DESIGN AND DEVELOPMENT OF PRIVACY-PRESERVING METHOD AND ANOMALY DETECTION METHOD USING MACHINE LEARNING FOR IOT-BASED HEALTHCARE

Dr. Shipra Kumari
Utkal University, Vani Vihar Bhubaneswar
Email ID: shipra.kumari.iitd@gmail.com

Dr. Prafulla Kumar Behera
Utkal University, Vani Vihar Bhubaneswar
Email ID: pkbehera.cs@utkaluniversity.ac.in

ABSTRACT: Sharing electronic health data helps clinicians detect some diseases and create appropriate treatment options, especially in the medical industry. Data sharing helps the medical business as a whole develop to some level. However, sharing resources is now difficult because of the sensitivity of medical data. This medical data is hesitant to divulge patient medical information because doing so could put them in violation of privacy ethics or even cause them to lose out financially, such as by publicizing information on patients or the active ingredients in patented therapeutic treatments. The primary goal of this investigation is to develop and design a privacy-preserving method and an anomaly detection method for IOT-based healthcare. This study uses machine learning to put this technique into practice. An Intrusion Detection System (IDS) and an Event Detection System (EDS) are both proposed in this article for use in a smart hospital IoT system in order to simultaneously detect both network intrusions and proceedings of interest concerning the health along with environment of patients. It has been demonstrated that utilizing a single system for the supervision of network infrastructure along with monitoring of Electronic Health Records (EHR) can both maximize the use of available resources and ensure the dependability of the system. As a result, decisions about patients' care and the adaption of their settings are made with a higher degree of precision. The edge deployment of the low latency allows for processing to take place in close proximity to the data sources. The suggested ADS is put into action and tested with the help of a simulator, and the EHR event detection is accomplished with the use of a realistic data-set investigation. According to the findings, the accuracy of detection is quite high for both HER-related events and intrusions into IoT networks.

Keywords: *IOT; Healthcare sector; Privacy Preserving techniques; Anomaly Detection; Machine learning; Deep learning.*

INTRODUCTION:

Electronic health records are capable of storing a wide variety of electronic medical data. The size of the electronic health record (EHR) market within India is anticipated to grow at a compound annual growth rate (CAGR) of 7 per cent from its projected value of $3850 crore in

2022 to an anticipated value of $6630 crore in 2030[1]. AI is enabling novel possibilities in healthcare by enhancing medicine in previously unthinkable ways and addressing a few major global healthcare concerns. For instance, the protein folding concern that prevented significant improvements in biology along with medicine in the last 50 years was fixed by AlphaFold, a new AI-powered protein structure prediction tool [1,2]. Pharmaceutical businesses may mimic clinical trials for drug development on larger population models having greater flexibility and fewer budget constraints because of innovations like InSilico Trailing, enabling them to develop top-notch pharmaceutical goods. AI has a wide range of applications in the medical field. Major medical specialities where researchers are striving to use AI-based digital solutions are shown in Fig. 1 [3]. The potential abuse of these advancements raises ethical questions, which is a drawback. Drug finding AI identified 40,000 potentially fatal compounds and the most potent nerve poisons in six hours. Nevertheless, there is unanimity regarding the advances AI could bring to healthcare, including improving quick diagnosis, customizing care, along with minimizing redundant outpatient sessions that possibly will save the economy billions [4,5].

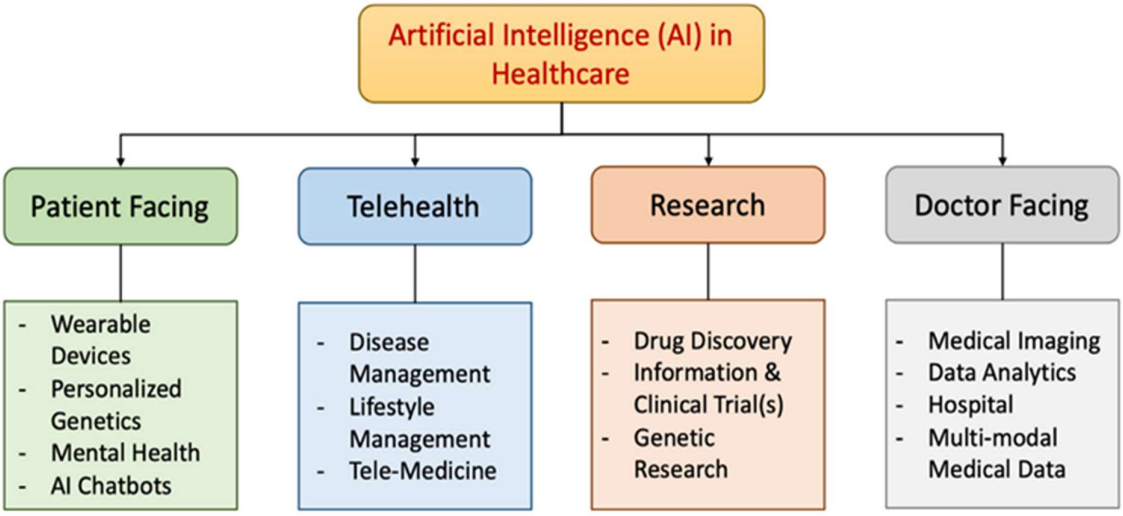

Figure 1: Illustration of artificial intelligence technology landscape in healthcare [3].

In order for machine learning (ML)-based AI algorithms to achieve human-level efficiency in pattern-matching tasks, a substantial quantity of high-quality data is necessary. The fact that data powers these algorithms raises serious issues with data privacy, particularly when the data needed for AI training contains private and sensitive patient data [5]. Patients, healthcare providers, and software suppliers could all suffer severe harm as a result of data leaks or exploitation of any kind.

[1] https://www.medicalbuyer.co.in/ehr-a-long-way-to-go-2/

Figure 2: Challenges towards building privacy-preserving AI in healthcare [3].

Despite performing better than humans, data privacy along with security concerns must be adequately addressed if AI is to fully realize its potential in the healthcare industry. The next section outlines the earlier literature on this topic.

LITERATURE REVIEW:

Table 1: Literature review

AUTHORS AND YEAR	METHODOLOGY	FINDINGS
Chatterjee & Ahmed (2022) [5]	The classification of IoT anomaly detection algorithms is covered after a brief overview of the detection techniques and applications. Following that, the study examined papers selected based on our search criteria in order to find unique application fields from recent publications.	This study found a lack of IoT anomaly detection approaches, such as when integrating systems with different sensors, dealing with data and idea drifts, and augmenting data in the absence of Ground Truth data.
Zhang et al., (2022) [8]	The security study demonstrated that the suggested solution fulfils data privacy. Theoretically, the costs of computation and communication are also examined.	The results of the experiments reveal that, in comparison to other existing systems, the proposed scheme obtained promising outcomes while guaranteeing that privacy was preserved.
AbaOud et al.,	Through the deployment of	As a result, the promise of a new

(2023) [9]	privacy-preserving federated learning models, the author of this study proposed an original technique that was aimed to overcome these issues.	frontier in collaborative healthcare informatics was confirmed by the findings of the study, which highlighted the potential of utilizing collective intelligence in healthcare while providing the highest possible level of privacy protection.
Das & Namasudra (2023) [10]	A privacy-preserving reciprocal authentication approach for IoT-enabled healthcare systems is suggested for lightweight and effective network device authentication. This authentication system uses lightweight cryptographic primitives including XOR, concatenation, and hash operation to allow IoT device processing.	Unauthorized devices cannot access healthcare systems using the proposed scheme's secure session across an authorized device along with a gateway. The security and performance analyses compare the proposed authentication method to well-known techniques.
Xu et al., (2023) [10]	This study introduced a data-driven approach for anomaly and intrusion detection, where data is filtered and processed utilising various algorithms. Utilizing mutual information along with the Synthetic Minority Oversampling Technique (SMOTE) algorithm, the training dataset's quality is raised.	The resulting approach outperforms the current algorithms by a substantial margin, correctly resolving a multi-class classification problem with 99.7% accuracy.

However, the data along with the privacy details are saved on the cloud server, which means there is still the possibility of an attack happening. This is extremely evident from the previous literature, which can be found here and here. As a result, it is essential to suggest an IoT-enabled healthcare system that protects patients' privacy in order to guarantee the continued safety of their data. Therefore, this investigation's primary objective is to carry out a study on the design and development of a privacy-preserving method and an anomaly detection method using machine learning for internet-of-things-based healthcare.

METHODOLOGY:

This study makes a proposal for a system that would integrate hospital network infrastructure monitoring with electronic health record monitoring. Following research into already existing EHR systems as well as various Internet of Things (IoT) systems, a proposed IoT architecture

has been developed to ensure an effective solution for anomaly detection. SVM has been selected as the detection approach for data analysis and processing based on comparative works using the most recent state-of-the-art research. Also, it was evaluated using two distinct types of data sets: those pertaining to EHR on the one hand, and those pertaining to network infrastructure on the other. To guarantee minimal notification latency and the security of sensitive data, a placement plan that involves a centralised solution right at the network's edge has been selected. This decision was made to maximize security. After that, a prototype is developed and put through its paces using a variety of scenarios involving the detection of EHR event failures or intrusions in order to evaluate the architecture that was proposed and the algorithm that was selected. The study focused on two specific use cases, namely fire detection along with heart attack detection, in the context of Electronic Health Record (EHR) event detection. In order to assess the effectiveness of network infrastructure intrusion detection, three different types of assaults were simulated and afterwards analysed for their detection accuracy. A tool called a wireless sensor network simulator is utilized to assess the proposed system's performance as well as to determine whether or not it is scalable. The results of the experiments demonstrate that this system is effective in anomaly detection accuracy and its quick latency from the sensing stage to the decision-making stage.

RESULTS AND DISCUSSIONS:

The system that is proposed to detect anomalies in system behaviour does so by first determining the normal behaviour of the system and then employing this normal behaviour as a baseline. Because of this, it considers anything that deviates from the norm to be an oddity. Because of this, the support vector machine (SVM), which is an algorithm for machine learning, was utilized to solve classification issues. The non-linear support vector machine allows for the capture of complex correlations between distinct factors without the need to execute difficult transformations. This is made possible by the use of a variety of parameters. The basic notion of SVM is as follows: In order to partition the data into their respective categories, the algorithm draws a line or a hyperplane. In order to reformat the data, it employs a mathematical function that is referred to as the kernel. The SVM algorithm will then construct an optimal boundary between the labels once this transformation is complete. Primarily, it applies a series of transformations in an effort to discover a way to segment the data in accordance with the labels or outputs that have been specified. The selection of this detection technique, which illustrates the efficacy of SVM in wireless sensor network (WSN) anomaly detection in contrast to various ML algorithms for anomaly detection, is the reason for this choice. Anomaly detection can be handled by other machine learning methods that are currently available; however, this work's focus is on the evaluation along with experimentation of the SVM technique.

A smart hospital infrastructure serves as the setting for the presentation of anomaly detection scenarios. There are two different scenarios that have been proposed: one pertains to the detection of network intrusions, and the other refers to the detection of EHR events. In order to guarantee the accuracy of the assessment of the proposed procedure for intrusion detection, along with the three assaults that are most frequently seen in IoT networks were selected to test the suggested solution. The following is a description of each of these three attacks: The Rank along with Version Number Modification Attack Scenarios, and the Flooding Attack Scenario. The architecture of the smart hospital is mostly made up of two categories of sensors:

Environmental sensors and body sensors. However, environmental sensors, such as temperature and humidity sensors, are dispersed around the building in various rooms. Human body temperature, along with oxygen level, and also heart rate sensors are examples of the types of body sensors that control the essential functions of the human body. As a result, to put the process of event detection to the test, two scenarios have been presented for each kind of sensor, and they are as follows: The Anomaly Scenario of the Environment Sensors and the Anomaly Scenario of the Body Sensors.

A metric known as the Anomaly Detection Rate (ADR) was used in this study to evaluate the effectiveness of the suggested system. This metric determines the frequency of abnormal occurrences and observations over a given amount of time. So, this measure is considered to be one of the most vital ones due to the fact that one of the major challenges that wireless networks must overcome is ensuring that the data that is being delivered is valid. Within the scope of this discussion, ADR enables the evaluation of the precision of the identification of abnormalities that are associated with data or even networks. The ADR is determined by the equation that is shown below (1).

$$ADR = \frac{\text{Abnormal events detected by SVM}}{\textit{Total data amount}} \quad \text{----------(1)}$$

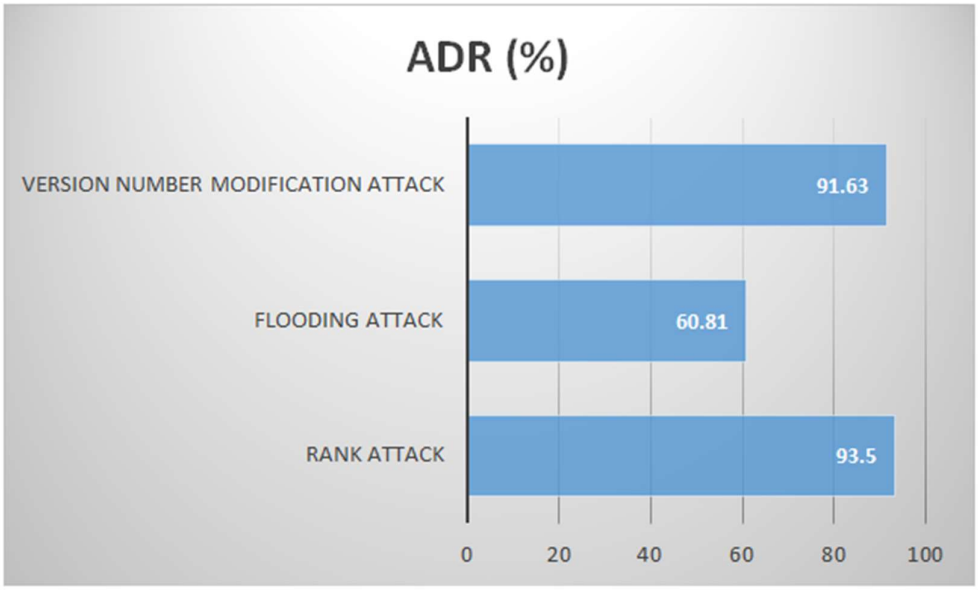

Figure 4: Anomaly detection rates for different attacks.

The results of simulations of assaults on networks are displayed in the figure located above. IDC (Intrusion Detection Component) is able to identify two distinct kinds of routing attacks having a high ADR (the Rank attack along with the Version number modification assault both have an ADR of more than 90%), as can be seen in the figure below. An ADR that is low is sufficient to detect a flooding attack. After the flooding attack has been completed, the countermeasures that are utilized in the simulator in an effort to restore it to its normal state may provide an explanation for this phenomenon.

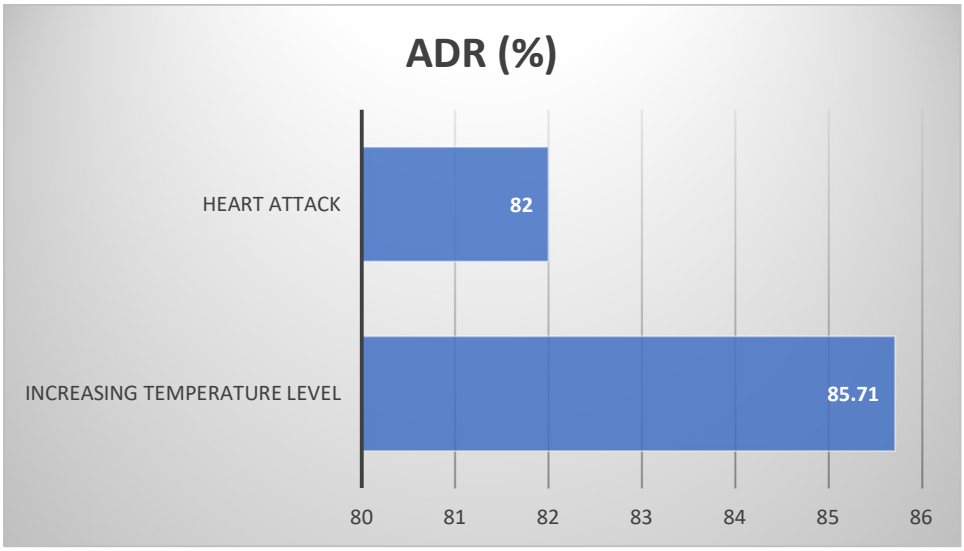

Figure 5: Anomaly detection for event scenario

In order to replicate anomalous behaviour, such as a fire, the temperature level is made higher, and the humidity level is made lower. A high ADR is used by EDC (Event Detection Component) to perceive any changes in the surrounding environment. In addition, the EDC identifies an abnormal bodily state together with a high ADR while dealing with the situation of a heart attack. Both EDC and IDC make use of a single-class support vector machine (SVM) that has a non-linear kernel (RBF) and has the following set of parameters: nu = 0.1, kernel = "rbf," and gamma = 0.1. Python version 3.6 is utilized as the programming language, and the ML library scikit-learn is taken into consideration when attempting to develop One-class SVM.

CONCLUSION:

To conclude, since EHR data have to be efficiently transferred, reliability in IoT is an extremely important topic to discuss. An Anomaly Detection System, abbreviated as "ADS," is proposed for use in smart hospital infrastructures in this article. The system would consist of two modules: IDC, which would be used to detect network abnormalities and assaults, and EDC, which would be used to detect HER-related events. Both the IDC and EDC modules work together well in a single, unified system, which results in a simplification of the administrative process and a reduction in the amount of money spent on system management. The ADS placement technique is carried out at the edge router in a completely centralized fashion. The sensitivity of the data, which could be stolen or corrupted if it were transported to the cloud, was a driving factor in the decision to use edge computing. In addition, processing data in close proximity to its sources greatly lessens decision latency and increases the network bandwidth. The innovation of this study is in its integration of two pre-existing anomaly detection systems, namely EHR monitoring along with infrastructure supervision, under a unified platform. This feature has the advantage of enhancing the overall dependability of the system and, as a result, delivering correct decision-making on EHR. Integration of subsystems also improves resource management and optimizes its use.

REFERENCES:

1. Milana, C., & Ashta, A. (2021). Artificial intelligence techniques in finance and financial markets: a survey of the literature. *Strategic Change*, *30*(3), 189-209.

2. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., ... & Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583-589.
3. Khalid, N., Qayyum, A., Bilal, M., Al-Fuqaha, A., & Qadir, J. (2023). Privacy-preserving artificial intelligence in healthcare: Techniques and applications. *Computers in Biology and Medicine*, 106848.
4. Sharma, S., Chen, K., & Sheth, A. (2018). Toward practical privacy-preserving analytics for IoT and cloud-based healthcare systems. *IEEE Internet Computing*, *22*(2), 42-51.
5. Fahim, M., & Sillitti, A. (2019). Anomaly detection, analysis and prediction techniques in iot environment: A systematic literature review. *IEEE Access*, *7*, 81664-81681.
6. Chatterjee, A., & Ahmed, B. S. (2022). IoT anomaly detection methods and applications: A survey. *Internet of Things*, *19*, 100568.
7. Zhang, L., Xu, J., Vijayakumar, P., Sharma, P. K., & Ghosh, U. (2022). Homomorphic encryption-based privacy-preserving federated learning in iot-enabled healthcare system. *IEEE Transactions on Network Science and Engineering*.
8. AbaOud, M., Almuqrin, M., & Khan, M. F. (2023). Advancing Federated Learning through Novel Mechanism for Privacy Preservation in Healthcare Applications. *IEEE Access*.
9. Das, S., & Namasudra, S. (2023). Lightweight and efficient privacy-preserving mutual authentication scheme to secure internet of things-based smart healthcare. *Transactions on Emerging Telecommunications Technologies*, e4716.
10. Xu, H., Sun, Z., Cao, Y., & Bilal, H. (2023). A data-driven approach for intrusion and anomaly detection using automated machine learning for the Internet of Things. *Soft Computing*, 1-13.