Journal of Cardiovascular Disease Research

ISSN: 0975-3583,0976-2833 VOL14, ISSUE 10, 2023

# LOGISTIC REGRESSION AND RANDOM FOREST CLASSIFIER FOR ATTACK DETECTION IN IOT SENSOR DATA

**S.Sunil Kumar[1], K. Vyshnavi[2], M. Pavana sri[2], K. Bharathi[2], N. Vyshnavi[2]**

[1] *Assistant Professor, Department of Information Technology, Mallareddy Engineering College for Women, (UGC-Autonomous), Hyderabad, India, suv.sunilkumar@gmail.com.*

[2] *Student, Department of Information Technology, Mallareddy Engineering College for Women, (UGC-Autonomous), Hyderabad, India.*

## Abstract

The Internet of Things (IoT) connects a vast array of devices, ranging from home appliances to industrial sensors, creating an interconnected network of smart devices. IoT applications generate large volumes of sensor data, which are highly susceptible to security breaches and attacks. Cyber-criminals may exploit vulnerabilities in the IoT ecosystem to manipulate sensor data, leading to disastrous consequences such as unauthorized access, data falsification, and service disruption. In addition, IoT-based attacks can lead to severe consequences such as data manipulation, privacy breaches, and economic losses. One of the major challenges is detecting and preventing attacks on the valuable sensor data collected by IoT devices. Traditional security methods designed for conventional networks may not be suitable for the complex and distributed nature of IoT systems. To address this concern, there is a need for specialized techniques tailored for IoT sensor data to protect these systems and their users. Further, the proposed work aims to contribute to the field of cybersecurity and foster more resilient and secure IoT implementations. This work introduces a comprehensive and practical solution to enhance IoT security. Here, Logistic Regression, and Random Forest classifiers are employed for attack detection from the IoT sensor data, where the first one is a straightforward yet powerful technique for binary classification, enabling the detection of simple intrusion attempts. Meanwhile, the Random Forest Classifier excels at handling complex patterns and interactions in data, making it effective in identifying sophisticated attacks with multiple variables and dependencies. By leveraging the strengths of these algorithms, the proposed approach provides a robust and advanced system for detecting a wide range of attacks in IoT sensor data.

## 1. Introduction

The general idea of the Internet of Things (IoT) is to allow for communication between human-to-thing or thing-to-thing(s). Things denote sensors or devices, whilst human or an object is an entity that can request or deliver a service [1]. The interconnection amongst the entities is always complex. IoT is broadly acceptable and implemented in various domains, such as healthcare, smart home, and agriculture. However, IoT has a resource constraint and heterogeneous environments, such as low computational power and memory. These constraints create problems in providing and implementing a security solution in IoT devices. These constraints further escalate the existing challenges for IoT environment. Therefore, various kinds of attacks are possible due to the vulnerability of IoT devices. IoT-based botnet attack is one of the most popular, spreads faster and create more impact than other attacks. In recent years, several works have been conducted to detect and avoid this kind of attacks [2]–[3] by using novel approaches. Hence, a plethora of relevant of relevant models, methods, and etc. have been introduced over the past few years, with quite a reasonable number of studies reported in the research domain. Many studies are trying to protect against these botnet attacks on the IoT environment. However, there are many gaps still existing to develop an effective detection mechanism. An intrusion detection system (IDS) is one of the efficient ways to deal with attacks. However, the traditional IDSs are often not able to be deployed for the IoT environments due to the resource constraint problem of

ISSN: 0975-3583,0976-2833 VOL14, ISSUE 10, 2023

these devices. The complex cryptographic mechanisms cannot be embedded in many IoT devices either for the same reason. There are mainly two kinds of IDSs: the anomaly and misuse approaches. The misuse-based, also called the signature-based, approach, is based on the attacks' signatures, and they can also be found in most public IDSs, specifically Suricata [4]. Formally, the attacker can easily circumvent the signature-based approaches, and these mechanisms cannot guarantee to detect the unknown attacks and the variances of known attacks. The anomaly-based systems are based on normal data and can support to identify the unknown attacks. However, the different nature of IoT devices is being faced with the difficulty of collecting common normal data. The machine learning-based detection can guarantee detection of not only the known attacks and their variances. Therefore, we proposed a machine learning-based botnet attack detection architecture. We also adopted a feature selection method to reduce the demand for processing resources for performing the detection system on resource constraint devices. The experiment results indicate that the detection accuracy of our proposed system is high enough to detect the botnet attacks. Moreover, it can support the extension for detecting the new distinct kinds of attacks.

## 2. Literature Survey

Soe et al. [5] adopted a lightweight detection system with a high performance. The overall detection performance achieves around 99% for the botnet attack detection using three different ML algorithms, including artificial neural network (ANN), J48 decision tree, and Naïve Bayes. The experiment result indicated that the proposed architecture can effectively detect botnet-based attacks, and also can be extended with corresponding sub-engines for new kinds of attacks.Ali et al. [6] outlined the existing proposed contributions, datasets utilised, network forensic methods utilised and research focus of the primary selected studies. The demographic characteristics of primary studies were also outlined. The result of this review revealed that research in this domain is gaining momentum, particularly in the last 3 years (2018-2020). Nine key contributions were also identified, with Evaluation, System, and Model being the most conducted. Irfan et al. [7] classified the incoming data in the IoT, contain a malware or not. In this research, this work under sample the dataset because the datasets contain imbalance class. After that, this work classified the sample using Random Forest. This work used Naive Bayes, K-Nearest Neighbor and Decision Tree too as a comparison. The dataset that has been used in this research are from UCI Machine Learning Depository's Website. The dataset showed the data traffic from the IoT Device in a normal condition and attacked by Mirai or Bashlite.

Shah et al. [8] presented a concept called 'login puzzle' to prevent capture of IoT devices in a large scale. Login puzzle is a variant of client puzzle, which presented a puzzle to the remote device during the login process to prevent unrestricted log-in attempts. Login puzzle is a set of multiple mini puzzles with a variable complexity, which the remote device is required to solve before logging into any IoT device. Every unsuccessful log-in attempt increases the complexity of solving the login puzzle for the next attempt. This paper introduced a novel mechanism to change the complexity of puzzle after every unsuccessful login attempt. If each IoT device had used login puzzle, Mirai attack would have required almost two months to acquire devices, while it acquired them in 20 h. Tzagkarakis et al. [9] presented an IoT botnet attack detection method based on a sparsity representation framework using a reconstruction error thresholding rule for identifying malicious network traffic at the IoT edge coming from compromised IoT devices. The botnet attack detection is performed based on small-sized benign IoT network traffic data, and thus we have no prior knowledge about malicious IoT traffic data. We present our results on a real IoT-based network dataset and show the efficacy of proposed technique against a reconstruction error-based autoencoder approach. Meidan et al. [10] proposed a novel network-based anomaly detection method for the IoT called N-BaIoT that extracts behavior snapshots of the network and uses deep autoencoders to detect anomalous network traffic from compromised IoT

ISSN: 0975-3583,0976-2833 VOL14, ISSUE 10, 2023

devices. To evaluate the method, this work infected nine commercial IoT devices in our lab with two widely known IoT-based botnets, Mirai and BASHLITE. The evaluation results demonstrated the proposed methods ability to detect the attacks accurately and instantly as they were being launched from the compromised IoT devices that were part of a botnet.

Popoola et al. [11] proposed the federated DL (FDL) method for zero-day botnet attack detection to avoid data privacy leakage in IoT-edge devices. In this method, an optimal deep neural network (DNN) architecture is employed for network traffic classification. A model parameter server remotely coordinates the independent training of the DNN models in multiple IoT-edge devices, while the federated averaging (FedAvg) algorithm is used to aggregate local model updates. A global DNN model is produced after several communication rounds between the model parameter server and the IoT-edge devices. The zero-day botnet attack scenarios in IoT-edge devices are simulated with the Bot-IoT and N-BaIoT data sets. Hussain et al. [12] produced a generic scanning and DDoS attack dataset by generating 33 types of scans and 60 types of DDoS attacks. In addition, this work partially integrated the scan and DDoS attack samples from three publicly available datasets for maximum attack coverage to better train the machine learning algorithms. Afterwards, this work proposed a two-fold machine learning approach to prevent and detect IoT botnet attacks. In the first fold, this work trained a state-of-the-art deep learning model, i.e., ResNet-18 to detect the scanning activity in the premature attack stage to prevent IoT botnet attacks. While, in the second fold, this work trained another ResNet-18 model for DDoS attack identification to detect IoT botnet attacks.

Abu et al. [13] proposed an ensemble learning model for botnet attack detection in IoT networks called ELBA-IoT that profiles behavior features of IoT networks and uses ensemble learning to identify anomalous network traffic from compromised IoT devices. In addition, this IoT-based botnet detection approach characterizes the evaluation of three different machine learning techniques that belong to decision tree techniques (AdaBoosted, RUSBoosted, and bagged). To evaluate ELBA-IoT, we used the N-BaIoT-2021 dataset, which comprises records of both normal IoT network traffic and botnet attack traffic of infected IoT devices. Alharbi et al. [14] proposed Gaussian distribution used in the population initialization. Furthermore, the local search mechanism was followed by the Gaussian density function and local-global best function to achieve better exploration during each generation. Enhanced BA was further employed for neural network hyperparameter tuning and weight optimization to classify ten different botnet attacks with an additional one benign target class. The proposed LGBA-NN algorithm was tested on an N-BaIoT data set with extensive real traffic data with benign and malicious target classes. The performance of LGBA-NN was compared with several recent advanced approaches such as weight optimization using Particle Swarm Optimization (PSO-NN) and BA-NN.

Ahmed et al. [15] proposed a model for detecting botnets using deep learning to identify zero-day botnet attacks in real time. The proposed model is trained and evaluated on a CTU-13 dataset with multiple neural network designs and hidden layers. Results demonstrated that the deep-learning artificial neural network model can accurately and efficiently identify botnets.

## 3. Proposed System Design

Activity diagram is another important diagram in UML to describe the dynamic aspects of the system as shown in Figure 1. Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. As the name suggests, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of

ISSN: 0975-3583,0976-2833 VOL14, ISSUE 10, 2023

predictions, and it predicts the final output. The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

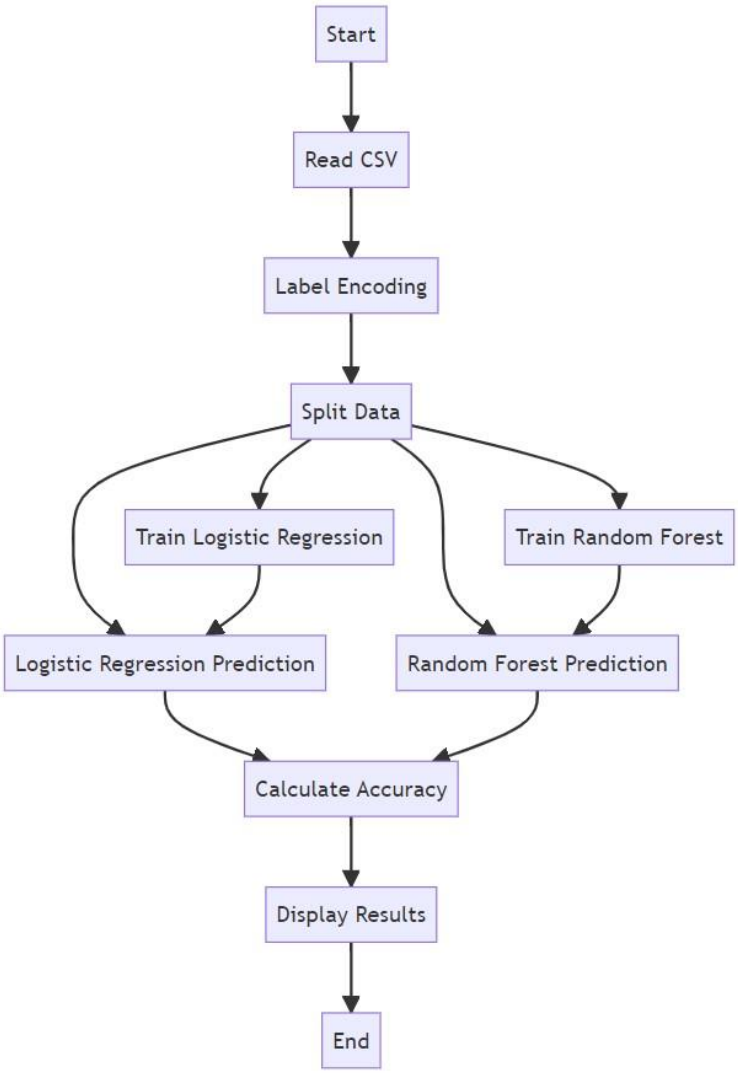


Figure 1. Proposed system design.

### 3.1 Random Forest Algorithm

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. As the name suggests, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output. The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

ISSN: 0975-3583,0976-2833 VOL14, ISSUE 10, 2023

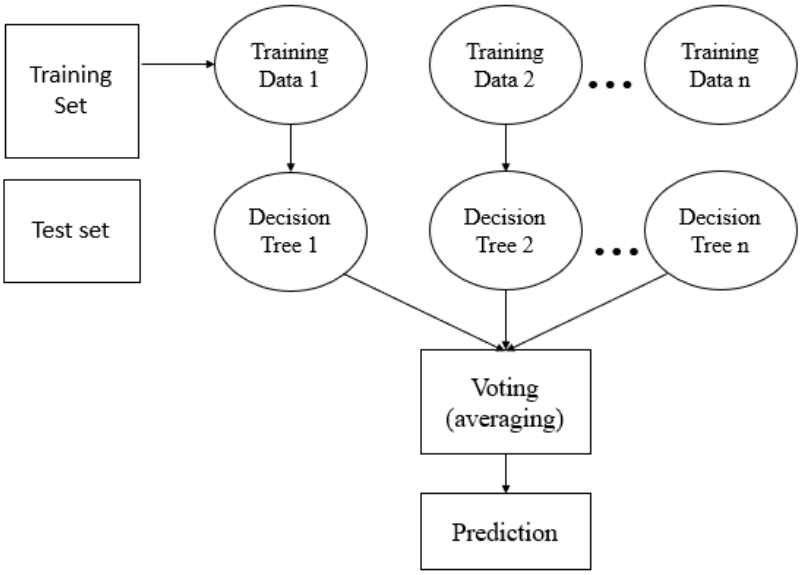


Figure 2: Random Forest algorithm.

## 4. Results and Description

The figure below shows a representation of a portion of the dataset used for detecting attacks in IoT sensor data. It displays rows and columns of sample data.

| | ts | device | co | humidity | light | lpg | motion | smoke | temp | Type |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1.590000e+09 | b8:27:eb:bf:9d:51 | 0.004956 | 51.000000 | False | 0.007651 | False | 0.020411 | 22.700000 | 0 |
| 1 | 1.590000e+09 | 00:0f:00:70:91:0a | 0.002840 | 76.000000 | False | 0.005114 | False | 0.013275 | 19.700001 | 0 |
| 2 | 1.590000e+09 | b8:27:eb:bf:9d:51 | 0.004976 | 50.900000 | False | 0.007673 | False | 0.020475 | 22.600000 | 0 |
| 3 | 1.590000e+09 | 1c:bf:ce:15:ec:4d | 0.004403 | 76.800003 | True | 0.007023 | False | 0.018628 | 27.000000 | 0 |
| 4 | 1.590000e+09 | b8:27:eb:bf:9d:51 | 0.004967 | 50.900000 | False | 0.007664 | False | 0.020448 | 22.600000 | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 404950 | 1.600000e+09 | b8:27:eb:bf:9d:51 | 0.005877 | 48.500000 | False | 0.008654 | False | 0.023284 | 22.300000 | 0 |
| 404951 | 1.600000e+09 | 00:0f:00:70:91:0a | 0.003745 | 75.300003 | False | 0.006247 | False | 0.016437 | 19.200001 | 0 |
| 404952 | 1.600000e+09 | b8:27:eb:bf:9d:51 | 0.005882 | 48.500000 | False | 0.008660 | False | 0.023301 | 22.200000 | 0 |
| 404953 | 1.600000e+09 | 00:0f:00:70:91:0a | 0.003745 | 75.300003 | False | 0.006247 | False | 0.016437 | 19.200001 | 0 |
| 404954 | 1.600000e+09 | b8:27:eb:bf:9d:51 | 0.005914 | 48.400000 | False | 0.008695 | False | 0.023400 | 22.200000 | 0 |

404955 rows × 10 columns

Figure 3: sample dataset used for attack detection in iot sensor data

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 404955 entries, 0 to 404954
Data columns (total 10 columns):
 #   Column    Non-Null Count   Dtype
---  ------    --------------   -----
 0   ts        404955 non-null  float64
 1   device    404955 non-null  object
 2   co        404955 non-null  float64
 3   humidity  404955 non-null  float64
 4   light     404955 non-null  bool
 5   lpg       404955 non-null  float64
 6   motion    404955 non-null  bool
 7   smoke     404955 non-null  float64
 8   temp      404955 non-null  float64
 9   Type      404955 non-null  int64
dtypes: bool(2), float64(6), int64(1), object(1)
memory usage: 25.5+ MB
```

Figure 4: Dataset summary of IOT sensor data

ISSN: 0975-3583,0976-2833 VOL14, ISSUE 10, 2023

| | ts | device | co | humidity | light | lpg | motion | smoke | temp | Type |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 1.590000e+09 | 2 | 0.004956 | 51.000000 | 0 | 0.007651 | 0 | 0.020411 | 22.700000 | 0 |
| **1** | 1.590000e+09 | 0 | 0.002840 | 76.000000 | 0 | 0.005114 | 0 | 0.013275 | 19.700001 | 0 |
| **2** | 1.590000e+09 | 2 | 0.004976 | 50.900000 | 0 | 0.007673 | 0 | 0.020475 | 22.600000 | 0 |
| **3** | 1.590000e+09 | 1 | 0.004403 | 76.800003 | 1 | 0.007023 | 0 | 0.018628 | 27.000000 | 0 |
| **4** | 1.590000e+09 | 2 | 0.004967 | 50.900000 | 0 | 0.007664 | 0 | 0.020448 | 22.600000 | 0 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **404950** | 1.600000e+09 | 2 | 0.005877 | 48.500000 | 0 | 0.008654 | 0 | 0.023284 | 22.300000 | 0 |
| **404951** | 1.600000e+09 | 0 | 0.003745 | 75.300003 | 0 | 0.006247 | 0 | 0.016437 | 19.200001 | 0 |
| **404952** | 1.600000e+09 | 2 | 0.005882 | 48.500000 | 0 | 0.008660 | 0 | 0.023301 | 22.200000 | 0 |
| **404953** | 1.600000e+09 | 0 | 0.003745 | 75.300003 | 0 | 0.006247 | 0 | 0.016437 | 19.200001 | 0 |
| **404954** | 1.600000e+09 | 2 | 0.005914 | 48.400000 | 0 | 0.008695 | 0 | 0.023400 | 22.200000 | 0 |

404955 rows × 10 columns

Figure 5: dataset after applying label encoding

```
array([[-0.64516935,  0.96913684,  0.25353646, ..., -0.03452063,
         0.28074365,  0.09209913],
       [-0.64516935, -1.41430784, -1.43863794, ..., -0.03452063,
        -1.46526998, -1.02015731],
       [-0.64516935,  0.96913684,  0.26959006, ..., -0.03452063,
         0.29636678,  0.05502391],
       ...,
       [ 1.54998063,  0.96913684,  0.99437351, ..., -0.03452063,
         0.98775919, -0.09327699],
       [ 1.54998063, -1.41430784, -0.71522289, ..., -0.03452063,
        -0.69167164, -1.20553343],
       [ 1.54998063,  0.96913684,  1.02014181, ..., -0.03452063,
         1.01187356, -0.09327699]])
```

Figure 6: Features of a dataset after preprocessing

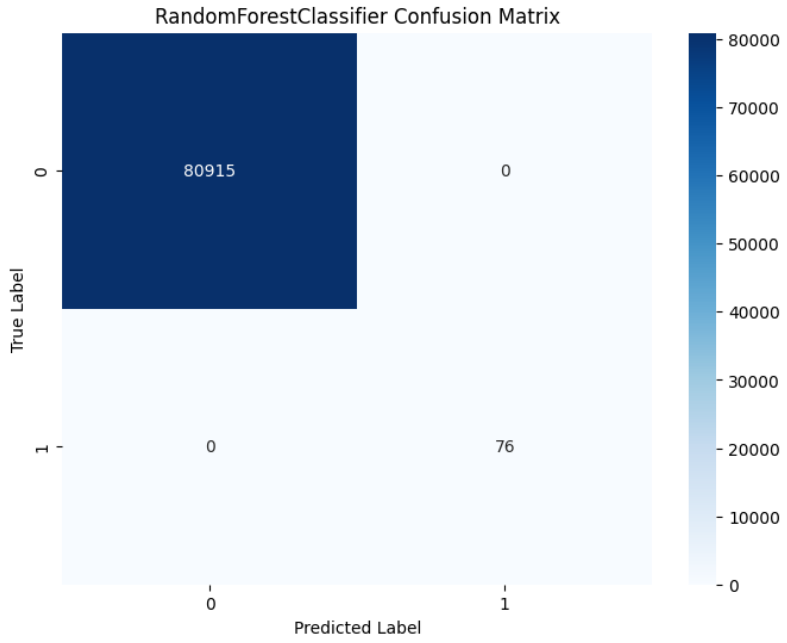


Figure 7: confusion matrix of random forest classifier.

ISSN: 0975-3583,0976-2833 VOL14, ISSUE 10, 2023

## 5. Conclusion

In conclusion, the task of attack detection in IoT sensor data is crucial for safeguarding the security and reliability of IoT systems Collect, clean, and preprocess the IoT sensor data, including feature engineering and data normalization. Divide the dataset into training, validation, and testing sets while ensuring a balanced distribution of normal and attack instances. Train and evaluate machine learning models like LRC and RFC on the training and validation sets. Assess model performance using appropriate metrics, which shows the superiority of RFC model over LRC. Finally, the selected model used to predict attacks in IoT sensor data.

## References

[1] S. Dange and M. Chatterjee, "Iot botnet: The largest threat to the iot network" in Data Communication and Networks, Cham, Switzerland:Springer, pp. 137-157, 2020.

[2] J. Ceron, K. Steding-Jessen, C. Hoepers, L. Granville and C. Margi, "Improving IoT botnet investigation using an adaptive network layer", Sensors, vol. 19, no. 3, pp. 727, Feb. 2019.

[3] Y. Meidan, M. Bohadana, Y. Mathov, Y. Mirsky, A. Shabtai, D. Breitenbacher, et al., "N-baiot-network-based detection of iot botnet attacks using deep autoencoders", IEEE Pervas. Comput., vol. 17, no. 3, pp. 12-22, 2018.

[4] Shah, S.A.R.; Issac, B. Performance comparison of intrusion detection systems and application of machine learning to Snort system. Futur. Gener. Comput. Syst. 2018, 80, 157–170.

[5] Soe YN, Feng Y, Santosa PI, Hartanto R, Sakurai K. Machine Learning-Based IoT-Botnet Attack Detection with Sequential Architecture. Sensors. 2020; 20(16):4372. https://doi.org/10.3390/s20164372

[6] I. Ali et al., "Systematic Literature Review on IoT-Based Botnet Attack," in IEEE Access, vol. 8, pp. 212220-212232, 2020, doi: 10.1109/ACCESS.2020.3039985.

[7] Irfan, I. M. Wildani and I. N. Yulita, "Classifying botnet attack on Internet of Things device using random forest", IOP Conf. Ser. Earth Environ. Sci., vol. 248, Apr. 2019.

[8] Shah, T., Venkatesan, S. (2019). A Method to Secure IoT Devices Against Botnet Attacks. In: Issarny, V., Palanisamy, B., Zhang, LJ. (eds) Internet of Things – ICIOT 2019. ICIOT 2019. Lecture Notes in Computer Science(), vol 11519. Springer, Cham. https://doi.org/10.1007/978-3-030-23357-0_3

[9] C. Tzagkarakis, N. Petroulakis and S. Ioannidis, "Botnet Attack Detection at the IoT Edge Based on Sparse Representation," 2019 Global IoT Summit (GIoTS), Aarhus, Denmark, 2019, pp. 1-6, doi: 10.1109/GIOTS.2019.8766388.

[10] Y. Meidan et al., "N-BaIoT—Network-Based Detection of IoT Botnet Attacks Using Deep Autoencoders," in IEEE Pervasive Computing, vol. 17, no. 3, pp. 12-22, Jul.-Sep. 2018, doi: 10.1109/MPRV.2018.03367731.

[11] S. I. Popoola, R. Ande, B. Adebisi, G. Gui, M. Hammoudeh and O. Jogunola, "Federated Deep Learning for Zero-Day Botnet Attack Detection in IoT-Edge Devices," in IEEE Internet of Things Journal, vol. 9, no. 5, pp. 3930-3944, 1 March1, 2022, doi: 10.1109/JIOT.2021.3100755.

[12] F. Hussain et al., "A Two-Fold Machine Learning Approach to Prevent and Detect IoT Botnet Attacks," in IEEE Access, vol. 9, pp. 163412-163430, 2021, doi: 10.1109/ACCESS.2021.3131014.

ISSN: 0975-3583,0976-2833 VOL14, ISSUE 10, 2023

[13] Abu Al-Haija Q, Al-Dala'ien M. ELBA-IoT: An Ensemble Learning Model for Botnet Attack Detection in IoT Networks. Journal of Sensor and Actuator Networks. 2022; 11(1):18. https://doi.org/10.3390/jsan11010018

[14] Alharbi A, Alosaimi W, Alyami H, Rauf HT, Damaševičius R. Botnet Attack Detection Using Local Global Best Bat Algorithm for Industrial Internet of Things. Electronics. 2021; 10(11):1341. https://doi.org/10.3390/electronics10111341

[15] Ahmed, A.A., Jabbar, W.A., Sadiq, A.S. et al. Deep learning-based classification model for botnet attack detection. J Ambient Intell Human Comput 13, 3457–3466 (2022). https://doi.org/10.1007/s12652-020-01848-9