


Anomaly Detection in IoMT Environment Based on Machine Learning: An Overview*

Research Article

Peyman Vafadoost Sabzevar¹, Hamidreza Rokhsati² , Alireza Chamansara³, Ahmad Hajipour⁴

DOI: [10.22067/cke.2024.89572.1127](https://doi.org/10.22067/cke.2024.89572.1127)

Abstract In today's era, the Internet of Things has become one of the important pillars in organizations, hospitals, and research circles and is recognized as an integral part of the Internet. One of the important areas that require online monitoring is medical imaging equipment, whose functional information is transmitted through the Internet of Things. Server security and intrusion prevention, along with anomaly detection, are critical requirements for these networks. The purpose of anomaly detection is to develop methods that can detect attackers' attacks and prevent them from happening again. Algorithms and methods based on statistics play an important role in predicting and diagnosing anomalies. In this article, the isolation forest algorithm was used for training on 80% of the dataset related to the data of the Internet of Medical Things network, and then this model was tested and evaluated on the remaining 20%. The results show 90.54% accuracy in detecting anomalies in the received data, which confirms the effective performance of this method in this field.

Keywords *Index Terms*— Anomaly, Detection Anomaly, Internet of Medical Things, Isolation Forest, Unsupervised Learning.

1. Introduction

The Internet of Things (IoT) has become one of the most prominent and transformative phenomena in the world of technology and communication since the beginning of 2000. This concept has widely been able to create a significant improvement in the way of communication and interaction between different objects and devices and thus, improve efficiency, efficiency, and communication in various systems. The IoT, which refers to the communication and interaction between various types of objects including sensors, home devices, and sophisticated gadgets, allows us to collect, send, and receive information to facilitate various processes in various societies and industries such as improve agriculture, healthcare, and

smart homes [1].

Currently, one of the fields that has been deeply affected by the Internet of Things is the field of medicine and health. The Internet of Medical Things (IoMT), which specifically relates to the use of sensors, wearable devices, smart medical devices, and other advanced technologies, has been able to help improve disease monitoring and management and improve the quality of healthcare services. This technology not only helps in early diagnosis of diseases and better management of patients' condition, but also generally increases the quality and accuracy of health services.

Also, the rapid growth of IoT devices, which includes smart home appliances and various industrial technologies, has led to increased communication between objects and improved efficiency and flexibility in various industries and daily life. Statistics and surveys show that the connection of devices to the IoT is increasing dramatically [2]. This rapid growth, shown in Table 1, shows the importance and necessity of paying attention to security and management issues in this area, because with the increase in the number of connected devices, there is a need for effective solutions to maintain security and data quality.

Table 1 Internet Use and World Population [2].

	2010	2015	2020
World Population	6.8 billion	7.2 billion	7.6 billion
Connected Devices	12.5 billion	25 billion	50 billion
Device for Each Person	1.84	3.47	6.58

* Manuscript received 2024 July 15, Revised, 2024 October 4, accepted 2024 November 6.

¹ Master of Biomedical Engineering, Electrical and Computer Engineering Department, Hakim Sabzevari University, Sabzevar, Iran.

² Corresponding Author. Master of Engineering in Computer Science, Department of Computer, Control and Management Engineering, Sapienza University of Rome, Rome, Italy. **Email:** rokhsati.1960699@studenti.uniroma1.it

³ PhD in Biomedical Engineering, Department of Biomedical Engineering, Materials and Energy Research Center, Tehran, Iran.

⁴ Assistant Professor, Electrical and Computer Engineering Department, Hakim Sabzevari University, Sabzevar, Iran.

Data available on the internet, especially medical data, are of particular importance. Medical data includes sensitive and vital information such as disease history, test results, and medical center device parameters. Breach of security or abnormality in this data can lead to serious problems, including unauthorized access to information, misuse, malfunction of monitoring systems, and even illegal sale of information. These issues can have serious consequences for the privacy and security of healthcare facilities and potentially affect public trust in health and medical systems [3, 4]. Figure 1 shows the IoMT architecture. This architecture includes various layers, including sensors and IoMT devices layer, network and communication layer, and data processing and analysis layer, which are connected to each other. This layered structure enables the communication and exchange of information between different IoMT devices and enables the processing and analysis of collected data.

Hence, the field of IoMT has become a complex and fully integrated ecosystem that includes interconnected medical devices and systems. This ecosystem is specifically designed to enhance the quality of patient care, optimize healthcare practices, and ultimately improve overall health outcomes [5, 6]. Considering the high importance of medical data and the complexities related to their security, it is essential to develop and implement effective solutions to protect these data and ensure their accuracy and security.

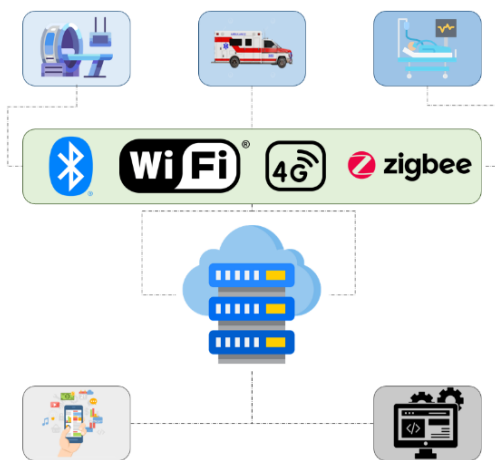


Figure 1. IoMT Reference Architecture [7].

Due to the emerging phenomenon of the IoT and its combination with new methods, the activities and researches conducted in this field are expanding rapidly. This innovative combination has not only increased researchers' interest in studying and developing the IoT, but also led to significant improvements in various applications of this technology, especially in medical fields.

Evidence shows that the number of scientific articles published in this field is increasing day by day. In particular, articles published in prestigious journals such as ScienceDirect show a significant growth as shown in Figure 2 and Figure 3, and pay a lot of attention to topics

related to IoT and IoMT. This growth shows that researchers are seriously investigating and developing this technology, working on its various aspects from improving the security and efficiency of systems to expanding new applications in clinical and health care.

The quick adoption of the IoMT, in the healthcare industry has transformed how patients are cared for. Has also brought about security hurdles to overcome. The application of machine learning (ML) for anomaly detection shows promise in recognizing and addressing these risks. This article gives an outline of existing ML methods, for detecting anomalies in IoMT settings. Emphasizes how they could bolster the security and dependability of networks. Our goal is to help tackle these obstacles and aid in the advancement of robust IoMT systems.

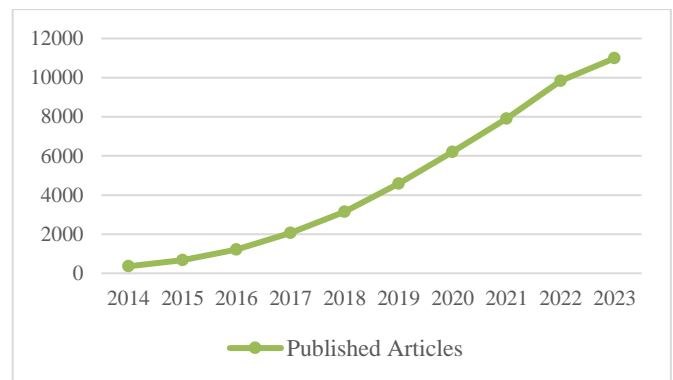


Figure 2. The Number of Articles Published in The Field of IoT

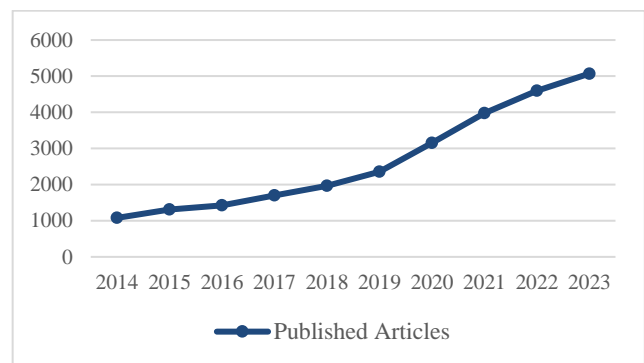


Figure 3. The Number of Articles Published in The Field of IoMT

A. IoT security challenges

In the field of IoT, due to its complex nature and architecture, this technology has many security challenges, each of which can create serious risks for the systems. Developers of this technology should avoid these damages by using different techniques. One of the most important security challenges in this field is the wireless nature of IoT networks [8]. This feature exposes the network to the risk that it can gain access to sensitive data and compromise or alter it. Another important security challenge is network topology. Networks can compromise the entire system by sending fake data to the network and

disrupt its core functionality. This type of operation can have serious consequences, especially in sensitive environments such as medicine and healthcare, because in these systems it can contribute to life-threatening injuries and serious problems for patients. In addition, data heterogeneity is one of the other main challenges of the IoT. Data generated by different sources, which are created separately and with different patterns, make the management and analysis of this data more complex. This diversity in data can cause IoT systems to face more challenges in maintaining security and integrity [9, 10].

Due to these challenges, the need to develop multilayered and comprehensive IoT security in and implement these systems to the users to ensure their integrity, security, and efficiency against various threats. These predictions are especially important in fields such as medicine that deals with human lives [11].

In this regard, exploiting vulnerabilities in IoT-based networks can lead to specific attacks that put these networks at risk. For example, the CyberMDX report [12] shows that nearly half of IoMT devices are vulnerable to anomalies. These vulnerabilities can lead to malicious attacks that not only disrupt the performance of systems, but can also have serious consequences for the health of patients.

IoMT systems are fundamentally different from other systems because they directly affect patients' lives. Serious privacy concerns arise if patients' identities and sensitive information are exposed. Additionally, the high value of healthcare data on the black market adds to these concerns. In particular, the average cost of healthcare-related data is approximately 50 times higher than credit card information, making this information a very valuable target for attackers [13].

These conditions show that providing security in IoMT systems is not only necessary for data protection and privacy, but also for protecting patients' lives and preventing financial abuses. Therefore, paying special attention to the security of these systems and implementing advanced solutions to deal with possible threats is an inevitable necessity [14].

Based on these cases, first the researches and works done in this field and similar methods are discussed. Then, in the next section, the research methodology, the algorithms used, and the available data are examined in detail. In the fourth part, the results obtained from the analyzes and experiments will be evaluated and analyzed. Finally, conclusion will be presented and the achievements of this research will be comprehensively discussed.

2. Literature review

So far, a lot of research has been done in the field of anomaly detection in order to prevent cyber-attacks in this field. These researches have developed advanced algorithms and models that are able to identify unusual behaviors and prevent security threats in IoT systems and especially IoMT.

Zachos et al. [15] propose an efficient and effective anomaly-based intrusion detection (AIDS) system for IoMT networks. The proposed idea aims to use host-based and network-based techniques to reliably collect log files from IoMT and gateway devices, as well as traffic from

the IoMT edge network, while considering the computational cost. The proposed idea relies on ML techniques to detect anomalies in the collected data and thereby detect malicious events in the IoMT network, taking into account the computation overhead. Tamara et al. [16] addresses this issue by building a Kalman filter and Cauchy clustering algorithm for anomaly detection and applying them to authenticate nodes in IoMT using the Extreme ML classifier. For anomaly detection, a critical aspect in various applications including security, healthcare, and network monitoring, Alsaman [17] has introduced FusionNet, an innovative ensemble model that leverages the strengths of multiple ML algorithms, namely random forest, K-nearest neighbor (KNN), It combines support vector machine (SVM) and multilayer perceptron (MLP) for advanced anomaly detection. The FusionNet architecture uses a variety of these algorithms to achieve high precision and accuracy. And the evaluation results on two sets of data show the high accuracy of this architecture compared to classical methods.

Wang et al. [18] addressed that IoMT devices lack security authentication mechanisms and trust between these devices is highly dependent on centralized third-party services. To solve this problem, they used blockchain technology as a secure interactive environment for IoMT. However, security issues are also increasingly raised in the IoMT-Blockchain environment. Cyber-attacks targeting IoMT-Blockchain not only compromise the security of IoT devices, but can also seriously affect the overall security of the Internet. Therefore, abnormal traffic detection becomes especially important in the IoMT-Blockchain environment. In their proposed method, an abnormal traffic detection model using deep neural network is designed to detect abnormal traffic in the IoMT-Blockchain environment. Their proposed algorithm uses Residual Learning to perform more optimally in identifying abnormal traffic.

On the other hand, Wagan et al. [19] used a variety of customized security tools and frameworks to counter several common attacks, including botnet-based distributed denial-of-service (DDoS) attacks and zero-day network attacks. They proposed a new approach called Duo-Secure IoMT, which uses data from multimodal sensory signals to distinguish between attack patterns and routine data from IoMT devices. Their proposed model employs a combination of two techniques including C-Means dynamic fuzzy clustering and a customized version of the Bi-LSTM technique. This approach securely processes sensory medical data and detects attack patterns in the IoMT network.

In their research, Awotunde and his colleagues [20] used a model based on swarm neural networks to detect intruders in IoMT data-driven systems. Their proposed model is capable of detecting intruders during data transmission and can provide efficient and accurate analysis of healthcare data at the edge of the network. The performance of this system was tested using the NF-ToN-IoT real-time dataset, which is designed for IoT applications and includes telemetry data, operating systems and network data. The results of these tests showed the high accuracy of the model in identifying anomalies in the system.

In another research, Kumar et al. [21] proposed an approach to combat cloud architecture-based cyber-attacks in IoMT networks based on the 2018 Ransomware cyber-attack on the Indiana hospital system. Their method is a combination of three classification algorithms: decision tree, naive bayes, and random forest, which is used to identify anomalies. In the next step, the results obtained from the classifications are processed by the XGBoost algorithm to accurately identify normal samples and samples under attack. This multilayered combination of machine learning techniques helps to improve the accuracy and efficiency of anomaly detection in IoMT networks.

Al-Qatf et al. [34] used the KDDTrain dataset and using a hybrid approach including deep learning and SVM to detect infiltration in the network. Their proposed method has provided higher accuracy in intrusion detection compared to the classical SVM method. By introducing an intelligent intrusion detection system based on automatic encoder (AE), Ieracitano et al. [35] have presented a new statistical analysis. The salient feature of their proposed method is the use of data analysis for more optimal and robust feature extraction, which improves the accuracy and efficiency of the intrusion detection system. Javid et al. [36] have also proposed a deep learning approach for developing an intrusion and anomaly detection system. In this method, Self-Taught Learning (STL), which is an advanced technique in the field of deep learning, is used on the NSL-KDD dataset.

The motivation for writing this article is to address the growing concerns that have formed around cyber attacks on healthcare networks. Considering the high sensitivity of medical data and their vital importance in providing services to patients, any security flaw or cyber attack can have irreparable consequences for patients and health organizations. This issue not only threatens the security of information, but can lead to the disclosure of sensitive data, interruption of medical services, and even endangering the lives of patients. Therefore, investigating and identifying anomalies in the traffic data of these networks is very important in order to prevent such attacks.

Focusing on these challenges, this article tries to identify and analyze anomalies in the traffic data of healthcare networks using statistical methods. The proposed system in this paper specifically addresses one of the important problems in IoMT data security. These data are always exposed to security threats due to their high importance in controlling and monitoring the condition of patients and medical equipment. Threats that, if not detected in time, can disrupt the performance of health systems and cause irreparable damage.

For this reason, the proposed system not only identifies anomalies in traffic data, but also tries to help improve security and reduce risks associated with these threats by providing advanced solutions. Using advanced statistical techniques, this system is able to identify anomalies that may indicate a cyber-attack with acceptable accuracy. This identification allows organizations to take the necessary steps to prevent potential attacks and maintain the security of medical data. Finally, this article, with a detailed review and comprehensive analysis, not only solves one of the

basic problems in IoMT data security, but also helps to provide solutions for the overall improvement of security in healthcare networks. In this study, isolated forest machine learning algorithm was used. This algorithm, unlike other machine learning methods, categorizes data based on the detection and isolation of abnormal data. This feature makes the isolated forest algorithm work faster and more accurately than other methods and has better efficiency in isolating and identifying anomalies.

The following topics will be covered in the following articles:

- First, the NSL-KDD dataset is examined.
- Then, the proposed algorithm of this research will be introduced.
- After that, the results of this research are analyzed.
- then this method is compared with other existing methods.
- At the end, the conclusion of the article will be presented.

3. Methodology

In this section, we begin by providing a detailed description of the dataset utilized in this research. This includes an overview of its structure, the types of data it contains, and the relevance of these data points to our study. Understanding the dataset is crucial as it forms the foundation upon which our analysis and anomaly detection efforts are built. Following this, we delve into the theoretical framework of the Isolation Forest algorithm. This involves a comprehensive examination of its initial formulation, including the key mathematical principles and assumptions that underpin the model. We will explore how the Isolation Forest algorithm identifies anomalies, its unique approach to data partitioning, and why it is particularly well-suited for detecting outliers in large datasets. By first establishing a clear understanding of the dataset and then the methodology, this section sets the stage for the subsequent analysis and findings presented in this research.

A. NSL-KDD Dataset

The NSL-KDD dataset is a modified and improved version of the KDD Cup 99 dataset, which is used for research purposes and the development of intrusion detection systems [22]. This dataset has become one of the main sources of research in the field of anomaly detection and network traffic analysis due to its special features, including reducing redundant data repetition and improving the quality of samples [23]. So far, many researchers have used this data set to conduct extensive research in the field of anomaly detection and have tried to develop effective intrusion detection systems using various techniques and tools. Deep analysis of this data set using different machine learning algorithms and advanced data mining methods has been done in numerous researches and their results have led to a significant improvement in identifying security threats and improving the efficiency of cyber security systems [24-30].

In this research, among the files available in NSL-KDD dataset, the KDDTrain+.TXT file has been selected as the desired dataset. This file contains the complete set of data in CSV format, and in each record there are 41 different

attributes that describe different aspects of the data flow. These features provide important information about network activities and communications. In addition, each record has a label that specifies the type of attack or designates it as normal data. Table 3 contains the names and data types of all 41 features in the NSL-KDD dataset.

These tags are critical for anomaly detection and analysis, as they help distinguish normal data from potential attacks. Table 2 in this research shows the amount of data in this dataset as well as the distribution of data related to each class, which includes normal data and different types of attacks. This table helps to better understand the composition of data and how they are scattered among different classes, and is the basis for evaluating the performance of anomaly detection models.

The data recorded in this dataset are generally divided into four main categories [23]:

- DoS (Denial of Service): This type of attack is designed to occupy the victim's system resources by sending a large volume of fake requests. This large volume of requests makes the system unable to provide services to real users and as a result the service is disrupted. DoS attacks can be carried out in different ways, but the ultimate goal of all of them is to disable the target system.
- Probing: The goal of this attack is to probe and examine the victim's system to discover vulnerabilities and sensitive information such as open ports, running services, and connection duration. By using the information obtained from these scans, the attacker can make a more detailed plan for his next attacks. This type of attack allows the attacker to gain sufficient knowledge of the target environment before performing more serious attacks.
- U2R (User to Root): In this type of attack, the attacker enters the victim's system through a normal user account and then tries to gain higher level privileges such as root or administrator privileges by exploiting existing vulnerabilities. This type of attack allows the attacker to take full control of the system and execute commands as an elevated user.
- R2L (Remote to Local): This type of attack involves unauthorized remote access to the victim's system. The attacker enters the victim's system by guessing or breaking the password, and after gaining access, he tries to carry out destructive operations or steal information. This attack usually involves logging into the system through the network without the need for the attacker to be physically present at the victim's location.

Table 2 Details of The Normal and Anomalous Data in the KDDTrain+.TXT Dataset [23].

Dataset Type	Record	Normal Class	Dos Class	Probe Class	U2R Class	R2L Class
KDDTrain+	125,973	67,343	45,927	11,656	52	995
		53.39%	36.65%	9.09%	0.04%	0.79%

Table 3 NSL-KDD Dataset Features [26].

No	Features	Type	No	Features	Type
0	duration	int64	21	is_guest_login	int64
1	protocol_type	object	22	count	int64
2	service	object	23	srv_count	int64
3	flag	object	24	serror_rate	float64
4	src_bytes	int64	25	srv_serror_rate	float64
5	dst_bytes	int64	26	rerror_rate	float64
6	land	int64	27	srv_rerror_rate	float64
7	wrong_fragment	int64	28	same_srv_rate	float64
8	urgent	int64	29	diff_srv_rate	float64
9	hot	int64	30	srv_diff_host_rate	float64
10	num_failed_logins	int64	31	dst_host_count	int64
11	logged_in	int64	32	dst_host_srv_count	int64
12	num_compromised	int64	33	dst_host_same_srv_rate	float64
13	root_shell	int64	34	dst_host_diff_srv_rate	float64
14	su_attempted	int64	35	dst_host_same_src_port_rate	float64
15	num_root	int64	36	dst_host_srv_diff_host_rate	float64
16	num_file_creations	int64	37	dst_host_serror_rate	float64
17	num_shells	int64	38	dst_host_srv_serror_rate	float64
18	num_access_files	int64	39	dst_host_rerror_rate	float64
19	num_outbound_cmds	int64	40	dst_host_srv_rerror_rate	float64
20	is_host_login	int64			

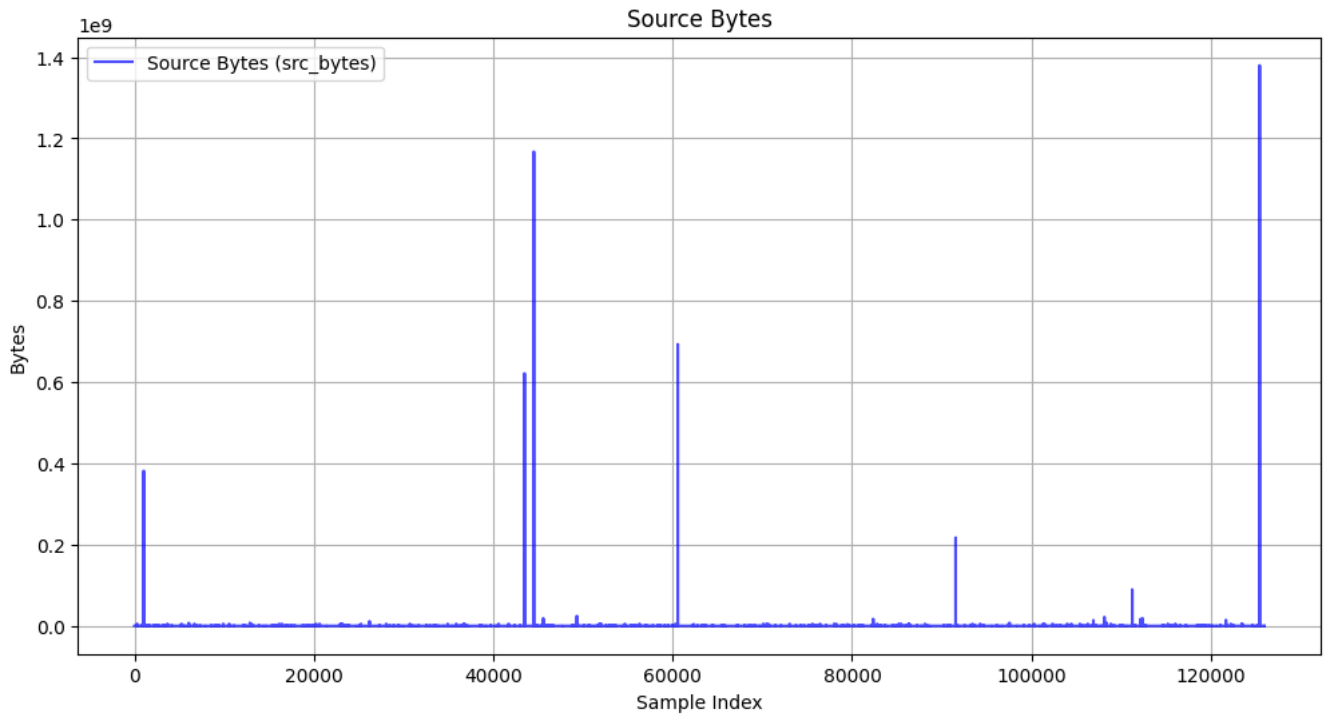


Fig 4. Data Traffic in Bytes from Source to Destination.

B. Isolation Forest

ML algorithms are divided into three main categories: supervised learning (SL), unsupervised learning (UL), and reinforcement learning (RL). Each of these categories have different uses and are designed to solve specific problems.

In SL, the model is trained using data, each of which is assigned a specific label or output. These labels indicate the correct answer or the correct classification of the data. The model learns from this labeled data how to relate inputs to the correct outputs and then uses this learning to evaluate and test on new data. UL is used where the data is not labeled. In this method, algorithms try to divide the data into different clusters or categories without having a specific output. This clustering is done based on various criteria and parameters that algorithms use to identify patterns and relationships in the data. Finally, RL is based on the trial and error of an agent in the environment. In this method, by performing various actions in the environment and receiving feedback (reward or punishment) from these actions, the agent gradually learns to adopt the best policies to achieve its goals. This type of learning is especially effective in situations where decisions are made in successive stages and long-term results are important [30].

The algorithm used in this research is called Isolation Forest as a branch of UL that is used to detect anomalies and patterns that have different characteristics from normal examples. Detecting abnormalities in this particular instance could mean an attacker could launch an attack on medical and healthcare systems that would threaten people's health [31].

Most ML-based methods for anomaly detection first train the model using normal samples. Then, using this trained model, they evaluate and test the new data. In this process, data that do not conform to normal patterns are

identified as anomalies. However, using these methods has two major problems [32, 33].

The first problem is that the anomaly detection may not achieve the desired results and the system will face an increase in error, so that a small number of anomalies are correctly detected. This issue can lead to unfavorable performance of the system in identifying security threats. The second problem is related to the high computational complexity of some of these methods, which makes these methods limited to small data and the learning process faces serious limitations. This can greatly affect the performance of the model, especially when there is a need to process a large amount of data.

In the proposed method, an approach that separates anomalies from normal data is used. This method identifies each abnormality separately by using the tree structure, the limited amount of abnormal data and the large difference between their values and normal data. This scheme not only reduces the computational complexity, but also increases the accuracy of anomaly detection and enables the system to detect anomalies more reliably.

Unlike many algorithms that first model normal behavior and then detect anomalies, Isolation Forest directly isolates abnormal points, thereby increasing detection accuracy and reducing processing time. Its special binary structure, called *iTree*, allows the algorithm to detect anomalous data at shallower depths, as abnormal points are quickly isolated at lower depths. In addition, this method is insensitive to the data scale due to the reliance on relative comparisons and does not require normalization or rescaling of the data, so changes in the data scale will not significantly affect the results. The random feature of feature selection in the separation forest also provides the possibility of effective identification of anomalies in multidimensional data with a high number of

features, and there is no need to reduce the dimensions or select specific features. In addition, the simple setting of the parameters of this algorithm has made it a favorable option for fast and extensive applications and has made it popular in various fields, especially anomaly detection and data security.

In order to build the iTree, having the data set $X = \{x_1, x_2, \dots, x_n\}$, the algorithm works recursively. This process starts by randomly selecting a feature named q and a threshold value p . These random selections are used to segment the data. Segmentation is done by dividing the data into two parts: those in which the feature value q is less than or equal to p , and those in which the value q is greater than p . This segmentation process continues recursively until one of the stopping conditions is met:

The node contains only one point: in other words, the data set has shrunk so much that only one data remains.

All node data have the same values: in the sense that no further partitioning is possible because all data is the same.

When the iTree is fully grown, each point of the data set X is placed in one of the external nodes (leaves). The length of the path traveled by each point in the tree to reach a leaf is determined as a criterion.

According to this algorithm, the points with shorter path length in the tree are identified as unusual and anomalous points. This is because anomalous data tends to be separated in the early stages of segmentation and, therefore, takes a shorter path to reach the leaf. On the other hand, normal data is more likely to travel longer paths because it is grouped with more similar data. The path length of a point in the tree is denoted by $h(x_i)$. This value is the number of sides that the point x_i travels from the root node to reach the leaf node. In other words, $h(x_i)$ indicates the depth of the node where the point x_i is located.

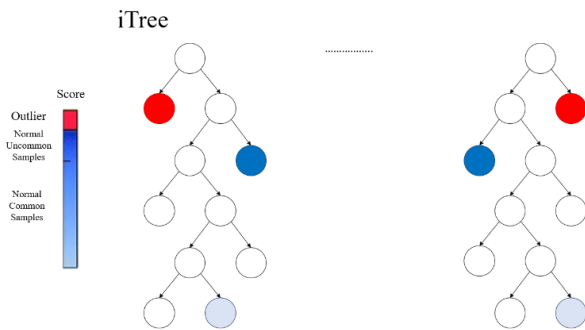


Figure 5. Isolation Forest Structure.

Abnormality detection generally consists of two main steps. In the first step, the training data set is used to build iTree trees. In this step, trees are created based on random feature partitions and threshold values to form a suitable tree structure for the data. In the second step, the test data is entered into the iTree trees, and each data sample passing through the trees is assigned an abnormality score. This abnormality score is calculated based on the length of the path that each sample takes in the tree.

To diagnose abnormality, calculating the abnormality score is a critical step. In methods that use a tree structure such as iTree, this score is calculated based on the length of the path that each data sample takes in the tree. Since the iTree has the same

structure as the binary search tree (BST), the estimation of the average path length $h(x)$ for external nodes is done in the same way as in the failed search in BST. In this analysis, the average iTree path length to terminate at external nodes is estimated using binary search tree features. This analysis, based on BST features, helps researchers estimate the average iTree path length and assign an anomaly score to each data sample based on that.

$$c(n) = 2H(n - 1) - \frac{2(n - 1)}{n} \tag{1}$$

Concepts such as harmonic number $H(i)$ and average path length $c(n)$ are used in calculating the abnormality score. The harmonic number $H(i)$ is one of the important tools in these calculations and is defined as follows:

$$H(i) = \ln(i) + \gamma \tag{2}$$

γ is equivalent to Euler-Mascheroni constant.

Equations 1 and 2 can be used for normalization and estimate the abnormality score for a given sample.

$$s(x, n) = 2 \frac{E(h(x))}{c(n)} \tag{3}$$

Here $E(h(x))$ represents the average path length $h(x)$ in a set of iTree trees.. In equation (3):

- If the value of s is close to 1, it can be concluded that x is an anomalous point with high probability. This means that the point x behaves differently and unusually from other data points and is identified as an anomaly.
- On the other hand, the proximity of the value s to 0.5 indicates the norm or normality of the point. In this case, the point x has the same characteristics as other points in the dataset and is likely not anomalous.
- If for all values of a random sample, the score or value s is close to 0.5, it can be concluded that all points in the data set behave normally and there is probably no anomaly in the data set. This result shows that the data follows a stable pattern and no significant deviation from this pattern is observed.

4. Results

Before starting the clustering operation on the data, we perform two pre-processing steps so that the data is fully prepared for the clustering process. These pre-processing steps play a key role in the efficiency of the clustering algorithm, because clean and standardized data can lead to better and more meaningful results. In the first step, the protocol_type, service and flag batch features, which are part of the batch features, are converted into numbers. Also, if there are text values in each column, they are converted into numbers.

Data standardization is critical in algorithms that are sensitive to the distance or scale of features, such as distance-based algorithms. This process ensures that all features are considered equally in the analysis and predictions and that each feature has an equal impact on the final result. Standardization is calculated using the following equation:

$$X_{scaled} = \frac{X - \mu}{\sigma} \quad (4)$$

where X is the original value of the feature. μ is the mean of that feature in the data. σ is the standard deviation of that feature. Using this relationship, all features are converted to a common scale (mean zero and standard deviation one), which improves the performance of scale-sensitive algorithms.

Also, in this process, 80% of the data is considered as training data, while the remaining 20% is used as testing data. The training data is used to train the model, while the test data is used to evaluate the performance of the model. This segmentation helps us to test the model on new data that has not been trained in the process and evaluate its accuracy and efficiency in real conditions.

Also, the following relationship is used to calculate the accuracy of the model:

$$Accuracy (\%) = \left(\frac{TP + TN}{TP + TN + FP + FN} \right) \times 100 \quad (5)$$

where TP is the number of positive samples that are correctly predicted as positive. TN is the number of negative samples correctly predicted as negative. FP is the number of negative samples that are falsely predicted as positive. FN is the number of positive samples that are falsely predicted as negative.

To evaluate this method, in addition to the main criterion, three other criteria have also been used:

Precision measures the ratio of the number of correct positive predictions to the total number of positive predictions (both true and false).

$$Precision = \left(\frac{TP}{TP + FP} \right) \quad (6)$$

Recall calculates the ratio of the number of correct positive predictions to the total number of true positive samples and indicates the rate of correct identification of true positive samples.

$$Recall = \left(\frac{TP}{TP + FN} \right) \quad (7)$$

F1-Score is the harmonic mean of Precision and Recall and is generally used to evaluate models that require a balance between Precision and Recall.

$$F1 - Score = 2 \times \left(\frac{Precision \times Recall}{Precision + Recall} \right) \quad (8)$$

Table 4 depicts the performance of the model in terms of data learning and clustering. This table examines and evaluates the model based on four key parameters and also pays attention to the analysis of the relationships between these parameters. In this evaluation, all the different aspects of the model's performance have been carefully examined in order to achieve a more comprehensive understanding of the model's efficiency in identifying and classifying data.

Table IV Model Performance Results.

	Accuracy	Precision	Recall	F1-Score
Performance	90.54%	94.53%	88.51%	91.2%

Also, the table below shows the correct and incorrect predictions of the samples in the evaluation process.

Table V Model Predictions in the Evaluation Process.

	TP	FP	TN	FN
Test Data	11,097	1,614	11,714	770

In this research, we evaluated our proposed method using four key criteria and compared its results with the methods presented in various articles that worked on the same dataset. This comparison includes the accuracy of the criteria tested on this data set. The results of these evaluations show the significant and improved performance of our proposed method compared to other existing methods. In other words, our proposed model has been able to properly provide predictions in terms of accuracy and quality compared to previous methods, and this issue is well evident in the comparisons and analyzes performed.

Table VI Performance Comparison with Other Approaches.

	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
[34]	84.96	96.23	76.57	85.28
[35]	84.21	87	80.37	81.98
[36]	88.39	85.44	95.95	90.4
Suggested Method	90.54	94.53	88.51	91.2

5. Conclusion

In this article, considering the increasing importance of data in the context of IoMT, we investigated the anomalies and security threats in the traffic of this field. For this purpose, we used the NSL-KDD dataset to identify and analyze intrusions, attack risks, and the probability of failure of healthcare systems. Our proposed method consists of several main steps. First, pre-processing and normalization of the data is done in order to prepare them to enter the model. In this step, the raw data is converted to the appropriate format and scale so that the model is able to accurately analyze them. After this step, we used the cluster method called isolated forest for clustering. This method is used to identify anomalies and unusual data and is especially useful in identifying unusual patterns that may be related to security attacks. In this approach, anomalous data are presented to the model as training samples, and the model performs the clustering process based on these samples. The isolated forest model has a high ability to identify abnormal points among the data and is especially useful in complex and noisy environments such as medical network traffic. Our experimental results and its comparison with existing methods and previous studies conducted by other researchers show significant

accuracy in predicting and correct clustering of samples. These results clearly indicate the model's ability to identify abnormal network traffic patterns that deviate from the normal state, thus contributing to a more accurate analysis of security threats. However, it should be noted that the characteristics of intrusion samples may appear differently in different datasets. Therefore, to achieve optimal results, we need detailed investigations and the use of different tools to work with different data. In addition, the effectiveness of this method should be tested and evaluated in real conditions and in medical environments, including hospitals and medical centers. These tests and evaluations will help to prove the actual application and basic functionality of the model in these environments and confirm its capabilities to improve the security and efficiency of healthcare systems.

6. Authorship contribution statement

Peyman Vafadoost: basic assistance in the concept or design of the article; Implementation of article design, analysis or interpretation of article data; Alireza Chamansara: revision of the article for important intellectual content; Hamidreza Rokhsati: writing the article, validating the results of the article; Ahmad Hajipour: revision of the article for important intellectual content; validating the results of the article.

7. References

- [1] M. Chun, S. Weber and H. Tewari, "A Lightweight Encryption Scheme for IoT Devices in the Fog," *Proceedings of the Future Technologies Conference*, 2022, pp. 147-161.
- [2] Singh, Davinder, "Internet of things," *Factories of the Future: Technological Advancements in the Manufacturing Industry*, 2023, ch. 9, pp. 195-227.
- [3] M. M. Salim et al. and J. H. Park. (2021, Sep.). Homomorphic encryption based privacy-preservation for iomt. *Applied Sciences*. [Online]. 11(18), p. 8757. Available: 10.3390/app11188757
- [4] H. Mazi, F. N. Arsene and A. M. Dissanayaka, "The influence of black market activities through dark web on the economy: a survey," In *The Midwest Instruction and Computing Symposium.(MICS)*, Milwaukee School of Engineering and Northwestern Mutual, Milwaukee, Wisconsin, 2020.
- [5] S. Gandhi, T. Poongodi and K. S. Kumar. (2024). Original Research Article Enhancing data security of cardiac patients in IoMT with Twin-Shield Encryption. *Journal of Autonomous Intelligence*. [Online]. 7(2). Available: 10.32629/jai.v7i2.1322
- [6] K. T. Kadhim, A. M. Alsahlany, S. M. Wadi, H. T. Kadhum. (2020, May.). An Overview of Patient's Health Status Monitoring System Based on Internet of Things (IoT). *Wireless Personal Communications*. [Online]. 114(3), pp. 2235-2262. Available: 10.1007/s11277-020-07474-0
- [7] A. Rghioui and A. Oumnad. (2018, Oct.). Challenges and Opportunities of Internet of Things in Healthcare. *International Journal of Electrical & Computer Engineering*. [Online]. 8(5), pp. 2753-2761. Available: 10.11591/ijece.v8i5.pp2753-2761
- [8] P. Kasinathan et al. and M. A. Spirito, "An IDS framework for internet of things empowered by 6LoWPAN" *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*, 2013, pp. 1337-1340.
- [9] M. Xie. S. Han, B. Tian and S. Parvin. (2011, Jul.). Anomaly detection in wireless sensor networks: A survey. *Journal of Network and computer Applications*. [Online]. 34(4), pp. 1302-1325. Available: 10.1016/j.jnca.2011.03.004
- [10] S. Rajasegarar, C. Leckie and M. Palaniswami. (2008, Aug.). Anomaly detection in wireless sensor networks. *IEEE Wireless Communications*. [Online]. 15(4), pp. 34-40. Available: 10.1109/MWC.2008.4599219
- [11] M. Behniafar, A. Nowroozi and H. R. Shahriari. (2018, Jul.). A Survey of Anomaly Detection Approaches in Internet of Things. *IseCure*. [Online]. 10(2). Available: sid.ir
- [12] CyberMDX, "2020 Vision: A Review of Major IT & Cyber Security Issues Affecting" Healthcare, 2020.
- [13] W. MADDOX, "Why Medical Data is 50 Times More Valuable Than a Credit Card," Healthcare, 2020.
- [14] A. Ghubaish, T. Salman, M. Zolanvari, D. Unal, A. Al-Ali and R. Jain. (2020, Dec.). Recent advances in the internet-of-medical-things (IoMT) systems security. *IEEE Internet of Things Journal*. [Online]. 8(11), pp. 8707-8718. Available: 10.1109/IIOT.2020.3045653
- [15] G. Zachos et al. and J. Rodriguez. (2021, Oct.). An anomaly-based intrusion detection system for internet of medical things networks. *Electronics*. [Online]. 10(21), p. 2562. Available: 10.3390/electronics10212562
- [16] T. S. Mohamed, S. Aydin, A. Alkhayyat and R. Q. Malik. (2022, Jul). Kalman and Cauchy clustering for anomaly detection based authentication of IoMTs using extreme learning machine. *IET Communications*. [Online]. Available: 10.1049/cmu2.12467
- [17] D. Alsalman. (2024, Jan.). A Comparative Study of Anomaly Detection Techniques for IoT Security using AMoT (Adaptive Machine Learning for IoT Threats). *IEEE Access*. [Online]. 12, pp. 14719-14730. Available: 10.1109/ACCESS.2024.3359033
- [18] J. Wang et al. and K. Zhong. (2022, Dec.). Anomaly detection in Internet of medical Things with Blockchain from the perspective of deep neural network. *Information Sciences*. [Online]. 617, pp. 133-149. Available: 10.1016/j.ins.2022.10.060
- [19] S. H. Wagan et al. and D. R. Shin. (2023, Jan.). A fuzzy-based duo-secure multi-modal framework for IoMT anomaly detection. *Journal of King Saud University-Computer and Information Sciences*. [Online]. 35(1), pp. 131-144. Available: 10.1016/j.jksuci.2022.11.007
- [20] J. B. Awotunde et al. and R. G. Jimoh, "A deep learning-based intrusion detection technique for a secured IoMT system," *International Conference on Informatics and Intelligent Applications*, Cham: Springer International Publishing, 2021, pp. 50-62.

- [21] P. Kumar, G. P. Gupta and R. Tripathi. (2021, Jan.). An ensemble learning and fog-cloud architecture-driven cyber-attack detection framework for IoMT networks. *Computer Communications*. [Online]. 166, pp. 110-124. Available: 10.1016/j.comcom.2020.12.003
- [22] M. Tavallaee, E. Bagheri, W. Lu and A. A. Ghorbani, "A detailed analysis of the KDD CUP 99 data set," *IEEE symposium on computational intelligence for security and defense applications*, Ottawa, ON, Canada, 2009, pp. 1-6.
- [23] L. Dhanabal and S. P. Shantharajah. (2015, Jun.). A study on NSL-KDD dataset for intrusion detection system based on classification algorithms. *International journal of advanced research in computer and communication engineering*. [Online]. 4(6), pp. 446-452. Available: 10.17148/IJARCCCE.2015.4696
- [24] T. Su, H. Sun, J. Zhu, S. Wang and Y. Li. (2020, Feb.). BAT: Deep learning methods on network intrusion detection using NSL-KDD dataset. *IEEE*. [Online]. 8, pp. 29575-29585. Available: 10.1109/ACCESS.2020.2972627
- [25] S. Kavitha and N. U. Maheswari. (2021, Mar.). Network anomaly detection for NSL-KDD dataset using deep learning. *Information Technology*. [Online]. 9(2), pp. 821-827.
- [26] W. Xu and F. Sabrina. (2021, Sep.). Improving performance of autoencoder-based network anomaly detection on nsl-kdd dataset. *IEEE*. [Online]. 9, pp. 140136-140146. Available: 10.1109/ACCESS.2021.3116612
- [27] F. Türk. (2023, Jun.). Analysis of intrusion detection systems in UNSW-NB15 and NSL-KDD datasets with machine learning algorithms. *Bitlis Eren Üniversitesi Fen Bilimleri Dergisi*. [Online]. 12(2), pp. 465-477. Available: 10.17798/bitlisfen.1240469
- [28] K. S. Bhuvaneshwari et al. and P. Prabu. (2022, Jul.). Improved Dragonfly Optimizer for Intrusion Detection Using Deep Clustering CNN-PSO Classifier. *Computers, Materials & Continua*. [Online]. 70(3). Available: 10.32604/cmc.2022.020769
- [29] S. S. Kaushik and P. R. Deshmukh. (2011). Detection of attacks in an intrusion detection system. *International Journal of Computer Science and Information Technologies (IJCSIT)*. [Online]. 2(3), pp. 982-986. Available: citeseerx.ist.psu.edu
- [30] B. Mahesh. (2020, Jan.). Machine learning algorithms-a review. *International Journal of Science and Research (IJSR)*. [Online]. 9(1), pp. 381-386. Available: 10.21275/ART20203995
- [31] F. T. Liu, K. M. Ting and Zhi-Hua Zhou, "Isolation forest," *eighth ieee international conference on data mining*, Pisa, Italy, 2008, pp. 413-422.
- [32] N. Abe, B. Zadrozny and J. Langford, "Outlier detection by active learning," *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2006, pp. 504-509.
- [33] Z. He, X. Xu and S. Deng. (2003, Jun.). Discovering cluster-based local outliers. *Pattern recognition letters*. [Online]. 24(9-10), pp. 1641-1650. Available: 10.1016/S0167-8655(03)00003-5
- [34] M. Al-Qatf, Y. Lasheng, M. Al-Habib and K. Al-Sabahi. (2018, Sep.). Deep learning approach combining sparse autoencoder with SVM for network intrusion detection. *Ieee Access*. [Online]. 6, pp. 52843-52856. Available: 10.1109/ACCESS.2018.2869577
- [35] C. Ieracitano et al. and A. Hussain. (2020, Apr.). A novel statistical analysis and autoencoder driven intelligent intrusion detection approach. *Neurocomputing*. [Online]. 387, pp. 51-62. Available: 10.1016/j.neucom.2019.11.016
- [36] A. Javaid, Q. Niyaz, W. Sun and M. Alam, "A deep learning approach for network intrusion detection system," *Proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies (formerly BIONETICS)*, 2016, pp. 21-26.