Journal of Autonomous Intelligence (2024) Volume 7 Issue 5
doi: 10.32629/jai.v7i5.1492

ORIGINAL RESEARCH ARTICLE

A machine learning based approach to identifying malicious activity to improve privacy in IoT-based intelligent healthcare monitoring system

Sanjeev Kumar, Sukhvinder Singh Deora*

Department of Computer Science & Application, Maharshi Dayanand University, Rohtak, Haryana 124001, India

*** Corresponding author:** Sukhvinder Singh Deora, Sukhvinder.dcsa@mdurohtak.ac.in

ABSTRACT

As the use of the Internet of Things (IoT) grows exponentially in the medical field, one of the biggest concerns is the safety of patients' personal health information. Leveraging IoT technology in a modern healthcare environment facilitates precise handling of data and patient monitoring. Healthcare systems are susceptible to security hazards and attacks. The primary objective of malicious operations targeting these systems is to compromise privacy and obtain unauthorized access to internal processes. Consequently, advanced analytics can strengthen IoT security as a whole by facilitating the detection, mitigation, and prevention of such intrusions. The fulfilment of security requirements is crucial for improving the current healthcare system with IoT technologies, and real-world applications can benefit greatly from Machine learning (ML) applications running on authentic datasets. This paper provides framework for detecting malicious activities occurred during data transmission in IoT based health monitoring system using ML approach. In proposed framework we enhanced decision tree algorithm by utilizing oversampling and fine-tuning during training of model. The proposed framework has been analysed using real dataset that contains IoT device data transmission activities that may contain activities generated by malicious nodes. The proposed mechanism achieved an accuracy of 99.6% from the perspective of other compared ML approaches.

Keywords: IoT, healthcare; cyber-attacks; ML; security and privacy

ARTICLE INFO

Received: 15 December 2023
Accepted: 15 January 2024
Available online: 15 May 2024

1. Introduction

Innovations in technology in recent times have facilitated the implementation of more precise diagnostic methods, more efficacious patient treatments, and technologies that elevate the standard of living for all. Due to the accelerated advancement of precise medical sensors, IoT-enabled devices and applications have contributed to the pervasive and intelligent nature of healthcare systems[1]. Implantable and wearable medical devices that are capable of collecting, storing, and analysing a vast array of physiological data while the patient is performing normal daily activities are included in IoT-based healthcare[2]. Medical issues can potentially be averted or detected earlier through the use of IoT devices that establish connections with adjacent devices or the cloud[3]. Consequently, IoT-based sensing devices assist healthcare organisations in meeting the growing demand for more efficient and error-free healthcare systems. IoT-based healthcare systems offer a gathering of benefits due to advancements in technology; however, they are also vulnerable to an extensive array of cyber threats. Some of the main threats to security that IoT healthcare are: device manipulation and tampering, denial of service (DoS) attacks, data

breaches and unauthorised access, etc. In an effort to compromise their performance or gather false data, hackers may try to tamper with or manipulate IoT devices. Modifying data from medical devices might have detrimental effects on treatment choices and patient safety. A common example of a security concern[4] is disabling the wireless link of a person's pacemaker to prevent hijacking. An attacker is capable of causing damage to medical procedures or altering existing ones[5].

Recently, active attacks against healthcare systems and services dependent on the IoT have commenced[6]. Additionally, cyberattacks targeting IoT-enabled healthcare systems can substantially impede or suspend medical services[7]. For example, consider a scenario in which an unauthorised individual gains access to an infusion device and alters its setting,s in order to manage an excessive quantity of insulin to a patient, potentially resulting in critical hypoglycemia[8,9]. The distinctive attributes exhibited by individual devices within an IoT framework render traditional security measures ineffectual in detecting threats. This significantly complicates the development of a security mechanism for IoTdevices[10]. In an effort to expedite the time required to bring their products to market, manufacturers are placing less emphasis on device security. Moreover, backdoors frequently incorporate by IoT device manufacturers enable malicious actors to exploit or gain remote access to the device[11]. A significant proportion of IoT devices utilised by end users are connected to the internet devoid of any security protocol. As a result, Internet of Things devices are susceptible to vulnerabilities[12]. As a consequence, safeguarding the integrity of the IoT is currently a significant concern. Although Intrusion Detection Systems (IDS) have been in existence for some time, they are currently unsuitable for use with IoT devices and networks due to their limited processing and storage capacities[13]. Consequently, it is critical to develop IDS that is IoT-enabled. For IDS training and testing, an effective IoT traffic dataset comprising both benign and malicious IoT traffic is required[14]. To evaluate and test IDSs that are supported by the Internet of Things (IoT), a limited number of researchers are endeavouring to compile a suitable IoTdataset[15]. Machine learning is a critical component for IoT-enabled healthcare systems to detect intrusions. ML models are capable of simulating the customary operations of medical equipment and systems. Machine learning can identify potential dangers in any deviation from this learned behaviour. They possess the capability to foresee potential vulnerabilities or entry points for attacks and implement preventive measures prior to their conscious awareness. Moreover, solutions powered by ML enable monitoring of IoT devices and network traffic in real time. Moreover, through the perpetual assimilation of real data and the identification of emerging hazards, these models exhibit the capacity to evolve and adapt. Machine learning is capable of analysing the behaviour of devices and users[16]. It is able to identify indicators of a security compromise, such as unusual device operation or user access patterns. The major contribution of paper is:

- To examine role of machine learning in the detection of attacks in patient monitoring healthcare system. It is essential for identifying attacks in patient monitoring healthcare systems. Machine learning algorithms, namely those utilising anomaly detection, have the capability to acquire knowledge about the typical functioning of patient monitoring systems. Machine learning is utilised in the development of sophisticated Intrusion Detection Systems specifically designed for healthcare system. These systems employ constant surveillance of network activity and possess the ability to promptly identify and address any dubious conduct or breaches in security.
- To propose a framework that uses enhanced decision tree approach for the detection of attacks from real dataset collected from Kaggle. Machine learning-based Intrusion Detection Systems (IDS) have the capability to adjust to changing attack methods, hence enhancing the system's capacity to identify new and unfamiliar security breaches.
- To analyse the performance of proposed framework with various performance metrices called accuracy, precision, recall and F1 score.

The paper is divided into five sections. The section 1 provides introduction of ML and IoT for patient

monitoring in healthcare environment, in section 2 review of literature has been presented, section 3 provides proposed framework, in section 4 results and discussion are presented, at last in section 5 conclusion and future research directions has been presented.

2. Review of literature

In this section review of literature on ML based attack detection in IoT based healthcare monitoring environment has been discussed. Here presents an outline to IoT applications in healthcare, focusing on how devices are integrated for patient monitoring, data collecting, and device-to-device communication. Furthermore, learn about the particular security issues in the Internet of medical things, including data privacy, integrity, and authentication, as well as the susceptibility of medical devices to cyberattacks. Moreover, emphasise how ML may be used to identify anomalies or cyberattacks in IoT health.Bharadwajet al.[17] present a novel security design for SHSs that uses ML to detect malicious activities called HealthGuard. By monitoring the patient's vital signs from many SHS-connected devices and comparing them, HealthGuard is able to identify changes in the patient's bodily processes and identify malicious or benign actions. Similarly, a methodology for creating context-aware security solutions for the IoT that can identify fraudulent traffic in IoT use cases was proposed by Hussain et al.[18]. An open-source IoT data generating tool called IoT-Flock is at the heart of the suggested framework. In order to build an Internet of Things use case with both benign and malevolent IoT devices, researchers can utilise the IoT-Flock tool to simulate traffic. Moreover, a IDS framework with a high detection rate and a more accurate false alarm rate was constructed by Iwendi et al.[19] using a feature optimisation approach that merged a Random Forest (RF) and a genetic algorithm. Similarly, a number of attacks were covered by Butt et al.[20], along with their effects on health monitoring systems and some recommendations based on their study. Furthermore, a fog-based attack detection (FBAD) system, suggested by Alrashdi et al.[21], suggests utilising an ML to effectively identify harmful actions.

Similarly, an analysis of smart healthcare within the framework of a smart city is provided by BahalulHaque et al.[22], encompassing current and pertinent research domains and their respective applications. The use of modern medical technology in early illness detection and emergency services has been the subject of much discussion. Concerns about privacy and security, as well as the difficulties presented by new technology like wearables and massive healthcare data, are also major themes. In order to create a safe smart healthcare system, Ambarkar and Shekokar[23] try to examine the architecture of the system, potential dangers, weaknesses, and security protocols. Furthermore, in order to identify ransomware, Iqbal et al.[24] suggest a hybrid approach that uses text, picture data, and application code to decipher encrypted or plain threat language. One of the best advantages for ransomware detection might be the ability to identify potentially harmful content. Moreover, Kalnoor and Gowrishankar[25] create an intelligent intrusion detection system (I-IDS) based on ML models. Data generated in an IoT smart environment is modelled taking both benign and harmful attacks into account. Furthermore, the goal of the model for smart healthcare service security proposed by Choi et al.[26] is to build it using the IoT. they present a paradigm for creating security zones for IoT services and apply it to smart healthcare services. We also summarise the security needs for IoT environment.

According to review of literature it has been observed that numerous studies have concentrated on examining Internet of things (IoT) applications, clarifying their executions, contrasting their contributions, and pinpointing unresolved issues. These papers provide valuable insights into crucial elements of IoT security within the healthcare domain. Given the emphasis on machine learning and deep learning, these methods encapsulate the difficulties that intelligent solutions encounter, such as the requirement for a comprehensive and authentic dataset.

3. Proposed framework

In this section, a proposed framework for detection of malicious profiles in IoT based healthcare system has been presented. In proposed framework real dataset is collected from Kaggle which is generated by the IoT-Flock tool. IoT-Flock is a freely available IoT traffic generating tool. It permits a user to construct an IoT use case, add customisedIoT devices to it, and produce regular and malicious IoT traffic. The dataset contains activities performed by normal IoT devices and malicious devices. In the generation of data using Flock, two IoT routing protocols are used: CoAP and MQTT. CoAP (Constrained Application Protocol) and MQTT (Message Queuing Telemetry Transport) are simple and effective methods intended for communication in restricted and IoT contexts, respectively. The dataset contains two types of data, such as patient monitoring and environment monitoring, that will be transmitted through MQTT-based devices. Further, the raw dataset is pre-processed using oversampling methodology in order to retrieve more relevant data as depicted in **Figure 1**. Preprocessing involves the act of refining and filtering the input data and prior to providing it to a machine learning model. Oversampling is a method employed to tackle imbalances between classes in classification tasks, when one class has a significantly lower number of cases compared to another. In actual datasets, certain classes may exhibit unusual or discrimination, resulting in biased models that exhibit worse performance on the minority class. Oversampling entails generating artificial examples of the underrepresented class to equalise the split of classes.

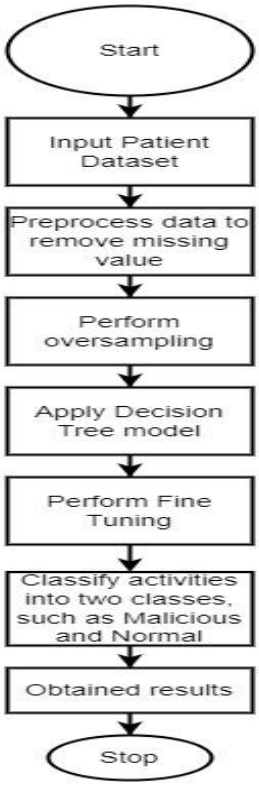

Figure 1. proposed framework.

After that ML models are applied to pre-processed data in order to classify normal and malicious activities. In the proposed framework, we apply fine tuning to the decision tree approach in order to achieve better results Fine-tuning refers to the process of optimising various parameters of a machine learning model in order to enhance its performance.

Additionally, fine-tuning may entail the implementation of regularisation methods to mitigate the risk of overfitting. Regularisation imposes a penalty on intricate models to prevent the inclusion of irrelevant information from the training data and enhance the ability to generalise to unfamiliar data.

4. Results and discussion

In order to implement the proposed mechanism, Python and the Google Collab platform are utilised. The method under consideration employs real datasets collected from Kaggle. The dataset comprises two distinct types of data: environmental monitoring data and patient monitoring data. The dataset comprises two distinct class types, namely malicious and normal.

4.1. Performance metrices used

- Accuracy: It is the fraction of correctly classified instances among all instances.
- Precision: It is defined as the proportion of accurately predicted positive observations to all expected positives.
- Recall: The ratio of accurately anticipated positive observations to all observations in the actual class is calculated by recall.
- F1 score: The harmonic mean of precision and recall is the F1 score. It takes into account both false positives and false negatives.
- Confusion matrix: It displays the true positive, true negative, false positive, and false negative prediction counts.

4.2. Discussion

The performance of proposed framework has been evaluated using four performance metrics such as accuracy, precision, recall and F1 score. **Tables 1–4** depicts the comparison of Accuracy, precision, recall and F1 score of proposed enhanced decision tree and other state of art models such as naïve bayes, and logistic regression model and support vector machine. Decision trees are a form of supervised machine learning technique employed for classification and regression applications. It utilises a sequence of inquiries to arrive at decisions. Every internal node in the tree reflects a choice made using a specific characteristic, while each leaf node represents the final conclusion or class label. Furthermore, naive bayes is a probabilistic method primarily employed for classification. The approach relies on Bayes' theorem and assumes that the characteristics are conditionally independent, given the class label. Despite its seemingly simplistic premise, it frequently demonstrates strong performance, particularly when used to textual data. Moreover, logistic regression is a linear model utilised for the purpose of binary and multiclass classification jobs. The logistic function is used to estimate the likelihood of an instance belonging to a specific class. Contrary to its title, logistic regression is employed for classification tasks rather than regression. The support vector machine (SVM) is a robust and flexible machine learning technique employed for both classification and regression problems. It is used to identify an optimal hyperplane that effectively divides the data points into distinct groups. When the data is not linearly separable, it can employ a kernel method to transform the data into a higher-dimensional space. This enables the identification of a hyperplane that effectively divides the different classes. Furthermore, **Figure 2** shows the accuracy comparison, **Figure 3** shows the precision comparison, **Figure 4** shows the recall comparison, and **Figure 5** depicts the F1 score comparison between proposed Decision tree and the other state of art models. According to results in the proposed framework, the rate of accuracy, precision, and F1 score is high from the perspective of other models. In proposed framework the accuracy is 99.69% which is high then other models. Whereas in naïve bayes accuracy is 79.67% which is low as compare to other models. In proposed framework the precision value is 99.37%, recall value is 99.47% and F1 score is 99.63% which is optimal in perspective of other models.

Table 1. Accuracy comparison of proposed approach and other state of art models.

Models	Accuracy %
Naïve bayes	79.67
Logistic regression	95.28
Support vector machine	97.71
Proposed enhanced decision tree	99.69

Table 2. Precision comparison of proposed approach and other state of art models.

Models	Precision %
Naïve bayes	99.7
Logistic regression	90.35
Support vector machine	94.24
Proposed enhanced decision tree	99.37

Table 3. Recall comparison of proposed approach and other state of art models.

Models	Recall
Naïve bayes	52.18
Logistic regression	99.5
Support vector machine	96.27
Proposed enhanced decision tree	99.47

Table 4. F-1 score comparison of proposed approach and other state of art models.

Models	F1 score
Naïve bayes	68.5
Logistic regression	94.7
Support vector machine	94.67
Proposed enhanced decision tree	99.63

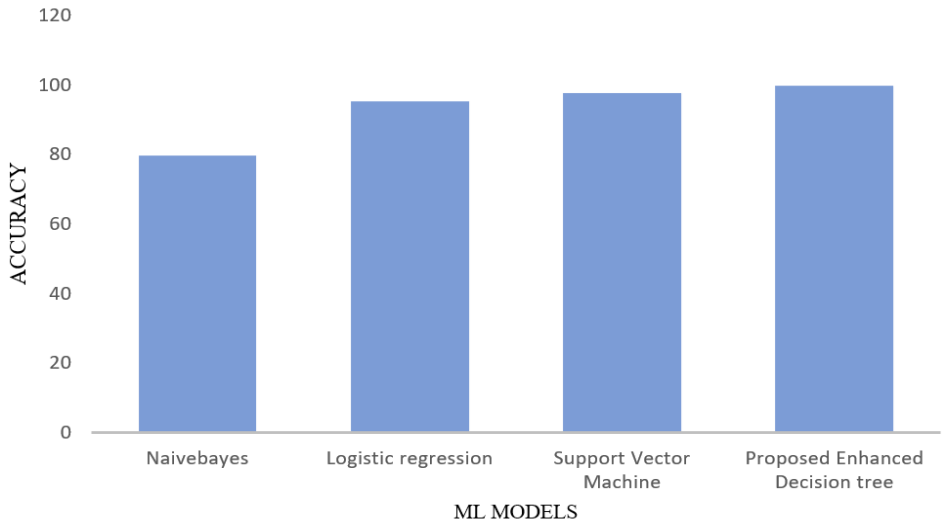

Figure 2. Accuracy comparison of ML approaches.

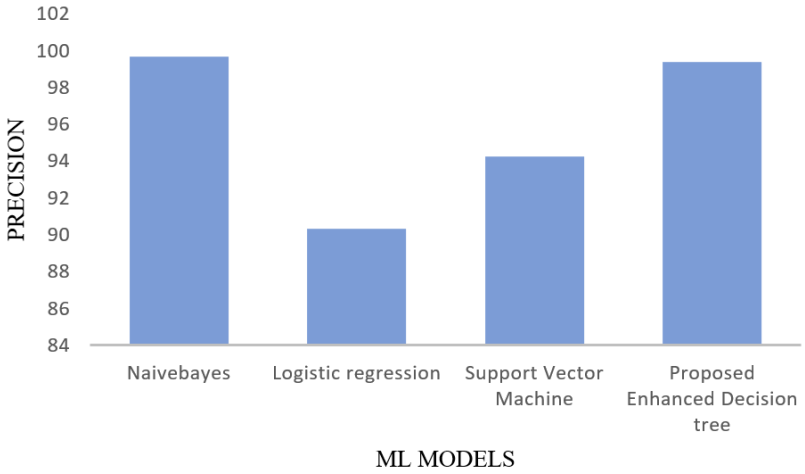

Figure 3. Precision comparison of ML approaches.

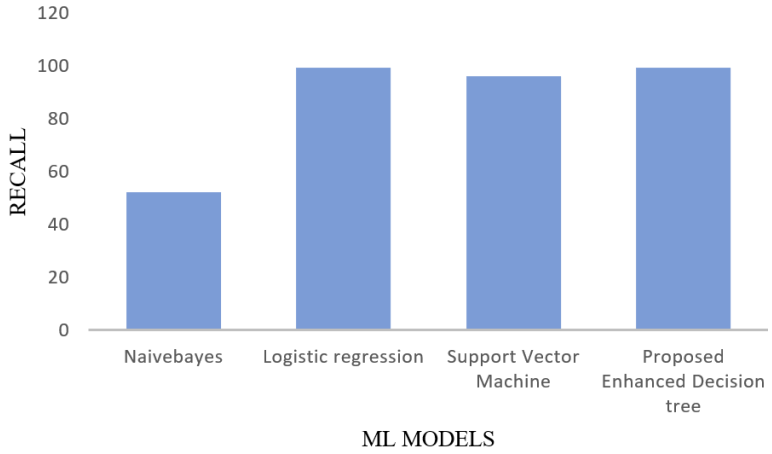

Figure 4. Recall comparison of ML approaches.

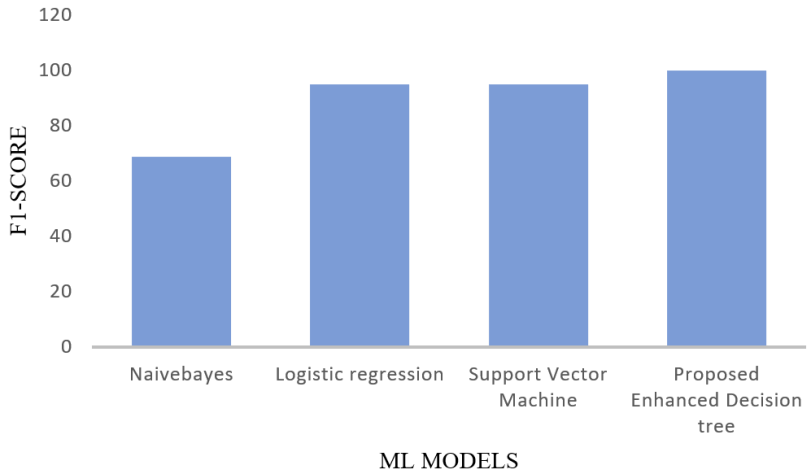

Figure 5. F1 score comparison of ML approaches.

5. Conclusion and future directions

Massive volumes of data are generated in IoT instances. With the use of machine learning (ML), this data may be analysed to find patterns that help distinguish between typical data gathered by trustworthy devices and abnormal or malicious data that might indicate an attack. ML methods are quite good at finding anomalies. They are able to recognise typical traffic patterns and issue a warning whenever variations or abnormalities take place. These may be signs of impending attacks or unusual behaviour from the device. In this paper, we suggest an enhanced decision tree model that uses oversampling and fine tuning to find attacks

in the normal traffic sent by IoT-enabled sensors that are used to keep an eye on patients and the healthcare environment. The dataset is collected from Kaggle. The proposed framework is analysed in Python. The results show that in enhanced DT, the accuracy is 99.6%, which is high compared to the other two models used for comparison. In future investigating privacy-preserving methodologies such as secure multi-party computation or homomorphic encryption in order to analyse data without compromising the privacy of individual users. Moreover, it is intended to develop a ML model that are specifically designed to resist adversarial attacks in the context of IoT healthcare systems.

Author contributions

Conceptualization, SD and SK; methodology, SK; software, SK; validation, SD; formal analysis, SD; investigation, SK; resources, SK; data curation, SK; writing—original draft preparation, SK; writing—review and editing, SD; visualization, SK; supervision, SD; project administration, SD; funding acquisition, SK. All authors have read and agreed to the published version of the manuscript.

Conflict of interest

The authors declare no conflict of interest.

References

1. Chouhan A, Tiwari A, Diwaker C, et al. Efficient Opportunities and Boundaries towards Internet of Things (IoT) Cost Adaptive Model. 2022 IEEE Delhi Section Conference (DELCON). Published online February 11, 2022. doi: 10.1109/delcon54057.2022.9753057
2. Kumar S, Deora SS. Comparative Analysis of Security Techniques in Internet of Things. 2022 Seventh International Conference on Parallel, Distributed and Grid Computing (PDGC). Published online November 25, 2022. doi: 10.1109/pdgc56933.2022.10053313
3. Otoum Y, Liu D, Nayak A. DL-IDS: A deep learning–based intrusion detection framework for securing IoT. Transactions on Emerging Telecommunications Technologies. 2019; 33(3). doi: 10.1002/ett.3803
4. Diwaker C, Tomar P, Sharma A. Future aspects and challenges of the internet of things for the smart generation. In ICCCE 2018: Proceedings of the International Conference on Communications and Cyber Physical Engineering 2018. Springer Singapore. 2019. pp. 599-606.
5. Kumar S, Deora SS. Security Challenges and Issues in IoT. 2021 6th International Conference on Signal Processing, Computing and Control (ISPCC). Published online October 7, 2021. doi: 10.1109/ispcc53510.2021.9609486
6. Lone AN, Mustajab S, Alam M. A comprehensive study on cybersecurity challenges and opportunities in the IoT world. SECURITY AND PRIVACY. 2023; 6(6). doi: 10.1002/spy2.318
7. Elshweikh AA, Mattar AM, Hussein M, et al. Literature Survey for Cybersecurity for Internet of Things (IoT). 2022 International Telecommunications Conference (ITC-Egypt). Published online July 26, 2022. doi: 10.1109/itc-egypt55520.2022.9855671
8. Mehla A, Deora SS. Use of Machine Learning and IoT in Agriculture. In IoT Based Smart Applications. Cham: Springer International Publishing. 2022. pp. 277-293.
9. Pires IM, Hussain F, Garcia NM, et al. Improving Human Activity Monitoring by Imputation of Missing Sensory Data: Experimental Study. Future Internet. 2020; 12(9): 155. doi: 10.3390/fi12090155
10. Al-Shareeda MA, Manickam S, Laghari SA, et al. Replay-Attack Detection and Prevention Mechanism in Industry 4.0 Landscape for Secure SECS/GEM Communications. Sustainability. 2022; 14(23): 15900. doi: 10.3390/su142315900
11. Goyal P, Deora SS. A Review: Trust Management Techniques Used for Cloud Computing. Proceedings of Data Analytics and Management: ICDAM 2021. 2022; 1: 117-132.
12. Ghazanfar, S.; Hussain, F.; Rehman, A.U.; Fayyaz, U.U.; Shahzad, F.; Shah, G.A. Iot-flock: An open-source framework for iot traffic generation. In Proceedings of the 2020 International Conference on Emerging Trends in Smart Technologies (ICETST); Karachi, Pakistan; 26–27 March 2020; pp. 1-6.doi: 10.1109/icetst49965.2020.9080732
13. Alhowaide A, Alsmadi I, Tang J. Ensemble Detection Model for IoT IDS. Internet of Things. 2021; 16: 100435. doi: 10.1016/j.iot.2021.100435
14. Sarhan M, Layeghy S, Moustafa N, et al. Feature extraction for machine learning-based intrusion detection in IoT networks. Digital Communications and Networks. Published online September 2022. doi: 10.1016/j.dcan.2022.08.012
15. Carta S, PoddaAS, Recupero DR, et al. A Local Feature Engineering Strategy to Improve Network Anomaly

Detection. Future Internet. 2020; 12(10): 177. doi: 10.3390/fi12100177
16. Corizzo, R.; Zdravevski, E.; Russell, M.; Vagliano, A.; Japkowicz, N. Feature extraction based on word embedding models for intrusion detection in network traffic. J. Surveillance, Secur. Saf. 2020, 1, 140–150.doi: 10.20517/jsss.2020.15
17. Bharadwaj HK, Agarwal A, Chamola V, et al. A Review on the Role of Machine Learning in Enabling IoT Based Healthcare Applications. IEEE Access. 2021; 9: 38859-38890. doi: 10.1109/access.2021.3059858
18. Hussain F, Abbas SG, Shah GA, et al. A Framework for Malicious Traffic Detection in IoT Healthcare Environment. Sensors. 2021; 21(9): 3025. doi: 10.3390/s21093025
19. Iwendi C,Anajemba JH, Biamba C, et al. Security of Things Intrusion Detection System for Smart Healthcare. Electronics. 2021; 10(12): 1375. doi: 10.3390/electronics10121375
20. Butt SA, Diaz-Martinez JL, Jamal T, et al. IoT Smart Health Security Threats. 2019 19th International Conference on Computational Science and Its Applications (ICCSA). Published online July 2019. doi: 10.1109/iccsa.2019.000-8
21. Alrashdi I, Alqazzaz A, Alharthi R, et al. FBAD: Fog-based Attack Detection for IoT Healthcare in Smart Cities. 2019 IEEE 10th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON). Published online October 2019. doi: 10.1109/uemcon47517.2019.8992963
22. BahalulHaque AKM, Bhushan B, Nawar A, et al. (2022). Attacks and countermeasures in IoT based smart healthcare applications. In: Recent Advances in Internet of Things and Machine Learning: Real-World Applications. pp. 67-90.
23. Ambarkar SS, Shekokar N. Toward smart and secure IoT based healthcare system. Internet of things, smart computing and technology: A roadmap ahead. 2020; 283-303.
24. Iqbal MJ, Aurangzeb S, Aleem M, et al. RThreatDroid: A Ransomware Detection Approach to Secure IoT Based Healthcare Systems. IEEE Transactions on Network Science and Engineering. 2023; 10(5): 2574-2583. doi: 10.1109/tnse.2022.3188597
25. Kalnoor G, Gowrishankar S. IoT-based smart environment using intelligent intrusion detection system. Soft Computing. 2021; 25(17): 11573-11588. doi: 10.1007/s00500-021-06028-1
26. Choi J, Choi C, Kim S, et al. Medical Information Protection Frameworks for Smart Healthcare based on IoT. Proceedings of the 9th International Conference on Web Intelligence, Mining and Semantics.Published online June 26, 2019. doi: 10.1145/3326467.3326496