International Journal of Innovative Engineering Applications

Journal homepage: https://dergipark.org.tr/ijiea

DECISION TREE BASED INTRUSION DETECTION METHOD IN THE INTERNET OF THINGS

***Rojbin Tekin*[1], *Orhan Yaman*[*1], *Turker Tuncer*[1]**

[1]*Firat University, Technology Faculty, Department of Digital Forensics Engineering, Elazig, Turkey*

Abstract

Original scientific paper

Developments in computer and network technologies have also positively affected internet technology. With the development of the Internet, the concept of IoT (Internet of Things) has been invented. Nowadays, IoT devices provide convenience in many areas, and the positive effects of IoT-based systems increase people's quality of life. People want to remotely monitor and manage smart cities, smart homes, and other platforms. However, IoT systems have many vulnerabilities and thus have become the target of attackers. Detecting such attacks and preventing security vulnerabilities will further increase the rate of use of IoT technology. In this work, an intelligent intrusion detection system (IDS) for IoT devices has been suggested. The presented intelligent IDS for IoT devices have been developed on a big attack dataset and this dataset contains 3,668,443 observations. In prior works which used this dataset, researchers worked on a binary classification problem (attacked and normal). However, this research aims to classify the attack types, hence, nine categories have been used. To propose a prompt responded IDS model, a fast classifier which is a decision tree (DT) has been employed. Our proposal attained 97.43% classification accuracy on this dataset using 10-fold cross-validation. This accuracy rate frankly demonstrates the classification ability of our proposed IDS model for IoT devices.

Keywords: *DDoS, decision tree, DoS, internet of things, intrusion detection.*

NESNELERİN İNTERNETİNDE KARAR AĞACI TABANLI SALDIRI TESPİT YÖNTEMİ

Özet

Orijinal bilimsel makale

Bilgisayar ve ağ teknolojilerindeki gelişmeler internet teknolojisini de olumlu yönde etkilemiştir. İnternetin gelişmesiyle birlikte IoT (Nesnelerin İnterneti) kavramı ortaya çıkmıştır. Günümüzde IoT cihazları birçok alanda kolaylık sağlamakta ve IoT tabanlı sistemlerin olumlu etkileri insanların yaşam kalitesini artırmaktadır. İnsanlar akıllı şehirleri, akıllı evleri ve diğer platformları uzaktan izlemek ve yönetmek istemektedir. Ancak IoT sistemleri birçok güvenlik açığına sahiptir ve bu nedenle saldırganların hedefi haline gelmiştir. Bu tür saldırıları tespit etmek ve güvenlik açıklarını önlemek, IoT teknolojisinin kullanım oranını daha da artıracaktır. Bu çalışmada, IoT cihazları için akıllı bir saldırı tespit sistemi (IDS) önerilmiştir. IoT cihazları için sunulan akıllı IDS, büyük bir saldırı veri seti üzerinde geliştirildi ve bu veri seti 3.668.443 örnek içermektedir. Bu veri setini kullanan önceki çalışmalarda, araştırmacılar ikili sınıflandırma problemi (Atak ve Normal) üzerinde çalışmışlardır. Ancak bu çalışmada saldırı türlerini sınıflandırmayı amaçladığından dokuz kategori kullanılmıştır. Hızlı yanıt veren bir IDS modeli önermek için karar ağacı (DT) olan hızlı bir sınıflandırıcı kullanılmıştır. Önerimiz, 10 kat çapraz doğrulama kullanarak bu veri setinde %97,43 sınıflandırma doğruluğu elde edilmiştir. Bu doğruluk oranı, IoT cihazları için önerilen IDS modelimizin sınıflandırma yeteneğini açıkça göstermektedir.

Anahtar Kelimeler: *DDoS, DoS, Karar ağacı, Nesnelerin İnterneti, Saldırı tespiti.*

1 Introduction

The development of the Internet contributes to us in almost every aspect of our lives. (Television, dishwasher, smart home systems, vehicles, cameras, etc.). This increase in the number of devices connected to the Internet has led to the emergence of the concept of IoT [1,2]. IoT connects physical objects and integrates both physical and digital objects to improve our daily tasks. However, today's developing internet and devices connected to the internet have become the target point of attackers. Attacks on IoT devices show that personal data is in danger [3,4]. IoT systems; consist of object components, data components, network components, cloud components, and analysis components. There are different security vulnerabilities for each component. Preventing these security vulnerabilities will contribute to the further development of IoT technology. Today, IoT technology has been widely used in many areas such as buildings and homes, industry, the health sector, transportation, and

* Corresponding author.

E-mail address: orhanyaman@firat.edu.tr (O. Yaman)

Received 12 July 2021; Received in revised form 07 February 2022; Accepted 14 March 2022

2587-1943 | © 2022 IJIEA. All rights reserved.

Doi: https://doi.org/10.46460/ijiea.970383

agriculture. While IoT technology increases the quality of life of people in social areas, it increases the quality of production in the field of manufacturing such as industry. IoT application areas and distribution according to 2020 data are shown in Figure 1.

As seen in Figure 1, IoT is mostly used in manufacturing/industry. Transportation, energy, and other fields follow it. Attacks that may occur in applications that require precision, such as manufacturing, cause production to stop, decrease in quality, and material damage. Any attack on IoT platforms used in the field of transportation may disrupt transportation and accidents. Thus, IoT security has become a hot-topic research issue to prevent attacks and various defense models have been presented in the literature for IoT security. There are many IoT attacks. Some of those; are physical attacks, encryption attacks, DoS attacks, firmware hijacking, botnets, man-in-the-middle attacks, ransomware, and brute force attacks.

The fact that objects are in constant communication and network connections allow cyber-attacks [6] [7]. Although there is not enough work in the field of security in the Internet of Things, the production of IoT devices is increasing. In the coming years, many applications will be developed to eliminate security vulnerabilities in the field of IoT [10]. Okegbile et al. In their studies proposed a model for DDOS (Denial of Service) attacks that occur on IoT devices. This model characterizes the behavior of attackers in the system. The model trust list table further improves the detection of malicious nodes, creating a precaution for future attacks [3]. Deniz examined security vulnerabilities in his study. He proposed a new model for the security of nodes on IoT platforms [8].

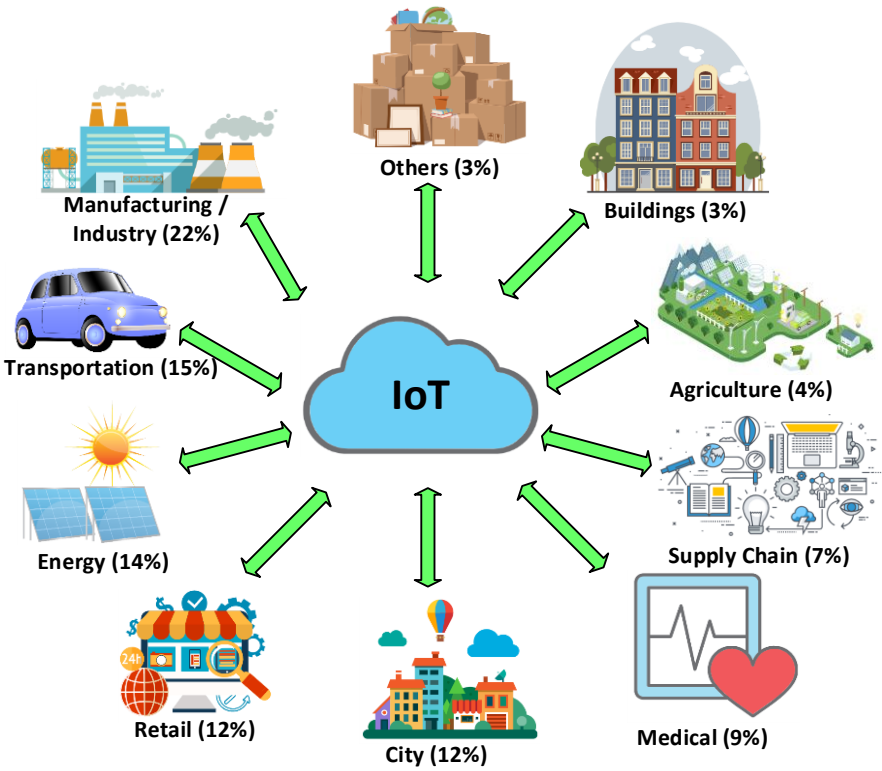

Figure 1. IoT application areas and distribution according to 2020 data [5].

Machine learning and deep learning methods are used to detect attacks on IoT platforms. Before classifying the attacks, preprocessing and feature extraction are performed. Feature extraction is important for the success of classification. Shafiq et al. In this study, a bijective soft set was applied for feature selection and then a new CorrACC feature selection metric approach was proposed. To evaluate the proposed approaches, four different machine learning classifiers were used in the BoT-IoT dataset and over 95% accuracy was calculated with the algorithm [11]. Mohammadi et al. [12] A comprehensive overview of the use of deep learning in the field of IoT is presented. Big data analysis and IoT flow data analysis were performed for IoT data. Emerging deep learning techniques for IoT data analytics are discussed and their challenges are presented [12]. In Yonem's study, artificial neural networks were used for artificial bee colony algorithms and time series. The realized model was used in the analysis of the data. This study showed that the artificial bee colony algorithm can be used in the Internet of Things [13]. Rathore et al. [14] proposed an intrusion detection method to detect IoT attacks. They collected the KDD dataset by creating an IoT platform. They calculated 86.53% accuracy on this dataset with machine learning methods. Xiao et al. [15], examined the types of attacks commonly used in IoT platforms. They explained the types of DoS attackers, Jamming, Spoofing, Man-in-the-middle attack, Software attacks, and Privacy leakage. They presented machine learning-based methods developed to detect these attack types. Kotenko et al. [16], presented a machine learning and big data approach for IoT infrastructure. They have detected attacks on IoT devices. The datasets created on IoT platforms are large. Therefore, they proposed a distributed machine learning-based model. They compared the proposed distributed model with the local model. They showed that the

performance results are higher in the distributed model. Vu et al. [17], proposed a deep transfer learning method with data collected from multiple IoT devices. They performed attack detection by collecting nine different datasets.

Zhang et al. [18], a lightweight defense algorithm is proposed for DDoS attacks over IoT network environments. They have been tested against various scenarios to study the interactive communication between different network nodes. In Yavuz study, a deep learning-based security system is presented. The dataset to be used in deep learning has been prepared with the Cooja simulator. Cooja IoT simulator has been used to generate high-quality attack data on IoT networks ranging up to 1000 nodes. Approximately 99% accuracy was computed with the trained dataset [19]. Koroniotis et al [20] have proposed a new dataset called Bot-IoT, which combines various types of attacks as well as legitimate and simulated IoT network traffic. They evaluated the reliability of the BoT-IoT dataset compared to other datasets by using different statistical and machine learning methods for forensic purposes. The best 10 features were selected from this large dataset. The features are classified using SVM. They compared results using the top 10 features and all features.

In this study, attack types are classified using the Bot-IoT dataset. In the literature, the Bot-IoT dataset has been used in attack detection. (Normal- DDoS HTTP, Normal-DDoS TCP, Normal- DDoS UDP etc.). In the proposed method, the results of "attack", "category" and "subcategory" are combined. Thus, nine classes have been created. DT (Decision Tree) algorithm has been used for classification.

2 Material and Method

The Bot-IoT dataset has been used in this study [20–22]. The Bot-IoT dataset was created in a laboratory environment. Pcap files were collected and edited with the Argus tool. Many features have been obtained with the Bot-IoT dataset. Koroniotis et al. [20] Bot-IoT selected and used the top 10 features on the dataset. In this study, the top 10 features selected on the Bot-IoT dataset have been used. The top 10 features and their descriptions are tabulated in Table 1.

Table 1. Features and descriptions of Bot-IoT dataset [20].

Feature	Description
srate	Source-to-destination packets per second
drate	Destination-to-source packets per second
rate	Total packets per second in transaction
max	Maximum duration of aggregated records
state_number	Numerical representation of feature state
mean	Average duration of aggregated records
min	Minimum duration of aggregated records
stddev	Standard deviation of aggregated records
flgs_number	Numerical representation of feature flags
seq	Argus sequence number

The top 10 features on the Bot-IoT dataset presented in Table 1 have been used. The block diagram of the proposed method can be illustrated in Figure 2.

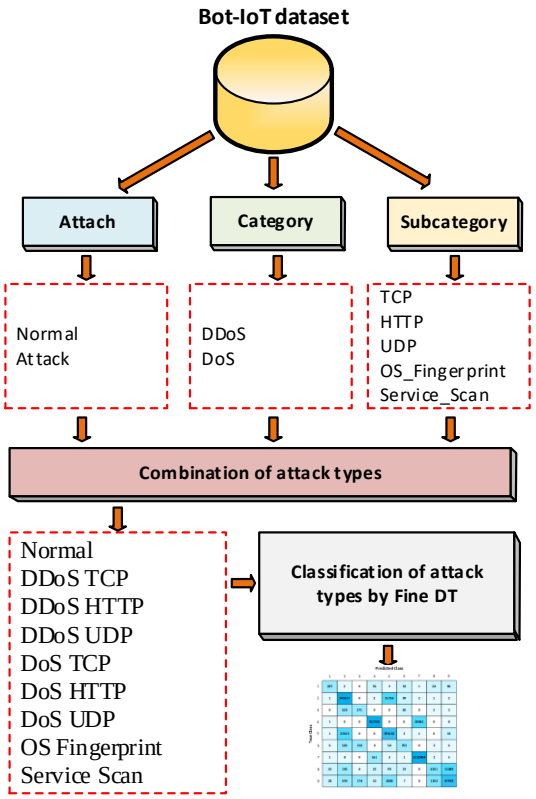

Figure 2. Block diagram of the proposed method for Bot-IoT dataset.

As can be viewed in Figure 2, the top 10 features are classified according to the decision tree. There are "attack", "category" and "subcategory" results on the Bo-IoT dataset. There are normal and attack classes in the "attack" category. In the "category", there are DDoS and DoS classes. In the "subcategory", there are TCP, HTTP, UDP, OS_Fingerprint, and Service_Scan classes. In the literature, the dataset is classified separately according to the results of "attack", "category" and "subcategory". In this study, nine classes were formed and classified by combining the results of "attack", "category", and "subcategory". The class types and sample numbers obtained are demonstrated in Table 2.

Table 2. New classes obtained by combining “attack”, “category” and “subcategory” results.

Number of Class	Class Type	Number of Sample
1	Normal	477
2	DDoS TCP	977380
3	DDoS HTTP	989
4	DDoS UDP	948255
5	DoS TCP	615800
6	DoS HTTP	1485
7	DoS UDP	1032975
8	OS Fingerprint	17914
9	Service Scan	73168

As can be show in Table 2, most examples belonging to the "DoS UDP" class. The fewest examples belong to the “Normal” class. These samples are 3668443 in total. Decision Tree has been used to classify attack types. The large size of the dataset affects the classification process. For this reason, a Decision Tree is preferred rather than

classifications such as KNN or SVM. The Decision Tree algorithm is fast compared to other classifiers [23]. For this reason, the Fine DT algorithm used has been compared with the Medium DT, Coarse DT, Ensemble Boosted Trees (EBT), and Linear Discriminant (LD) algorithms. Comparison results are displayed in Figure 3.

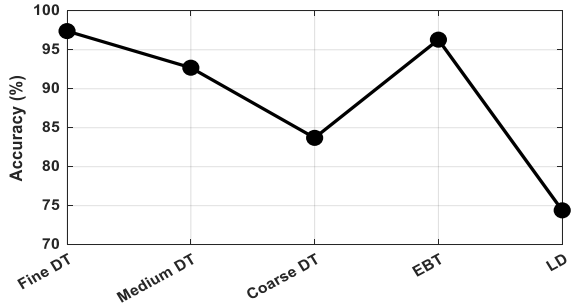

Figure 3. Comparison of Fine DT algorithm with other classifiers.

3 Experimental Results

In this study, Classification Learner Toolbox in MATLAB program has been used to classify the Bot-IoT dataset. The large size of the dataset affects the testing time of the proposed method. For this reason, the Decision Tree algorithm, which is fast and has low computational complexity, is preferred instead of algorithms such as deep learning. In the Fine DT classifier used in this study, the "Split Criterion" parameter was selected as "gdi" and the "Max Number of Splits" parameter as "100". The Confusion Matrix obtained as a result of the classification is shown in Figure 4.

ROC curves and AUC values of all classes can be seen in Figure 5.

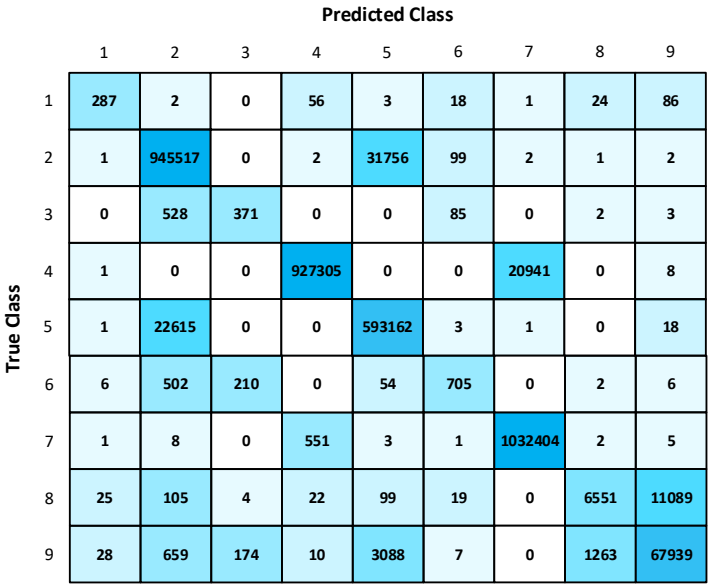

Figure 4. Confusion Matrix results obtained with Fine DT classifier.

As can be seen in Figure 5, the best results have been calculated for the "DDoS TCP", "DDoS UDP", "DoS TCP" and "DoS UDP" classes. Accuracy, Precision, Recall, Geometric Mean, and F1-Score values have been computed by running 100 iterations of the proposed method. Maximum, Minimum, Average, and Standard Deviation values of Accuracy, Precision, Recall, Geometric Mean, and F1-Score values are given in Table 3.

As can be illustrated in Table 3, the best 97.43% accuracy has been calculated with the Fine DT classifier. These results have been obtained using 10 Fold-CV. In addition, in the proposed method, Fold-wise accuracy values can be computed and displayed in Figure 6.

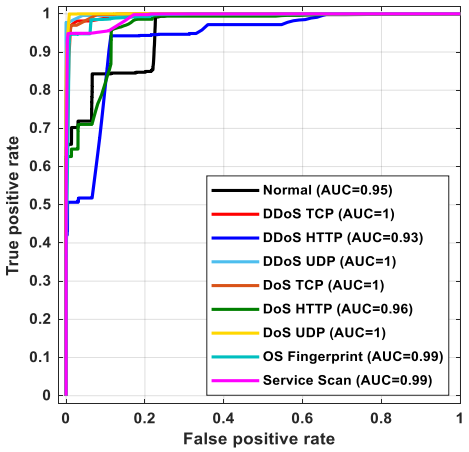

Figure 5. ROC Curve of all class.

Table 3. Performance values of the proposed method with 100 iterations.

	Accuracy	Precision	Recall	Geometric Mean	F1-Score
Max	97.43	85.29	73.99	68.56	79.23
Min	97.41	84.98	73.92	68.48	79.08
Mean	97.42	85.06	73.93	68.49	79.11
Std	0.001	0.08	0.01	0.01	0.04

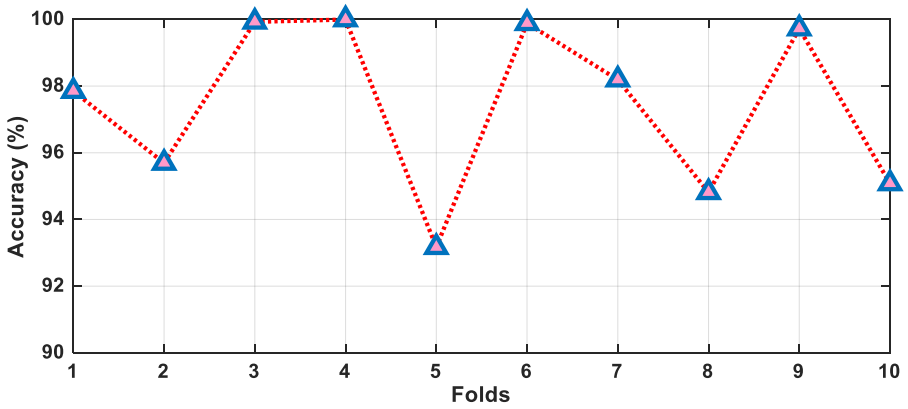

Figure 6. The calculated fold-wise accuracies employing Fine DT classifier.

As can be viewed in Figure 6, the highest results have been computed with Fold3, Fold4, Fold6, and Fold9 (>99.5% accuracy). The lowest result has been calculated at 93.16% with Fold5. Class-wise results of the proposed method can be illustrated in Figure 7.

As can be seen in Figure 7, the highest results have been computed for the "DDoS TCP", "DDoS UDP", "DoS TCP" and "DoS UDP" classes. Accuracy is 96.73% for "DDoS TCP", 97.79% for "DDoS UDP", 96.32% for "DoS TCP" and 99.94% for "DoS UDP". The lowest results have been calculated in the "DDoS HTTP", "DoS HTTP" and "OS Fingerprint" classes. When the sample numbers in the Bot-IoT dataset are compared with the results, it can be demonstrated that the success rate is related to the sample numbers. The accuracy value has been computed high because the number of samples in the "DDoS TCP", "DDoS UDP", "DoS TCP", and "DoS UDP" classes are high.

The proposed Fine DT algorithm has been applied to the Bot-IoT dataset and has achieved high classification accuracy. Our results are compared to other state-of-art methods and comparative results are listed in Table 4.

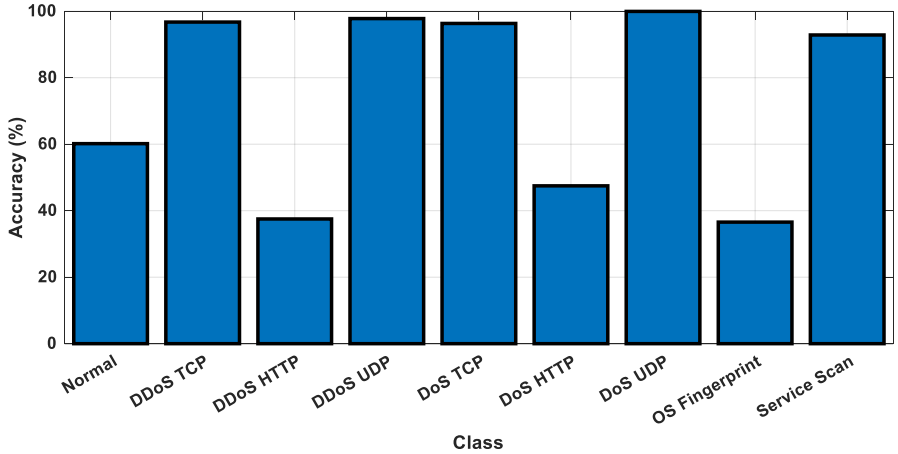

Figure 7. The calculated class-wise accuracies employing Fine DT classifier.

Table 4. Summary of comparison with other state-of-the-art methods using Bot-IoT dataset.

References	Methods	Number of Features	Number of Classes	Results (%)
Leevy et al. [24], 2021	Decision Tree	37	-	AUC=96.16
	Logistic Regression			AUC=97.37
	Naive Bayes			AUC=95.59
	Random Forest			AUC=97.18
Zeeshan et al. [25], 2022	Deep Learning, LSTM	26	3	Acc=96.32
Swarna et al. [26], 2020	KNN	-	-	Acc=92.29
	LSTM	-	-	Acc=97.28
Bhuvaneswari et al. [27], 2020	Deep Learning	10	5	Acc=93.44
Our Method	Fine DT	10	9	Acc=97.43

As can be listed in Table 4, the Bot-IoT dataset is widely used in the literature. This dataset is usually classified into two categories such as "Normal", "Attack" or three categories such as "Normal", "DDoS", "DoS". Leevy et al. [24] used 37 features in the Bot-IoT dataset and calculated 95% accuracy. Zeeshan et al. [25] selected 26 features in the Bot-IoT dataset and computed 96.32% accuracy with deep learning for three categories. Bhuvaneswari et al. [27] computed 93.44% accuracy using deep learning for 10 features and five categories. In the proposed method, the categories and subcategories in the Bot-IoT dataset have been combined and nine classes

have been obtained. For the 10 selected features, both fast and high accuracy have been calculated using the Fine DT algorithm.

The Random Sampling Reduction Technique has been used to show the accuracy of the results of the imbalance Bot-IoT dataset. 477 samples have been randomly selected from each class belonging to the dataset. Thus, 4293 samples have been selected for 9 classes. Randomly selected samples have been classified with Fine DT. As a result of the classification, 97.4% accuracy has been computed and the confusion matrix is displayed in Figure 8.

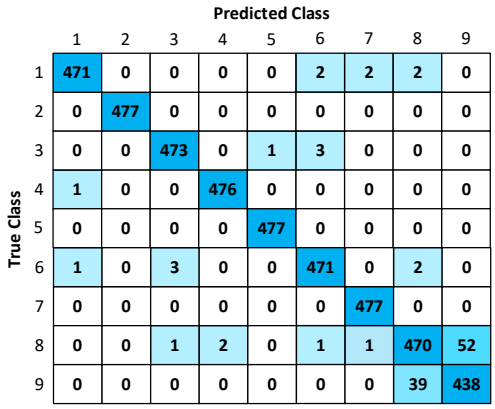

Figure 8. Fine DT classification result obtained from selected samples using Random Sample Reduction Technique.

4 Conclusions

In this study, attack types are classified using the Bot-IoT dataset. In the literature, the Bot-IoT dataset has been evaluated in binary classes (Normal- DDoS HTTP, Normal- DDoS TCP, Normal- DDoS UDP, etc.). SVM, RNN, and LSTM based methods have been proposed to classify these binary classes. Due to the large size of the Bot-IoT dataset, high-spec computers are needed to propose/develop an intelligent IDS model. We aimed to propose a lightweight model to test this dataset using a simple configured personal computer. Therefore, a lightweight method has been developed for the Bot-IoT dataset. In addition, the results have been calculated by creating nine classes instead of the binary classes in the literature. 97.43% accuracy has been computed with the Fine DT classifier. In future studies, the laboratory environment will be developed and new datasets will be collected. Real-time attack detection methods will be developed on these datasets.

Declaration

Ethics committee approval is not required.

References

1. Ertam, F., Kilincer, I. F., Yaman, O., & Sengur, A. (2020, September). A new IoT application for dynamic WiFi based wireless sensor network. In 2020 *International Conference on Electrical Engineering (ICEE)* (pp. 1-4). IEEE.
2. Hasan, M., Islam, M. M., Zarif, M. I. I., & Hashem, M. M. A. (2019). Attack and anomaly detection in IoT sensors in IoT sites using machine learning approaches. *Internet of Things*, *7*, 100059.
3. Okegbile, S. D., & Ogunranti, O. I. (2020). Users emulation attack management in the massive internet of things enabled environment. *ICT Express*, 6(4), 353-356.
4. Ashraf, J., Keshk, M., Moustafa, N., Abdel-Basset, M., Khurshid, H., Bakhshi, A. D., & Mostafa, R. R. (2021). IoTBoT-IDS: A novel statistical learning-enabled botnet detection framework for protecting networks of smart cities. *Sustainable Cities and Society*, 72, 103041.
5. IoT Anaytics. *Market insights for the Internet of Things*. Retrieved 2 June, 2021 from https://iot-analytics.com/
6. Gupta, K., & Shukla, S. (2016, February). Internet of Things: Security challenges for next generation networks. In 2016 *International Conference on Innovation and Challenges in Cyber Security (ICICCS-INBUSH)* (pp. 315-318). IEEE.
7. Kumar, P., Braeken, A., Gurtov, A., Iinatti, J., & Ha, P. H. (2017). Anonymous secure framework in connected smart home environments. *IEEE Transactions on Information Forensics and Security*, 12(4), 968-979.
8. Deniz, E. (2019). *Nesnelerin İnternetinde Gizlilik Ve Güvenlik Yönetimi*. (Master's dissertation, Ankara University).
9. D'angelo, G., Palmieri, F., Ficco, M., & Rampone, S. (2015). An uncertainty-managing batch relevance-based approach to network anomaly detection. *Applied Soft Computing*, 36, 408-418.
10. Ring, M., Wunderlich, S., Scheuring, D., Landes, D., & Hotho, A. (2019). A survey of network-based intrusion detection data sets. *Computers & Security*, 86, 147-167.
11. Shafiq, M., Tian, Z., Bashir, A. K., Du, X., & Guizani, M. (2020). IoT malicious traffic identification using wrapper-based feature selection mechanisms. *Computers & Security*, 94, 101863.
12. Mohammadi, M., Al-Fuqaha, A., Sorour, S., & Guizani, M. (2018). Deep learning for IoT big data and streaming analytics: A survey. *IEEE Communications Surveys & Tutorials*, 20(4), 2923-2960.
13. Yönem, E. (2019). *Nesnelerin Internetinde Veri Analizi İçin Tekrarlayıcı Sinir Ağları Yönetiminin Yapay Arı Koloni Algoritması İle Eğitilmesi*. (Master's dissertation, Erciyes University).
14. Rathore, S., & Park, J. H. (2018). Semi-supervised learning based distributed attack detection framework for IoT. *Applied Soft Computing*, 72, 79-89.
15. Xiao, L., Wan, X., Lu, X., Zhang, Y., & Wu, D. (2018). IoT security techniques based on machine learning: How do IoT devices use AI to enhance security?. *IEEE Signal Processing Magazine*, 35(5), 41-49.
16. Kotenko, I., Saenko, I., Kushnerevich, A., & Branitskiy, A. (2019, February). Attack detection in IoT critical infrastructures: a machine learning and big data processing approach. In 2019 *27th Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP) (pp. 340-347)*. IEEE.
17. Vu, L., Nguyen, Q. U., Nguyen, D. N., Hoang, D. T., & Dutkiewicz, E. (2020). Deep transfer learning for IoT attack detection. *IEEE Access*, 8, 107335-107344.
18. Zhang, C., & Green, R. (2015, April). Communication security in internet of thing: preventive measure and avoid DDoS attack over IoT network. In *Proceedings of the 18th symposium on communications & networking* (pp. 8-15).
19. Yavuz, F.Y. (2018). *Deep Learning in Cyber Security for Internet of Things*. (Master's dissertation, Istanbul City University).
20. Koroniotis, N., Moustafa, N., Sitnikova, E., & Turnbull, B. (2019). Towards the development of realistic botnet dataset in the internet of things for network forensic analytics: Bot-iot dataset. *Future Generation Computer Systems*, 100, 779-796.

21. Koroniotis, N., & Moustafa, N. (2021). The Bot-IoT Dataset. *UNSW Canberra ADFA*.
22. Koroniotis, N. (2020). *Designing an effective network forensic framework for the investigation of botnets in the Internet of Things* (Doctoral dissertation, University of New South Wales, Sydney, Australia).
23. Yaman, O., Yetis, H., & Karakose, M. (2020, October). Decision Tree Based Customer Analysis Method for Energy Planning in Smart Cities. In *2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI)* (pp. 1-4). IEEE.
24. Leevy, J. L., Hancock, J., Khoshgoftaar, T. M., & Peterson, J. (2021, December). Detecting Information Theft Attacks in the Bot-IoT Dataset. In *2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA)* (pp. 807-812). IEEE.
25. Zeeshan, M., Riaz, Q., Bilal, M. A., Shahzad, M. K., Jabeen, H., Haider, S. A., & Rahim, A. (2021). Protocol-Based Deep Intrusion Detection for DoS and DDoS Attacks Using UNSW-NB15 and Bot-IoT Data-Sets. *IEEE Access*, 10, 2269-2283.
26. Sugi, S. S. S., & Ratna, S. R. (2020, December). Investigation of machine learning techniques in intrusion detection system for IoT network. In *2020 3rd International Conference on Intelligent Sustainable Systems (ICISS)* (pp. 1164-1167). IEEE.
27. Bhuvaneswari, B.A., & Selvakumar, S. (2020). Anomaly detection framework for Internet of things traffic using vector convolutional deep learning approach in fog environment. *Future Generation Computer Systems*, 113, 255-265.