ULUSLARARASI 3B YAZICI TEKNOLOJİLERİ
VE DİJİTAL ENDÜSTRİ DERGİSİ

INTERNATIONAL JOURNAL OF 3D PRINTING
TECHNOLOGIES AND DIGITAL INDUSTRY

ISSN:2602-3350 (Online)

URL: https://dergipark.org.tr/ij3dptdi

INTERNET OF THINGS BOTNET DETECTION VIA ENSEMBLE DEEP NEURAL NETWORKS

Yazarlar (Authors): Yağız Onur KOLCU*, Ahmet Haşim YURTTAKAL, Berker BAYDAN

Bu makaleye şu şekilde atıfta bulunabilirsiniz (To cite to this article): *Kolcu Y.O., Yurttakal A.H., Baydan B.,"Internet Of Things Botnet Detection Via Ensemble Deep Neural Networks" Int. J. of 3D Printing Tech. Dig. Ind., 7(2): 191-197, (2023).*

DOI: 10.46519/ij3dptdi.1293277

Araştırma Makale/ Research Article

Erişim Linki: (To link to this article): https://dergipark.org.tr/en/pub/ij3dptdi/archive

INTERNET OF THINGS BOTNET DETECTION VIA ENSEMBLE DEEP NEURAL NETWORKS

Yağız Onur KOLCU[a]*, Ahmet Haşim YURTTAKAL[b], Berker BAYDAN[c]

[a]Afyon Kocatepe University, Institute of Science and Technology, TÜRKİYE
[b]Afyon Kocatepe University, Engineering Faculty, Computer Engineering Department, TÜRKİYE
[c]HAVELSAN, Ankara, TÜRKİYE

* *Corresponding Author: y.onurkolcu@gmail.com*

(**Received:** 06.05.23; **Revised:** 21.06.23; **Accepted:** 10.07.23)

ABSTRACT

The widespread use of the Internet of Things (IoT) and the rapid increase in the number of devices connected to the network bring both benefits and many problems. The most important of these problems is cyber attacks. These cyber attacks cause financial losses as well as loss of reputation and time. Intrusion detection systems (IDS) and intrusion prevention systems (IPS) are used to eliminate or minimize these losses. IDS are designed to be signature-based or anomaly-based, and are currently being developed using anomaly-based systems as machine learning methods. The aim of this study is to detect whether there is an attack on your network, with a high success rate, by considering botnet as one of the attack types. In order to develop this system, it is aimed to use Ensemble Deep Neural Networks (DNN), which is one of the machine learning methods, and to search for solution methods for the most accurate result. In the study, N-BaIoT dataset in the UCI Machine Learning library was used for scientific research. The data consists of 1 benign network stream and 9 malicious network streams carried by 2 botnets. Stacked ensemble of DNN networks has been used from the classification stage. The proposed method has achieved %99 accuracy and the results are encouraging for future studies.

Keywords: Botnet, Internet of Things, Ensemble, Deep Neural Network, Cyber Threats.

1. INTRODUCTION

The Internet of Things (IoT) is the interconnection of physical devices and objects through networking [1]. The most important benefit is that it provides remote sensing and control of objects. It also provides economy and efficiency by minimizing human intervention. IoT consists of internet-connected devices such as smart home appliances, water meters, security cameras. These devices are small computers with processor and IP address running on Linux devices [2]. While there are at least 10 billion active IoT devices in 2021, it is predicted that this number will exceed 25.4 billion in 2030 [3].

IoT devices generate, process and exchange significant amounts of data during their operations. IoT devices are prone to a variety of physical, network, and application layer attacks that can lead to business disruptions, privacy breaches, and even physical injury. Therefore, security requirements come to the same level of priority as product innovation [4]. Network security, which is the biggest problem of our age, is now a big problem not only for computers, but also for every device connected to the internet. Millions of new devices are connecting to the internet every day. However, this great development brings with it great risks. Every day, cyber crimes are spreading rapidly and cybercriminals are carrying out various attacks [5]. For this reason, increasing the performance of intrusion detection systems is among the most important goals of information technologies [6]. Botnets are one of the threats that can most exploit IoT vulnerabilities [7]. A botnet is a collection of devices connected to the internet and infected with malware. These bots can be used for denial of service and various spam attacks [8]. Also, bots goal to distribute false information from illegal sources, to obtaine identity, password and financial data and to processe data to crack the password for

access to additional hosts [9]. Therefore, the detection and elimination of botnets are significant cybersecurity challenges. Reliable and inexpensive Botnet detection models are essential to detect and warn of risks without corrupting transmission data [10]. Detection of botnets is different from malware detection systems or anomaly detection systems. Because while other attacks represent an individual pattern, botnet attacks are part of a large attack network [11].

Artificial intelligence and machine learning algorithms learn complex patterns on data that cannot be noticed by the human eye, and provide fast, accurate, human-independent results [12]. Artificial intelligence and machine learning have achieved successful results in health [13-14], agriculture [15] and safety [16-17]. With the developments in hardware and processing capacity technology in recent years, machine learning and artificial intelligence algorithms make network security studies more successful [18]. Stevanovic and Pedersen (2016) emphasized supervised learning methods such as artificial neural networks and unsupervised learning methods such as hierarchical clustering in their botnet identifying studies [19]. Verma and Ranga (2020) found that ensemble methods are successful for detecting Denial of Service (DoS) attacks in IoT networks [20]. Altunay and Albayrak (2021) implemented an attack detection application based on feature selection using a Convolutional Neural Networks (CNN) to prevent cyber attacks. The success of detecting data as a threat was 98.7% for Brute Force, 98.5% for DoS, 98.9% for Botnet and 99.1% for SQL Injection [21].

In this study, the detection of botnet attacks on IoT devices is classified with Ensemble Deep Neural Network. The network traffics of 10 different IoT devices, including 1 benign and 9 malignant, belonging to 2 botnet attacks, were classified with 99% accuracy. The proposed method is fast, secure and automatic with high accuracy.

2. MATERIAL AND METHODS

2.1. Dataset

The dataset used in the study, N-BaIoT, created by Meidan et al.[22], is a general dataset used to detect botnet attacks on IoT devices with open access for academic studies. Network traffic belongs to 10 IoT devices, 9 attack classes and 1 benign, carried by 2 botnets. IoTs consist of a thermostat, a baby monitor, a webcam, two different doorbells, and four different inexpensive security cameras. Devices that wereDanmini (Doorbell), Ecobee (Thermostat), Ennio(Doorbell), Philips_B120N10 (Baby_Monitor), Provision_PT_737E (Security_Camera), Provision_PT_838 (Security_Camera), Samsung_SNH_1011_N (Webcam), SimpleHome_XCS7_1002_WHT (Security_Camera), SimpleHome_XCS7_1003_WHT(Security_Camera), were malignant.This private network was created by infecting one of the security cameras with a real sample of Mirai botnet malware [23]. The traffic information of the dataset is given in Table 1.

Table 1. Traffic Information [24]

ID	Name	Device Type	Traffic
0	benign	All Devices	555932
1	Danmini	Door bell	968750
2	Ecobee	Thermostat	822763
3	Ennio	Door bell	316400
4	Philips_B120N10	Baby Monitor	923437
5	Provision_PT_737E	Security camera	766106
6	Provision_PT_838	Security camera	738377
7	Samsung_SNH_1011_N	Webcam	323072
8	SimpleHome_XCS7_1002_WHT	Security camera	816471
9	SimpleHome_XCS7_1003_WHT	Security camera	831298

Mirai is malware that mostly targets networked smart home and consumer devices and can turn them into a zombie network of remote bots [25]. The largest distributed denial of service (DDoS) attack in 2016 was carried out by the Mirai botnet. Mirai botnet detection is important when it is predicted that the number of devices using IoT will gradually increase [26]. The main features based on the classification are information that summarizes the flow traffic, information that summarizes the jitter of the traffic, time frame information, the number of items that have appeared recently, and the

variances of the two flows. While the dataset consists of 165645 samples, the number of features that are the basis for classification is 115.

2.2. Ensemble Deep Neural Network

Machine learning is a useful artificial intelligence technique that automatically finds useful information from large datasets. Deep learning is a branch of machine learning [27]. Since there are many security areas, machine learning is important for this area. Bots generate different flows than normal flows. In this way, machine learning (ensemble classifier algorithms) can classify flows with the highest accuracy [10]. It seems more rational to use machine learning to detect botnet [28]. Deep neural networks (DNN) learn the parameters that provide the best approximation by mapping the convergence of a function f with input I to the value O. The information flows through the calculated function starting from I, passes through the f function with intermediate calculations and reaches the output value of O. The output of the model has no feedback links fed into it. Model f(I)=f3(f2(f1(I))) can consist of functions linked like a chain. In this case, f1 constitutes the first layer of the network and f2 constitutes the second layer of the network. The length of the entire chain gives the depth of the pattern. This is where the concept of depth in a deep neural network comes from. The layers between the input and output layers of the neural network are hidden layers. The dimensions of the hidden layers give the width of the model. The last layer of the deep neural network gives the output layer [29].

Ensemble learning is learning that allows building the model with more than one learner. It aims that the models will make more accurate decisions in solving the problem. There are different ensemble learning techniques such as Boosting, Bagging, Stacking, Voting. Stacking ensemble was used in the study [30]. Stacking involves training a different classifier by combining the predictions of several classifiers. In the first stage, the existing data is trained with classifiers. It is then trained with a final meta-learner algorithm, using the predictions of the first-stage algorithms as additional inputs [31].

3. EXPERIMENTAL RESULTS

All codes of the study were developed in open source Python environment. 70% of the dataset is split for training and 30% for testing. The proposed ensemble model was developed on the basis of the model that was successful in diabetes detection before [32]. In the first stage, output prediction values were taken from 2 DNN models. The neuron numbers of the model used at this stage are 128-64-32-10. While the activation function of the intermediate layers is ReLu, the activation function of the output layer is Softmax. Activation function is necessary fort his model structure because this neural network is need to learn nonlinear situation as well. If the activation function is not used, the output signal becomes a simple linear function. Linear functions are only polynomials of odd degree. A neural network without activation function will act like a linear regression with limited learning power. In the DNN structure used in the Meta Learner stage, two layers, 32-10, were used. While the number of learnable parameters of the models used in the first stage is 25514, the total number of learnable parameters of the whole model is 52030. The architectural structure of the proposed model is given in Figure 1.

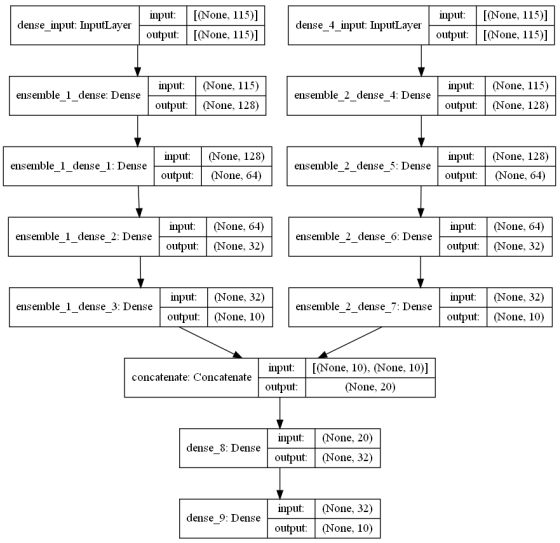

Figure 1. Proposed Model.

While the networks in the first layer were run 500 epochs, Stacked Ensemble model 300 epochs were run. Adam Optimizer was used as the optimizer. Adam Optimizer was preferred because it gave more successful results. Other hyperparameters are given in Table 2.

Table 2. Hyperparameters

Parameters	Value
Learning Rate	0.001
Beta 1	0.9

Beta 2	0.999
Epsilon	1e-07

The loss function is the categorical cross entropy. Because our problem to be classified is multiclass. Figure 2 shows the accuracy epoch graph for the training process.

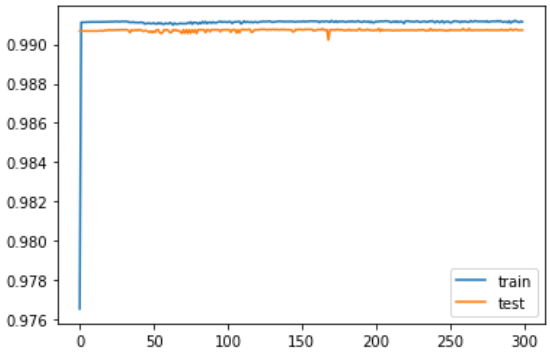

Figure 2. Accuracy- Epoch Graph.

According to the figure, the training processes of the training and test groups are compatible with each other. This shows that there is no over-fitting in the training of the model and that the proposed model has a high generalization capacity. In Figure 3, the confusion matrix of the test set is given.

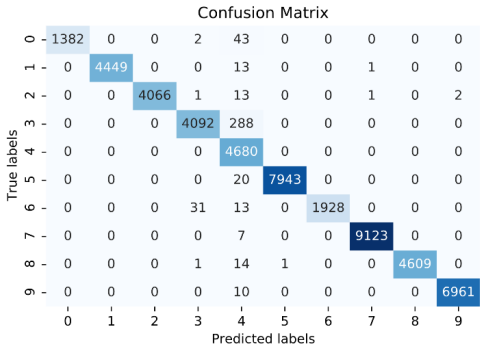

Figure 3. Test set confusion matrix.

According to the figure, most of the samples were estimated correctly. Accuracy, precision, recall, F1 score performance metrics are 0.99, 0.991, 0.986, 0.988, respectively. Performance metric values calculated separately for each class are given in Table 3.

Table 3. Performance Metrics.

Class	Precision	Recall	F1-Score	Number of Samples
0	1.00	0.97	0.98	1427
1	1.00	1.00	1.00	4463
2	1.00	1.00	1.00	4083
3	0.99	0.93	0.96	4380
4	0.92	1.00	0.96	4680
5	1.00	1.00	1.00	7963
6	1.00	0.98	0.99	1972
7	1.00	1.00	1.00	9130
8	1.00	1.00	1.00	4625
9	1.00	1.00	1.00	6971

According to the results obtained, the lowest 0.96 F1 score was obtained in the malignant network traffic of the doorbell coded as 3 and the baby monitor coded as 4. Figure 4 shows the ROC curve for each label. The results show that the AUC value for each label is close to 100%.

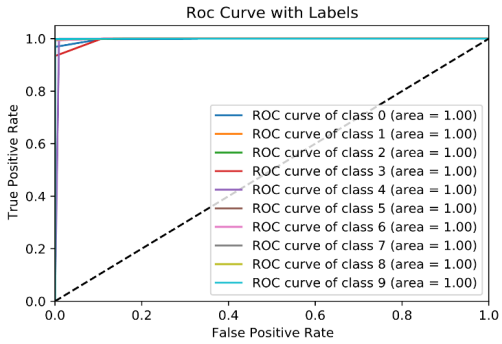

Figure 4. ROC Curve.

4. DISCUSSION AND CONCLUSION

In the era of digital transformation, IoT is undoubtedly one of the leading technologies that revolutionize the lives of both institutions and individuals. IoT is the communication of physical devices that have an IP address, connect to the internet and share data with each other. Although this technology is the key point of the industrial revolution, it is also frequently used in smart home, smart agriculture and smart health applications. IoT devices communicating with each other can pose a security threat to networks. Even a seemingly simple device can pose serious dangers when compromised by attackers. For example, a baby camera captured by a hacker can be used for spying. Attackers can disrupt network services, steal data, cause physical harm or even harm people. Botnet is one of the biggest threats to IoT devices. For this reason, many studies have been conducted on artificial intelligence-based botnet detection. Botnet is one of the biggest threats to IoT devices. For this reason, many studies have been conducted on artificial intelligence-based botnet detection. Ahmetoğlu and Daş (2019) tried to detect benign, FTP patator, SSH-

patator, DoS, Heartbleed, Brute Force, Web Attack–SQL Injection, DDOS, Port attack, Botnet attack types with a fully connected artificial neural network [6]. Wai et al. (2018) proposed a botnet traffic detection technique based on machine learning. Their research relies on multi-layer perceptrons and decision trees to analyze network traffic for automatic traffic detection [33]. It can be seen that deep learning methods are effective in creating behavioral analysis of the large amount of data created in the IoT network. It has been determined that deep learning mechanisms outperform other contingency solutions in multiple domains using unstructured and heterogeneous data.

The problem with traditional machine learning algorithms is that while they can run well in a self-created environment, data increases as more devices are included in the network, giving these models a display of wear. Since deep learning algorithms learn more from more data, there are studies in this area where deep learning solves this problem. In addition, the state of art algoritms which aredecision tree and artificial neural networks, are used for botnet detection. Regarding to comparision of these algorithms, CTU-13 that contained network trafic dataset, was used in this study. Neural network and decision tree have accuries 0.91 and 0.98, respectively. Our proposed solution's F1 score which is 0.99, is more accurate than these two algorithms [26]. In another study, Cyber Clean Center dataset which containes traffic packets 6667 as port number used for IRC and 80 as port number http. According to comparision of F1-score of botnet detection, our proposed algorithm is better than ELM (91.6%), CNN(92.56%), SVM(93.36%) and Ensemble Classifier Algorithm with Stacking Process (ECASP) (94.08%) [34].In Meidan and friends' IoT botnet detection, deep autoencoders was used in their proposed method. The detection of IoT botnet attacks N-BaIot dataset was used in this study. Mirai and Bashlite which are most popular IoT based botnets, were infected for each device. Meidan and friends' model structure were occured by four hidden layers of encoders which decreasing sizes of 75%, 50%, 33% and 25% of the input layer's dimension. The other next layers contained decoders via similar size as encoders. This structure of their model's true positive rate is 100% similar with our proposed solution of true positive rate. Local Outlier Factor (LOF), One-Class SVM, and Isolation Forest were also other evaluated algorithms in Meidan and friends' study. Their proposed solution of TPRis similar with LOF and SVM and more accurate than Isolation Forest [22].

In another study Algelal and friends' botnet detection methods was trained and tested with CTU-13 dataset and 10 fold cross validation. The accuracy of proposed model is 99.84%. Ensemble classifiers in this study consist of AdaBoost and Jrip algorithms. IoT botnet detection via ensemble deep neural networks has almost similar accuracy with AdaBoost and Jrip algorithms. The rest of method which are Clustering (98.39%), Neural Network (89.38%), Recurrrent Neural Network (83.09%), K-medoids, L-means, LSTM, decision trees, has less accuracy detection than proposed model for botnet detection [10]. The another study which was used ISCX dataset, analyzed botnet traffic with Ensemble of classifie algorithm This dataset contained normal traffic and botnet traffic. This study also showed that ensemble classifier algoritms which were Ada-Boost with Decision Tree (94.78%) and Soft Voting of KNN & Decision Tree (96.41%), increased the accuracy of botnet detection as our proposed model's accuracy [35]. Recently, real time automatic botnet detection tool was developed for large network bandwidths. Since the duration botnet detection was really important, they developed ultra fast network analysis tool by using their proposed new machine learning model. Although their processing time was very fast (0.007 ms), their F1 score was less (0.926) than F1 score of our proposed model [36]. In another recent study was related with economic system to detect IoT botnets with deep learning model. Their proposed model asserted that decrease the implementation budget and supplied used effiectient low cost development structure for their model. However, the F1 score was really low, especially test class 2 (0.41), class 3 (0.77) and class 5 (0) [37].

In this study, the network traffic of 10 devices, 9 malignant and 1 benign, belonging to the N-BaIoT dataset was classified with an Ensemble DNN-based approach. According to the results obtained, the proposed method works with 99% accuracy. The problem, which is a complex and multi-classification problem due to its structure, works quickly, independently of the user, and with high accuracy thanks to the proposed

method. It resolves concerns by providing solutions to individual and corporate security problems. The results are promising and encouraging for future studies.

REFERENCES

1. Elkhodr, M., Shahrestani S. and Cheung, H. "The Internet of Things: Vision & Challenges", IEEE 2013 Tencon-Spring, Pages 218-222, Sydney, 2013.

2. Barrera, D., Molloy, I. and Huang, H. "IDIoT: Securing the Internet of Things like it's 1994," arXiv preprint arXiv:1712.03623, 2017.

3. Huyghue, B.D. "Cybersecurity, Internet of Things, and Risk Management for Businesses", Diss. Utica College, Utica, NY, 2021.

4. Skorin-Kapov, N. et al. "Physical-Layer Security in Evolving Optical Networks." IEEE Communications Magazine, Vol. 54, Issue 8, Pages 110-117, 2016.

5. Gantz J. and David, R. "The digital universe in 2020: Big Data, Bigger Digital Shadows and Biggest Growth in the Far East." IDC iView: IDC Analyze the future 2007, Pages 1-16, 2012

6. Ahmetoğlu, H. and Daş, R., "Derin Öğrenme ile Büyük Veri Kumelerinden Saldırı Türlerinin Sınıflandırılması", IDAP, Pages 455-463, Malatya, Türkiye, 2019.

7. Bezerra, V.H. et al, "IoTDS: A One-Class Classification Approach To Detect Botnets in Internet of Things Devices." Sensors, Vol. 19, Issue 14, 2019.

8. Bertino E. and Islam, N. "Botnets and Internet of Things Security." Computer, Vol. 50, Issue 2, Pages 76-79, February 2017.

9. Grizzard J.B. et al, "Peer-to-Peer Botnets: Overview and Case Study," HotBots, Vol. 7, Pages 1-8, 2007,

10. Algelal, Z. et al, "Botnet Detection Using Ensemble Classifiers of Network Flow", International Journal of Electrical and Computer Engineering (IJECE), Vol. 10, Issue 3, Pages 25-43, 2020.

11. Geer, D., "Malicious Bots Threaten Network Security." Computer, Vol. 38, Issue 1, Pages 18-20, January 2005.

12. El Naqa, I. and Murphy, M. J. "What is Machine Learning?" Machine Learning in Radiation Oncology, Pages 3-11, Springer, Cham, 2015.

13. Yurttakal, A. H., & Erbay, H. "Segmentation of Larynx histopathology images via convolutional neural networks" In Intelligent and Fuzzy Techniques: Smart and Innovative Solutions: Proceedings of the INFUS 2020 Conference, Istanbul, Turkey, July 21-23, Pages 949-954. Springer International Publishing, 2021.

14. Çinarer, G., Emiroğlu, B. G., & Yurttakal, A. H. "Predicting 1p/19q chromosomal deletion of brain tumors using machine learning" Emerging Materials Research, Vol. 10, Issue 2, Pages 238-244, 2021

15. Yurttakal, A. H. "Extreme gradient boosting regression model for soil thermal conductivity" Thermal Science, Vol. 25, Issue 1, Pages 1-7, 2021

16. Arslan, R. S., & Yurttakal, A. H. "K-nearest neighbour classifier usage for permission based malware detection in android". Icontech Internatıonal Journal, Vol. 4, Issue 2, Pages 15-27, 2020.

17. Horasan, F., & Yurttakal, A. H. Darknet Web Traffic Classification via Gradient Boosting Algorithm. International Journal of Engineering Research and Development, Vol. 14, Issue 2, Pages 794-798, 2022.

18. Lu, Y., "Artificial Intelligence: A Survey on Evolution, Models, Applications and Future Trends", Journal of Management Analytics, Vol. 6, Issue 1, Pages 1-29, 2019.

19. Stevanovic M. and Pedersen, J.M. "On the Use of Machine Learning for Identifying Botnet Network Traffic." Journal of Cyber Security and Mobility, Vol. 4, Issue 2 & 3,Pages 109-128, 2016.

20. Verma, A. and Ranga, V., "Machine learning Based Intrusion Detection Systems for IoT Applications", Wireless Personal Communications, Vol. 111, Issue 4, Pages 2287-2310, 2020.

21. Altunay, H. C. and Albayrak, Z., "Network Intrusion Detection Approach Based on Convolutional Neural Network." Avrupa Bilim ve Teknoloji Dergisi, Vol. 26, Pages 22-29, 2021.

22. Meidan, Y. et al. "N-Baiot—Network-Based Detection of IoT Botnet Attacks Using Deep Autoencoders." IEEE Pervasive Computing, Vol. 17, Issue 3, Pages 12-22, 2018.

23.Mirsky, Y. et al, "Kitsune: An Ensemble of Autoencoders for Online Network Intrusion Detection." arXiv preprint arXiv:1802.09089, 2018.

24. Rey, V., Sánchez, P. M. S., Celdrán, A. H., & Bovet, G. “Federated learning for malware detection in iot devices”. Computer Networks, Vol. 204, 2022.

25. Antonakakis, M., et al. "Understanding The Mirai Botnet." 26th USENIX Security Symposium, Pages 1093-1110, Berkeley, CA, USA, 2017.

26. Ryu, S. and Yang, B., "A Comparative Study of Machine Learning Algorithms and Their Ensembles for Botnet Detection," Journal of Computer and Communications, Vol. 6, Pages 119-129, 2018.

27. Liu, H., & Lang, B. “Machine learning and deep learning methods for intrusion detection systems: A survey”. Applied Sciences, Vol. 9, Issue 20, 2019.

28.Rezai, A. "Using Ensemble Learning Technique for Detecting Botnet on IoT," SN Computer Science, Vol. 2, Issue 2, Pages 148, 2021.

29. Goodfellow, I. Yoshua B. and Aaron, C. “Deep learning”, MIT Press, Cambridge, MA, 2016.

30. Rokach, L. “Ensemble-based classifiers.” Artificial Intelligence Review, Vol. 33, Issue 1, Pages 1-39, 2010.

31. Wolpert, D. H. "Stacked Generalization." Neural Networks, Vol. 5, Issue 2, Pages 241-259, 1992.

32. Yurttakal A.H. and Baş, H. "Possibility Prediction of Diabetes Mellitus at Early Stage Via Stacked Ensemble Deep Neural Network." Afyon Kocatepe Üniversitesi Fen ve Mühendislik Bilimleri Dergisi, Vol. 21, Issue 4, Pages 812-819, 2021.

33. Wai, F.K. et al, "Automated Botnet Traffic Detection Via Machine Learning.", TENCON 2018-2018 IEEE Region 10 Conference, Pages 38-43, Jeju Island, Korea, 2018.

34. Srinivasan S. and Kumar, D. "Enhancing the Security in Cyber-World by Detecting the Botnets Using Ensemble Classification Based Machine Learning", Measurement: Sensors, Vol. 25, Pages 2023.

35.Bijalwan, A. et al, “Botnet Analysis Using Ensemble Classifier” Perspectives in Science, Vol. 8, Pages 502-504, 2016.

36. Velasco-Mata, J., González-Castro, V., Fidalgo, E. et al., “Real-time botnet detection on large network bandwidths using machine learning” Sci Rep, Vol. 13, Pages 4282, 2023.

37. Elsayed N. et al., “IoT botnet detection using an economic deep learning model” AIIoT, 2023.