A Smart Machine Learning Architecture for Protecting IOT against DDOS Assaults

S.Diwakaran[1], M.Tamilselvi[2], S.Radhika[3], S. Scinthia Clarinda[4], Shamitha.C[5]

[1]Associate Professor, Department of Electronics and Communication Engineering, Kalasalingam Academy of Research and Education, Krishnankoil, Srivilliputhur.
[2]Senior Lecturer, Department of Mechatronics Engineering, T.S.Srinivasan Centre for Polytechnic College and Advanced Training, Chennai.
[3]Associate Professor, Department of CSE, Saveetha School of Engineering, SIMATS, Chennai.
[4]Assistant Professor, Department of IT,Jeppiaar Institute of Technology, Chennai.
[5]Assistant Professor, Department of ECE, Manipal Institute of Technology, MAHE, Bengaluru Campus,Bengaluru.

[1]s.divakaran@klu.ac.in, [2]tamilselvivlsi@gmail.com, [3]radhikas.sse@saveetha.com,[4]scinthiaclarinda@jeppiaarinstitute.org,
[5]shamithac2@gmail.com

Abstract.

The Internet of Things (IoT) is a collection of millions and millions of systems that are capable of interacting with one another while requiring little or no involvement from the user. The IoT is one of the most rapidly expanding fields of computing; nevertheless, the fact is that in the very adverse domain of information technology, the IoT is subject to a wide range of intrusions of various forms. As a result, realistic remedies to protect IoT networks, such as network anomaly detection, must be developed in order to address this issue. Regardless of the fact that assaults cannot be completely avoided in perpetuity, early identification of an attack is critical for effective defense in the real world. A standard high-end security solution for an IoT system is not suited since IoT devices have limited storage space and computing power, and hence do not provide enough protection. Furthermore, IoT devices are now capable of remaining connected for extended periods of time without the need for human involvement. Consequently, in order to counteract this danger, intelligent infrastructure authentication mechanisms, such as artificial intelligence methods, will need to be developed. In spite of the fact that numerous researches have been conducted in recent times on the use of ML solutions to attack identification challenges, little consideration has been devoted to the identification of assaults especially in Internet of Things (IoT) networks. Through the analysis of numerous network classification models that may be employed in order to quickly and effectively detect attacks on the Internet of Things, this study hopes to make a significant addition to the research. The Bot-IoT dataset, which is a new addition to the collection, is used to assess different detection techniques. A total of seven distinct machine learning algorithms were utilized throughout the project execution, with the majority of them achieving good efficiency. The Bot-IoT database was searched for extra characteristics that might be used during the deployment. These additional innovations were compared to existing studies in the field and were found to be more effective.

Keywords. *Identification of system anomalies, computer vision, the Internet of Things (IoT), assaults, bot-IoT collection*

1. INTRODUCTION

Problems with network security, confidentiality, and protection are becoming more prevalent throughout the world, and cyber security has emerged as a need as a result of the increased use of digital Smartphone and business. Following the rise in the number of Internet-based applications and the introduction of cutting-edge technology like as the Internet of Things, new initiatives to penetrate information infrastructure have been initiated. In computing, the Internet of Things (IoT) is a collection of interconnected devices that have the capability of connecting with one another without the need for individual involvement. The Internet of Things allows many devices that have detectors in domains such as healthcare, agriculture, transportation, and other sectors to communicate with one another across a network of wireless connections. The Internet of Things (IoT) applications are transforming our business and personal life by saving us resources and time. There are also various positives, as well as innumerable chances for the exchange of information, innovation, and progress that come as a result of globalization. As the Internet of Things (IoT) has grown in popularity, these systems have become more vulnerable to cyber-attacks that may compromise their protection, confidentiality, and availability [1, 2].

In part because the Internet is at the heart and centre of the Internet of Things, every security threat which occurs on the Web is also accessible through the Internet of Items, and vice versa. The Internet of Things (IoT) nodes have minimal facilities and infrastructure as compared to other conventional systems, and they do not have human controls. In addition, the fast expansion and widespread acceptance of Internet of Things devices in everyday life makes IoT security challenges more problematic, necessitating the development of network-based security solutions. While modern systems are capable of recognizing certain types of assaults, it is still difficult to detect other types of attacks. In order to keep pace with the growing number of network attacks and the huge increase in total of details provided in connections, quicker and more accurate methods for detecting attacks are necessary [2] [12], and there is no reason to suspect that there is room for more accelerated approach to enhance network security. For example, we might consider ML as one of the most successful computational models for providing embedded intelligence in the Internet of Things (IoT) ecosystem in this situation. For a range of connected security applications, like deep packet inspection, penetration testing, and malware classification, and many others, neural network models are being used successfully. The most difficult difficulty in building a stable network communication system is determining how to identify and prevent hostile network activities in an efficient manner. A Network Intrusion Detection System (NIDS) may be used as a defensive mechanism against cyber assaults [3][4] [13], according to the authors.

It is possible to define Machine Learning as a smart device's capacity to change or regulate a knowledge-based state or behavior. Machine Learning is regarded to be a vital component in the development of an Internet of Things solution. A machine learning algorithm's capacity to infer useful information from data supplied by devices or people is shown by the fact that it is employed in tasks such as classification and regression problems. In the same way, machine learning may be utilized to offer security services in an Internet of Things network. The use of neural networks to assault detection issues is becoming an increasingly popular research topic, and ML is being employed in a growing number of various purposes in the realm of cybersecurity [14]. Despite the fact that numerous studies in the literature have employed machine learning approaches to determine the most effective methods to find assaults, only a small amount of research has been done on effective investigative techniques that are ideal for IoT contexts.

Unusual case information security and handwriting computer hackers are two methods in which reinforcement learning could be applied to the task of identifying cyber-attacks. Using certain traffic characteristics in known attacks, signature-based approaches are intended to identify and prevent known assaults from occurring. One of the approaches' most important characteristics is its capacity to correctly detect all terrorist vulnerabilities and avoiding the generation of an overwhelming loss in accuracy. [4] [16] employed four different data mining methods as starting tools to learn about the features of numerous well-known intrusions, for illustration, in the field of deep packet inspection. It was also possible to detect infected devices using signature-based approaches, which included recognizing botnet network traffic patterns. The two most significant disadvantages of signature-based techniques are that they need regular human modifications of attack traffic signatures in order to be effective, and that they are incapable of detecting previously undisclosed assaults. The second kind of detection approach is anomaly-based detection, which is described below. This class represents regular network activity, and anything out of the ordinary is regarded as an assault. The capacity of this class to identify unknown assaults makes it a desirable tool for security professionals to use. When using unusual case techniques, the much more major issue to solve is the possibility of high false positive frequencies (FARs), because completely undiscovered (although if permitted) activities may be labeled anomalous. To build a hybrid technique, the characteristic and adversarial learning techniques may be utilized in combination to identify anomalies. As per the researchers, a mixed technique illustration is utilized to increase the recognition rate of common threats while concurrently slowing the rate of false positives (FP) for attack types.

In this research, we provide a contribution to the field by exploring the usefulness of employing machine learning algorithms to identify IoT network assaults as part of a defense against IoT attack behavior. The detection techniques are based on the evaluation of a current collection, Bot-IoT, whose comprises actual and generated IoT network activity, as well as numerous types of attacks [5] [15]. A random forest regression approach was used to extract characteristics from this dataset, which were then analyzed. During the development process, seven different ML algorithms were applied, and excellent results were obtained in terms of performance. The following is a list of the ml algorithms that we used in this research, in alphabetical order: Learning algorithms such as K-nearest neighbours, ID3, principal components assessment, Randomized Forests, AdaBoost, Multi - layer perceptron (MLP), and Naïve Bayes classifier are all examples of those that are used.

Through this study, we may make the following contributions to society:

- By monitoring the characteristics of ml algorithms on a current IoT dataset, it is possible to make improvements in threat detection in IoT networks.
- Generate new information from the data and choose the most relevant features for use in improving the computational efficiency in ML algorithm.
- Make a contribution to the Internet of Things literature. Because there have only been a few research conducted using the Bot-IoT dataset, working with this database may be seen as a potential important addition to the literature.

In the following sections, you will find an outline of the paper: After reviewing similar work and discussing the history in this area, Section III shows our suggested strategy, followed by technical details. Section IV concludes with a discussion of the topic. Section IV presents the results of the experiment together with assessments, and Section V serves as a conclusion to this work by providing a summary of the findings.

2. RELATED STUDY

Because of the rising widespread use of The internet of Things (IoT), hackers can easily are ready to take advantage of the known vulnerabilities that several devices are constructed with from the outset of their development. No reasonable expectation can be placed on users to cope with this threat through their own, and several of the current infrastructure solution providers are either unavailable or impossible to be used for the average user, creating a gap that must be bridged in the future. In [6], the paper proposed an effective handwriting technique to measure, analyze, and recognize potentially dangerous visitors for Network environments in the household electrical communication network, which makes use of surveillance smelling methodologies as well as a fog implementation to supervise aberrant behavior. The suggested solution is focused on two attack and dissemination pathways used by the well-known Mirai botnet, namely DNS and Telnet, which are the emphasis of the suggested system.

Due to the rapid deployment of the Internet of Things (IoT) in a variety of fields, the restricted abilities of such devices pose considerable security risks, noting their connection with Software Defined Networks (SDNs) to offer more flexible services. The author [7] investigated effective threat detection algorithms for application Web of Things (SD-IoT) systems in the context of these infrastructures. First of all and primarily, they are imitations of commonly utilized attack techniques. In the next chapter, we analyze the impacts of multiple attribute values just on prediction performance for multiple attacks, with such a specific focus placed on Randomized Forests (RF) face detection algorithms in significant. Particularly explored is the influence of various RF configurations (wild size and trees thickness, for instance) on the correctness rates and operated overhead expenses. Along with the information we collected, two well-known Iot devices collections were also used in the analysis.

Whenever the given model parameters are used in conjunction with the examined attacks, the data suggest how RF can deliver superior detection performance for the assaults under investigation. It is also worth noting that the classification performance of RF is only significantly reduced with lower woodland widths, which enables for significant savings in walk outgoings to be realized. This demonstrates the viability of the solutions under discussion in Distributed systems with fewer money, as seen by the example.

With the passage of time, an increasing number of devices are being linked to a particular network. If just one of the devices in the network is hacked, then all of the devices in the network will be subject to attack. This makes Intrusion Detection more difficult in any given network. It is virtually hard to identify and intervene manually in this situation. As a result, it is critical to be able to identify diverse forms of assaults with greater certainty while requiring less processing complexity and time to do so. There has already been a significant amount of study done in this field, where the assaults have been evaluated independently. The creator of [8] concentrates on identifying intrusions, such as Internet of Things botnet assaults and other sorts of network threats. In order to do this, they develop a multiclass classification system that incorporates supervised learning models as well as the dimensionality reduction approach. A large number of researches on ML-based IDS have made use of the KDD dataset or improved versions of the KDD dataset. Specifically, we employed a novel dataset, the IoT network Intrusion Detection dataset, for this investigation.

To safeguard home Wi-Fi networks, the author created and developed "A System for Preventing IoT Device Attacks on Home Wi-Fi Router" (SPIDAR) in [9], which was published in the journal Computer Science. Home users may benefit from this system, which comprises of an SPIDAR home Wi-Fi router, an SPIDAR Raspberry Pi, and an SPIDAR web application that both prevents attacks and displays attack data. Moreover, it helps save money by avoiding the need to purchase and install costly intrusion protection software and hardware at home. The company offers two types of prevention methods: the signature-based method, which uses Snort software, and the behavior-based method, which learns and analyses the behavior of IoT devices by using either the benchmark or machine learning in order to improve the system's growth by enhancing the system's performance.

There has been a lack of attention dedicated to the detection of harmful attacks in the setting of Internet of Things networks. The author [10] offers an effective intrusion detection system (IDS) to identify unanticipated Internet of Things assaults by using multiple bagging and boosting ensemble techniques as well as a feed forward artificial neural network to achieve this goal. They utilized a newly released dataset, UNSW-NB15, which included simulated IoT sensor data, to evaluate the performance of the suggested models using a 5-fold cross validation approach, which they found to be effective.

3. METHODOLOGY

A short overview of the database that was utilized, as well as our suggested technique for detecting attacks in IoT networks, is provided in this section. Machine learning approaches are used to discover abnormalities in our suggested strategy, which includes a variety of pre-processing steps as well as real applications [17] [18]. CICFlowMeter was used to extract flow-based characteristics from the raw dataset in the first step (CFM). Initial and foremost, in this first step, the condition characterized technique was performed, followed by the division of the information into 2 segments: learning and assessment. To transform the information into a form which can be utilized by ml techniques, it is important to do information which was before on the information. Following these procedures, the feature selection stage determines which attributes will be employed by the algorithms and which will not. Finally, the deployment of ml algorithms brings our strategy to a conclusion. Figure 1 depicts a high-level overview of the suggested technique. Figure 1.

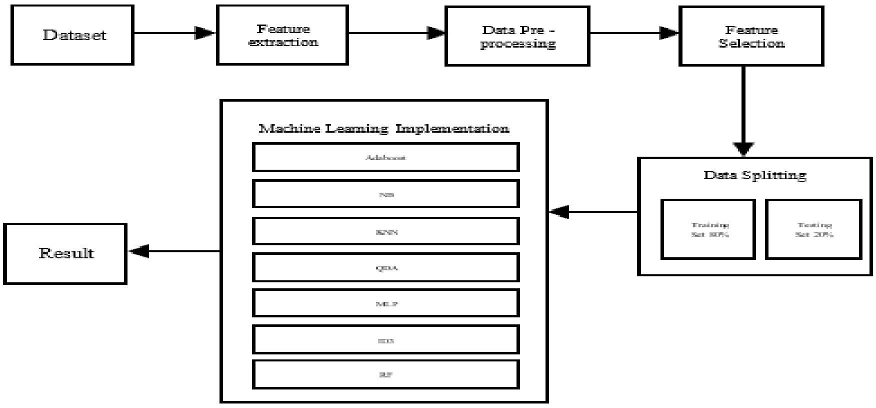

Figure 1: Proposed Approach Overview.

In order to conduct the tests, we used the Bot-IoT database, which was chosen for its frequent releases, broad attack variety, integration of IoT-generated traffic, and capacity to build new features from the input data. There really are three basic sorts of attacks in this collection, all of which have been predicated on botnet situations, and which together includes sniffing, disruption of services, and computer hacking as its objectives. CFM was utilized to extract circulation features from significant traffic samples, which would then be subjected to further examination. Network traffic flow generator (CFM) is a software package offered by CIC that generates up to 84 different network traffic characteristics.

According to what others have said in the preceding paragraphs, the fundamental purpose of the research is to evaluate the efficiency of different methods in the identification of threats on the Internet of Things system architecture. Throughout this section, we will describe the information, the techniques and algorithms that we used, and the steps we followed to put our results into action.

A. Database

Because ml algorithms are used in the designed for various network security activities, big databases are required in order to assess network flows and discriminate between regular and problematic traffic patterns. Numerous studies have been carried out to build network datasets throughout the course of the years. The bulk of computer science has assessed its conclusions there against generated or actual data from networking. Some of the datasets, such as DARPA 98, KDD99, and N-BaIoT, have been publicly accessible, despite the fact that a significant number of them remain secret, mostly owing to security concerns. The production of genuine IoT and internet traffic databases that incorporate novel Botnet situations is still in its early stages, despite the fact that multiple datasets have already been developed. More crucially, some databases do not include IoT-generated traffic, while others do not develop any new characteristics as a result of the IoT. A number of the assaults were unsuccessful because the test bed employed was not genuine. In other situations, the assault situations were not varied either. Despite the fact that this dataset is reasonably big and clean, it is uneven, with the proportion of normal data being much smaller than the proportion of assault data. The Bot-IoT dataset includes both actual and artificial IoT network traffic, as well as a variety of different forms of assaults.. Bot-IoT data attacks may be divided into three categories: probing attacks, denial of service attacks, and data breaches.

B. ML Algorithms

The Bot-IoT database was used to test seven well-known ml methods: KNN, ID3, Random Forest, AdaBoost, Quadratic discriminant analysis, Multilayer perceptron, and Naive Bayes. The classifiers were evaluated using the Bot-IoT dataset. When selecting these classifiers, the emphasis is on putting together prominent methods with a variety of properties in a single package. The algorithms that were employed in this context are quickly discussed in the next section.

- **K-Nearest Neighbours:**KNN is one of the easiest and most successful algorithms for guided learning. It's used to find comparable data points in the given dataset and link new ones to them. In contrast to the KNN approach, this works well enough on high dimensionality and therefore is speedy as during testing period, it is slow when it comes to producing predictions.
- **Quadratic discriminant analysis (QDA):**Supervised classification issues are well-suited to the QDA algorithm. Assigning samples to one of several categories is done using discriminant analysis, a statistical method. QDA may be used in situations when a category lacks a lot of data. In order to use QDA, more data must be seen than there are categories in the dataset.
- **ID3:**In order to build a decision tree from a dataset, ID3 is utilized. Using a tree-like decision structure, decision trees may be used to classify data. Displaying a technique that has just control instructions is one method. When constructing criteria, characteristics are utilized as "nodes" and the "leaves" are class variables, which are a record's assigned class values. This algorithm, ID3, is utilized in the fields of ML and NLP.
- **Random Forest:**Decision trees are used in RF, a ML technique. A "forest" is generated by combining a variety of various decision tree structures, each of which is constructed in a slightly different manner, according to this approach. When compared to other approaches, this method has numerous benefits, including the capacity to execute effectively on large datasets, its small weight, and its resilience against noisy and anomalies.
- **Adaptive Boosting:**It is a machine learning methodology that relies on classification problems and aims to improve poor classifications. The AdaBoost algorithm's most essential feature is its capacity to handle incomplete data in a database.
- **Multilayer Perceptron:**A feedforward artificial neural network, MLPs is a subset of this kind. ANNs are a predictive model that mimics the natural mind's ability to learn and synthesize new knowledge. The input, output, and hidden layers all exist in an MLP. For its first training, MLP makes use of a method known as back-propagation, which is a kind of supervised learning.
- **Naïve Bayes:**One of the most extensively used supervised algorithms; the NB is well-known for the clarity of its underlying assumptions. Because the traffic classification characteristics employed may be reliant on each other in some way yet are handled separately by the NB classifier, since it requires fewer samples and is simpler to implement, NB has several advantages for users. The connection and interactions between features are not accessible to NB, on the other hand, since it treats each feature as a standalone entity.

C. Implementation Steps

The five key stages in our method are image enhancement, information or before, refers to the things, image segmentation, as well as the application of ml algorithms (multiple linear regression).

- **Extraction of features:**Flow-based characteristics were obtained from the original network traffic data using CFM. Networking existing traffic converter CFM is a traffic flow characteristic generator supplied by CIC that generates 84 different networking characteristics. There is an option to download a CSV file of the datasets, as well as a visual summary of the characteristics derived from the file. Classifiers' prediction ability may be improved by extracting additional dataset characteristics from this approach.
- **Data pre-processing:**In order to make the database acceptable for machine learning, pre-processing methods are employed. As a result, this stage also involves deleting unnecessary or damaged data samples, which makes it more effective and precise.
- **Splitting Data:**Data are essential to the machine learning model because they allow it to learn. Test data are also needed to assess the algorithm's effectiveness, so that we can observe how well it performs. About 80% of the Bot-IoT database was deemed training data, while the other 20% was deemed testing data in our research.

- **Selecting Features:**In order to develop a secure environment for IoT systems that is both lightweight and suitable for training and testing methods, it is essential to reduce the number of features and only employ those that are absolutely necessary. As a means of selecting relevant characteristics, we implemented the Random Forest Regressor method. Many studies have shown how successful random forest regression is in shrinking a dataset. More than 80 network traffic characteristics may now be reduced to seven, making the model train and react faster. Figure 2 depicts the whole dataset's relevance weights for the various attributes.

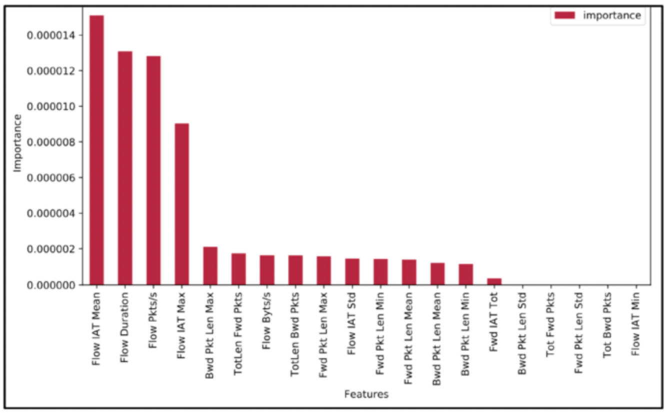

Figure 2: Feature Importance of dataset

- **Implementation of ML Algorithms:**Using Python deep learning packages, we conducted all of our tests. So every incursion in the group of statistics was studied in three phases: independently on every invasion, with a set of best aspects for each invasion combined, on the entire database with the series of great attributes combined, and finally on the database server with both the seven great attributes acquired in this showcase sampling stage. The complete list of possible attacks are shown in TABLE I.

Flow IAT Mean	Flow Duration	Flow Pkts/s
Flow IAT Max	FwdPkt_Len_Max	TotLenFwd Pkts
Fwd Pkt Len Mean	Tot Bwd Pkts	Fwd IAT Tot
Flow IAT Std	Flow Bytss	Tot Fwd Pkts
Flow IAT Min		

TABLE I. A COMPLETE LISTING OF AVAILABLE FEATURES FOR ALL POSSIBLE ATTACKS

4. Results and Discussions

A. Evaluation Metrics

Measures that are relevant and applicable to the job at hand are critical for evaluate the effectiveness of machine-learning models. The following are the most important performance indicators for correctness, sharpness, f-measure, as well as remember, as represented by the formulas following table:

$$Precision = \frac{TP}{TP + FP} \qquad (1)$$

$$Recall = \frac{TP}{TP + FN} \qquad (2)$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \qquad (3)$$

$$F - measure = \frac{2}{\frac{1}{Recall} + \frac{1}{Precision}} \qquad (4)$$

B. Results

The assessment of ML algorithms on the dataset was divided into three steps, as described in the preceding section. Each assault in the dataset is treated as a distinct case for applying machine learning algorithms in phases one and two. In phases three, the top seven features from the feature selection stage are used to apply ML methods across the whole dataset. Following are the outcomes of all the experiments. The arithmetic means of the 10 performance assessment methodologies for each ML algorithm are shown in the tables.

Attack Names	F-Measures						
	NB	QDA	RF	ID3	AB	MLP	KNN
HTTP	0.72	0.85	0.96	0.96	0.96	0.95	0.96

UDP	0.73	0.92	0.98	0.98	0.98	0.97	0.98
TCP	0.71	0.85	0.99	0.99	1.0	0.74	0.99
HTTP	0.72	0.82	0.95	0.96	0.95	0.95	0.96
UDP	0.72	0.83	0.97	0.97	0.98	0.98	0.97
DOS_TCP	0.64	0.74	1.0	1.0	1.0	0.78	0.99
Data exfiltration	0.72	0.76	0.96	0.97	0.97	0.94	0.97
Keylogging	0.72	0.82	0.95	0.95	0.95	0.91	0.98
Service_Scan	0.73	0.83	0.95	0.95	0.95	0.94	0.94
OS_Scan	0.72	0.76	0.94	0.97	0.98	0.97	0.99

TABLE II. Distribution of Results Based on the Kind of Attack.

Phase 1: Machine learning methods may be used to each individual assault. Table II shows the results of seven different ml methods employed to 10 distinct attack types. if the F-measure is equal to zero, for the purpose of excluding equivalence, the various definitions are examined: clarity, reliability, recollection, as well as duration.

If one looks at Table II, it can be seen that all of the algorithms, with the exception of the Naïve Bayes classifier and the Algebraic approach (QDA), were able to detect more than 90 percent of the different types of assaults tested.With the best score in six of the ten trials, the ID3 algorithm proved to be the most effective in the battle against DDoS attacks. In fact, ID3's top score is shared by at least one other method in every challenge. Its minimal processing time, however, puts it ahead of the others. Naive Bayes, the algorithm with the lowest F-measure, was utilized in all jobs. It got a poor score, in part because of the DOS TCP assault. The speed of Naive Bayes was much superior than that of the other methods, despite its lower accuracy. Also worth mentioning is the QDA, which had a dismal efficiency ranking of #2 among the algorithms tested in this study.

Phase 2: The best aspects of each assault were integrated using machine learning methods on the complete dataset. The whole collection of data is tapped into at this point. We employed seven distinct ml algorithms to analyze the full dataset, and each assault had its own feature set. As shown in Table III, 13 characteristics were retrieved for each assault.

ML Algorithm	**Accuracy**	**Precision**	**Recall**	**F-Measure**	**Time**
NB	0.78	0.84	0.78	0.75	5.056
QDA	0.88	0.89	0.88	0.87	6.1964
RF	0.98	0.98	0.98	0.98	27.0328
ID3	0.99	0.99	0.99	0.99	19.3447
Adaboost	1.0	1.0	1.0	1.0	308.9403
MLP	0.84	0.88	0.84	0.83	1011.5001
KNN	0.99	0.99	0.99	0.99	2052.1801

TABLE III. Assessment of Results from Phase 1

Table III shows that AdaBoost had the best showing, closely by KNN and ID3, with AdaBoost coming out on top. For this feature, ID3 takes priority over KNN, which is much slower. Naive Bayes had the lowest score of any algorithm, at 0.75. NB and QDA were the quickest in terms of speed. Even while KNN scored well in terms of effectiveness, it was still much slower than the competition.

Phase 3: Implementing machine learning techniques to the whole information using the seven most important structural interventions during the characteristic evaluation phase.

ML Algorithm	**Accuracy**	**Precision**	**Recall**	**F-Measure**	**Time**
NB	0.79	0.85	0.79	0.77	4.0472
QDA	0.87	0.89	0.87	0.86	4.4056
RF	0.97	0.97	0.97	0.97	28.9246
ID3	0.97	0.97	0.97	0.97	17.0899
Adaboost	0.97	0.97	0.97	0.97	238.8618
MLP	0.84	0.87	0.84	0.86	949.6977
KNN	0.99	0.99	0.99	0.99	1615.9852

TABLE IV. Use of the RF for all datasets to execute the features that were discovered.

The algorithms' F-measure effectiveness did not change much, however the running times of all the methods were significantly decreased. The procedure in Table IV uses 13 characteristics, but just 7 attributes are utilized in Table III, which results in a faster execution time. In order to speed up the machine learning model, the number of features was lowered.

Table V shows the final outcomes of the execution in comparison to a previous research. The research done by Ferrag et al. [11] in 2019 was used as a basis for this comparison. Why? Because we employed two machine learning approaches that were inspired by previous work using the same dataset. Random Forest and Naive Bayes are two related machine learning methods. The set of features employed in our work differs significantly from that of theirs. Our set of features was derived from CFM, whereas theirs was based on the source. An assessment criteria was established based on the rate of recall (Recall). Table V compares the findings of the two investigations. The Random Forest method employed in our research is superior than that used in [11], and the same can be shown for most types of attack with the NB approach, when the results are compared. We can observe that the additional characteristics employed in our study improved the efficiency of both methods.

Attack Names	Ferrag et al		Our Work	
	RF	NB		RF
HTTP	82.26%	50.78%	96%	71%
TCP	88.28%	78.67%	99%	70%
UDP	55.26%	78.50%	98%	72%
HTTP	82.20%	68.68%	95%	71%
TCP	81.77%	65.56%	100%	63%
UDP	82.99%	100%	97%	71%
Data exfiltration	86.55%	66.55%	96%	71%
Key_logging	70.12%	65.62%	95%	71%
OS_Scan	82.20%	68.68%	94%	70%
Service_Scan	69.82%	65.21%	95%	72%

TABLE V. Two Algorithms are Compared for their Performance.

5. CONCLUSION AND FUTURE SCOPE

In this post, we describe how we employed ml approaches to discover IoT network attacks. The Bot IoT was used as a database in this instance due to its daily reports, extensive attack diversification, and wide range of network elements. Following the analysis of the raw traffic traces, we used CFM to obtain a range of relevant flow-based properties that might be used in other applications. The 84 network traffic characteristics described by CFM serve to characterize network flow. It was decided which attributes would be employed in the machine learning techniques using the Randomized Forests Linear regression methodology, and the importance of weight calculations was determined by deploying the technique and determining the importance of value calculations. The calculations for this project were completed by merging two different methodologies. For the purpose of determining relevance values, two techniques were used: the first way assigned a different relative importance to every charge kind, while being in the second technique the relevance scales for all assaults were pooled together again and calculated as a single group. As a final step, data was subjected to testing using a diverse range of machine learning techniques. According to the F-measure, these techniques and their performance ratios are as follows: The F-measure ranged from 0 to 1, as did the Naive Bayes, QDA, Random Forest, ID3, AdaBoost, MLP, and K Nearest Neighbors scores. Seven supervised algorithms were tested in this study. Performance evaluation of certain unsupervised methods would be an exciting scenario effort. In addition, we used a variety of machine learning techniques on their own. Ultimately, we want to build a multi-layered model that incorporates many ml algorithms.

6. REFERENCES

[1] M. A. Mahmood and A. M. Zeki, "Securing IOT against DDOS attacks using machine learning," 3rd Smart Cities Symposium (SCS 2020), 2020, pp. 471-476, doi: 10.1049/icp.2021.0905.

[2] H. Alsheakh and S. Bhattacharjee, "Towards a Unified Trust Framework for Detecting IoT Device Attacks in Smart Homes," 2020 IEEE 17th International Conference on Mobile Ad Hoc and Sensor Systems (MASS), 2020, pp. 613-621, doi: 10.1109/MASS50613.2020.00080.

[3] S. -J. Wang, C. X. Cai, Y. -W. Tseng and K. S. -M. Li, "Feature Selection for Malicious Traffic Detection with Machine Learning," 2020 International Computer Symposium (ICS), 2020, pp. 414-419, doi: 10.1109/ICS51289.2020.00088.

[4] C. -Y. Chen, L. -A. Chen, Y. -Z. Cai and M. -H. Tsai, "RNN-based DDoS Detection in IoT Scenario," 2020 International Computer Symposium (ICS), 2020, pp. 448-453, doi: 10.1109/ICS51289.2020.00094.

[5] N. Koroniotis, N. Moustafa, E. Sitnikova, and B. Turnbull, “Towards the development of realistic botnet dataset in the internet of things for network forensic analytics: Bot-iot dataset,” Future Generation Computer Systems, vol. 100, pp. 779–796, 2019.

[6] M. Hammoudeh et al., "Network Traffic Analysis for Threat Detection in the Internet of Things," in IEEE Internet of Things Magazine, vol. 3, no. 4, pp. 40-45, December 2020, doi: 10.1109/IOTM.0001.2000015.

[7] Y. Zhang et al., "Efficient and Intelligent Attack Detection in Software Defined IoT Networks," 2020 IEEE International Conference on Embedded Software and Systems (ICESS), 2020, pp. 1-9, doi: 10.1109/ICESS49830.2020.9301591.

[8] M. G. Desai, Y. Shi and K. Suo, "IoT Bonet and Network Intrusion Detection using Dimensionality Reduction and Supervised Machine Learning," 2020 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), 2020, pp. 0316-0322, doi: 10.1109/UEMCON51285.2020.9298146.

[9] V. Visoottiviseth, P. Sakarin, J. Thongwilai and T. Choobanjong, "Signature-based and Behavior-based Attack Detection with Machine Learning for Home IoT Devices," 2020 IEEE REGION 10 CONFERENCE (TENCON), 2020, pp. 829-834, doi: 10.1109/TENCON50793.2020.9293811.

[10] M. Shorfuzzaman, "Detection of cyber attacks in IoT using tree-based ensemble and feedforward neural network," 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2020, pp. 2601-2606, doi: 10.1109/SMC42975.2020.9283443.

[11] R.Bharathi, T.Abirami,” Energy efficient compressive sensing with predictive model for IoT based medical data transmission”, Journal of Ambient Intelligence and Humanized Computing, November 2020, https://doi.org/10.1007/s12652-020-02670-z

[12] G. Anitha, P. Nirmala, S. Ramesh, M. Tamilselvi and G. Ramkumar, "A Novel Data Communication with Security Enhancement using Threat Management Scheme over Wireless Mobile Networks," 2022 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), 2022, pp. 1-6, doi: 10.1109/ACCAI53970.2022.9752584.

[13] P. Nirmala, S. Ramesh, M. Tamilselvi, G. Ramkumar and G. Anitha, "An Artificial Intelligence enabled Smart Industrial Automation System based on Internet of Things Assistance," 2022 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), 2022, pp. 1-6, doi: 10.1109/ACCAI53970.2022.9752651.

[14] S. Ramesh, M. Tamilselvi, G. Ramkumar, G. Anitha and P. Nirmala, "Comparison and analysis of Rice Blast disease identification in Greenhouse Controlled Environment and Field Environment using ML Algorithms," 2022 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), 2022, pp. 1-5, doi: 10.1109/ACCAI53970.2022.9752538.

[15] A. G, S. K. M, M. Ayyadurai, S. K. C and G. Ramkumar, "Design of Miniaturized Single Bit MEMS Phase Shifter using MEMS Switches," 2021 5th International Conference on Trends in Electronics and Informatics (ICOEI), 2021, pp. 235-239, doi: 10.1109/ICOEI51242.2021.9453063.

[16] G. Ramkumar and M. Manikandan, "Uncompressed digital video watermarking using stationary wavelet transform," 2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies, 2014, pp. 1252-1258, doi: 10.1109/ICACCCT.2014.7019299.

[17] G. Ramkumar, et al, “Experimental analysis of Brain Tumor detection system using Machine learning approach”, Materials Today: Proceedings, 2021, ISSN 2214-7853,https://doi.org/10.1016/j.matpr.2021.01.246.

[18] Benisha.M, Rubala.R, Anisha.M, Thandiah Prabu.R, Ponmozhi Chezhiyan. (2020). An IoT Secured System Design for Real-Time Health Monitoring of Post-Chemotherapeutic Effects. International Journal of Advanced Science and Technology, 29(05), 10193 - 10201.