Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

A Machine Learning-Based Intelligent Framework With Two-Step Process For Leveraging Cybersecurity In Internet Of Things Use Cases

Masarath Saba [1] , Soumya Terala[2] , Dr. Imtiyaz Khan [3] , Nagendra Babu Rajaboina[4] , My Lapalli Kanthi Rekha [5] , Dr. D.Shravani[6]

[1]assistant Professor, Cse (Ai&Ml), Cvr College Of Engineering, Jntuh Hyderabad, Tg, India Masarathsaba.Be@Cvr.Ac.In
[2] Assistant Professor, Department Of Cse, Mjcet, Ou, Hyderabad, Tg, India Sowmya@Mjcollege.Ac.In
[3]professor, Department Of Cse, Shadan College Of Engineering And Technology Jntuh Hyderabad, Ts India Imtiyaz.Khan.7@Gmail.Com
[4]assistant Professor, Department Of Cse, Koneru Lakshmaiah Education Foundation Vaddeswaram Ap India Rnagendrababu@Kluniversity.In
[5]assistant Professor, Department Of It, Andhra Loyola Institute Of Engineering And Technology ,Vijayawada A.P, India Kaantirekha.Mylapalli@Gmail.Com
[6]associate Professor, Department Of Adce, Scetw, Ou, Hyderabad, Ts, India Drdasarishravani@Stanley.Edu.In

With the growing use of Internet of Things (IoT) applications, it's crucial to prioritize security, as IoT devices are limited in resources and susceptible to various attacks. Currently, global security standards for IoT architectures are not widely enforced. Traditional security approaches are not feasible due to the limited computational power and energy of IoT devices. Artificial intelligence (AI) has emerged as a technology that innovates solutions in various domains. Machine learning (ML) is widely used to solve real-world problems in AI. It is a learning-based approach that incrementally gains knowledge to enhance cybersecurity. Existing literature on IoT application security mainly focuses on attack detection and classification. In a recent paper, we proposed an ML framework with a two-step process to improve IoT application security. In the first step, we utilized ML models for attack detection. In the second step, we used the best-performing model to classify attacks by leveraging labels in the process. Our proposed algorithm, Learning-based Cyberattack Detection and Classification (LbCDC), was tested using the UNSW-NB15 dataset. The experimental results showed that our system could detect and classify cyberattacks, achieving 89.55% accuracy for attack detection and 95.97% accuracy for attack classification. This ML framework can be integrated into organizations' existing security platforms.

Keywords: Cyber security, Internet of Things, Machine Learning, Artificial Intelligence, IoT Security

1. Introduction

Internet of Things (IoT) technology has made unprecedented applications involving seamless integration of things and digital devices. In other words, bright and intelligent applications, which were only possible after the invention of IoT, are realized with IoT. However, there are specific security challenges as the technology involves heterogeneous devices, applications, and protocols that do not have global standards. In other words, IoT use cases are vulnerable to various kinds of attacks [1]. Since IoT devices are resource-constrained, the traditional security primitives are not directly suitable for IoT applications. Since IoT-integrated applications are sensor-based and produce large volumes of data (big data), it is essential to have a learning-based approachthat incrementally gains knowledge for known and unknown attacks. With the emergence of artificial intelligence (AI), learning-based approaches became significant. For instance, ML-based approaches for cybersecurity enhancement in IoT applications are used in [2], [3], and [4], to mention a few. ML models are widely used for improving security in networks and information systems, as revealed by prior works reviewed in this paper. Neural networks and advanced neural networks like deep learning models are also used to improve IoT security.

Many researchers contributed towards ML-based security to IoT applications. Farhan et al. [5] addressed IoT cyber security, proposing a deep learning method to effectively detect software piracy and malware threats. Adi et al. [6] examined challenges in IoT data processing for ML and proposed a framework for adaptive learning, fostering intelligent IoT applications. Kamran et al. [7] Explored the evolving cyberspace vulnerabilities and the inadequacy of conventional security systems. It focuses on the critical role of ML in cyber security, discussing challenges, applications, datasets, and evaluation metrics. Mohanta et al. [8] explored IoT's rapid growth, emphasizing security challenges. It discusses CIA (confidentially integrity availability) concerns and layer-wise issues and addresses them with ML, AI, and Blockchain technologies, outlining research challenges. Mohamed et al. [9] introduced the Edge-IIoT (industrial Internet of things) set, a realistic cyber security dataset for IoT and IIoT applications, enabling the evaluation of ML-based intrusion detection systems in centralized and federated learning(FL) modes. From the literature review, it is understood that the existing methodology for cyber-attack detection and classification can be improved by using attack detection results in classification along with the knowledge gained from training data.

The literature reveals challenges, including a lack of sufficient training samples, a weak pre-processing methodology, and an inability to detect new, previously unknown attacks. Motivated by these issues, this paper aims to create an ML-based intelligent framework with a two-step process for leveraging cybersecurity in IoT use cases. Our contributions to this paper are as follows.

1. We proposed an ML framework with a two-step process that leverages state of the art to improve the security of IoT applications. The novel approach of using attack detection results in training models in classification could enhance performance in attack classification.

2. We proposed an algorithm known as learning-based Cyberattack Detection and Classification (LbCDC). This algorithm helps realize our framework by facilitating attack detection and classification procedures.
3. An empirical study is made with a prototype using the UNSW-NB15 dataset which has scope for comprehensive cybersecurity research about IoT use cases. Our system achieved the highest accuracy, 89.55% for attack detection and 95.97% accuracy for attack classification.

The remainder of the paper is structured as follows: Section 2 reviews the literature on existing methods based on ML models for cyberattack detection and classification. Section 3 presents our methodology, which follows a two-step process for automatically detecting attacks and classifying them with improved efficiency. Section 4 presents observations made in our empirical study with attack detection and classification results. Section 5 discusses the significance of our approach and the performance benefits. Section 6 concludes our work and provides scope for future research.

2. Related Work

This section reviews the literature on existing methods based on ML for intrusion detection in IoT applications. Fan et al. [1] discussed the impact of IoT and MLon cyber security and cyber-physical systems (CPS)/IoT, emphasizing benefits, vulnerabilities, and malicious uses. Bagaa et al. [2] highlight IoT security challenges and introduce an AI-based framework using software-defined networking (SDN), network function virtualization (NFV),and ML for threat detection. Fatima et al. [3] proposed an overlapped spectrum sensing approach for cognitive radio networks, prioritizing the secondary source, enhancing relay assistance, and improving quality of services (QoS)requirements.Liang et al. [4] addressed IoT security challenges like attacks and proposed ML-based solutions for authentication, access control, and malware detection. Farhan et al. [5] addressed IoT cyber security, suggesting a deep learning method to detect software piracy and malware threats effectively. Adi et al. [6] examined challenges in IoT data processing for ML and proposed a framework for adaptive learning, fostering intelligent IoT applications.Kamran et al. [7] Explored the evolving cyberspace vulnerabilities and the inadequacy of conventional security systems. It focuses on the critical role of ML in cyber security, discussing challenges, applications, datasets, and evaluation metrics.Mohanta et al. [8] explored IoT's rapid growth, emphasizing security challenges. It discusses CIA concerns and layer-wise issues and addresses them with ML, AI, and Blockchain technologies, outlining research challenges.Mohamed et al. [9] introduced Edge-IIoTset, a realistic cyber security dataset for IoT and IIoT applications, enabling the evaluation of ML-based intrusion detection systems in centralized and FL modes. Elsis et al. [10] introduced a novel IoT architecture using XGBoost for online monitoring of gas-insulated switchgear GIS, effectively detecting defects and cyber-attacks.Muhammad et al. [11] tracked malicious communications as necessary for IoT security. Feature selection problems hamper ML models. A new framework, CorrAUC, addresses this, which achieves enormous accuracy.

Mahmudul et al. [12] addressed IoT security, comparing ML models logistic regression (LR), support vector machine (SVM), decision tree (DT), Random Forest (RF), and artificial neural

network (ANN) for attack prediction.RFoutperforms, but further research on real-time data is needed. Ali et al. [13] discussed the rapid growth of IoT, its security challenges, and the need for enhanced measures using ML/deep learning (DL)methods. Ahzamet al. [14] addressed IoT security vulnerabilities and proposed solutions by integrating deep learning and big data technologies for enhanced efficiency and effectiveness, backed by a comprehensive survey and thematic taxonomy. Donglianget al. [5] Tran et al. [16] presented an IoT architecture utilizing ML to monitor induction motor status, emphasizing fault detection and cyber-attack suppression. Experimental scenarios confirm its effectiveness. Laizhonget al. [17] Discussed the integration of ML in IoT applications, emphasizing recent advancements in applications like traffic profiling, security, and network management. Challenges and open issues are also addressed. Josaet al. [18] presented a brilliant IoT security architecture blending CEP and ML for real-time attack detection validated in a healthcare network. Kuzluet al. [19] Explored the surge in IoT use, the integration of AI in cyber security, and the dual role of AI in cyber-attacks. Bechoo et al. [20] tackled the drawbacks of using several security repositories and manual vulnerability assessment in IoT. It suggests a cognitive cybersecurity technique that uses machine learning to enhance the analysis and accuracy of data. This technique will be improved upon in the future, and its effects on vulnerability identification and security assessments will be assessed.

Dilara et al. [21] reviewed ML methods for cyber security intrusion detection, focusing on recent deep learning approaches. It includes analysis of benchmark datasets, aiding researchers in ML and DL for cyber security applications. Sarker et al. [22] explored MLalgorithms, emphasizing their application across real-world domains like cyber security, healthcare, and more. It outlines challenges and research directions, serving as a reference guide. Rahman et al. [23] proposed semi-distributed and distributed intrusion detection systems for resource-constrained IoT devices, enhancing responsiveness and accuracy. Experimental results demonstrate promising performance. Syeda et al. [24] highlight the expanding influence of IoT on daily life, emphasizing ML's role in addressing evolving security challenges. Ali et al. [25] discussed methods for improving IoT security against flaws and assaults using machine learning (ML) and deep learning (DL). The efficacy of techniques like graph neural networks and AdaBoost is highlighted in its evaluation of current research. Further research should concentrate on improving these methods and looking at fresh ML/DL approaches to increase accuracy. Eklaset al. [26] explored big data and ML applications in the IoT-based smart grid, emphasizing challenges, security concerns, and future research directions. Selvan et al. [27] Proceedings of ICOECA 2022, held in Bangalore, India, highlight intelligent computing applications, featuring 57 selected high-quality research works. Zacharias et al. [28] introduced a FL model for IoT network attack detection, ensuring privacy without compromising performance. Martin et al. [29] proposed integrating SCARGC, an extreme verification latency algorithm, into an IoT intrusion detection system, addressing non-stationary environments and concept drift challenges for sustainable security in IoT. The approach shows promising results in real-world IoT datasets, demonstrating its effectiveness against cyber-attacks. Future work includes exploring other extreme verification latency (EVL) methods and expanding implementations with neural networks and deep learning models for enhanced security. Muhammad et al. [30] Emphasized ML for the timely detection

of cyber threats and incident response using Mitre attacks. RF achieves the highest accuracy. Future work involves diverse log sources and real-time detection. From the literature review, it is understood that the existing methodology for cyber-attack detection and classification can be improved with the usage of attack detection results in classification along with the knowledge gained from training data.

Emanuel et al. [31] examined ML methods such as logistic regression, naïve Bayes, perceptron, and k-nearest neighbors—for stopping cyberattacks on Colombian IoT devices. Fatima et al. [32] demand for ML-based threat detection systems has been fuelled by worries about cybersecurity arising from the rise of IoT. Zhiyanet al. [33] examined ML techniques for IoT network security, with an emphasis on advanced persistent threat (APT) assaults, intrusion detection system (IDS) varieties, and detection difficulties. Yawei et al. [34] examined the possibilities of deep learning for IoT security, focusing on device identification and profiling for better defense.Amin et al. [35] highlighted the influence of IoT and ML on urban efficiency and livability while discussing the possibilities of these technologies in smart cities.

Iqbal et al. [36] highlighted the promise of ML for proactive defense while discussing how to use it for cybersecurity in the digital age. Yakub et al. [37] investigated the use of ML to identify intrusions in IoT networks, attaining excellent performance and accuracy using several methods. Anand et al. [38] created new security issues by revolutionizing communication with networked devices. To stop assaults, an IPS that links manufacturer Serial Number internet protocol (IP) addresses is suggested. Iqbal et al. [39] investigated how AI, in particular machine and deep learning, improves Internet of Things security by utilizing raw data analysis to thwart cyberattacks. It criticizes conventional security measures as insufficient and recommends that future studies concentrate on developing AI strategies to counter new threats. The research intends to guide IoT security to cybersecurity specialists. Rashid et al. [40] compared to single models, this research suggests ensemble ML techniques for assault detection. Cybersecurity threats exist for IoT-powered intelligent cities.

Alwahedi et al. [41] examined the use of machine learning to improve IoT security, emphasizing emerging trends, difficulties, and open problems. It highlights the need for more investigation and comprehensive approaches, suggesting future studies using massive language models and generative artificial intelligence to improve IoT cyber threat identification and resistance.Algethami and Alshamrani [42] combined the architectures of ANN, BLSTM, and GRU to present a hybrid deep learning model for IoHTcybersecurity. There are very few false positives and 100% accuracy. With the evolution of IoT devices and cyber threats, future research should concentrate on adaptive models and unsupervised learning to identify new risks. Gongada et al. [43] enhanced cyber threat detection in big data and process mining by the application of machine learning and new metrics. Though its accuracy is quite good, it has issues with industry generalization and model validation. To improve the resilience and applicability of the model in different real-world circumstances, future work should incorporate cross-validation and more extensive testing. MOHAMED et al. [44] investigated the use of federated deep learning for Internet of Things cybersecurity,

looking at its applicability in a variety of fields and its connection to intrusion detection systems and blockchain. When it comes to privacy and threat detection, federated learning works better than centralized techniques. Subsequent research endeavorsshould tackle detected weaknesses and investigate more advancements in federated learning methodologies. Karimy et al. [45] analyzed machine learning technologies for anomaly detection and discusses IoT security issues. It draws attention to the utilization of datasets and current models while stressing the necessity for more robust security measures. To improve IoT device security and solve highlighted difficulties, future work will concentrate on building sophisticated machine learning-based intrusion detection algorithms.

Chen et al. [46] discussed the use of machine learning techniques to identify cyber threats, particularly advanced persistent threats (APTs) in Internet of Things networks. The limitations of APTs in terms of limited datasets and rare traffic emphasize problems. To improve detection and address unresolved challenges, the study classifies intrusion detection systems and explores potential future research routes. Mazhar et al. [47] examined how to use deep learning and machine intelligence to improve IoT security in the face of changing cyber threats. It draws attention to the shortcomings of conventional techniques and investigates fresh AI-driven options for threat identification. Upcoming projects will focus on creating sophisticated models and solving problems to improve IoT security.

3. Proposed System

We proposed a system that exploits a two-step process to enhance cybersecurity in IoT use cases. Unlike using ML models for attack classification, we followed a different approach consisting of two steps leading to better classification performance. More details are provided in the subsequent sections.

3.1 Problem Definition

Provided network flows in IoT network, automatic detection of cyberattacks and classifying them to leverage state of the art is the problem considered. There are many existing systems using ML to detect cyberattacks. However, in this paper, we considered the problem of using attack detection results as labels for the classification of attack categories.

3.2 Our Framework

We proposed an ML-based framework, as presented in Figure 1, for efficient detection and classification of cyberattacks. The proposed system is designed differently than traditional supervised learning-based intrusion detection systems.An intrusion detection system such as the one discussed in [4] detects intrusions based on supervised learning. Many approaches are found similar in the literature. The novelty of the proposed system is that it makes use of attack detection results (labels) in the process of classification. Therefore, the whole procedure is done with a two-step process. UNSW-NB15 [48] dataset is used for empirical study. The dataset is subjected to pre-processing, which takes care of missing values and requires data transformation that is suitable for a learning-based approach. Rowswith missing values are discarded.The dataset is split into training and test sets with an 80:20 ratio. Training data is

used for the ML process. The ML models used in this study are known as the HistGradientBoosting (HGB) classifier [49], RF [50], Multilayer Perceptron [51], and GradientBoosting Classifier [52]. These four classifiers are trained using the training data. The ML models on training result in learned models that will persist for reuse in the future. Each model is tested with test instances with binary classification with 0 for standard and 1 for attack flows. The attack detection process results in the labeling of test instances. This is part of the first step in a two-step process. The result of the best-performing model involved in the detection process (class labels) is used in the attack classification process.

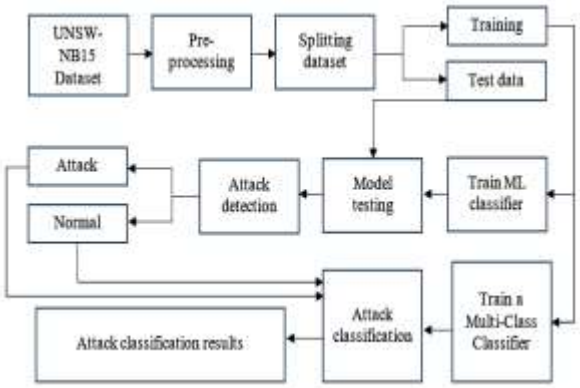

Figure 1: Proposed for attack detection and classification

Once the attack detection models are evaluated, the best-performing model is identified, and its results (attack or normal labels) are used in the multi-class classification process. The multi-class classifier is trained with training data, and the best-performing ML models produce the attack classification labels. Attack classification is a multi-class classification with nine labels (10, including the normal category). The proposed framework exploits four ML models, as mentioned earlier. Histogram-based gradient boosting (GB) is one of the ML models that use gradient-boosting trees. HGB has the potential to accelerate the training process, leading to faster execution. It is achieved by the model by exploiting histograms and data structures based on integers, unlike traditional models, which are based on GB. RF is another model used in this paper that follows an ensemble approach combining many decision trees leading to corresponding predictions over randomly sampled subsets of given training. Eventually, RF makes a final decision that will be more accurate besides reducing the risk of overfitting. The third model used in our empirical study is known as MLP, which is a kind of artificial neural network based on multiple perceptrons. It has an input layer to take input and an output layer to provide results or predictions. Between the two layers, there might be several hidden layers that process data and enable the output layer to render predictions. MLP is used as a supervised learning-based model for detecting intrusions automatically. In the training process, the model's biases and weights are adjusted to minimize error. It makes use of backpropagation to adjust biases or weights. The fourth model used in our framework is known as GB. It is made up of many weak prediction models with an ensemble approach. The prediction models are in

the form of decision trees. It is a gradient-boosting approach using a decision tree as a weak learner. Predictions from multiple learner models are used to form a final prediction model. Combining many constituent weak prediction models results in a better prediction model.

3.3 Algorithm Design

We proposed an algorithm known as LbCDC. The UNSW-NB15 dataset was used for our empirical study. Experimental results revealed that the proposed system can detect and classify cyberattacks.

Algorithm 1: Learning-based Cyberattack Detection and Classification

Algorithm: Learning-based Cyberattack Detection and Classification (LbCDC)
Inputs:
UNSW-NB15 dataset D
Machine learning pipeline P (HistGradientBoosting Classifier, Random Forest, Multilayer Perceptron, GradientBoosting Classifier)

Output:
Attack detection results R1, attack classification results R2, performance statistics R3

1. Begin
2. Initialize attack detection results map RM1
3. Initialize results vector R (to hold predictions of best performing model)
4. D'←PreProcess(D)
5. (T1, T2)←SplitDataset(D')
   Attack Detection
6. For each model m in P
7. Train m with T1
8. R1←TestModel(m, T2)
9. R3←EvauatePerformance(R1, ground_truth)
10. Update RM1 with the m and R1
11. Display R1 //attack detection results
12. Display R3 //attack detection performance statistics
13. End For
    Find Results of Best Performing Model
14. R←FindResultsOfBestModel(RM1)
    Attack Classification
15. For each model m in P
16. Train m with T1 and R //using labels of attack detection results
17. R2←TestModel(m, T2)
18. R3←EvauatePerformance(R2, ground_truth)
19. Display R2 //attack classification results
20. Display R3 //attack classification performance statistics
21. End For

22. End

As presented in Algorithm 1, it takes the UNSW-NB15 dataset and ML pipeline P HGB Classifier, RF, Multilayer Perceptron, and GB (Classifier) as inputs. It generates outputs like attack detection results, attack classification results, and performance statistics. In step 4, pre-processing is carried out in terms of dealing with missing values and required data transformation suitable for a learning-based approach. In step 5, the algorithm divides the dataset into two parts, namely the training set (T1) and the test set (T2). Step 6 through Step 13, there is an iterative process meant for attack detection. The four ML models in the pipeline are used for attack detection. Each model is trained with T1 and evaluated with T2. The attack detection results and performance statistics are observed. Then, in step 14, the results (classification labels) of the best-performing model are obtained from the map RM1, which holds key and value pairs in the form of the model and its attack prediction results. Step 15 through step 21, there is another iterative process that trains each ML model with T1 and also attack detection results (R) of the best-performing model. Finally, the algorithm provides results of attack detection, classification, and performance statistics.The labels in the detection process can help in going for multi-class classification. A non-zero label indicates a class that needs to be determined.

3.4 Dataset Details

UNSW-NB15 [31] is the dataset used in the empirical study. It is widely used in cybersecurity studies linked to IoT use cases as it has IoT network traffic data consisting of all kinds of attacks and expected flows. It has 256,673 instances in total, including 175,341 instances for training and 82,332 cases for testing. Figure 2 shows different attack flows found in the dataset.The class labels and their attack category name are provided in Table 1. best

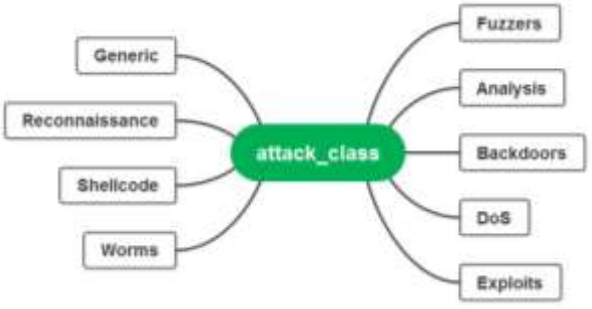

Figure 2: Different attack classes found in the UNSW-NB15 dataset

The dataset reflects an IoT network with 47 features consisting of 2.5 million data points. It has two target variables. The first variable holds 1 (attack) or 0 (standard), which is used to

determine whether there is an attack. The second variable has ten values from 0-9, reflecting different kinds of attacks, as shown in Table 1.

Table 1: Shows attack class values and corresponding attack names

Attack Class Value	Attack Name
0	Normal
1	Fuzzers
2	Analysis
3	Backdoors
4	DoS
5	Exploits
6	Generic
7	Reconnaissance
8	Shellcode
9	Worms

The attack classes are used to classify attacks. They are particularly useful when generating a confusion matrix for all classes.

3.5 Performance Evaluation Methodology

Since we used supervised learning models, the ground truth of each test instance is compared against the predicted algorithm's predictions. This will result in the identification of a number of true positives (TPs), true negatives (TNs), false positives (FPs), and false negatives (FNs). The meaning of these four cases is illustrated in Figure 3.

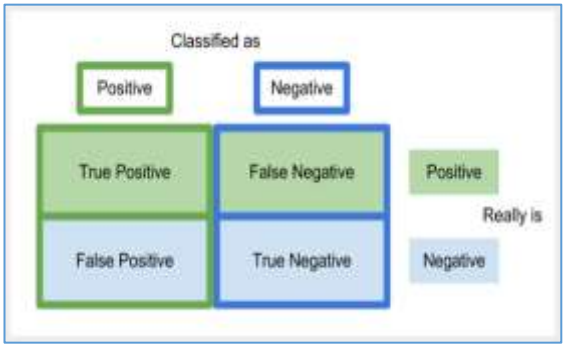

Figure 3: Confusion matrix

After observing algorithm predictions and comparing them with ground truth, four performance metrics are computed. These metrics are mathematically expressed in Equation 1, Equation 2, Equation 3, and Equation 4.

$$\text{Precision (p)} = \frac{TP}{TP+FP} \quad (1)$$

$$\text{Recall (r)} = \frac{TP}{TP+FN} \quad (2)$$

$$\text{F1-Score} = 2 * \frac{(p * r)}{(p+r)} \quad (3)$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

These metrics are widely used in ML applications to evaluate model performance. Each metric results in a value between 0 and 1. This means that these measures cannot have a value higher than 1, as 1 indicates 100% performance.

4. Experimental Results

We built a Python-based application to implement and evaluate the proposed system. The environment used for execution of the application includes a system with Windows 11 OS and an Intel Core i5-1335U processor with 16 GB RAM. UNSW-NB15 [31] dataset is used for empirical study. The results of the experiments in terms of exploratory data analysis and performance evaluation are provided in this section. The proposed system uses four ML models as HGB classifier [32], RF [33], Multilayer Perceptron [34], and GradientBoostingClassifier [35]. All are the ML models that could perform well in the attack detection process.

HGB classifier is used with a learning rate of 0.1, maximum features of 1.0, and early stopping "auto."RF model is configured with criterion "gini," minimum samples split 2, max features "sqrt." MLP model is configured with a learning rate of "constant," activation "real," initial learning rate of 0.0001, and max iterations 200. The gradientBoosting model is used with a learning rate of 0.1, several estimators of 100, a loss function of "log loss," and a max depth of 3.

Figure 4 visualizes the partial dependency of the attack detection process on different features. The results are generated using the HGB model.

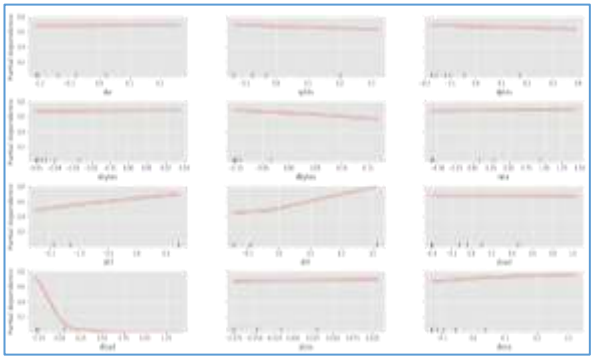

Figure 4: Dependence of attack detection (attack or average) on features of the dataset using HGB

As presented in Figure 5, partial dependency of the attack type classification process on different features is visualized. The results are generated with the model HGB for features such as ct_state_ttl and ct_dst_ltm using a one-way approach and a two-way approach.

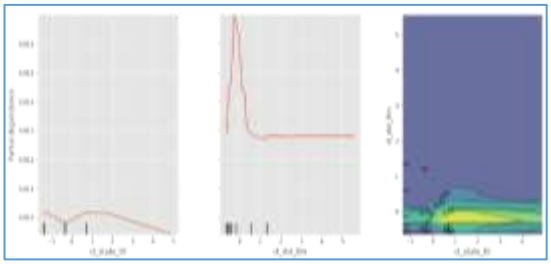

Figure 5: Dependence of attack type on features of dataset like ct_state_ttl and ct_dst_ltm using HGB

As presented in Figure 6, partial dependency of the attack type classification process on different features is visualized. The results are generated with the model HGB for features such as rate, state, and data.

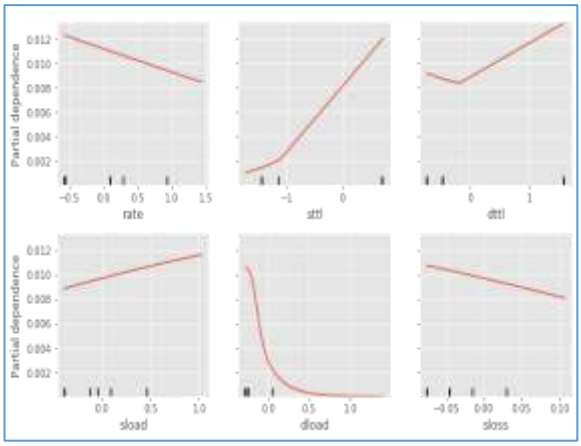

Figure 6: Dependence of attack type classification on features of the dataset using HGB

As presented in Figure 7, the partial dependency of the attack-type classification process on different features is visualized. The results are generated with the model HGB for features such as loss, sinks, dinpkt and others.

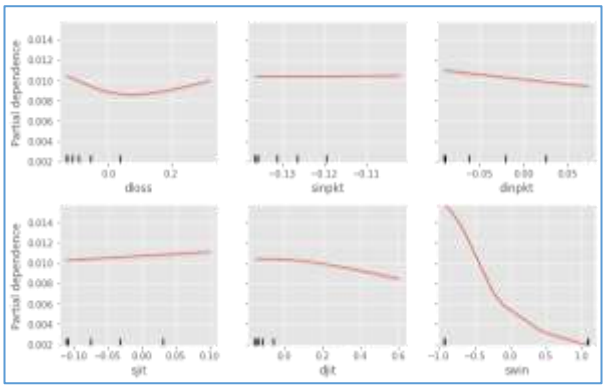

Figure 7: Dependence of attack type classification on features of the dataset (cotd…) using HGB

As presented in Figure 8, the partial dependency of the attack type classification process on different features is visualized. The results are generated with the model HGB for features such as dur, spots, dpkts and others.

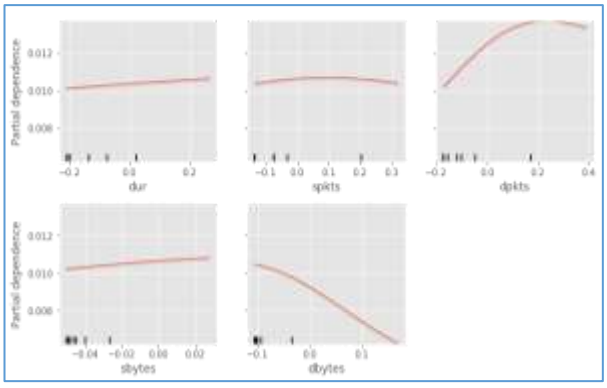

Figure 8: Dependence of attack type classification on features of the dataset (contd…) using HGB

As presented in Figure 9, attack detection results are provided. Each model is evaluated for its performance in attack detection. Four metrics are used in performance evaluation. A higher value observed for each metric indicates better capability of a model in attack detection. The GB model has the least precision, 84.89%, MLP 88.86%, RF 89.54%, and HGB 89.55% precision, which is the highest. Concerning recall, the lowest performance is exhibited by RF with 89.41%, while the highest performance is shown by GB with 91.82% recall. MLP and RF achieved 89.98% and 89.41% recall respectively. GB achieved the lowest F1-Score, with 88.22%, while HGB showed the highest F1-Score, with 90.26%. MLP and RF exhibited 89.41% and 89.47% F1-Score respectively. Concerning accuracy, GB showed least

performance with 86.24%, MLP 88.08%, RF 88.21% and HGB 89.55%. HGB achieves the highest accuracy in attack detection with 89.55%.

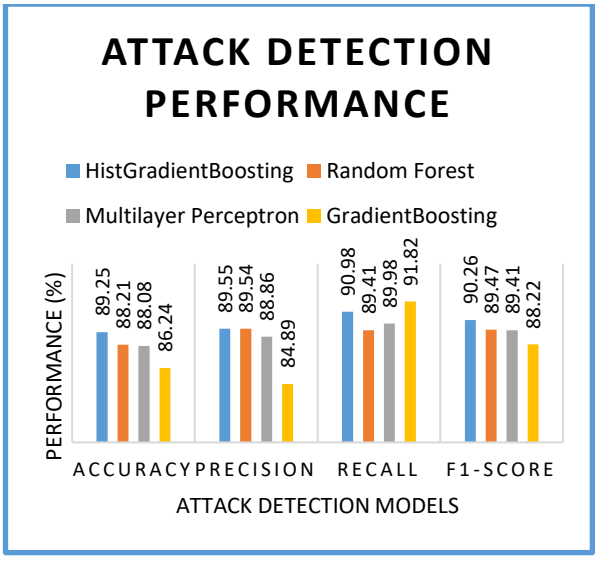

Figure 9: Attack detection performance comparison

As presented in Figure 10, attack classification results are provided. Each model is evaluated for its performance in attack classification. Four metrics are used in performance evaluation.

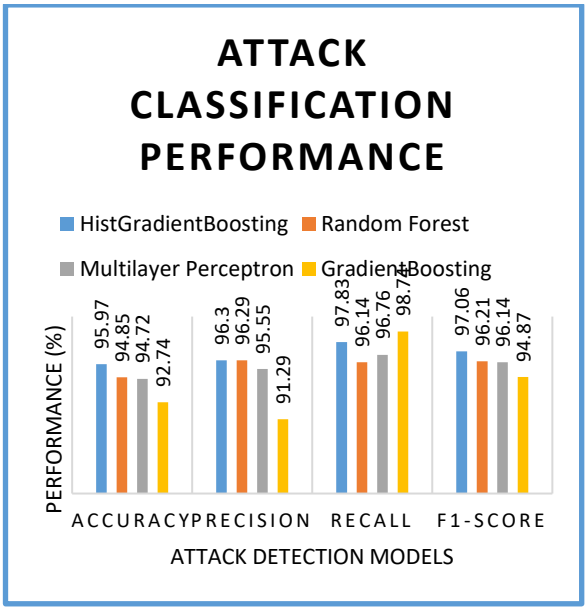

Figure 10: Attack classification performance comparison

A higher value observed for each metric indicates better capability of a model in attack classification. The lowest precision is exhibited by the GB model with 91.29%, MLP with 95.55%, RF with 96.29%, and HGB with 96.30% precision, which is the highest. Concerning recall, the lowest performance is exhibited by RF with 86.14%, while the highest performance is shown by GB with 98.74% recall. MLP and RF achieved 96.76% and 96.14% recall, respectively. GB achieved the lowest F1-Score with 94.87%, while HGB showed the highest F1-Score with 97.06%. MLP and RF exhibited 96.14% and 96.21% F1-Score respectively. Concerning accuracy, GB showed the most minor performance with 92.74%, MLP 94.72%, RF 94.85%, and HGB 95.97%. HGB achieves the highest accuracy in attack classification, with 95.97%. The rationale behind the achievement of higher accuracy by all models in attack classification when compared with attack detection is the novelty of our approach in attack classification, where each model is trained with a training set and also attack detection results of the best-performing model, which happened to be HGB. The best model is determined based on evaluation metrics. In other words, the higher accuracy in attack classification of the models is due to our two-step approach used in the proposed framework.

5. Discussion

This research aims to develop a machine learning (ML)-based intelligent framework with a two-step process to enhance cybersecurity in Internet of Things (IoT) applications. Due to various reasons, such as the absence of security standards and the limited resources of devices, IoT applications are vulnerable to a variety of attacks. Many researchers have worked on ML-based approaches for attack detection and classification in IoT systems. However, the approach proposed in this paper introduces a novel two-step process that utilizes cutting-edge methods to enhance the security of IoT applications. In the first step, ML models are used for attack detection, and in the second step, the results from the best-performing model are used for attack classification by incorporating the labels into the classification process. This approach allows the classification models to gain additional insights from the attack detection results, leading to improved performance in the attack classification process.The key findings in this paper include the utility of machine learning models in protecting network flows from intrusions and enhanced cyber security. Another important observation is that the process involved in the proposed system with two important steps could help in improving detection performance. The proposed system has set a limitation as discussed in section 5.1.

5.1 Limitations of the Study

The system proposed in this paper has some limitations. The dataset used in the empirical study has a limited number of instances, which may hinder the generalization of the findings unless diversified datasets are used. Additionally, the system is comprised solely of machine learning models, and utilizing neural networks and extended neural networks could enhance its performance further. It is also important to consider implementing novel feature selection models to improve training accuracy and attack detection performance, a step thatshould have been taken in this paper.

6. Conclusion and Future Work

In this paper, we proposed an ML-based system for leveraging security in IoT applications. The system is built with a two-step process. Unlike existing systems where a single model detects and classifies cyberattacks, the proposed system does two steps to enhance attack classification accuracy. In the first step, we use ML models for attack detection. Each model is evaluated with training and an attack detection process. This step results in the identification of the best model among the models used in the first step. Then, in the second step, the results ofthe best-performing model identified in the first step are used for attack classification by exploiting labels in the attack classification process. We proposed an algorithm known as LbCDC. The UNSW-NB15 dataset is used for our empirical study. Experimental results revealed that the proposed system can detect and classify cyberattacks. Our system achieved the highest accuracy,89.55% for attack detection and 95.97% accuracy for attack classification. The research carried out in this paper reveals that multi-class classification, which takes attack detection results as input, is able to perform better. Our work has many limitations. First, we did not tune hyperparameters using UNSW-NB15. Hyperparameter tuning could improve performance. Second, we should have explored deep learning models. Deep learning models have depth in the training process and could leverage performance further. Third, feature selection methods have yet to be explored in this paper. In the future, we improve our framework to address these limitations.

References

1. Fan L, William GH, Weixian L, Weichao G, Wei Y. Machine Learning for Security and the Internet of Things: the Good, the Bad, and the Ugly. IEEE Access. 2019003B 1–1. http://doi:10.1109/access.2019.2948912
2. Bagaa M, Taleb T, Bernabe JB, Skarmeta A. A Machine Learning Security Framework for Iot Systems. IEEE Access. 2020; 8:114066–114077. http://doi:10.1109/access.2020.2996214
3. Fatima H, Rasheed H, Syed Ali H, Ekram H. Machine Learning in IoT Security: Current Solutions and Future Challenges. IEEE Communications Surveys & Tutorials. 2020; 1–1. http://doi:10.1109/COMST.2020.2986444
4. Liang X, Xiaoyue W, Xiaozhen L, Yanyong Z, Di W. IoT Security Techniques Based on Machine Learning: How Do IoT Devices Use AI to Enhance Security? IEEE Signal Processing Magazine. 2018; 35(5):41–49. http://doi:10.1109/MSP.2018.2825478
5. Farhan U, Hamad N, Sohail J, Shehzad K, Ahsan LM, Fadi AT, Leonardo M. Cyber Security Threats detection in Internet of Things using Deep Learning approach. IEEE Access. 2019; 1–1. http://doi:10.1109/ACCESS.2019.2937347
6. Adi E, Anwar A, Baig Z, Zeadally S. Machine learning and data analytics for the IoT. Neural Computing and Applications. 2020. http://doi:10.1007/s00521-020-04874-y
7. Kamran S, Suhuai L, Vijay V, Ibrahim AH, Min X. A Survey on Machine Learning Techniques for Cyber Security in the Last Decade. IEEE Access. 2020; 8:222310–222354. http://doi:10.1109/access.2020.3041951
8. Mohanta BK, Jena D, Satapathy U, Patnaik S. Survey on IoT Security: Challenges and Solution using Machine Learning, Artificial Intelligence and Blockchain Technology. Internet of Things. 2020; 100227. http://doi:10.1016/j.iot.2020.100227
9. MOHAMED AF, OTHMANE F, DJALLEL H, LEANDROS M, HELGE J. Edge-IIoTset: A New Comprehensive Realistic Cyber Security Dataset of IoT and IIoT Applications for Centralized and Federated Learning. IEEE. 2022; 10:40281-40306. http://DOI:10.1109/ACCESS.2022.3165809

10. Elsisi M, Tran MQ, Mahmoud K, Mansour DEA, Lehtonen M, Darwish MMF. Towards Secured Online Monitoring for Digitalized GIS Against Cyber-Attacks Based on IoT and Machine Learning. IEEE Access. 2021. http://doi:10.1109/ACCESS.2021.3083499
11. Muhammad S, Zhihong T, Ali KB, Xiaojiang D, Mohsen G, CorrAUC: a Malicious Bot-IoT Traffic Detection Method in IoT Network Using Machine Learning Techniques. IEEE Internet of Things Journal. 2020; 1–1. http://doi:10.1109/JIOT.2020.3002255
12. Mahmudul H, Islam M, Ishrak I, Hashem MMA. Attack and Anomaly Detection in IoT Sensors in IoT Sites Using Machine Learning Approaches. Internet of Things. 2019; 100059–. http://doi:10.1016/j.iot.2019.100059
13. Ali AGM, Mohamed A, Abdulla AA, Xiaojiang D, Ali I, Mohsen G. A Survey of Machine and Deep Learning Methods for Internet of Things (IoT) Security. IEEE Communications Surveys & Tutorials. 2020; 1–1. http://doi:10.1109/COMST.2020.2988293
14. Ahzam MA, Ariyaluran RAH, Fariza Hanum N, Abdullah G, Ejaz A, Abdul SMN, Nazihah A, Imran M. Deep learning and big data technologies for IoT security. Computer Communications. 2020. http://doi:10.1016/j.comcom.2020.01.016
15. Dongliang C, PawelW, Zhihan Lv. Cyber Security in Smart Cities: A Review of Deep Learning-based Applications and Case Studies. Sustainable Cities and Society. 2020; 102655–. http://doi:10.1016/j.scs.2020.102655
16. Tran MQ, Elsisi M, Mahmoud K, Liu MK, Lehtonen M, Darwish MMF. Experimental Setup for Online Fault Diagnosis of Induction Machines via Promising IoT and Machine Learning: Towards Industry 4.0 Empowerment. IEEE Access. 2021. http://doi:10.1109/access.2021.3105297
17. Laizhong C, Shu Y, Fei C, Zhong M, Nan L, Jing Q. A survey on the application of machine learning for Internet of Things. International Journal of Machine Learning and Cybernetics. 2018. http://doi:10.1007/s13042-018-0834-5
18. JosÃ R, Juan BP, JosÃ LM, Guadalupe O. Integrating complex event processing and machine learning: An intelligent architecture for detecting IoT security attacks. Expert Systems with Applications. 2020; 149:113251. http://doi:10.1016/j.eswa.2020.113251
19. Kuzlu M, Fair C, Guler O. Role of Artificial Intelligence in the Internet of Things (IoT) cybersecurity. Discover the Internet of Things. 2021. http://doi:10.1007/s43926-020-00001-4
20. Bechoo L, Ravichandran S, Kavin R, Anil KN, Dibyahash B. IOT-based cyber security identification model through machine learning technique. Measurement: Sensors, 2023; 27:1-8.
21. Dilara G, Tulay Y, Angelo G, Fabio S. A Comprehensive Survey of Databases and Deep Learning Methods for Cybersecurity and Intrusion Detection Systems. IEEE Systems Journal. 2020; 1–15. http://doi:10.1109/JSYST.2020.2992966
22. Sarker IH. Machine Learning: Algorithms, Real-World Applications and Research Directions. SN Computer Science. 2021. http://doi:10.1007/s42979-021-00592-x
23. Rahman A, Asyharia AT, Leong LS, Satrya GB, Tao MH, Zolkipli MF. Scalable Machine Learning-Based Intrusion Detection System for IoT-Enabled Smart Cities. Sustainable Cities and Society. 2020; 102324–. http://doi:10.1016/j.scs.2020.102324
24. Syeda MT, Hadis K, Petros S. Machine learning based solutions for the security of Internet of Things (IoT): A survey. Journal of Network and Computer Applications. 2020; 161: 102630–. http://doi:10.1016/j.jnca.2020.102630
25. Ali G, Nasim J, Samira P, Nahide D. Securing internet of things using machine and deep learning methods: a survey. Cluster Computing, 2024; 1-25.

26. Eklas H, Imtiaj K, Fuad UN, Sarder SS, Samiul HS. Application of Big Data and Machine Learning in Smart Grid, and Associated Security Concerns: A Review. IEEE Access. 2019; 1–1. http://doi:10.1109/ACCESS.2019.2894819
27. Selvan C, Jenifer GGC, Aruna M, Sridhar S. Healthcare Application System with Cyber-Security Using Machine Learning Techniques. Springer. 2022; 127-141. https://doi.org/10.1007/978-981-19-2500-9_9
28. Zacharias A, Konstantinos P, Terpsi V, Stavroula B, Artemis V, Dimitrios S, Antonis G, Theodore Z. Enhancing Cyber Security in IoT Systems using FL-based IDS with Differential Privacy. IEEE. 2022; 1-5. http://DOI:10.1109/GIIS56506.2022.9936912
29. Martin ML, Sicong S, Salim H, Soheil S. Machine Learning for Intrusion Detection: Stream Classification Guided by Clustering for Sustainable Security in IoT. ACM. 2023; 691–696. https://doi.org/10.1145/3583781.3590271
30. Muhammad I, Hafeez RS, Ali R, Muhammad AR, Furqan R, Imran A. A performance overview of machine learning-based defence strategies for advanced persistent threats in industrial control. Elsevier. 2023; 134:1-12. https://doi.org/10.1016/j.cose.2023.103445
31. Emanuel OR, Juan RB, Juan AS, Javi. Machine Learning Techniques for Cyberattack Prevention in IoT Systems A Comparative Perspective of Cybersecurity and Cyber. MDPI. 2024; 1-24.
32. Fatima A, Alyazia A, Mohamed AF, Ammar B, Norbe. Machine learning techniques for IoT security: Current research and future vision with generative AI and large language m. Internet of Things and Cyber-Physical Systems. 2024; 4:167–185.
33. Zhiyan C, Jinxin L, Yu S, Murat S, Burak K, Hussein TM, Petar D. Machine Learning-EnabledIoTSecurity: OpenIssues and ChallengesUnderAdvancedPersistentThreats. ACM Computing Surveys. 2022; 1-35.
34. Yawei Y, Shancang L, Phil L, Fuzhong L. Deep Learning-Based Security Behaviour Analysis in IoT Environments: A Survey. Hindawi Security and Communication Networks. 2021; 1-13.
35. Amin U, Syed MA, Jianqiang L, Lubna N, Tariq M. Smart cities the role of Internet of Things and machine learning in realizing a data-centric smart environment. Complex & Intelligent Systems. 2024; 10:1607–1637.
36. Iqbal HS. Machine Learning for Intelligent Data Analysis and Automation in Cybersecurity: Current and Future Prospects. Annals of Data Science. 2023; 10(6):1473–1498.
37. Yakub KS, Aremu IA, Sanjay M, Monica K. A machine learning-based intrusion detection for detecting internet of things network attacks. Alexandria Engineering Journal. 2022; 61:9395-9409.
38. Anand K, Dharmesh D, Pankaj A, Nagender A, Pankaj D. Cyber-Internet Security Framework to Conquer Energy-Related Attacks on the Internet of Things with Machine Learning Tech. Hindawi Computational Intelligence and Neuroscience. 2022; 1-13.
39. Iqbal HS, Asif IK, Yoosef BA, Fawaz A. Internet of Things (IoT) Security Intelligence A Comprehensive Overview, Machine Learning Solutions and Research Direction. 2022; 1-18.
40. Rashid MM, Kamruzzaman J, Hassan MM, Imam T, Gordon S. Cyberattacks Detection in IoT-Based Smart City Applications Using Machine Learning Techniques. International Journal of Environmental Research and Public Health. 2020; 17(24):1–21. doi:10.3390/ijerph17249347
41. Fatima A, Alyazia A, Mohamed AF, Ammar B, Norbert. Machine learning techniques for IoT security Current research and future vision with generative AI and large language mo. Elsevier, 2024; 1-25.
42. Sarah AA, Sultan SA. A Deep Learning-Based Framework for Strengthening Cybersecurity in Internet of Health Things (IoHT) Environments. MDPI, 2024; 1-15.
43. Taviti NG, Amit A, Kathari S, Vijaya. Leveraging Machine Learning for Enhanced Cyber Attack Detection and Defence in Big Data Management and Process Mining. (IJACSA) International Journal of Advanced Computer Science and Applications, 2024; 15(2):629-638.

44. MOHAMED AF, OTHMANE F, LEANDROS M, HELGE J. Federated Deep Learning for Cyber Security in the Internet of Things: Concepts, Applications, and Experimental Analysis. IEEE Access, 2021; 9:1-34.
45. Aziz U, Chandrasekhar RP. Securing the Internet of Things: A Study on Machine LearningBased Solutions for IoT Security and Privacy Challenges. ZKG international, 2023; 8(2):1-36.
46. Zhiyan C, Jinxin L,Yu S, Murat S, Burak K, HusseinT.M, Petar D. Machine Learning-Enabled IoT Security: Open Issues and Challenges Under Advanced Persistent Threats. ACM Computing Surveys, 2022; 1-35.
47. Tehseen M, Dhani BT, Tamara AS, Yazeed YG. Analysis of IoT Security Challenges and Its Solutions Using Artificial Intelligence. MDPI, 2023; 1-30.
48. Moustafa N, Slay J. UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). Military Communications and Information Systems Conference (MilCIS). 2015; 1-6, doi: 10.1109/MilCIS.2015.7348942.
49. Mbunge E. et al. Implementation of ensemble machine learning classifiers to predict diarrhoea with SMOTEENN, SMOTE, and SMOTETomek class imbalance approaches. Conference on Information Communications Technology and Society (ICTAS). 2023; 1-6, doi: 10.1109/ICTAS56421.2023.10082744.
50. Amos O, Joseph M, Ugochukwu A. Using Machine Learning Techniques Random Forest and Neural Network to Detect Cyber Attacks. 2023; 1-11.
51. Balamuthukumar P. Detecting Attacks to Computer Networks Using a MultiLayer Perceptron Artificial Neural Network. INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN TECHNOLOGY. 2020; 6(12):673-676.
52. Swati PG, Sudhir KM. A Comparative Study of Cyber Attack Detection & Prediction Using Machine Learning Algorithms. Research square. 2023; 1-20.
53. Prasanthi, B., Suresh Pabboju, and D. Vasumathi. "QUERY ADAPTIVE HASH BASED IMAGE RETRIEVAL IN INTENT IMAGE SEARCH." Journal of Theoretical & Applied Information Technology 93.2 (2016).
54. Prasanthi, B., Pabboju, S. & Vasumathi, D. A Novel Indexing and Image Annotation Structure for Efficient Image Retrieval. Arab J Sci Eng 43, 4203–4213 (2018). https://doi.org/10.1007/s13369-017-2827-1
55. Prasanthi, B., Suresh, P., Vasumathi, D. (2017). Index-Based Image Retrieval-Analyzed Methodologies in CBIR. In: Vishwakarma, H., Akashe, S. (eds) Computing and Network Sustainability. Lecture Notes in Networks and Systems, vol 12. Springer, Singapore. https://doi.org/10.1007/978-981-10-3935-5_24
56. D. Vasumathi, S. Pabboju and B. Prasanthi, "Specific query semantic signatures in web page re-ranked image retrieval," 2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Chennai, India, 2016, pp. 1-8, doi: 10.1109/ICCIC.2016.7919710.
57. Prasanthi, B & Pabboju, Suresh & Devara, Vasumathi. (2021). Feature Selection based Reduction in Dimensions and Indexing of Images for Efficient Image Retrieval. 456-461. 10.1109/ESCI50559.2021.9397054.