and Distributed Computing. vol. 134, pp. 198-206, 2019.
[28] M. Mahmudul Hasan, M. Islam, M. I. Islam Zarif, M. M. A. Hashem, “Attack and anomaly detection in IoT sensors in IoT sites using machine learning approaches,” Internet of Things. vol. 7, pp. 1-16, 2019.
[29] C. Lyu, X. Zhang, Z. Liu, C.H. Chi, “Selective authentication based geographic opportunistic routing in wireless sensor networks for Internet of Things against DoS attacks,” IEEE Access, vol. 7, pp. 31068-31082, 2019.
[30] O. Brun, Y. Yin, E. Gelenbe, Y.M. Kadioglu, Augusto-Gonzalez, J. and Ramos, M., “Deep learning with dense random neural networks for detecting attacks against IoT-connected home environments,” In International ISCIS Security Workshop, pp. 79-89, 2018. Springer, Cham.
[31] M. Ahmad, Q. Riaz, M. Zeeshan, H. Tahir, S.A. Haider, and M.S. Khan, "Intrusion detection in internet of things using supervised machine learning based on application and transport layer features using UNSW-NB15 data-set." EURASIP Journal on Wireless Communications and Networking vol. 2021, no. 1, pp. 1-23, 2021.
[32] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification." IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) vol. 42, no. 2, pp. 513-529, 2011.

Authors’ Profiles

Ms. Laiby Thomas working as an Asst. Professor in Naipunnya Institute of Management and Information Technology, Pongam, Koratty, Kerala and Research Scholar in Srinivas University, Mangalore. Her area of research is Machine Learning and she had 10 years of teaching experience various colleges.

Dr. Anoop B. K. working as a Professor in Srinivas Institute of Technology Mangaluru. He published more than 50 research articles in the field of Artificial Intelligence and Machine Learning. His Area of research is AI and ML, Computer Communication and Image Processing. He had 14 years of teaching experience in different parts of the world.

How to cite this paper: Laiby Thomas, Anoop B. K., "An Efficient IoT Based Intrusion Detection System Using Optimization Kernel Extreme Learning Machine", International Journal of Computer Network and Information Security(IJCNIS), Vol.17, No.2, pp.72-87, 2025. DOI:10.5815/ijcnis.2025.02.05