REPUBLIC OF TURKEY
YILDIZ TECHNICAL UNIVERSITY
GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

ENHANCING SECURITY LEVEL OF INDUSTRIAL INTERNET OF THINGS DEVICES BASED ON BoTNeT DETECTİON AND FEATURE SELECTİON

Weam Husham Abdulwahhab AL-JABBARI

MASTER THESIS
Department of Computer Engineering
Computer Engineering Program

Supervisor
Prof. Dr. Hasan Hüseyin BALIK

1
GENERAL INTRODUCTION

1.1 Internet of Things (IoT)

The Internet of Things (IoT) is a network of computers and other gadgets that can interact and communicate with one another. It has arisen in the modern period and is driving the creation of innovative business process technologies [1]. The networked, IP-enabled objects are referred to as "things" (both physical and virtual). Telematics boxes, self-driving vehicles, printers, video surveillance, tablet devices, smartphones, ultra-wideband (UWB), infrared data association (IrDA), ZigBee, NFC data centers, and mobile and Wi-Fi networks are a few examples of things that could be utilized. The Internet of Things (IoT) is thought of as a network of countless physical things with all of its subtechnologies (24.15 billion devices in 2019 increasing to 76.45 billion computing devices in 2026) [2]. By 2025, the IoT could have a financial impact on the worldwide economy of $3.9 to $11.1 trillion [3]. These gadgets are equipped with various additional critical technologies in addition to Internet Protocol (IP), such as radio-frequency identification (RFID) technology, sensors, actuators, GPS services, nanotechnologies, near field communication (NFC), and cloud computing.

1.2 Industrial Internet of Things (IIoT)

The Industrial Internet of Things (IIoT) is a revolutionary concept for a completely connected, transparent, automated, and intelligent industrial setup that boosts production processes and efficiency. Industrial IoT (IIoT) or Industry 4.0 (I4.0) is a principle for a cutting-edge, intelligent, completely integrated factory that uses emerging innovation (such as IoT, cloud computing, artificial intelligence, etc.) and creative technologies (such as IIoT, automation, monitoring, etc.) to enhance the manufacturing environment with reduced costs, adaptability, quality, remote operational processes, etc. [4]

Digitization and the spread of smart gadgets or devices have penetrated critical industrial and infrastructure sectors. Healthcare, water, power, as well as other components of crucial systems can run more effectively, profitably, and dependably due to devices' having wireless and wired connectivities [5]. Such installations are referred to as the Industrial Internet of Things (IIoT), in which industrial resources are becoming fully capable of automatically responding to and modifying their activities depending on information obtained through the vast networked control loop spanning from the Edge and Cloud to the Enterprise layers [6]. Yet, this significant integration between Operational Technology (OT) and Information Technology (IT) systems, which is at the core of IIoT technologies, has extended the cyber threats and raised the likelihood that cyberattacks will be undertaken against such vital systems [7].

1.3 Differences between IoT and IIoT

The Industrial Internet of Things (IIoT) refers to industrial sectors such as energy and production control, as well as interconnected sensors, instruments, and other computer-connected devices [8]. IIoT is for the industrial sector, while IoT is for the commercial sector. The next industrial revolution, also known as the Industrial Internet of Things, will begin within the time period known as Industry 4.0 [9]. Industry 4.0 [10] is regarded as being part of the 4th Industrial Revolution. In Industry 4.0, factories feature machines with internet communication and sensors that are connected to a computer that can manage the entire production line and draw significant conclusions on their own [11]. With the support of the Internet of Services, both inner and intra-organizational facilities are supplied to users in order for them to utilize the production series. Cybersecurity, cloud and edge computing, 3D printers, industrial automation, big data, and the Internet of Things are all facilitated by the IIoT.

Briefly, IoT and IIoT rely on hardware like sensors, internet connectivity, and embedded systems. However, IIoT devices tend to be more expensive than IoT ones because of their need for greater precision in comparison to those of IoT devices. IIoT employs more advanced technology for greater precision. Because IIoT operates in key business areas like manufacturing, machinery monitoring, and so on, IIoT services are more sophisticated than IoT systems; as technology breakthroughs improve, so does complexity. Thus, the complexity of IIoT applications exceeds that of IoT applications. IoT end requirements prioritize user comfort, whereas IIoT end requirements prioritize return on investment [12].

1.4 Security Threats on IIoT

There are many similarities between IoT and IIoT in terms of security, privacy, and integrity because many of the technological solutions for IIoT are built on IoT. While IIoT-enabled devices make people's and businesses' lives easier, their privacy may suffer as a result [13]. IIoT devices have hardware and software that may be used to track user activity, so it is essential to develop regulations and technical solutions to guarantee that users' privacy, security, and autonomy are constantly protected. The IIoT creates a wide platform for sophisticated cyberattacks as more and more devices are linked to the Internet nearly every day. Data leaks, loss of privacy, and the potential for misuse due to unauthorized access that could seize control of the devices are common worries with regard to IIoT devices [14]. Distributed denial of service (DDoS) cyberattacks using various communication protocols, data breaches by keylogging and data leakage, tracking via fingerprinting, and network scans for open ports are a few frequent threats that affect IIoT devices and networks. BotNets are often used to carry out many of these cyberattacks on IIoT systems and networks [15, 16].

1.5 BotNet

BotNets are networks made up of host computers that have been made to serve as slaves under the control of one or more hackers known as botmasters, in order to carry out destructive actions. Malware that can turn a computer into a zombie or bot is distributed by hackers using a variety of creative tactics. When a hacker instead of the user is in charge of a computer, suspicious actions on the internet are carried out without the user's knowledge. In other words, BotNets are groups of computers that work together to carry out illegal operations using malicious software. Bots are frequently used by attackers to infect a large number of machines. These machines collectively make up the BotNet. Such zombies can be used to propagate infections, attack servers, send spam emails, perform various forms of fraud, and engage in other online crimes. BotNet size is an unpredictable quantity that can be modest or enormous. It depends on how complex and sophisticated the employed bots are. There are tens of thousands of zombies in a huge BotNet. A smaller BotNet, on the other hand, was made up of just a few thousand zombies.

The owner of the computer that has turned into a zombie is unaware that the affected machine and all of its capabilities are now being controlled remotely, enslaved, and utilized for malicious purposes by one or more malware runners who use Internet Relay Chat (IRC) as a key tool for their illegal deeds. Numerous varieties of malware, malicious software, and programs have already and are still entangling the internet. While simpler types of bots lack such capabilities, large bots employ their internal spreaders to propagate infections [17].

1.6 BotNet and IIoT

The network of compromised machines expands as the BotNet spreads to new IIoT devices, giving the BotNet greater computational power and the ability to launch more intense attacks. Additionally, because of the widespread use of IIoT in important industries and businesses, these entities are now more vulnerable to cybercrimes where foreign adversaries attempt to bypass security measures [18]. The threats associated with the potential takeover of IIoT devices are immense. Hacking risks include stealing private information, violating personal rights, and occasionally even carrying out cyberattacks that put people in danger of dying, such as sabotaging medical equipment. For IIoT systems used in Industry 4.0, this can mean production and service disruptions, trade secret theft, and sensitive company data leakage, all of which could result in significant financial losses. IIoT attacks, particularly those launched by BotNets, have significantly increased in frequency in recent years. It becomes more and more challenging to protect the IIoT networks and equipment due to the wide variety of attacks that can be made on different protocols and devices [19].

1.7 Security of IIoT in Accordance to BotNet Detection

The technological age, in which IIoT plays a vital role, has profoundly impacted our lives. Moreover, the IIoT's phenomenal expansion is a major cause of numerous and serious cybersecurity vulnerabilities. Hence, there has recently been a lot of interest in both academic and industry in identifying and mitigating potential cyberattacks on IIoT networks. Establishing an IIoT BotNet, as previously stated, is a major attack; typically, enterprises use a variety of security controls, such as intrusion detection and threat intelligence, to detect and block IIoT

BotNets. These strategies could be somewhat useful, but they are incapable of detecting the emergence of zero-day IIoT BotNets with no known patterns. This is why researchers and businesses alike are concentrating on methods for detecting BotNets in the IIoT. Typically, the goal is to identify the source of an attack and minimize the traffic it causes. Analyzing how BotNet structures arise in IIoT systems with the assistance of both business and academics should make it easier to improve security measures for spotting both known and emerging BotNets [20].

Research into the unique characteristics of IIoT BotNets is helping to improve defenses against these threats. In an effort to combat this widespread issue, the field of BotNet attacks detection is seeing progress thanks to machine learning. It is generally agreed that BotNets present one of the greatest threats to IIoT networks. There are now a wide variety of solutions and services for intrusion detection on the market. These products and services offer varying degrees of security across IIoT devices. Recent developments in machine learning have shown some impressive performance in the recognition and classification of various types of attacks [21].

1.8 IIoT Datasets and Features Selection

Machine Learning (ML), a subset of Artificial Intelligence (AI), has attracted more attention in recent years. This is largely because ML models and algorithms have grown more potent and are now integrated with more data and computing capacity, boosting their effectiveness. An emerging approach for spotting network traffic that indicates malicious behavior on a target network is ML-based intrusion detection systems. Researchers have effectively applied ML to the creation of more sophisticated intrusion detection systems technologies, obtaining a reasonable attack detection accuracy [22]. From labelled data samples, supervised ML seeks to understand and learn about sophisticated security events. However, due to a number of logistical and privacy concerns, collecting labeled datasets from real production networks is quite difficult. As a result, researchers have developed simulated intrusion detection system datasets, which are often produced in controlled test conditions in which data labels for both malicious and benign traffic can be consistently and quickly inserted.

Feature selection is a critical step in building an efficient ML-based network attack detection model. A smaller set of features can be expected to result in a more efficient collection and storage, which is necessary for a high-speed IIoT network environment [22]. In other words, the performance of ML-based intrusion detection systems depends on the quality and integrity of the data used to train and evaluate the ML models. Utilizing Features Selection will significantly cut down on the amount of time required to finish the detection, as well as reduce the amount of CPU and main memory (RAM) consumption, and in several cases improve accuracy.

1.9 Literature Review

The literature review is divided into two sections: the literature review for the IIoT datasets utilized in the BotNet detection field is included in the first section, and the literature review for features selection methods is included in the second.

1.9.1 IIoT Datasets Literature Review

In this section, a literature review is carried out on the research that has been undertaken over the past years using the datasets pertaining to IIoT BotNet detection, as well as the outcomes of these studies.

Three distinct deep learning-based models were employed by Mohammed Mudassir et al. [23] (2022) to categorize different IIoT BotNet attacks affecting a variety of communication channels and devices. It was demonstrated that all of the models performed well. The study's conclusion was that deep learning algorithms could be utilized to detect IIoT malware, particularly in the setting of emerging threats that frequently evade traditional techniques. The IIoT BotNet assault dataset created by [24] was used by the researchers. It includes a variety of attacks, such as DoS and DDoS, thievery, and reconnaissance. There are three different protocols used in DoS and DDoS attacks, including HTTP, TCP, and UDP. Theft includes data theft and keylogging. Operating system fingerprinting and a check of open ports are two examples of reconnaissance. About 37 features are included. There are roughly 16 Gigabytes in the entire labeled dataset. The researchers only considered 5% of this material for the study. In terms of future work, the researchers stated that federated learning for IIoT networks, as well as unique techniques of sharing security information between devices and networks, might be investigated further. Additionally, various sorts of IIoT datasets could be combined to form a comprehensive IIoT system dataset, which can be utilized for training ML and DL models as well as providing security through federated learning and edge computing.

P.L.S. Jayalaxmi et al. [25] (2022) demonstrated a DL model for BotNet identification for industrial network activity called DeBot in their study. The Cascade Forward Back Propagation Neural Network (CFBPNN) framework was employed by DeBot. Five well-known bot datasets were used by the researchers for a comprehensive set of experiments: NF-UNSW-NB15, NFToN-IoT, NF-BoT-IoT, NF-CSE-CIC-IDS2018, and ToN-IoT-Windows. According to the researchers, an assault scenario and general network infrastructure with consistent data properties are necessary. [26] suggests NIDS datasets with 1048576 rows with 14 features based on Net-Flow, including both multi-class and binary attack categorization in different modules (V1,V2). The research team chose the Net-Flow (NF) V1 module produced from the NF-UNSW-NB15, NF-BoT-IoT, NF-ToN-IoT, and NF-CSECIC-IDS2018 benchmark NIDS datasets [27]. These datasets' raw packet capture files are transformed into Net-Flow format using a standard feature set. Both attack and normal data are present in each dataset. The Cyber Range and IoT Labs at the School of Engineering and Information Technology (SEIT), UNSW Canberra @ the Australian Defense Force Academy (ADFA) are the source of these extensive network datasets [28]. Both IoT and IIoT sensors generate telemetry-processed data. This covers detailed information about Windows 7, Windows 10, Ubuntu 14, as well as other IoT device types. With other already-built neural network models, they contrasted CFBPNN. They noticed that CFBPNN in DeBot exhibits great accuracy across all datasets with subgroup evaluation and achieves the best F1-score and zero precision.

To protect IIoT infrastructure against fatal and complex multi-variant BotNet attacks, Tooba Hasan et al. [29] (2022) suggested a hybrid intelligent Deep Learning (DL)-enabled technique. The suggested mechanism was carefully tested using available datasets, standard and expanded performance evaluation measures, and existing deep learning benchmark algorithms. By reaching a high detection rate, the proposed approaches outperform in accurately recognizing multi-variant sophisticated bot attacks. The researchers used the newly revised N-BaIoT [30] IoT dataset to train the suggested method. The collection contains benign and recent IoT malware (Gafgyt and Mirai), which are two malware from the BotNet family that are particularly created to attack IoT devices. The dataset includes network traces from Gafgyt and Mirai executions across nine different IoT products (i.e. webcams, doorbells, security cameras, baby monitors, and thermostats).

By using machine-learning approaches, S. Latif et al. [31] (2020) anticipated various cyber attacks on IIoT devices. A unique lightweight random neural network (RaNN)-based prediction model has been put out in their article as a way to forecast these attacks. Several evaluation criteria, including accuracy, precision, recall, and F1 score, were calculated and compared with the conventional support vector machine (SVM), artificial neural network (ANN), and decision tree (DT) in order to assess the effectiveness of the RaNN-based prediction model. The evaluation study revealed the excellent accuracy that the suggested RaNN model produces. Kaggle [32] provided an open-source dataset referred to as DS2OS. This is one of the most modern versions of the IIoT datasets used to assess the accuracy and effectiveness of various machine and deep learning-based cybersecurity systems. This dataset was contributed by Pahl et al. [33]. It encompasses cyberattacks on sensors and applications, so it gives information on a variety of attacks and abnormalities in IIoT applications, such as smart factories, smart buildings, smart homes, and so on. There are 13 attributes and 357952 samples in the dataset. It has 8 classes, 347935 typical data entries, and 10017 abnormal data values.

Using data from TCP/IP packets, Joseph Bamidele Awotunde et al. [34] (2021) suggested a methodology for detecting damaging operations in IIoT networks. To provide a vivid explanation of typical network architectures for unsupervised learning, it used unsupervised DL algorithms, which were mixed rule-based with automated dimensionality reductions. In comparison to other previous research methodologies, the proposed model obtains the highest identification rate of 99.0 percent with the fewest false alerts of 1.0 percent. To adequately test their suggested technique, the researchers employed a novel dataset entitled UNSW-NB15. It contains current synthetic attack behaviors and reflects actual current typical behaviors [35]. It has 257,673 recordings in total (93,000 normal and 164,673 violent attacks), each of which has 41 attributes as well as a classification value. Backdoors, Fuzzers, examination, standard, DoS, vulnerabilities, reconnaissance, worm, and shellcode are among the ten distinct class names, one standard and nine attacks. By employing only acceptable attributes for class classification throughout the datasets, the usage of mixed rule-based feature collection increased the consistency of the current proposal. The future investigation would include the usage of real-world data collected by the IIoT scheme to establish the efficiency of its own operations in various contexts. Furthermore, the suggested model will be modified in future work to accept alternative protocols.

1.9.2 Feature Selection Literature Review

When working with a large dataset, feature selection plays a crucial role in filtering out unnecessary or duplicate features. It is a preprocessing procedure for vast amounts of data that can be used to select a subset of features or a collection of attributes, which aids in the construction of an effective model for characterization of the selected subset. It also has several additional goals, such as reducing dimensionality, decreasing the size of the data needed for the learning process, improving predictive accuracy, and expanding the developed models. This section examines the literature pertaining to feature selection in concentration with intrusion detection.

By combining the Particle Swarm Optimization (PSO) approach with feature selection (FS), Nilesh Kunhare et al. [36] (2020) discussed improving the accuracy and detection rate of intrusion detection systems (IDSs). In an effort to eliminate extraneous and noisy features that would have a detrimental effect on the system's performance, researchers only employed 10 features from the NSL-KDD dataset. The basic goal of the FS approach is to simplify the dataset by lowering its dimensionality and identifying the best subset of characteristics. The best 10 features out of 41 were chosen using the RF algorithm. To obtain the optimum output, the PSO algorithm was employed with a separate number of iterations and fixed particles for the 10 features that were chosen. The quantity of particles was set at 2800, and the iteration count varied from 20 to 28. At 28 iterations and 2800 particles, the greatest accuracy and detection rate were shown. On the dataset's training and test sets, the results were contrasted with those obtained by the SVM, DT, NB, LR, and k-NN algorithms. The observed performance metrics, including accuracy, were superior to those produced by other algorithms. This approach was also compared to existing methods that use the FS method with the identical dataset; the results demonstrated that this algorithm outperforms other algorithms in terms of accuracy with only 10 features.

Abdullah Alharbi et al. [37] (2021) suggested a Local-Global Best Bat Algorithm for Neural Networks (LGBA-NN) to choose feature subsets and hyperparameters for effective BotNet attack detection, derived from 9 commercial IoT devices affected by two BotNets: Gafgyt and Mirai. The local-global best-based inertia weighting was used by the proposed Bat Algorithm (BA) to update the velocity of the bat in the swarm. Researchers suggested using a Gaussian distribution for population initialization to deal with BA swarm diversity. In order to improve exploration throughout each generation, the local-global best function and Gaussian density function were added after the local search process. Enhanced BA was also used to identify ten different BotNet attacks, including an extra benign target class using neural network hyperparameter adjustment and weight optimization. The proposed LGBA-NN method was evaluated on an N-BaIoT dataset that included a large amount of real-world traffic with both malicious and benign class labels. The effectiveness of LGBA-NN was evaluated against a number of contemporary sophisticated techniques, including BA-NN and PSO-NN for weight optimization. The test results indicated LGBA-NN's superiority over other variants, which had a 90% accuracy rate. Further studies are aimed at expanding the use of the bat method for neural network optimization to alternative evolutionary models, including differential evolution, particle swarm optimization, and genetic.

By enhancing the functionality of the Gorilla Troops Optimizer (GTO) with the algorithm for bird swarms (BSA), also known as the GTO-BSA method, Saif S. Kareem et al. [38] (2022) attempted to provide a novel feature selection (FS) strategy. With the aid of BSA, which is effective at locating the regions that yield the optimal solution, GTO's effectiveness was greatly improved. The performance of the proposed GTO-BSA approach was evaluated on four IoT-IDS datasets: NSL-KDD, CICIDS-2017, UNSW-NB15, and BoT-IoT, and was compared with that of other outstanding algorithms. The study revealed that, when compared to a variety of existing metaheuristic techniques, including the conventional GTO and BSA, MVO, HGS, PSO, and HHO, the suggested GTO-BSA technique provided better results. In the NSL-KDD, CICID2017, UNSW-NB, and BoT-IoT datasets, respectively, it attained accuracy of 95.5%, 98.7%, 81.5%, and 81.5%. In the future, researchers suggested looking at how well the GTO-BSA technique works for a variety of problems, such as ML hyperparameter optimization, solving different kinds of multi-objective problems, and multilevel threshold segmentation.

Maria Habib et al. [39] (2020) introduced a novel method for transforming conventional IDSs into intelligent, adaptive, and multi-objective IDSs for IoT networks. Researchers also included a redesigned technique for IDSs that addresses the issue of feature selection. The foundation of the updated approach is the integration of MOPSO-Lévy, which stands for multi-objective particle swarm optimization with a Lévy flight randomization component. The updated MOPSO-Lévy has been tested on actual IoT network data taken from the UCI repository. When compared to cutting-edge evolutionary multi-objective algorithms, MOPSO-Lévy has shown better performance outcomes. Although research is a continual process, some study areas could still be expanded, according to researchers. Exploring other statistical distributions for the mutation operator that might have significant effects on the search process is one of the areas that needs further study. The integration of novel scalarizing techniques for the transformation functions, such as polynomial and exponential functions, is another possibility. Furthermore, with the resurgent concept of green computing, one of the limitations of wrapper-based techniques over filter-based approaches is the high computational cost. It would be fascinating to incorporate creative strategies to direct the design of wrapper-based FS methods toward Green FS methods.

Maha M. Althobaiti et al. [40] (2021) suggested a model that included preprocessing to eliminate noise from the data. The described model then adopted a feature selection strategy based on binary bacterial foraging optimization (BBFO) to pick an optimal subset of features. In addition, the gated recurrent unit (GRU) model was used to detect intrusions in the industrial cyber-physical systems (CPS) context. Finally, the Nesterov-accelerated Adaptive Moment Estimation (NADAM) optimizer was used to improve the detection rate by optimizing the hyperparameters of the GRU model. A series of tests were conducted utilizing data from industrial CPS to validate the performance of the BBFO-GRU model, and the results revealed the positive performance of the proposed method with an accuracy of 98.45%. Researchers suggested using lightweight cryptographic approaches in future studies to improve the security level of industrial CPS. Furthermore, anomaly detection methods could be used to remove undesired data from the dataset. Additionally, feature reduction algorithms could be created to enhance the entire classification process.

A health-intelligent intrusion detection system (HIIDS) based on machine learning and meta-heuristic methods was presented by Sohail Saif et al. [41] (2022). It was created and put into service for use in an IoT-based healthcare system. The hybrid framework has been used to combine machine learning algorithms like KNN and Decision Tree with meta heuristic algorithms like PSO, GA, and DE. The goal of this hybrid strategy was to achieve optimum accuracy with the fewest possible features. This method of reducing the number of features meant that there would be less time and storage needed, making it appropriate for time-sensitive applications like healthcare. According to research observations, the hybrid approach's GA-DT variation outperformed comparable state-of-the-art studies in regards to classification accuracy, reaching 99.88%, 86.40%, 95.39%, 96.90%, and 100% accuracy for the DoS, U2R, R2L, Probe, and Normal classes with the aid of 8–10 features. Using the GA-DT hybrid version technique, which produced the maximum classification accuracy, researchers looked into the best attributes for each class. The NSL-KDD training dataset was used to create a decision tree-based machine learning model. Additionally, experiments have been run to track execution time, CPU utilization, and memory consumption in order to assess the model. The results of the experiments showed that this model would accurately and quickly distinguish between attack and legitimate traffic. It was appropriate for implementation in resource-constrained IoT-based smart healthcare systems due to its comparatively low CPU and memory usage.

In order to locate attackers in the Iot setting, Mythili Henry Boopathi [42] (2022) developed an efficient intrusion detection technique using the suggested Competitive Swarm Henry Optimization (CSHO)-based Deep Maxout network. By incorporating the characteristics from both optimization methods into the classifier's training, the implemented approach successfully recognizes attacks by determining the most suitable overall solution with the best weight value for each iteration. The information is routed to the base station (BS) through the routing process, which is carried out using the Fractional Gravitational Grey Wolf Optimization (FGGWO) algorithm, and is accessible to the network's distributed nodes. When data is sent to BS, it goes through a phase called "feature selection". During this phase, the Tversky index is used to choose a subset of features. Eventually, the Deep Maxout network, which was trained using the newly created CSHO algorithm, achieves intrusion detection. By taking into account a no-attack situation, the implemented technique attained effective performance of 0.1610 J, 0.9001, 0.9052, and 0.8993 in terms of energy, F-measure, precision, and recall. In order to improve the effectiveness of detection, the researcher stated that hybrid feature selection methods would be taken into consideration in future research.

Raisa Abedin Disha and Sajjad Waheed [43] (2022) suggested a strategy that utilized multiple classification models for the intrusion detection system due to the efficiency of machine learning (ML) techniques. The UNSW-NB15 and Network TON IoT datasets were utilized for offline analysis to assess the performance of the models. When compared to the NSL-KDD dataset, both datasets are relatively more recent and better reflect modern attacks. Still, for the binary classification problem, the performance study was done by training and evaluating the Gated Recurrent Unit (GRU), Long-Short Term Memory (LSTM), Gradient Boosting Tree (GBT), Multilayer Perceptron (MLP), AdaBoost, and Decision Tree (DT). An ideal set of features was chosen using a Gini Impurity-based Weighted Random Forest (GIWRF) model as the integrated feature selection technique because the performance of IDS suffers with a high-dimensional feature vector. In order to help the learning algorithm comprehend the class distribution, this technique used the Gini impurity as the splitting criterion of trees and changed the weights for two different classes of the unbalanced data. Twenty features from UNSW-NB15 and ten features from the Network TON IoT dataset were chosen based on significance scores. The results of this study demonstrated that DT outperformed the other trained models in this study when using the feature selection strategy. Furthermore, the proposed GIWRF-DT performed better in terms of the F1 score than other current approaches examined in the literature. Unfortunately, multiclass classification and computational time analysis were not performed in this work; as a result, future work could implement a multiclass classification strategy for IDS that takes time complexity analysis into account.

Mohammed Otair et al. [44] (2022) proposed a method for identifying intrusions that combines Particle Swarm Optimization (PSO) and Grey Wolf Optimization (GWO) to tackle feature selection issues and apply the optimal value to update each grey wolf position's information. This method avoids the GWO algorithm entering a local optimum by maintaining the individual's best location information through the PSO algorithm. The NSL KDD dataset was utilized to assess the effectiveness of the suggested method. The accuracy, detection rate, false alarm rate, number of features, and execution time of the classification were evaluated using the k-means and SVM algorithms. The outcomes have demonstrated that, when utilizing K-means or SVM algorithms, the proposed technique achieved the essential enhancement of the GWO algorithm. The researchers recommended switching from K-means to the Bagging (Bootstrap Aggregating) method as a classifier in the future. It is one of the ensemble learning techniques that can raise the detection rate in the context of Wireless Sensor Networks (WSN), particularly IDS, by increasing the accuracy of regression and classification. The same issue might be resolved by different optimization procedures, such as the Arithmetic Optimization Algorithm (AOA). The suggested approach could be used to address several other optimization issues, including those related to data mining, task scheduling, wind energy, industrial engineering, benchmark functions, feature selection, image segmentation, and other challenges.

In their research, Aniss Chohra et al. [45] (2022) developed a novel approach that concentrated on determining the appropriate hyperparameters for ensemble methods in order to identify the essential features from a given communication dataset. The suggested method was tested on three benchmark datasets: NSL-KDD, UNSW-NB15, and IoT-Zeek, which contains connection logs from the Zeek network-based intrusion detection system. They generated the IoT-Zeek dataset by applying ensemble classification and deep learning methods to the Zeek connectivity logs of IoT devices, utilizing publicly available harmful and harmless threat intelligence. A swarm intelligence optimization method (PSO) was combined with ensemble methods to create the proposed approach. In their tests, they found that the suggested algorithm did a better job of finding the best solutions when it was adjusted using boosting (XGBoost) techniques instead of bagging (Random Forest). Additionally, they chose the proper collection of features on every validation dataset using the best solutions identified by the optimization algorithm. They developed and optimized an anomaly detection autoenocoder for each of these datasets using solely those features. According to the evaluation results, their anomaly detection algorithms surpassed the most effective state-of-the-art techniques used on these datasets. It also decreased and achieved acceptable training time delays. Their work did have some shortcomings, though, and those will need to be fixed in the future. The first one is that when utilizing Random Forests, they only utilized two hyperparameters for the optimization process (tree count and test size), while when using XGBoost, they used three (number of trees, test size, and learning rate). The researchers said they are looking at the prospect of adding (optimizing) more hyperparameters and that the feature selection (optimization) algorithm has to be more scalable (execution times). The latter, however, is not a problem because it only needs to be run once for each dataset and not on a regular basis. Additionally, they had not looked into how to set the PSO hyperparameters (c1, c2, and w) in a flexible way, which could have increased the effectiveness of the search. This involves using PSO variations like Adaptive Particle Swarm Optimization (APSO) to find the best settings for these three hyperparameters.

1.10. Motivation

In accordance with [46], developments in the Industrial Internet of Things (IIoT), networking, the continued growth of smart ecosystems, and the urgent demand for process automation have not only created new opportunities but also widened the terrain of cyber-threats and attacks. This has mostly been observed even as industries rush to conform to the objectives of the highly anticipated industry 4.0. In particular, it is now clear that the fusion and convergence of industrial operational technology (OT) and information technology (IT) leads to intelligence, complexity in the IIoT ecosystem, changes in society, and general changes in security and digital forensic investigation designs. As a result, as many industries try to reach their industry 4.0 goals, the digital forensic side of IIoT isn't really taken into account or integrated. This is because the cybersecurity threat landscape in this context is also getting more complicated because the attack surface is getting bigger and new system behaviors are becoming more different. According to the previously discussed literature, ML learning techniques have shown good prediction accuracy in identifying data from network traffic as safe or malicious. Devices net flow anomaly detection using a ML technique is a promising subject due to its power of understanding complex traffic on the network and detecting anomalies, even though there are not many solutions used today for larger IIoT contexts, including corporate networks. When deploying anomaly detection models in actual IIoT operational environments, the challenges for ML-based BotNet attack detection include scaling to larger sizes of network traffic; reducing computational power consumption; encompassing different attack types; classifying traffic aggregation as benign or malicious; and gaining high accuracy of ML model results by easing the fitted model to avoid additional investments. In contrast to other studies, the main target of this thesis is to use feature selection algorithms to obtain high performance of standard ML models with less resources consumption.

1.11. Objective of the Thesis

The most significant goal of this thesis is to improve the detection of BotNet attacks in the IIoT environment, which will ultimately result in enhancing the security of IIoT devices. The following will be the study's primary objectives in brief:

- To explore the IIoT BotNet attack detection as a multiclass classification problem utilizing the IIoT datasets.
- To assess and utilize the new X-IIoTID dataset [47], which is a specialized dataset in the IIoT cyberattacks field that has not been utilized previously (according to the best of our knowledge).
- To implement the new Meta Heuristic algorithm, Eurasian Oystercatcher Optimizer [48], as a feature selection algorithm.
- To apply the Eurasian Oystercatcher Optimizer algorithm as a features selection approach. This will beneficial in getting high system accuracy (less false data means improved system accuracy), decrease training time and overall system resources (because algorithms will train quicker with fewer resources since they are working with less data), and increase overfitting robustness (dealing with less redundant data means less opportunity to make decisions based on noisy data).
- To utilize a machine learning algorithm as a fitness for the Oystercatcher Optimizer algorithm. This guarantees that the Oystercatcher Optimizer algorithm's output features will all be trained. Because the Oystercatcher Optimizer algorithm will actually train on all the chosen features based on its fitness, this strengthens the algorithm.
- To evaluate the new dataset having the reduced features using state of the art ML algorithms.
- To compare the results with the literatures.

1.12. Thesis Outline

The thesis consisting of five chapters (including this chapter), Chapter Two provides >>>>>>>>>>>>. Chapter Three explains the proposed system and all of its details including system requirements and tools that are used to implement this system. Chapter Four contains the simulation of the proposed system and the discussion and analysis of the results that are obtained through out this work. Finally, Chapter Five includes the most important work conclusions and recommendations for future work.

REFERENCES

[1] I.M. El-Hasnony, R.R. Mostafa, M. Elhoseny, and S.I. Barakat, “Leveraging Mist and Fog for Big Data Analytics in IoT Environment,” Trans. Emerg. Telecommun. Technol. 2021, 32, e4057.

[2] Lee, “Internet of Things (IoT) Cybersecurity: Literature Review and Iot Cyber Risk Management,” Future Internet 2020, 12, 157.

[3] G.S. Kushwah, and V. Ranga, “Voting Extreme Learning Machine Based Distributed Denial of Service Attack Detection in Cloud Computing,” J. Inf. Secur. Appl. 2020, 53, 102532.

[4] Dhirani, L., and Newe, T., “Hybrid Cloud SLAs for Industry 4.0: Bridging the gap,” Ann. Emerg. Technol. Comput. 2020, 4, 41–60.

[5] M. Al-Hawawreh, E. Sitnikova, and F. den Hartog, “An efficient intrusion detection model for edge system in brownfield industrial internet of things,” in Proceedings of the 3rd International Conference on Big Data and Internet of Things. Melbourne, Australia: ACM, 2019, pp. 83–87.

[6] M. Al-Hawawreh, F. den Hartog, and E. Sitnikova, “Targeted ransomware: A new cyber threat to edge system of brownfield industrial internet of things,” IEEE Internet of Things Journal, vol. 6, no. 4, pp. 7137–7151, 2019.

[7] K. Tange, M. De Donno, X. Fafoutis, and N. Dragoni, “A systematic survey of industrial internet of things security: Requirements and fog computing opportunities,” IEEE Communications Surveys & Tutorials, 2020.

[8] Varshney, T., Sharma, N., Kaushik, I., and Bhushan, B., “Authentication & Encryption Based Security Services in Blockchain Technology,” International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), 2019. doi: 10.1109/icccis48478.2019.8974500.

[9] S. Zhao, S. Li, and Y. Yao, “Blockchain enabled industrial internet of things technology,” IEEE Transactions on Computational Social Systems, 2019.

[10] J. Al-Jaroodi and N. Mohamed, “Blockchain in industries: A survey,” IEEE Access, vol. 7, pp. 36 500–36 515, 2019.

[11] T. M. Fernández-Caramés and P. Fraga-Lamas, “A review on the application of blockchain to the next generation of cyber secure industry 4.0 smart factories,” IEEE Access, vol. 7, pp. 45 201–45 218, 2019.

[12] Tiwari, R., Sharma, N., Kaushik, I., Tiwari, A., and Bhushan, B., “Evolution of IoT & Data Analytics using Deep Learning,” International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), 2019. doi: 10.1109/icccis48478.2019.8974481.

[13] Wang Ren, Xin Tong, Jing Du, Na Wang, Shan Cang Li, Geyong Min, Zhiwei Zhao, Ali Kashif Bashir, “Privacy-preserving using homomorphic encryption in Mobile IoT systems,” Computer Communications, Volume 165, 2021, Pages 105-111, ISSN 0140-3664, https://doi.org/10.1016/j.comcom.2020.10.022.

[14] M. M. Ogonji, G. Okeyo, and J. M. Wafula, “A survey on privacy and security of Internet of Things,” Computer Science Review, vol. 38, article 100312, 2020.

[15] G. De La Torre Parra, P. Rad, K.-K. R. Choo, and N. Beebe, “Detecting Internet of Things attacks using distributed deep learning,” Journal of Network and Computer Applications, vol. 163, article 102662, 2020.

[16] I. Ali, A. I. A. Ahmed, A. Almogren, M. A. Raza, S. A. Shah, A. Khan, and A. Gani, “Systematic Literature Review on IoT-Based Botnet Attack,” IEEE Access, vol. 8, pp. 212220-212232, 2020, doi: 10.1109/ACCESS.2020.3039985.

[17] Umar Iftikhar, Kashif Asrar, Maria Waqas and Syed Abbas Ali, “BOTNETs: A Network Security Issue,” International Journal of Advanced Computer Science and Applications (IJACSA), 11(11), 2020. http://dx.doi.org/10.14569/IJACSA.2020.0111155.

[18] R. Kour, “Cybersecurity issues and challenges in Industry 4.0,” in Applications and Challenges of Maintenance and Safety Engineering in Industry 4.0, pp. 84–101, IGI Global, 2020.

[19] J. Prinsloo, S. Sinha, and B. von Solms, “A review of Industry 4.0 manufacturing process security risks,” Applied Sciences, vol. 9, no. 23, p. 5105, 2019.

[20] Majda Wazzan, Daniyal Algazzawi, Omaima Bamasaq, Aiiad Albeshri, and Li Cheng, “Internet of Things Botnet Detection Approaches: Analysis and Recommendations for Future Research," Applied Sciences 11, no. 12: 5713, 2021. https://doi.org/10.3390/app11125713

[21] L. Gupta, T. Salman, A. Ghubaish, D. Unal, A. K. Al-Ali, and R. Jain, “Cybersecurity of multi-cloud healthcare systems: A hierarchical deep learning approach,” Applied Soft Computing, vol. 118, p. 108439, 2022.

[22] Mohanad Sarhan, Siamak Layeghy, and Marius Portmann, “Feature Analysis for ML-based IIoT Intrusion Detection,” ArXiv, vol. abs/2108.12732, 2021.

[23] Mohammed Mudassir, Devrim Unal, Mohammad Hammoudeh, Farag Azzedin, “Detection of Botnet Attacks against Industrial IoT Systems by Multilayer Deep Learning Approaches,” Wireless Communications and Mobile Computing, vol. 2022, Article ID 2845446, 12 pages, 2022. https://doi.org/10.1155/2022/2845446.

[24] 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.

[25] P.L.S. Jayalaxmi, Gulshan Kumar, Rahul Saha, Mauro Conti, Tai-hoon Kim, Reji Thomas, “DeBot: A deep learning-based model for bot detection in industrial internet-of-things,” Computers and Electrical Engineering, Volume 102, 2022, 108214, ISSN 0045-7906, https://doi.org/10.1016/j.compeleceng.2022.108214.

[26] Sarhan Mohanad, Siamak Layeghy, Nour Moustafa, and Marius Portmann, “Netflow datasets for machine learning-based network intrusion detection systems,” Big data technologies and applications. Springer; 2020, p. 117–35.

[27] Portmann Marius, “Network flow data set,” Machine Learning-Based NIDS Datasets, 2020, http://staff.itee.uq.edu.au/marius/NIDS_datasets/.

[28] Nour Moustafa, “A new distributed architecture for evaluating AI-based security systems at the edge: Network TON_IoT datasets,” Sustainable Cities and Society, Volume 72, 2021, 102994, ISSN 2210-6707, https://doi.org/10.1016/j.scs.2021.102994.

[29] Tooba Hasan, Jahanzaib Malik, Iram Bibi, Wali Ullah Khan, Fahd N. Al-Wesabi, Kapal Dev, and Gaojian Huang, “Securing Industrial Internet of Things Against Botnet Attacks Using Hybrid Deep Learning Approach,” IEEE Transactions on Network Science and Engineering, VOL. 09, 2022, doi: 10.1109/TNSE.2022.3168533.

[30] Y. Meidan, M. Bohadana, Y. Mathov, Y. Mirsky, A. Shabtai, D. Breitenbacher, and Y. Elovici, “N-BaIoT-Network-based detection of IoT botnet attacks using deep autoencoders,” IEEE Pervasive Computing, vol. 17, no. 3, pp. 12–22, 2018.

[31] S. Latif, Z. Zou, Z. Idrees, and J. Ahmad, "A Novel Attack Detection Scheme for the Industrial Internet of Things Using a Lightweight Random Neural Network," in IEEE Access, vol. 8, pp. 89337-89350, 2020, doi: 10.1109/ACCESS.2020.2994079.

[32] M.-O. Pahl and F.-X. Aubet. (2018). Ds2Os Traffic Traces IoT Traffic Traces Gathered in a The Ds2Os IoT Environment. [Online]. Available: https://www.kaggle.com/francoisxa/ds2ostraf_ctraces.

[33] M.-O. Pahl and F.-X. Aubet, "All eyes on you: Distributed multidimensional IoT microservice anomaly detection," in Proc. 14th Int. Conf. Netw. Service Manage. (CNSM), Nov. 2018, pp. 72_80.

[34] Joseph Bamidele Awotunde, Chinmay Chakraborty, and Abidemi Emmanuel Adeniyi, "Intrusion Detection in Industrial Internet of Things Network-Based on Deep Learning Model with Rule-Based Feature Selection," Wireless Communications and Mobile Computing, vol. 2021, Article ID 7154587, 17 pages, 2021. https://doi.org/10.1155/2021/7154587.

[35] N. Moustafa and J. Slay, "UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)," in 2015 Military Communications and Information Systems Conference (MilCIS), pp. 1–6, Canberra, ACT, Australia, November 2015.

[36] Nilesh Kunhare, Ritu Tiwari, and Joydip Dhar, "Particle swarm optimization and feature selection for intrusion detection system," Sādhanā 45, 109 (2020). https://doi.org/10.1007/s12046-020-1308-5

[37] Abdullah Alharbi, Wael Alosaimi, Hashem Alyami, Hafiz Tayyab Rauf, and Robertas Damaševicius, "Botnet Attack Detection Using Local Global Best Bat Algorithm for Industrial Internet of Things," Electronics 10, no. 11: 1341., 2021, https://doi.org/10.3390/electronics10111341.

[38] Saif S. Kareem, Reham R. Mostafa, Fatma A. Hashim, and Hazem M. El-Bakry, "An Effective Feature Selection Model Using Hybrid Metaheuristic Algorithms for IoT Intrusion Detection," Sensors 22, no. 4: 1396., 2022, https://doi.org/10.3390/s22041396.

[39] Maria Habib, Ibrahim Aljarah, and Hossam Faris, "A Modified Multi-objective Particle Swarm Optimizer-Based Lévy Flight: An Approach toward Intrusion Detection in Internet of Things," Arabian Journal for Science and Engineering, 45:6081–6108, 2020, https://doi.org/10.1007/s13369-020-04476-9.

[40] Maha M. Althobaiti, K. Pradeep Mohan Kumar, Deepak Gupta, Sachin Kumar, and Romany F. Mansour, "An intelligent cognitive computing based intrusion detection for industrial cyber-physical systems," Measurement, Volume 186, 2021, 110145, ISSN 0263-2241, https://doi.org/10.1016/j.measurement.2021.110145.

[41] Sohail Saif, Priya Das, Suparna Biswas, Manju Khari, and Vimal Shanmuganathan, "HIIDS: Hybrid intelligent intrusion detection system empowered with machine learning and metaheuristic algorithms for application in IoT based healthcare," Microprocessors and Microsystems, 2022, 104622, ISSN 0141-9331, https://doi.org/10.1016/j.micpro.2022.104622.

[42] Mythili Henry Boopathi, "Henry MaxNet: tversky index based feature selection and competitive swarm henry gas solubility optimization integrated Deep Maxout network for intrusion detection in IoT," International Journal of Intelligent Robotics and Applications, 6, 365–383, 2022. https://doi.org/10.1007/s41315-022-00234-2.

[43] Raisa Abedin Disha, and Sajjad Waheed, “Performance analysis of machine learning models for intrusion detection system using Gini Impurity-based Weighted Random Forest (GIWRF) feature selection technique,” Cybersecurity 5:1, 2022. https://doi.org/10.1186/s42400-021-00103-8.

[44] Mohammed Otair, Osama Talab Ibrahim, Laith Abualigah, Maryam Altalhi, and Putra Sumari, “An enhanced Grey Wolf Optimizer based Particle Swarm Optimizer for intrusion detection system in wireless sensor networks,” Wireless Networks, 28, 721–744, 2022. https://doi.org/10.1007/s11276-021-02866-x.

[45] Aniss Chohra, Paria Shirani, ElMouatez Billah Karbab, and Mourad Debbabi, “Chameleon: Optimized feature selection using particle swarm optimization and ensemble methods for network anomaly detection,” Computers & Security, Volume 117, 2022, 102684, ISSN 0167-4048, https://doi.org/10.1016/j.cose.2022.102684.

[46] Victor R. Kebande, “Industrial internet of things (IIoT) forensics: The forgotten concept in the race towards industry 4.0,” Forensic Science International: Reports, Volume 5, 2022, 100257, ISSN 2665-9107, https://doi.org/10.1016/j.fsir.2022.100257.

[47] Muna Al-Hawawreh, Elena Sitnikova, Neda Aboutorab, July 30, 2021, "X-IIoTID: A Connectivity- and Device-agnostic Intrusion Dataset for Industrial Internet of Things", IEEE Dataport, 2021, Available at: https://dx.doi.org/10.21227/mpb6-py55.

[48] Ahmad Salim, Wisam K. Jummar, Farah Maath Jasim, and Mohammed Yousif, “Eurasian oystercatcher optimizer: New meta-heuristic algorithm,” Journal of Intelligent Systems, vol. 31, no. 1, 2022, pp. 332-344. https://doi.org/10.1515/jisys-2022-0017.