Research Article

Machine Learning Models and Explainable Artificial Intelligence Approaches for Intrusion Detection in IoT Networks

Asuman Besi Kütük[1*], Özlem Coşkun[2], Hikmet Kütük[3], İbrahim Kök[4]

[1] PTT Bilgi Teknolojileri A.Ş., Orcid ID: https://orcid.org/0000-0002-7600-8813 ,
e-mail: asuman.besi@pttteknoloji.com.tr
[2] PTT Bilgi Teknolojileri A.Ş., Orcid ID: https://orcid.org/0009-0000-9413-9414 ,
e-mail:ozlem.coskun@pttteknoloji.com.tr
[3]PTT Bilgi Teknolojileri A.Ş., Orcid ID: https://orcid.org/0009-0008-1455-5484 ,
e-mail:hikmet.kutuk@pttteknoloji.com.tr
[4] Ankara University, Orcid ID: https://orcid.org/0000-0001-9787-8079 ,
e-mail:ikok@ankara.edu.tr
[*] Correspondence: asuman.besi@pttteknoloji.com.tr; Tel.: (+90 506 146 64 50)

Received: 17 October 2024
Revised: 15 May 2025
Accepted: 27 May 2025
Published: 31 May 2025

Reference: Besi Kütük, A., Coşkun, Ö., Kütük, H., & Kök, İ. (2024). Machine learning models and explainable artificial intelligence approaches for intrusion detection in IoT networks. The European Journal of Research and Development, 5(1), 17-33.

Abstract

The rapid spread of Internet of Things (IoT) technologies and the rapidly increasing use of IoT devices offer technological transformation and innovative solutions in many areas from daily life to industrial processes. However, the resource constraints, simple operating systems, non-standard protocols and embedded software of IoT devices make them vulnerable to cyber-attacks. This makes IoT networks risky against malicious attacks and increases the size of security threats. Moreover, the complexity and heterogeneity of IoT networks render traditional security approaches inadequate and increase the need for advanced solutions. In this context, the need for methods for detecting and preventing attacks on IoT networks that are not only reliable and effective, but also understandable by users and security experts has become increasingly critical. This need for network security necessitates the development of strategies that will both secure

technical infrastructures and increase the trust of human elements interacting with these infrastructures. In this context, the need for more interpretable, explainable and transparent security approaches is increasing. In particular, machine learning (ML) and deep learning (DL) based intrusion detection systems offer effective solutions to security problems such as anomaly detection and attack classification. The comprehensibility of the decision mechanisms of the models used enables both security experts to manage the systems more effectively and users to have more confidence in the security measures taken. Explainable Artificial Intelligence (XAI) techniques make the decision processes of ML and DL models transparent, allowing to understand how and why attacks are detected. Accordingly, it has become a critical requirement for security systems not only to achieve high accuracy rates, but also to make the decisions taken interpretable. In this study, the effectiveness of artificial intelligence (ML and DL) techniques for the detection and classification of security threats in IoT networks is analysed. In addition, the applications of XAI methods such as SHapley Additive exPlanations (SHAP), Local Interpretable Model-Agnostic Explanations (LIME) and Explain Like I'm 5 (ELI5) for IoT security are investigated. It is shown how these methods make the decision processes of ML and DL models used in IoT networks more transparent and provide a better analysis. As a result, this study presents an approach that combines both performance and explainability in IoT security. By demonstrating the effectiveness of XAI-supported ML and DL models, it aims to contribute to future research and innovative security solutions for enhancing security in IoT networks.

Keywords: *Intrusion Detection, Machine Learning, Deep Learning, Explainable Artificial Intelligence*

1. Introduction

The Internet of Things (IoT), also known as the Internet of Things, is a technology that enables networked interactions of objects that we frequently encounter in daily life [1]. Offering new features and services, IoT has become a rapidly developing trend. According to estimates, the number of connected devices in use is expected to reach 75 billion by 2025 [2]. IoT integrates every object to interact through embedded systems, thus extending the internet and creating a distributed network of devices that enables it to communicate with other devices as well as with people [1]. Smart devices sense changes in their environment and interact with each other, then analyze the results using internet infrastructure and standards. The emergence of Radio Frequency Identification (RFID) and sensor network technologies has been cited as the beginning of the IoT paradigm [2]. The term was first used by Kevin Ashton in 1999 [3]. "The Internet of Things (IoT), as an emerging paradigm, enables objects to connect anytime, anywhere, with anything and anyone, ideally through every path, every network, and every service" [4]. IoT is widely

used in different application areas. These application areas can be listed as vehicles, health, finance-banking, manufacturing, information and communication technologies, education, pharmaceutical industry, logistics, energy, public security, military, transportation, food and agriculture, entertainment, aviation, sports, government, environment [3]. While IoT networks support modern life with the huge volume of data they generate, they also face serious security threats. The often limited computational capacity of these devices, their heterogeneous nature and the need for uninterrupted online connectivity make IoT networks vulnerable to cyber-attacks. In this context, the need for effective and reliable intrusion detection systems (Intrusion Detection-IDS) to ensure the security of IoT networks is increasing academically and technically.

Traditional rule-based methods used in intrusion detection are inadequate in the face of the dynamic and complex nature of IoT networks, leading to the rise of artificial intelligence-based approaches, especially machine learning (ML) and deep learning (DL) techniques. Machine learning offers a promising solution for intrusion detection in IoT networks thanks to its ability to detect unknown threats by learning patterns from large datasets. Deep learning, with its capacity to analyze complex data relationships, provides further success in this field. For example, Ge et al. demonstrated that deep learning-based intrusion detection models in IoT networks offer higher accuracy rates compared to traditional methods [5]. Similarly, Saheed et al. stated that ML techniques are effective in detecting attacks in IoT networks [6].

However, the complexity and "black box" nature of artificial intelligence models makes it difficult to understand the decision-making processes of these systems and puts their reliability into question. Explainable Artificial Intelligence (XAI) approaches, developed as a solution to this problem, have become a field that has attracted attention in recent years. XAI aims to both increase user confidence and strengthen the practical applicability of systems by presenting the decisions of models in a human-understandable way. However, the complexity and "black box" nature of artificial intelligence models makes it difficult to understand the decision-making processes of these systems and puts their reliability into question. Explainable Artificial Intelligence (XAI) approaches, developed as a solution to this problem, have become a field that has attracted attention in recent years. XAI aims to both increase user confidence and strengthen the practical applicability of systems by presenting the decisions of models in a human-understandable way. Sharma et al. stated that XAI-supported deep learning methods in IoT networks not only provide high accuracy but also transparency [7]. In another study, Kök et al. emphasized that XAI-based malicious traffic detection plays a critical role in protecting sensitive data in next generation IoT-based healthcare systems [8].

2. Internet of Things Architecture and Fundamentals

2.1. IoT Architecture

The layers that make up the IoT are defined by the function of the layer and the devices used in the layer. In the literature, there are IoT layer structures from three layers to seven layers according to the application needs. It is commonly seen that a three-layer architecture structure is used as Sensing Layer, Network Layer and Application Layer [9][10]. Accordingly, the tasks of the layers in the three-layer architecture are given below.

2.1.1. Perceptron Layer

This layer consists of heterogeneous devices (sensors, actuators, etc.) of different shapes and forms. The devices sense and transmit information about their environment to the network layer [9].

2.1.2. Network Layer

The network layer, which aims to transfer data between different networks, is responsible for routing and transmitting data over the internet to different IoT hubs and devices. In this layer, cloud platforms, internet gateways, switching-routing devices, etc. It works using technologies such as WiFi, Bluetooth, LTE, Zigbee, 3G, etc. Gateways collect and filter data from sensors and forward it to different sensors, thus acting as intermediaries between IoT nodes [11].

2.1.3. Application Layer

This is the layer where IoT applications are deployed through middleware. The application layer ensures the authenticity, integrity and confidentiality of the data. A smart environment is created at this layer where IoT achieves its purpose [9][10].

2.2. IoT Network Attacks

Despite the ever-evolving Internet security architecture, there are many types of attacks. Network-based attack types are classified under several specific headings [12].

1) Denial of Service (DoS) Attack: A DoS attack is an action that reduces or eliminates the ability of a function to occur on a network. This threat prevents or denies the user's resource from being created on the network by creating unwanted, unnecessary traffic

on the network. Hardware failures, resource exhaustion, software bugs, environmental causes can cause DoS [12][13].

2) Distributed Denial of Service (DDoS): This type of attack aims to overload the target system with large amounts of traffic from multiple networks or devices. These attacks are carried out simultaneously from many different devices, which makes the attack more challenging and difficult to detect [12].

3) Malicious Software: The term malware is created by combining the terms malicious and software. This term is used to refer to unwanted software. An attacker can exploit flaws in the firmware running on devices and run their software to disrupt the IoT architecture. Executable code is used by attackers to intercept and disrupt devices in the IoT network fabric [14]. It is defined by G. McGraw and G. Morrisett as "any code that is added to, modified, or removed from a software system in order to disrupt the intended function of the system or intentionally cause harm" [15]. This virus is referred to as "a general term encompassing viruses, trojans, spyware and other intrusive code" [16].

4) Data Breaches: The type of attack, a data breach incident, results in the confidentiality, integrity and availability of data being compromised. This attack involves attackers gaining unauthorized access to protected, sensitive or confidential data. This data consists of personally identifiable information such as individuals' health, business, and financial information [17].

5) Weakening Perimeters: The design of IoT network devices does not include commonly used security measures. Therefore, devices that prevent network threats do not have network security mechanisms [12].

2.3. Machine Learning and Deep Learning Models Used in Intrusion Detection

Machine learning models are divided into supervised and unsupervised learning [18]. Supervised learning re-feeds data and the results obtained from this data to the machine, allowing the machine to extract a function (a match between input data and result data) from this information. The machine thus learns the relationship between the data. It extracts a function from a set of training examples generated by labeled training data [19]. Through the supervised learning algorithm, the training data is analyzed and a function is created that can be used to match new examples [18]. Unsupervised learning is used for training data where the algorithm is not assigned a label or score in advance [20].

Deep learning is defined as a sub-branch of artificial intelligence and machine learning and is an approach that aims to model the learning mechanisms of the human brain. Deep

learning has the capacity to produce predictions with high accuracy rates by automatically extracting features from large datasets with multi-layered structures (neural networks). Deep learning models are built on systems of interconnected neurons. These networks are capable of learning complex patterns and relationships by processing data across layers. Deep learning requires large data volumes and high computational power to be implemented effectively [18].

2.3.1. Logistic Regression (LR)

Regression, which corresponds to the logit natural logarithm of a probability ratio, is the basic mathematical concept of logistics. It is expressed by considering the case of matching a binary outcome distribution with a binary prediction variable expressed in a binary form in a regression model. Logistic regression is a convenient model used to describe and test the relationship between a categorical outcome variable and one or more categorical variables or the relationship between a categorical outcome and a continuous prediction variable [21].Logistic regression is also referred to as modeling a chance that will result in the influence of individual characteristics. Since chance is also a ratio, it actually expresses the logarithm of the chance to be modeled [22].Logistic regression is classified as Single-Variate Logistic Regression and Multi-Variate Logistic Regression [23].

2.3.2. Decision Tree (DT)

Decision Tree classification, one of the possible approaches to any multi-stage decision making, breaks a complex decision into a combination of simple decisions and expects the resulting solution to be similar to the desired solution [24]. Once the initial state of the root node decision tree containing all the instances in the training set has been created, no further decision is needed if it is seen that all instances belong to the same class, the solution has been reached but if the node belongs to two or more classes, the splitting process is repeated for the intermediate nodes until a fully discriminative tree is obtained[25].

The advantages of decision trees include the fact that they are nonparametric and do not require any assumptions about the distribution of input data. They can also handle numeric and categorical inputs, can be used in non-linear relationships with classes and their attributes, and can accept missing values [26].

2.3.3 Random Forest (RF)

A random forest consists of many decision trees that have lower classification error compared to other traditional classification algorithms. One of the most important advantages of random forests is that the generated forests can be saved and referenced in the future [27]. In a random forest classifier, which is a combination of tree classifiers, the classifiers are generated independently from the input vector using a random vector and each tree is classified by the input vector by voting for the most popular class [28].

2.3.4. Support Vector Machine (SVM)

Support Vector Machines (SVM) is a widely used, general-purpose and practical classification algorithm; it is particularly effective in binary classification problems [29]. In the SVM algorithm, a hyperplane is used to separate positive and negative class variables, and this separation is based on the Structural Risk Minimization value [30]. SVM offers a strong generalization capability, is robust to local minima and is defined by a small number of parameters. The reason SVM is robust to local minima is that the algorithm focuses on finding a global solution that fits the entire dataset, rather than only looking for small and suboptimal solutions during the training process. This approach increases the generalization capacity of the model and allows it to achieve accurate results without getting bogged down in local minima. In addition, the SVM is represented by a small number of parameters because the model generates its parameters using only the "support vectors" that determine the classification decision, which makes it possible to obtain efficient and effective results with fewer computational resources [31].

2.3.5. Multi Layer Perceptron (MLP)

A multilayer perceptron, a deep learning method, is a modern feed-forward neural network consisting of fully connected neurons. In this network structure, neurons are organized in layers using non-linear activation functions. The most remarkable feature of this method is its ability to discriminate data that cannot be linearly decomposed [32].

2.3.6. Convolutional Neural Networks (CNN)

Convolutional Neural Networks are a sub-branch of deep learning and are particularly effective in analyzing visual information. These networks use a mathematical process called convolution to detect local patterns in input data and automatically extract salient

features. The basic components of Convolutional Neural Networks include convolution layers, activation functions and pooling layers. Thanks to this structure, they can achieve high accuracy rates in tasks such as image processing, classification and segmentation [33].

2.3.7. Recurrent Neural Network (RNN)

In a feedback neural network or back propagation algorithm, the connections between nodes form a sequential and coordinated graph structure. This structure allows feedback neural networks to exhibit dynamic behavior over time sequences. In this context, Recurrent Neural Network refers to a type of neural network that expands over time. Unlike traditional neural networks, in RNNs, data is transmitted to the next step in time instead of being passed to the next layer; this allows RNNs to work with time-dependent data [34].

RNNs were developed to recognize and process sequential data such as text or speech signals. The structure of these networks includes loops, which represent short-term memory. Thanks to these loops, the network can influence the output of future steps by memorizing information from previous time steps. Rather than a hierarchical structure, they offer a powerful learning capability on sequence data with their structure that progresses over time and iteratively processes information [35].

2.4. Explainable Artificial Intelligence (XAI)

Explainable Artificial Intelligence (XAI) is being explored because machine learning (ML) and deep learning (DL) models are opaque and difficult to understand, making it difficult to interpret model predictions. Explainable AI explains the predictions of the model, thus increasing the transparency and confidence of the model. The idea of Explainable AI is a new concept and involves the use of model annotation techniques to describe the generated models and explain the contributions of each feature on the prediction [36].

2.4.1. Local interpretable model-agnostic explanations (LIME):

LIME is a method that describes the predictions of any classifier. It provides the user with interpretations of the model and these interpretations explain the prediction made on a given example. With this model, a new dataset is created containing more understandable representations of the original data and a local explainable model is trained on this set. These models, often known as “white-boxes”, can be interpretable

structures such as linear models or decision trees. Thus, LIME makes the decisions of complex machine learning models more transparent and understandable [37].

2.4.2. Shapley additive explanations (SHAP):

SHAP is a method that describes the model using Shapley values based on feature importance. In the formula for extracting Shapley values, Equation (1) is used to calculate the contribution of feature j to the SHAP value for a data point [38]:

$$\varphi(i) = \Sigma S \subseteq F \not j((|S|!(|P| - |S| - 1)!) \not| P |!)(v(S \cup j) - v(S)) \quad \text{Equation (1)}$$

Here, P represents all features in the dataset and S represents the set of all features except feature j. v(x) is the contribution of subset x.

2.4.3. Explain Like Im 5 (ELI5)

ELI5 is an XAI method for providing attribute importance ranking, prediction annotations and model insights, which can work on different types of models such as linear models, decision trees and deep learning by supporting model-agnostic approaches. It provides visual tools and descriptions to explain model decisions in a simple and understandable way [39].

2.4.4. Integrated Gradients (IG)

Integrated gradients (IGs) are an attribution-based XAI method often used to identify words that strongly influence the target prediction. Citation-based methods use DNN (Deep Neural Network) predictions to estimate the importance of input features, revealing the reasons for model predictions [40].

3. Artificial Intelligence and Explainable Artificial Intelligence for Intrusion Detection in IoT Networks

Literature surveys have shown that intrusion detection in IoT networks has become more complex over time and has evolved to methods that provide higher accuracy. While rule-based systems were preferred in the early days, today, advanced technologies such as machine learning, deep learning and explainable artificial intelligence (XAI) stand out. These advancements provide more effective solutions against new security threats with the increase in the number and diversity of IoT devices.

In a study, a CNN-based Intrusion Detection (IDS) model is proposed. The model was used to detect intrusions in the network and tested on the KDD Cup 99 dataset. The results demonstrate the effective use of CNN in the field of network security [41]. In a study comparing deep learning models, the methods were tested on the open source UNSW-NB15 and NSL-KDD datasets. The study evaluated the performance of each model and stated that the DNN model provided a high accuracy rate [42]. A hybrid deep learning model was developed for DDoS attack detection in SDN-based IoT environments. It was reported that the model achieved 99% prediction accuracy. In this study, an effective detection system against DDoS attacks was proposed [43]. In another study, a deep learning-based intrusion detection model was proposed and evaluated on the NSL-KDD dataset. The model examined deep learning techniques to detect attacks in the network and the accuracy rate of the model was reported to be 78%. The study analyzed the performance of deep learning-based models on the NSL-KDD dataset [44]. A model combining CNN and RNN models was developed to detect DoS (Denial of Service) attacks. The model achieved approximately 99% accuracy on the KDD Cup dataset and 91% accuracy on the CICIDS2018 dataset. As a result of the study, it was seen that hybrid models provide high success in intrusion detection [45]. In another study, a DNN-based intrusion detection system was developed and tested on the UNSW-NB15 dataset. It was reported that the model achieved an accuracy rate of 77.16% with the multi-class classification method [46]. In a study using big data analytics and Random Forest (RF) model for real-time DDoS attack detection in IoT devices, the RF model applied together with big data analytics aimed to accurately detect DDoS attacks and demonstrated high accuracy with a success rate of over 99.5% [47]. Another study proposed a solution to detect network traffic anomalies in IoT networks using CNN-LSTM hybrid model. The study improved IoT security using deep learning-based hybrid approaches and achieved 97% accuracy [48]. A hybrid deep learning model was developed to detect DDoS attacks in SDN-based IoT environments. It is reported that the model provides 99% prediction accuracy [49]. In this study, which aims to detect attacks in IoT networks by proposing a feedforward neural network (FNN) based model, the model aims to detect anomalies in the network using FNN. The study was tested on the BoT-IoT dataset and achieved over 99% success. This result shows the effectiveness of deep learning methods for detecting network attacks in IoT devices [50]. We analyzed the performance of machine learning based intrusion detection systems in IoT networks. In tests using machine learning classifiers, it was observed that the GB classifier achieved 99.9944% success. The study examines how different machine learning algorithms perform in IoT security [51].

Although machine learning-based methods provide high accuracy in intrusion detection, the understandability of decision mechanisms is still an important research topic. For this reason, Explainable Artificial Intelligence (XAI) approaches have been used to make model decisions more transparent and to explain the susceptibility of IoT devices to network attacks. This paper examines XAI techniques for IoT security from the work done in this area. In particular, it addresses how explainable AI methods such as SHAP and LIME can be integrated into Intrusion Detection (IDS) systems. These methods increase the reliability and transparency of intrusion detection models used in IoT networks [52]. In another study, an intrusion detection system that preserves privacy in IoT networks using Explainable Artificial Intelligence (XAI) is proposed. This approach aims to securely analyze the data of IoT devices without sending it to a central server. It is emphasized that XAI makes decision processes explainable in intrusion detection [53]. In another recent study, a malicious network traffic detection and monitoring system was proposed in IoT-based healthcare systems using explainable artificial intelligence (XAI). The study aims to provide security, especially in next generation IoT health systems [8].

Tablo 1. Recent studies on IoT security

Reference	**Year**	**Research/ Problem Area**	**Machine Learning Model**	**XAI Approach**	**Dataset**
Xiao et al.	2019	Intrusion Detection (IDS)	CNN	-	KDD Cup 99
S.Nagisetty ve S. Gupta	2019	Deep Learning Comparison	DNN	-	UNSW-NB15, NSL-KDD
Liang et al.	2019	DDoS Attack Detection	Hybrid Deep Learning Model	-	NSL-KDD
Vinayakumar et al.	2019	Attack Detection	Deep Learning	-	NSL-KDD
Kim et al.	2020	DDoS Attack Detection	CNN-RNN Hybrid	-	KDD Cup, CICIDS2018

A. Kasongo ve H. Sun,	2020	Attack Detection	DNN	-	UNSW-NB15
Awan et al.	2021	DDoS Attack Detection	Random Forest	-	-
Sahu et al.	2021	IoT Security	CNN-LSTM Hybrid	-	-
Ge et al.	2021	DDoS Attack Detection	Hybrid Deep Learning Model	-	-
E. Sungur ve B. Bakır	2022	IoT Security	-	SHAP, LIME	-
Pehlivanoğlu et al.	2023	Intrusion Detection in IoT Network	FNN	-	BOT-IOT
Wang et al..	2023	IoT Security	GB	-	-
Zhang et al.	2023	IoT Security	Federated Learning	SHAP, LIME	-
Kök et al.	2023	IoT Health Security	KNN, RF, DT,NB,SVM, MLP, ANN	SHAP, LIME, ELI5, IG	Intensive Care Unit(ICU) dataset

4. Evaluation

The studies analyzed in this paper have shown that intrusion detection in IoT networks has become more complex over time and advanced technologies offer much more effective solutions in this field. Traditional rule-based approaches have been replaced by ML and DL methods. These techniques have been observed to perform better against new threats that have emerged with the increase in the number and variety of IoT devices. Research has shown that DL models can have high accuracy rates in terms of intrusion detection in IoT networks. In particular, studies using CNN, RNN and hybrid models

have made significant progress in the field of network security. CNN-based intrusion detection systems achieved impressive results on the KDD Cup 99 dataset, while hybrid models developed for SDN-enabled IoT environments achieved accuracy rates of up to 99%. In addition, hybrid models designed to detect DoS and DDoS attacks have also achieved remarkable success. The critical role of machine learning-based classification algorithms in IoT security has been clearly demonstrated. Big data analytics-oriented studies using the Random Forest model achieved a 99.5% success rate in detecting real-time DDoS attacks on IoT networks.

In addition, explainable artificial intelligence approaches, machine learning and deep learning models have been shown to offer great potential for making decision-making more transparent. XAI techniques such as SHAP and LIME have been effective in improving the transparency and reliability of intrusion detection systems in IoT environments. The focus of these methods is to make security more accessible by ensuring not only the accuracy of the models, but also the ability to comprehend the decisions of these models.

This study examines the importance of models for IoT security that offer both high accuracy and provide transparency and reliability with XAI, and suggests that future work should examine how these approaches can be used more efficiently in wider IoT systems.

5. Result

The study examines the effectiveness of machine learning and deep learning techniques for detecting security threats in IoT networks and compares the performance of both sets of methods on different datasets. The findings reveal that deep learning methods perform strongly, offering higher accuracy rates than traditional methods. It is also emphasized that the explainable artificial intelligence (XAI) methods used in the study play an important role in making the decision processes of machine learning and deep learning models applied in IoT networks transparent. XAI techniques such as SHAP, LIME and ELI5 used in the studies make the outputs of the models more understandable and enable security experts to make accurate and effective analyzes.

In conclusion, this study demonstrates the importance of developing models that are both highly accurate, transparent and explainable in IoT security. Explainable AI-powered models increase the effectiveness of security solutions by making decision processes meaningful. Future research should examine how these methods can be applied more effectively, especially in critical sectors such as healthcare. It is also concluded that the integration of Explainable Artificial Intelligence methods into security detection systems in IoT environments will contribute to more secure and efficient protection of IoT networks.

References

[1] Xia, F., Yang, L. T., Wang, L., & Vinel, A. (2012). *Internet of Things. International Journal of Communication Systems, 25(9), 1101–1102.* doi:10.1002/dac.2417

[2] Statista Research Department (2016), Internet of Things (IoT) connected devices installed base worldwide from 2015 to 2025, Available at https://www.statista.com/statistics/471264/ iot-number-of-connected-devices-worldwide/. Online; accessed 10 January 2025.

[3] Sezer, O. B., Dogdu, E., & Ozbayoglu, A. M. (2018). Context-Aware Computing, Learning, and Big Data in Internet of Things: A Survey. IEEE Internet of Things Journal, 5(1), 1–27.

[4] Perera, C., Zaslavsky, A., Compton, M., Christen, P., & Georgakopoulos, D. (2013). *Semantic-Driven Configuration of Internet of Things Middleware. 2013 Ninth International Conference on Semantics, Knowledge and Grids.* doi:10.1109/skg.2013.9

[5] Ge, M., Fu, X., Syed, N., Baig, Z., Teo, G., & Robles-Kelly, A. (2019). *Deep Learning-Based Intrusion Detection for IoT Networks. 2019 IEEE 24th Pacific Rim International Symposium on Dependable Computing (PRDC).* doi:10.1109/prdc47002.2019.00056

[6] Y. K. Saheed, A. I. Abiodun, S. Misra, M. K. Holone, and R. Colomo-Palacios, "A machine learning-based intrusion detection for detecting internet of things network attacks," *Alexandria Eng. J.*, vol. 61, pp. 9395–9409, 2022.

[7] *B. Sharma, L. Sharma, C. Lal, and S. Roy, "Explainable artificial intelligence for intrusion detection in IoT networks: A deep learning based approach," Expert Syst. Appl., vol. 238, p. 121751, 2024.*

[8] E. Gürbüz, Ö. Turgut and İ. Kök, "Explainable AI-Based Malicious Traffic Detection and Monitoring System in Next-Gen IoT Healthcare," *2023 International Conference on Smart Applications, Communications and Networking (SmartNets)*, Istanbul, Turkiye, 2023, pp. 1-6, doi: 10.1109/SmartNets58706.2023.10215896.

[9] K. Zhao and L. Ge, "A survey on the internet of things security," in *Int'l Conf. on Computational Intelligence and Security (CIS),* 663-667, 2013.

[10] L. Atzori, A. Iera, G. Morabito, and M. Nitti, "The social internet of things (siot)–when social networks meet the internet of things: Concept, architecture and network characterization," *Computer Networks,* vol. 56, 3594-3608, 2012.

[11] Leo, M., Battisti, F., Carli, M., & Neri, A. (2014). *A federated architecture approach for Internet of Things security. 2014 Euro Med Telco Conference (EMTC).*

[12] Hodo, E., Bellekens, X., Hamilton, A., Dubouilh, P.-L., Iorkyase, E., Tachtatzis, C., & Atkinson, R. (2016). Threat analysis of IoT networks using artificial neural network intrusion detection system. 2016 International Symposium on Networks, Computers and Communications (ISNCC). doi:10.1109/isncc.2016.7746067

[13] Wood, A. D., & Stankovic, J. A. (2002). *Denial of service in sensor networks. Computer, 35(10), 54–62.*

[14] Imtithal A. Saeed Ali Selamat Ali M. A. Abuagoub, A Survey on Malware and Malware Detection Systems, International Journal of Computer Applications (0975 – 8887) Volume 67– No.16, April 2013

[15] McGraw, G. and G. Morrisett, Attacking Malicious Code: A Report to the Infosec Research Council. IEEE Softw., 2000. 17(5): p. 33-41.

[16] Xufang, L., P.K.K. Loh, and F. Tan. Mechanisms of Polymorphic and Metamorphic Viruses. in Intelligence and Security Informatics Conference (EISIC), 2011 European. 2011.

[17] Sen, R., & Borle, S. (2015). *Estimating the Contextual Risk of Data Breach: An Empirical Approach. Journal of Management Information Systems, 32(2), 314–341*

[18] Liu, H., & Lang, B. (2019). *Machine Learning and Deep Learning Methods for Intrusion Detection Systems: A Survey. Applied Sciences, 9(20), 4396.*

[19] Wikipedia," Supervised learning," *Wikipedia*. [Online]. Available: https://en.wikipedia.org/wiki/Supervised_learning. [Access Date: 15 Mar. 2025].
[20] Wikipedia, " Unsupervised learning," *Wikipedia*. [Online]. Available: https://en.wikipedia.org/wiki/Unsupervised_learning. [Access Date: 15 Mar. 2025].
[21] Peng, C.-Y. J., Lee, K. L., & Ingersoll, G. M. (2002). *An Introduction to Logistic Regression Analysis and Reporting. The Journal of Educational Research, 96(1), 3–14.* doi:10.1080/00220670209598786
[22] Sperandei, S. (2014). *Understanding logistic regression analysis. Biochemia Medica, 12–18.* doi:10.11613/bm.2014.003
[23] Real Python, "Python Programming Tutorials," *Real Python*. [Online]. Available: https://realpython.com/. [Access Date: Mar. 20, 2025].
[24] Safavian, S. R., & Landgrebe, D. (1991). A survey of decision tree classifier methodology. IEEE Transactions on Systems, Man, and Cybernetics, 21(3), 660–674.
[25] Apté, C., & Weiss, S. (1997). Data mining with decision trees and decision rules. Future Generation Computer Systems, 13(2-3), 197–210. doi:10.1016/s0167-739x(97)00021-6
[26] Friedl, M. A., & Brodley, C. E. (1997). Decision tree classification of land cover from remotely sensed data. Remote Sensing of Environment, 61(3), 399–409. doi:10.1016/s0034-4257(97)00049-7
[27] Farnaaz, N., & Jabbar, M. A. (2016). Random Forest Modeling for Network Intrusion Detection System. Procedia Computer Science, 89, 213–217.
[28] Pal, M. (2005). *Random forest classifier for remote sensing classification. International Journal of Remote Sensing, 26(1), 217–222*
[29] W. Wang, X. Du, N. Wang, "Building a Cloud IDS Using an Efficient Feature Selection Method and SVM," *IEEE Access*, vol. 7, pp. 1345–1354, 2019.
[30] M. Al-Qatf, Y. Lasheng, M. Al-Habib, K. Al-Sabahi, "Deep Learning Approach Combining Sparse Autoencoder with SVM for Network Intrusion Detection," *IEEE Access*, vol. 6, pp. 52843–52856, 2018.
[31] W. Feng, J. Sun, L. Zhang, C. Cao, Q. Yang, "A support vector machine based naive Bayes algorithm for spam filtering," in *Proc. 2016 IEEE 35th Int. Perform. Comput. Commun. Conf. (IPCCC 2016)*, 2017
[32] Cybenko, G. 1989. Approximation by superpositions of a sigmoidal function *Mathematics of Control, Signals, and Systems*, 2(4), 303–314.
[33] Van den Oord, Aaron; Dieleman, Sander; Schrauwen, Benjamin (2013-01-01). Burges, C. J. C.; Bottou, L.; Welling, M.; Ghahramani, Z.; Weinberger, K. Q. (eds.). *Deep content-based music recommendation* (PDF). Curran Associates, Inc. pp. 2643–2651.
[34] Tealab, Ahmed (1 Aralık 2018). "Time series forecasting using artificial neural networks methodologies: A systematic review". *Future Computing and Informatics Journal* (İngilizce). **3** (2). ss. 334-340. doi:10.1016/j.fcij.2018.10.003 . ISSN 2314-7288.
[35] *Graves, Alex*; Liwicki, Marcus; Fernandez, Santiago; Bertolami, Roman; Bunke, Horst; *Schmidhuber, Jürgen* (2009). *"A Novel Connectionist System for Improved Unconstrained Handwriting Recognition"* (PDF). IEEE Transactions on Pattern Analysis and Machine Intelligence. **31** (5). ss. 855-868. *CiteSeerX 10.1.1.139.4502 $2*. *doi:10.1109/tpami.2008.137*. *PMID 19299860*
[36] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). ''Why should i trust you?'' Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1135–1144).
[37] Q. Sun, A. Akman ve B. W. Schuller, "Explainable Artificial Intelligence for Medical Applications: A Review," ACM Transactions on Computing for Healthcare, cilt 6, sayı 2, ss. 1-31, Şubat 2025. DOI: 10.1145/3709367
[38] S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

[39] E. S. G. Ribeiro, "ELI5: A Python package for machine learning model explanation," *ELI5 Documentation*, [Online]. Available: https://eli5.readthedocs.io/en/latest/overview.html. [Access Date: 2-Mar-2025].

[40] H. Moraliyage, G. Kulawardana, D. De Silva, Z. Issadeen, M. Manic, and S. Katsura, "Explainable Artificial Intelligence with Integrated Gradients for the Detection of Adversarial Attacks on Text Classifiers," *Appl. Syst. Innov.*, vol. 8, no. 1, p. 17, Jan. 2025, doi: 10.3390/asi8010017.

[41] H. Xiao, Y. Xing, J. Zhang, and F. Zhao, "A CNN-based IDS model for network intrusion detection," IEEE Access, vol. 7, pp. 156665-156675, 2019.

[42] S. Nagisetty and S. Gupta, "Comparison of deep learning models for IoT intrusion detection using open-source datasets," IEEE Transactions on Information Forensics and Security, vol. 14, no. 2, pp. 2201-2215, 2019.

[43] Y. Liang, J. Chen, and S. Wang, "DNN-based IDS for NSL-KDD dataset," IEEE Transactions on Network Science and Engineering, vol. 8, no. 1, pp. 25-37, 2019.

[44] R. Vinayakumar, K. Alazab, and M. Simic, "DNN-based intrusion detection: An evaluation on NSL-KDD dataset," IEEE Transactions on Information Forensics and Security, vol. 14, no. 2, pp. 2310-2319, 2019.

[45] H. Kim, J. Kim, D. Kim, J. Shim, and H. Choi, "CNN-RNN hybrid model for DoS attack detection," IEEE Transactions on Network and Service Management, vol. 17, no. 4, pp. 1342-1355, 2020.

[46] A. Kasongo and H. Sun, "Deep neural network-based intrusion detection system using UNSW-NB15 dataset," IEEE Transactions on Information Forensics and Security, vol. 15, no. 3, pp. 1517-1525, 2020.

[47] M. Awan, R. Khan, and H. Ali, "Real-time DDoS detection in IoT using big data analytics and random forest," IEEE Access, vol. 9, pp. 118234-118245, 2021.

[48] S. Sahu, P. Patel, and A. Sharma, "Hybrid CNN-LSTM model for IoT security: Anomaly detection in network traffic," IEEE Transactions on Emerging Topics in Computational Intelligence, vol. 6, no. 3, pp. 432-445, 2021.

[49] W. Ge, F. Syed, Y. Fu, Z. Baig, and A. Robles-Kelly, "A feedforward neural network model for intrusion detection," IEEE Internet of Things Journal, vol. 9, no. 2, pp. 1433-1444, 2021.

[50] E. Sungur and B. Bakır, "A hybrid deep learning model for DDoS attack detection in SDN-based IoT environments," IEEE Transactions on Information Forensics and Security, vol. 18, no. 5, pp. 2871-2882, 2023.

[51] H. Pehlivanoğlu, A. Demir, and Y. Kılıç, "Evaluation of ML-based IDS for IoT: Performance analysis with multiple classifiers," IEEE Transactions on Network and Service Management, vol. 20, no. 2, pp. 2156-2169, 2023.

[52] J. Kim and L. Wang, "Explainable AI for IoT security: SHAP and LIME in intrusion detection," IEEE Access, vol. 10, pp. 120456-120468, 2022.

[53] [14] F. Zhang, X. Li, and H. Chen, "Federated learning with explainable AI for privacy-preserving intrusion detection in IoT," IEEE Internet of Things Journal, vol. 9, no. 7, pp. 6111-6124, 2023.