Gazi Üniversitesi
Fen Bilimleri Dergisi
PART C: TASARIM VE TEKNOLOJİ

Gazi University
Journal of Science
PART C: DESIGN AND TECHNOLOGY

GU J Sci, Part C, 11(2): 321-328(2023)

A new Intrusion Detection System for Secured IoT/IIoT Networks based on LGBM

İlhan Fırat KILINÇER[1] Oğuzhan KATAR[2,*]

[1]*Firat University, Department of Informatics, 23100, Merkez/ELAZIG*

[2]*Firat University, Faculty of Technology, Department of Software Engineering, 23100, Merkez/ELAZIG*

Article Info:

Research article
Received: 09/09/2022
Revision: 01/12/2022
Accepted: 19/04/2023

Keywords

Internet of Things
LGBM
Cyber Security
Intrusion Detection

Makale Bilgisi

Araştırma makalesi
Başvuru:09/09/2022
Düzeltme: 01/12/2022
Kabul: 19/04/2023

Anahtar Kelimeler

Nesnelerin İnterneti
LGBM
Siber Güvenlik
Saldırı Tespiti

Graphical/Tabular Abstract (Grafik Özet)

In this study, a multi-class classification method was applied using various datasets from ToN_IoT and the Light Gradient Boosting Machine (LGBM) classifier, showing that it is an effective method in preventing cyber attacks on IoT/IIoT networks.

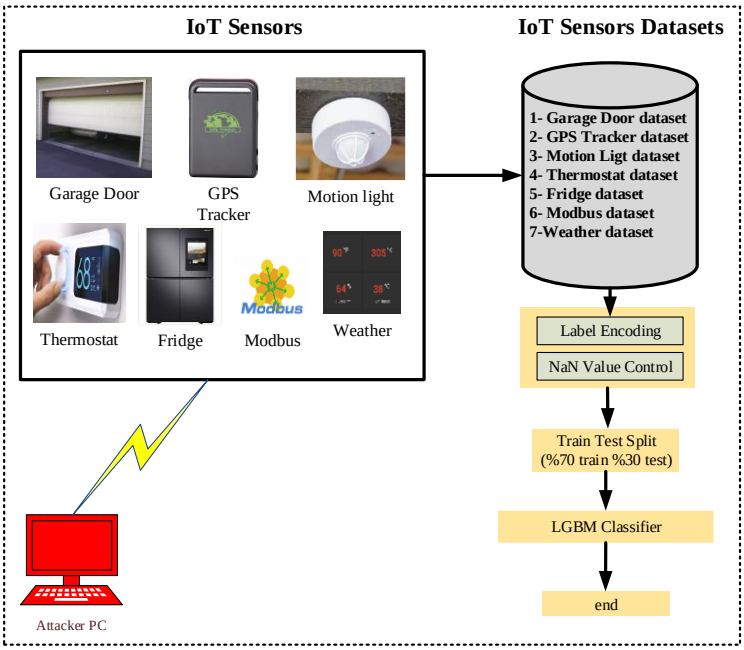


***Figure A**: Proposed method /**Şekil A:**.Önerilen yöntem*

Highlights (Önemli noktalar)

- *The proposed method is effective in preventing cyber attacks on IoT/IIoT networks./Önerilen yöntem, IoT/IIoT ağlarına yönelik siber saldırıların önlenmesinde etkilidir.*
- *The proposed method provided the highest classification performance in the literature./Önerilen yöntem literatürdeki en yüksek sınıflandırma performansını sağlamıştır.*
- *The study uses multiple datasets related to IoT sensors from ToN_IoT, such as fridge, garage door, GPS tracker, modbus, motion light, weather, and thermostat/Çalışma, buzdolabı, garaj kapısı, GPS izleyici, modbus, hareketli ışık, hava durumu ve termostat gibi ToN_IoT'den IoT sensörleriyle ilgili birden çok veri kümesi kullanır.*

***Aim (Amaç):** Automatic detection of cyber attacks against IoT networks./ IoT ağlarına yönelik siber saldırıların otomatik tespiti.*

***Orginality (Özgünlük):** This is the first study to employ multiple sensor datasets and the LGBM classifier./ Çoklu sensör veri setlerini ve LGBM sınıflandırıcısını kullanan ilk çalışmadır.*

***Results (Bulgular):** The proposed method has achieved over 90% accuracy in detecting attacks./Önerilen yöntem, saldırıları tespit etmede %90'ın üzerinde doğruluk elde etmiştir.*

***Conclusion (Sonuç):** The proposed method has shown promising results in automatic detection of cyber attacks against IoT/IIoT networks./ Önerilen yöntem, IoT / IIoT ağlarına yönelik siber saldırıların otomatik olarak tespit edilmesinde umut verici sonuçlar göstermiştir.*

*Corresponding author, e-mail: okatar@firat.edu.tr
DOI: 10.29109/gujsc.1173286

GU J Sci, Part C, 11(2): 321-328(2023)

Gazi Üniversitesi
Fen Bilimleri Dergisi
PART C: TASARIM VE TEKNOLOJİ

Gazi University
Journal of Science
PART C: DESIGN AND TECHNOLOGY

http://dergipark.gov.tr/gujsc

A new Intrusion Detection System for Secured IoT/IIoT Networks based on LGBM

İlhan Fırat KILINÇER[1] Oğuzhan KATAR[2,*]

[1]*Firat University, Department of Informatics, 23100, Merkez/ELAZIG*

[2]*Firat University, Faculty of Technology, Department of Software Engineering, 23100, Merkez/ELAZIG*

Article Info

Araştırma makalesi
Başvuru: 09/09/2022
Düzeltme: 01/12/2022
Kabul: 19/04/2023

Keywords

Internet of Things
LGBM
Cyber Security
Intrusion Detection

Abstract

The Internet of Things (IoT) is one of the technologies used in many fields today. Cyber attacks against IoT/Industrial IoT (IIoT) networks, which are increasingly used thanks to the convenience it provides, are constantly increasing. Detection of attacks against IoT/IIoT networks is one of the popular topics recently. The development of a dataset for IoT applications is essential for the intrusion detection in IoT networks. In this context, the ToN_IoT dataset created in the laboratory of UNSW Canberra (Australia) is one of the most comprehensive datasets that can be used to detect cyber attacks on IoT networks. In this study, fridge, garage door, GPS tracker, modbus, motion light, weather, thermostat datasets related to IoT sensors from ToN_IoT datasets were used. The datasets used were subjected to multi-class classification with the Light Gradient Boosting Machine (LGBM) classifier proposed in the study. The obtained results were compared with the literature and it was seen that the proposed method provided the highest classification performance in the literature. It has been determined that the proposed method is effective in preventing cyber attacks on IoT/IIoT networks.

Güvenli IoT / IIoT Ağları İçin LGBM Tabanlı Yeni Bir Saldırı Tespit Sistemi

Makale Bilgisi

Araştırma makalesi
Başvuru: 09/09/2022
Düzeltme: 01/12/2022
Kabul: 19/04/2023

Anahtar Kelimeler

Nesnelerin İnterneti
LGBM
Siber Güvenlik
Saldırı Tespiti

Öz

Nesnelerin İnterneti (IoT) günümüzde birçok alanda kullanılan teknolojilerden biridir. Sağladığı kolaylıklar sayesinde giderek daha fazla kullanılan IoT/Endüstriyel IoT (IIoT) ağlarına yönelik siber saldırılar sürekli artmaktadır. IoT / IIoT ağlarına yönelik saldırıların tespiti son zamanlarda popüler konulardan biridir. IoT uygulamaları için bir veri kümesinin geliştirilmesi, IoT ağlarında izinsiz giriş tespiti için gereklidir. Bu bağlamda UNSW Canberra (Avustralya) laboratuvarında oluşturulan ToN_IoT veri kümesi, IoT ağlarına yönelik siber saldırıları tespit etmek için kullanılabilecek en kapsamlı veri kümelerinden biridir. Bu çalışmada ToN_IoT veri setlerinden IoT sensörlerine ait buzdolabı, garaj kapısı, GPS takip cihazı, modbus, hareket ışığı, hava durumu, termostat veri setleri kullanılmıştır. Kullanılan veri kümeleri, çalışmada önerilen Light Gradyan Artırma Makinesi (LGBM) sınıflandırıcısı ile çok sınıflı sınıflandırmaya tabi tutuldu. Elde edilen sonuçlar literatür ile karşılaştırılmış ve önerilen yöntemin literatürde en yüksek sınıflandırma performansını sağladığı görülmüştür. Önerilen yöntemin IoT/IIoT ağlarına yönelik siber saldırıların önlenmesinde etkili olduğu belirlenmiştir.

1. INTRODUCTION (GİRİŞ)

Internet of Things (IoT) is a developing technology that is used in many areas such as smart transportation, smart health services, smart home, smart city. The Internet of Medical Things (IoMT), which is the adaptation of IoT to the health sector, and the Industrial Internet of Things (IIoT), which is the adaptation to industrial areas, have created a great revolution. IoT networks, which are formed as a result of connecting many smart devices such as sensors, actuators and smart modules, provide great convenience to users [1,2]. IoT networks and devices used in many areas such as SCADA systems, healthcare services, transportation services are vulnerable to cyber attacks [3].

Intrusion detection in IoT networks is one of the most important problems today. New methods based on artificial intelligence methods are being

*Corresponding author, e-mail: okatar@firat.edu.tr
DOI: 10.29109/gujsc.1173286

developed both in the literature and in the industry for intrusion detection on IoT networks [4]. Alsaedi et al. [5], proposed the ToN_IoT dataset for intrusion detection in IoT and IIoT networks. Their proposed dataset includes telemetry data, system logs and network traffic in IoT and IIoT networks. In addition, they used machine learning and deep learning methods to measure the performance of the ToN_IoT dataset they proposed in their study. Essop et al. [6], produced a new IoT/IIoT dataset using the Cooja simulator. Zachos et al. [7], proposed the Anomaly-based Intrusion Detection Systems (AIDS) system to detect anomalies in IoMT networks. In their proposed method, they used machine learning methods to detect attacks in IoMT networks.

Increasing the performance of machine learning methods, which are frequently used in the intrusion detection on IoT networks, has become an important issue today. Weinger et al. [8], applied their proposed data augmentation method to DS2OS and ToN_IoT datasets for intrusion detection on IoT networks. Bui et al. [9], established a toolchain called Configuration, REproduction, Multi-dataset, and Evaluation (CREME) to increase the intrusion detection capabilities of IDS, and measured both a new dataset and the quality of the dataset they created. Haider et al. [10], proposed the Fuzzy Gaussian Mixture-based Correntropy-Host Anomaly Detection Systems (FGMC-HADS) method based on the Fuzzy Rough Attribute Reduction (FRAR) method and the Gaussian Mixture Model (GMM). They used NGIDS-DS, KDD-98 and ToN_IoT Linux datasets to measure the intrusion detection capability of the proposed method.

In this study, a Light Gradient Boosting Machine (LGBM) based system has been developed that detects cyber attacks on IoT networks with high accuracy. In addition, the ToN_IoT dataset, which comprehensively addresses attacks on today's IoT networks, was used in the study. Other parts of the research are presented as follows. In the second part of the study, the material and method related to the proposed method are discussed. In the third part of the study, the performances of the proposed method in all datasets are presented in a table. Performance analysis, conclusion and future studies are examined in section 4 and 5, respectively.

2.MATERIALS AND METHODS (MATERYAL VE METOD)

In this study, an advanced intrusion detection system is proposed for the detection of attacks on IoT networks. In summary, the proposed method consists of data preprocessing, training and test data separation and classification. The flow chart of the proposed method is given in Figure 1.

According to the flowchart given in Figure 1, the proposed method consists of the following steps.

1. The IoT sensor datasets taken from the ToN_IoT data sets were first standardized by label encoding and NaN value check. If the datasets have NaN values, the NaN values are replaced by the column averages.
2. After the datasets were set to a certain standard, the datasets were separated as 70% training and 30% test data.
3. At the last stage, the datasets were classified with the proposed LGBM classifier. If the desired accuracy rate is achieved as a result of the classification, the process is finished. After the label encoding step, the classes under the "type" label in each dataset were encoded as in Table 1.

The amount of data received for each attack scenario is also listed in Table 1. The data amounts of the attack scenarios in each dataset are visualized in Figure 2.

2.1. Dataset (Veri Seti)

In the study, the ToN_IoT dataset was used to reflect the attacks in a real IoT network. ToN_IoT datasets were collected from data of IoT/IIoT networks. The dataset is generated from operating systems logs and IoT network traffic. The datasets were obtained from a realistic UNSW Canberra IoT lab consisting of cloud layer, edge layer and fog layer. In the datasets obtained, there is the label "Label", which indicates whether a feature is normal or an attack, and the "type" label, which indicates the subclasses of the attacks. Scanning, DoS, DDoS, ransomware, backdoor, data injection, Cross-site Scripting (XSS), password cracking attack and Man-in-The-Middle (MITM) attacks were made under the “type” tag for multi-classification [5,11,12]. In this study, fridge, garage door, gps tracker, motion light, modbus, thermostat and weather datasets obtained from IoT sensor data were studied. Unnecessary "date" and "time" features were removed from all datasets.

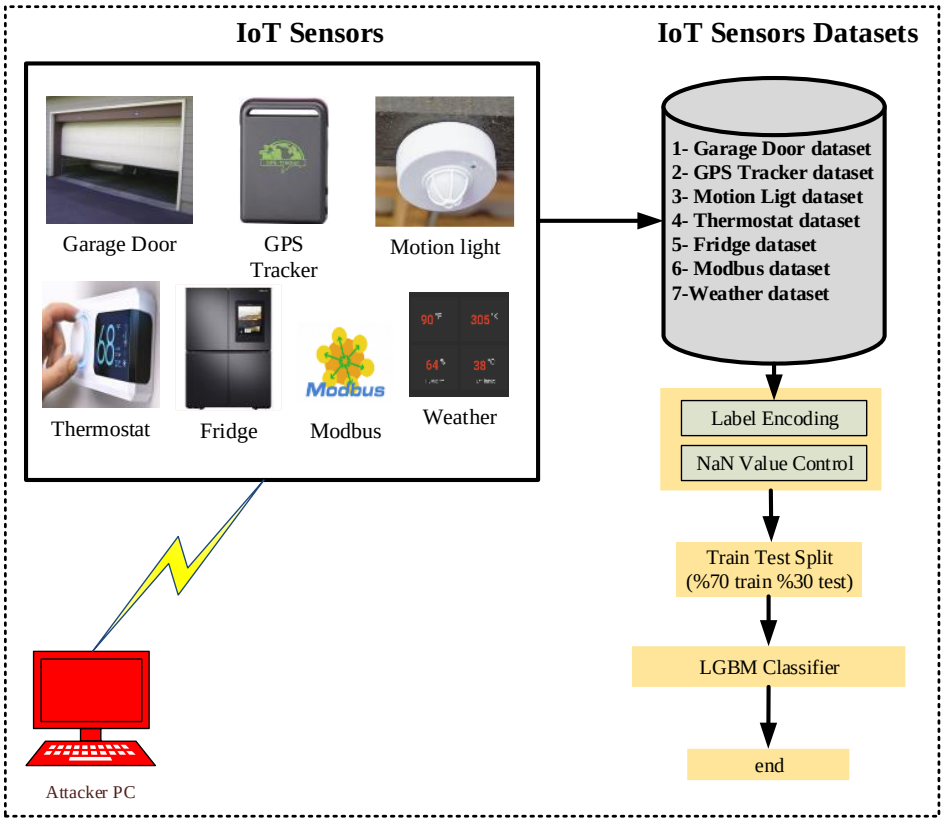
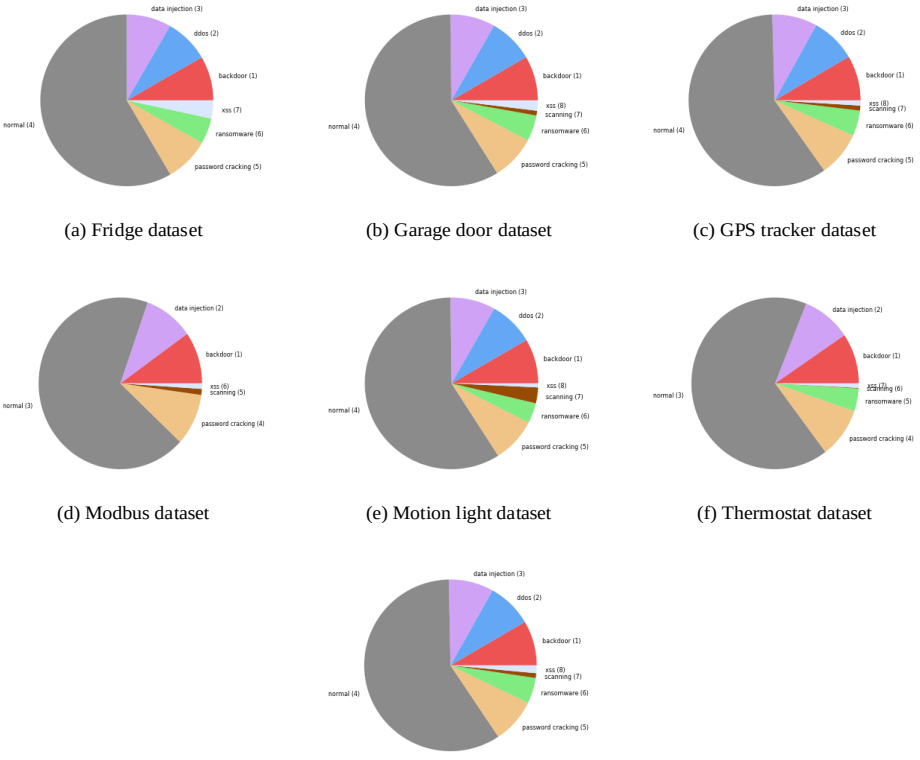


Figure 1. Proposed method flow diagram (Önerilen yöntem akış şeması)

Table 1. The encoded version of the classes in the "type" attribute in the datasets (Veri kümelerindeki "type" özelliğindeki sınıfların kodlanmış versiyonu)

Dataset	Type	Type Number After Label Encoding	Data Count
Fridge	backdoor	1	5000
	ddos	2	5000
	data injection	3	5000
	normal	4	35000
	password cracking	5	5000
	ransomware	6	2902
	xss	7	2042
Garage Door	backdoor	1	5000
	ddos	2	5000
	data injection	3	5000
	normal	4	35000
	password cracking	5	5000
	ransomware	6	2902
	scanning	7	529
	xss	8	1156
GPS Tracker	backdoor	1	5000
	ddos	2	5000
	data injection	3	5000
	normal	4	35000
	password cracking	5	5000
	ransomware	6	2833
	scanning	7	550
	xss	8	577
Modbus	backdoor	1	5000
	data injection	2	5000
	normal	3	35000

	password cracking	4	5000
	scanning	5	577
	xss	6	529
Motion Light	backdoor	1	5000
	ddos	2	5000
	data injection	3	5000
	normal	4	35000
	password cracking	5	5000
	ransomware	6	2264
	scanning	7	1775
	xss	8	449
Thermostat	backdoor	1	5000
	data injection	2	5000
	normal	3	35000
	password cracking	4	5000
	ransomware	5	2264
	scanning	6	61
	xss	7	449
Weather	backdoor	1	5000
	ddos	2	5000
	data injection	3	5000
	normal	4	35000
	password cracking	5	5000
	ransomware	6	2865
	scanning	7	529
	xss	8	866



(a) Fridge dataset (b) Garage door dataset (c) GPS tracker dataset

(d) Modbus dataset (e) Motion light dataset (f) Thermostat dataset

(g) Weather dataset

Figure 2. Class distributions in ToN_IoT datasets, a) Fridge dataset, b) Garage door dataset, c) GPS tracker dataset, d) Modbus dataset, e) Motion light dataset, f) Thermostat dataset, g) Weather dataset (ToN_IoT veri kümelerindeki sınıf dağılımları, , a) Fridge veri seti, b) Garage door veri seti, c) GPS tracker veri seti, d) Modbus veri seti, e) Motion light veri seti, f) Thermostat veri seti, g) Weather veri seti)

2.2. **LGBM Classifier** (LGBM Sınıflandırıcı)

The LGBM classifier, developed to improve the training time performance of the XGBoost algorithm, uses a leaf-wise tree growth strategy. The leaf-wise growth method used by LGBM is summarized in Figure 3 [13,14].

In the Leaf-wise growth strategy shown in Figure 3, the decision trees try to open the tree vertically as far as they can go, when the maximum depth is achieved, it starts to open the other branch vertically from the top. In this study, LGBM classifier was used for classification due to its high performance.

3. EXPERIMENTAL RESULTS (Deneysel BULGULAR)

In the study, sensor datasets from ToN_IoT datasets are discussed. Multi-class classification was made according to the "type" parameter in the considered datasets. The codes of the proposed method were written using the scikit-learn, matplotlib libraries in Python 3.7 environment. Accuracy, precision, recall, F-Score values were calculated for each dataset in the study. Performance metrics calculated as equations 1, 2, 3 and 4 respectively [11,15]. The results obtained are given in Table 2.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F-Measure = \frac{2TP}{2TP+FP+FN} \quad (4)$$

Confusion matrices obtained for all datasets are given in Figure 4. The numbers in the matrices show the encoded classes in Table 1.

4. DISCUSSION (TARTIŞMA)

In this section, the performance of the proposed method in the study is compared with the existing studies in the literature. The comparative analysis made is given in Table 3. Only the Accuracy values are compared in the table.

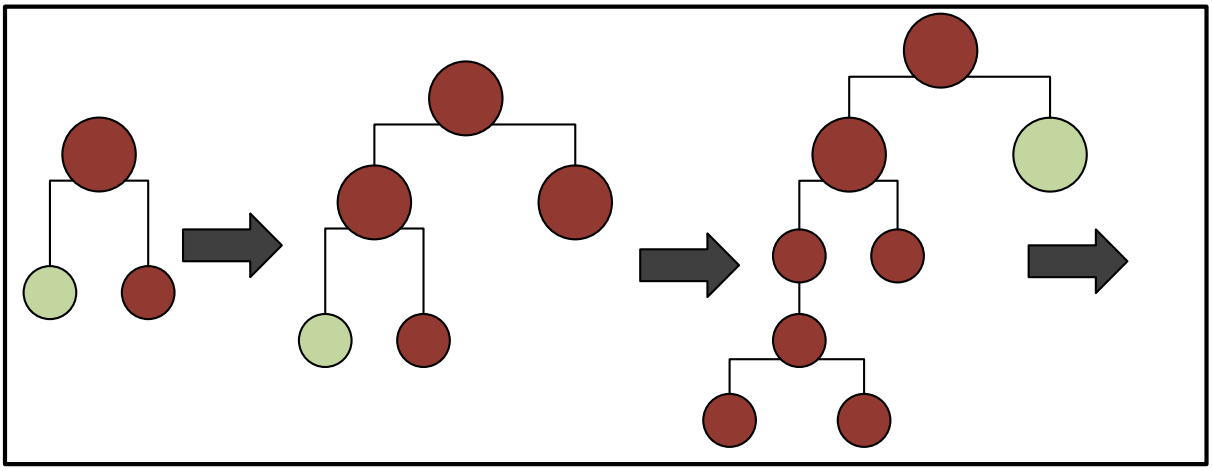


Figure 3. LGBM leaf-wise growth strategy (LGBM yaprak bazında büyüme stratejisi)

Table 2. Proposed method performance evaluation (%) (Önerilen yöntemin performans değerlendirmesi (%))

Datasets	Model	Accuracy	Precision	Recall	F-Score
Fridge	LGBM	99.68	100	100	100
Garage Door	LGBM	99.75	98	100	99
GPS Tracker	LGBM	99.977	100	100	100
Modbus	LGBM	100	100	100	100
Motion Light	LGBM	97.79	98	98	97
Weather	LGBM	99.971	100	100	100
Thermostat	LGBM	92.27	89	92	90

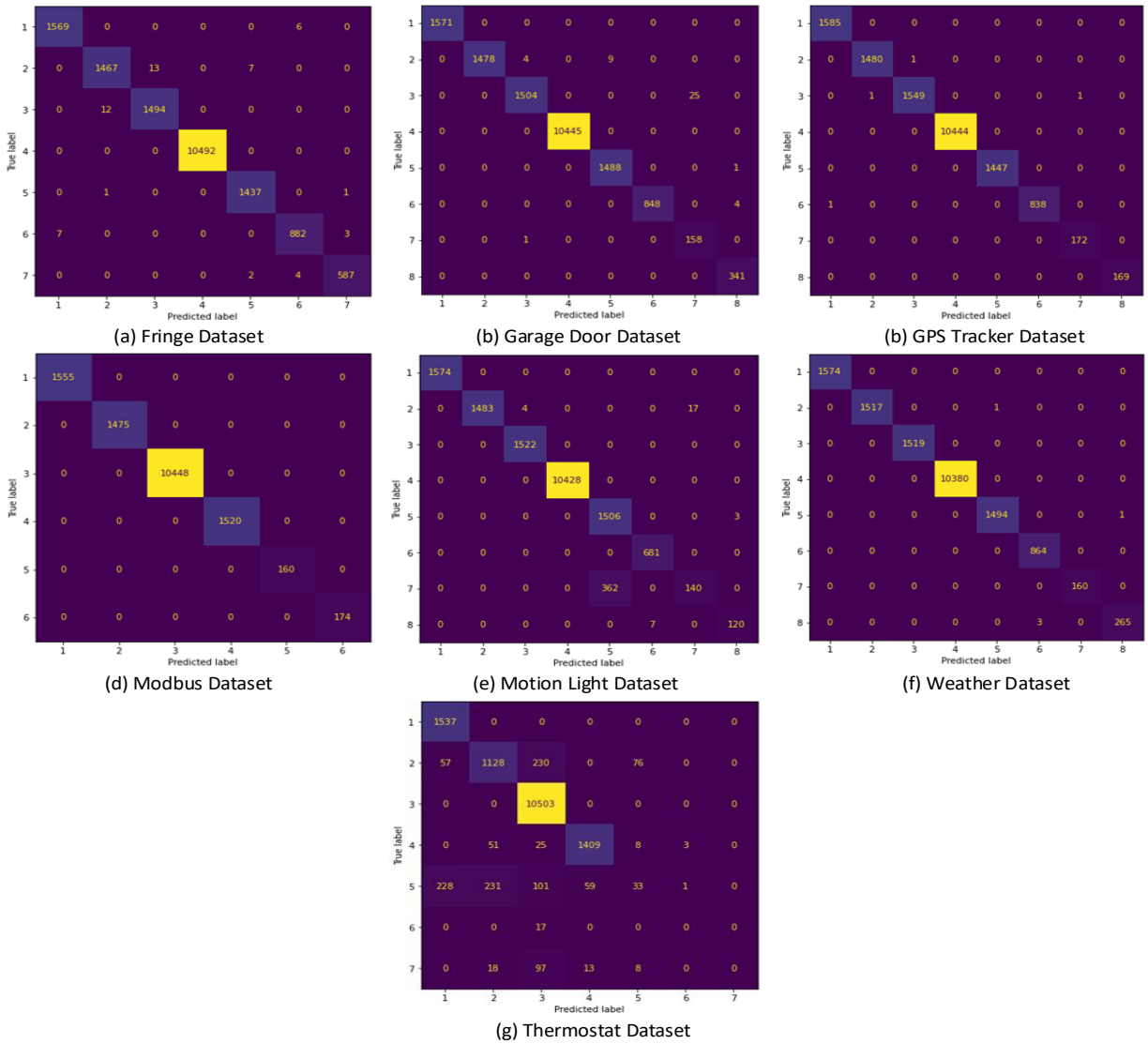


Figure 4. Confusion matrices for each results, a) Fridge dataset, b) Garage door dataset, c) GPS tracker dataset, d) Modbus dataset, e) Motion light dataset, f) Thermostat dataset, g) Weather dataset (Her sonuç için karışıklık matrisleri, a) Fridge veri seti, b) Garage door veri seti, c) GPS tracker veri seti, d) Modbus veri seti, e) Motion light veri seti, f) Thermostat veri seti, g) Weather veri seti)

Table 3. Proposed method and literature comparison (Önerilen yöntem ve literatür karşılaştırması)

References	Datasets	Model	Accuracy
[5]	Fridge	LSTM	100
	Garage Door	LSTM	100
	GPS Tracker	kNN	88
	Modbus	CART	98
	Motion Light	LSTM	59
	Weather	CART	87
	Thermostat	LSTM	66
Proposed Method	Fridge	LGBM	99.68
	Garage Door	LGBM	99.75
	GPS Tracker	LGBM	99.977
	Modbus	LGBM	100
	Motion Light	LGBM	97.79
	Weather	LGBM	99.971
	Thermostat	LGBM	92.27

The proposed method was applied to the "Train_Test_IoT_dataset" dataset, which is one of the ToN_IoT datasets. Since the number of studies with this dataset is limited in the literature, only one study could be compared. According to Table 2, the proposed method provided a performance similar to the literature in the fridge and garage door datasets with approximately 100% accuracy values. While an accuracy rate of 88% was obtained in the literature for the gps tracker dataset, the proposed method reached an accuracy of 99.977% for this dataset. While the values of 98%, 59%, 87% and 66% were obtained, respectively, in the literature for modbus, motion light, weather, and thermostat datasets, the proposed method reached 100%, 97.79%, 99.971% and 92.27% for these datasets, respectively. As a result, the proposed method provided high performance for all datasets.

5.CONCLUSIONS (SONUÇLAR)

In this study, a method has been proposed for the detection of cyber attacks on IoT/IIoT networks that we encounter in almost every field. The proposed method has been applied to the ToN_IoT dataset, which represents a realistic IoT/IIoT network. The datasets used include many attack vectors that are frequently encountered today, such as scanning, DoS, DDoS, ransomware, backdoor, data injection, XSS, password cracking attack and MITM. With the LGBM classifier suggested in the study, fridge, garage door, gps tracker, modbus, motion light, weather, thermostat datasets were classified according to the "type" parameter. Performance values of 99.68%, 99.75%, 99.97%, 100%, 97.79%, 99.97% and 92.27% were reached for the datasets, respectively. The proposed method has achieved very high performances in detecting attacks used in the specified datasets.

In future studies, it is planned to obtain a new IoT dataset using the Cooja simulator in the first stage. In the second stage, it is aimed to establish a new IoT laboratory and to create a new IDS dataset for IoT networks by applying various attack scenarios to this IoT laboratory to be established.

DECLARATION OF ETHICAL STANDARDS (ETİK STANDARTLARIN BEYANI)

The author of this article declares that the materials and methods they use in their work do not require ethical committee approval and/or legal-specific permission.

Bu makalenin yazarı çalışmalarında kullandıkları materyal ve yöntemlerin etik kurul izni ve/veya yasal-özel bir izin gerektirmediğini beyan ederler.

AUTHORS' CONTRIBUTIONS (YAZARLARIN KATKILARI)

***İlhan Fırat KILINÇER*:** He conducted the experiments, analyzed the results and performed the writing process.

Deneyleri yapmış, sonuçlarını analiz etmiş ve maklenin yazım işlemini gerçekleştirmiştir.

***Oğuzhan KATAR*:** He conducted the experiments, analyzed the results and performed the writing process.

Deneyleri yapmış, sonuçlarını analiz etmiş ve maklenin yazım işlemini gerçekleştirmiştir.

CONFLICT OF INTEREST (ÇIKAR ÇATIŞMASI)

There is no conflict of interest in this study.

Bu çalışmada herhangi bir çıkar çatışması yoktur.

REFERENCES (KAYNAKLAR)

[1] Nandy S., Adhikari M., Khan M. A., Menon V. G., Verma S., An intrusion detection mechanism for secured IoMT framework based on swarm-neural network, IEEE Journal of Biomedical and Health Informatics, 26 (2021) 1969-1976.

[2] Ahmad J., Shah S. A., Latif S., Ahmed F., Zou Z., Pitropakis N., DRaNN_PSO: A deep random neural network with particle swarm optimization for intrusion detection in the industrial internet of things, Journal of King Saud University-Computer and Information Sciences,(2022).

[3] Lu K. D., Zeng G. Q., Luo X., Weng J., Luo W., Wu Y., Evolutionary deep belief network for cyber-attack detection in industrial automation and control system, IEEE Transactions on Industrial Informatics, 17 (2021) 7618-7627.

[4] Campos E. M., Saura P. F., González-Vidal A., Hernández-Ramos J. L., Bernabe J. B., Baldini G., Skarmeta A., Evaluating Federated Learning for intrusion detection in Internet of Things: Review and challenges, Computer Networks,(2021).

[5] Alsaedi A., Moustafa N., Tari Z., Mahmood A., Anwar A., TON_IoT telemetry dataset: A new generation dataset of IoT and IIoT for data-driven intrusion detection systems, IEEE Access, 8 (2020) 165130-165150.

[6] Essop I., Ribeiro J. C., Papaioannou M., Zachos G., Mantas G., Rodriguez J., Generating datasets for anomaly-based intrusion detection systems in iot and industrial iot networks, Sensors, 21 (2021) 1528.

[7] Zachos G., Essop I., Mantas G., Porfyrakis K., Ribeiro J. C., Rodriguez J., An anomaly-based intrusion detection system for internet of medical things networks, Electronics, 10 (2021) 2562.

[8] Weinger B., Kim J., Sim A., Nakashima M., Moustafa N., Wu K. J., Enhancing IoT anomaly detection performance for federated learning, Digital Communications and Networks,(2022).

[9] Bui H. K., Lin Y. D., Hwang R. H., Lin P. C., Nguyen V. L., Lai Y. C., CREME: A toolchain of automatic dataset collection for machine learning in intrusion detection, Journal of Network and Computer Applications, 193 (2021) 103212.

[10] Haider W., Moustafa N., Keshk M., Fernandez A., Choo K. K. R., Wahab A., FGMC-HADS: Fuzzy Gaussian mixture-based correntropy models for detecting zero-day attacks from linux systems, Computers & Security, 96 (2020) 101906.

[11] Gad A. R., Nashat A. A., Barkat T. M., Intrusion detection system using machine learning for vehicular ad hoc networks based on on ToN-IoT dataset, IEEE Access, 9 (2021) 142206-142217.

[12] Idrissi I., Azizi M., Moussaoui, O., Accelerating the update of a DL-based IDS for IoT using deep transfer learning, Indones. J. Electr. Eng. Comput. Sci., 23 (2021) 1059-1067.

[13] Zhang Z., Zhang Y., Guo D., Song, M., A scalable network intrusion detection system towards detecting, discovering, and learning unknown attacks, International Journal of Machine Learning and Cybernetics, 12 (2021) 1649-1665.

[14] Al Daoud E., Comparison between XGBoost, LightGBM and CatBoost using a home credit dataset, International Journal of Computer and Information Engineering, 13 (2019) 6-10.

[15] Mohindru G., Mondal K., Banka, H., Different hybrid machine intelligence techniques for handling IoT-based imbalanced data, CAAI Transactions on Intelligence Technology, 6 (2021) 405-416.