International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Integrity Shield: Ensuring Real-time Data Integrity in Healthcare IoT with Isolation Forest Anomaly Detection

Dr. Sudhanshu Maurya[1], Yahya Al Balushi[2], Jyoti Kharade[3], Jagadeesh B N4, Dr. Pavithra G[5], Achyutha Prasad N[6]

Submitted: 10/12/2023 **Revised:** 16/01/2024 **Accepted:** 30/01/2024

Abstract: The healthcare sector has been greatly transformed by the Internet of Things (IoT) which brings opportunities, for monitoring and management of patient health. However, there are challenges in ensuring the reliability and authenticity of the amount of healthcare data transmitted through IoT devices. In this paper we suggest an approach called" Machine Learning Based Data Integrity Assurance for Healthcare IoT" to tackle these challenges. Our proposed algorithm utilizes machine learning techniques to detect anomalies and potential tampering attempts in time thus guaranteeing the trustworthiness and dependability of healthcare data collected from IoT devices. By establishing data profiles and continuously monitoring data streams our algorithm can adjust to evolving data patterns. Promptly identify any issues related to data integrity. Moreover, through trust-based data fusion our algorithm takes into account the trust level associated with each device in order to appropriately assess their contributions. With its adaptability, scalability and cost effectiveness our solution holds promise in enhancing the security and integrity of healthcare data, within IoT based healthcare systems.

Keywords: *Healthcare, Internet of things, Machine learning, SVM*

1. Introduction

The integration of the Internet of Things (IoT) into the healthcare sector has brought transformative advancements in patient monitoring and healthcare delivery. With wearable devices and sensors collecting vast amounts of health data from patients, the potential for personalized and remote healthcare has become a reality. However, with this vast volume of sensitive health data transmitted through IoT devices, ensuring its integrity and authenticity has become a critical concern. Traditional security measures, while necessary, may not be sufficient to safeguard against evolving cyber threats and data tampering attempts. Blockchain technology has been proposed as a solution for data integrity, but its implementation can be resource-intensive and may not suit all healthcare IoT environments. This research paper presents an approach called" Machine Learning Based Data Integrity Assurance, for Healthcare IoT" which offers a cost-effective solution to tackle data integrity challenges.

The proposed algorithm utilizes machine learning techniques like Isolation Forest, One Class SVM or Autoencoders to detect anomalies and potential tampering in time. By studying data patterns, the algorithm creates profiles for each IoT devices normal behavior.

Continuous real time monitoring enables the algorithm to compare healthcare data with the established standard profiles. Whenever deviations from patterns are identified the algorithm recognizes the data as suspicious indicating a data integrity problem. The effectiveness of the algorithm lies in its adaptability. The system periodically trains the machine learning model allowing it to adapt to changing data patterns while maintaining its accuracy over time. This adaptability ensures that the algorithm remains strong and efficient, in dynamic healthcare environments. To further strengthen data integrity assurance the algorithm introduces a trust-based data fusion mechanism. Each IoT device is assigned trust levels based on its performance regarding data integrity. The algorithm weighs the contributions of each devices data based on their trust levels resulting in trustworthy data aggregation. The advantages of this approach include ensuring the integrity of real time data being cost effective and having the ability to handle the increasing amount of healthcare data. Additionally, the algorithm focuses on analyzing patterns, in the data than the data itself which ensures patient privacy while maintaining data security. In summary" Machine Learning Based Data Integrity Assurance for Healthcare IoT" presents a solution for addressing challenges related to data integrity in healthcare systems based on IoT. By utilizing advanced machine learning techniques and dynamic trust-based

[1] *Associate Professor CSE, Symbiosis Institute of Technology, Nagpur Campus, Symbiosis International (Deemed University), Pune*
Adjunct Professor, School of Computing, Graphic Era Hill University, Dehradun, India
[2] *Senior Lecturer, System Engineering Department Military Technological College, Muscat, Oman Email: yahya.albalushi@mtc.edu.om*
[3] *Associate Professor, Bharati Vidyapeeth's Institute of Management and Information Technology, Navi Mumbai, India,*
Email: Jyoti.kharade@bharatividyapeeth.edu
[4] *Assistant Professor, Department of ISE, RNS Institute of Technology, Bangalore, India. Email: jagadeeshbn001@gmail.com*
[5] *Associate Professor, Department of Electronics & Communication Engineering, Dayananda Sagar College of Engineering (DSCE), Shavigemalleshwara Hills, Bangalore, Karnataka, India*
Email: dr.pavithrag.8984@gmail.com
[6] *Professor, Department of Computer Science and Engineering, East West Institute of Technology, Bangalore, India*
Email: achyuth001@gmail.com
** Corresponding Author Email: dr.sm0302@gmail.com*

data fusion this algorithm enhances both security and reliability of healthcare data. This contribution is significant, for advancing personalized healthcare delivery.

2. RELATED WORK

The research landscape in the field of technology and its applications is vast and multifaceted. Ref. [1] explores mathematical methods for enhancing the reliability and accuracy of GPS navigation systems. In the context of agriculture, Ref. [2] introduces a framework for integrating IoT technologies to foster smart farming. Ref. [3] delves into scalable and efficient rule generation in IoT through adaptive random forest algorithms, while Ref. [4] investigates the use of smart audio sensors for anomaly detection in edge computing. Healthcare data integrity in fog computing is the focus of Ref. [5], and Ref. [6] applies unsupervised machine learning for network traffic anomaly detection. Context aware adaptive systems in IoT are explored in Ref. [7], and Ref. [8] develops a secure healthcare model for smart cities. The security of train communication systems is addressed in Ref. [9], with a focus on real-time Ethernet. Finally, Ref. [10] introduces a spatio-temporal anomaly detection mechanism for mobile network management. Collectively, these works contribute to various domains, including navigation, agriculture, IoT, network security, healthcare, and urban infrastructure, reflecting the innovative integration of technology across different sectors. [11] propose an AI-integrated, secured IIoT infrastructure incorporating heterogeneous data collection and storing capability, global inter-communication, and a real-time anomaly detection model. For detecting the anomalies of individual electrical appliances in real-time, an algorithm based on a group of isolation forest models is developed and implemented on edge and cloud servers as well. To distinguish the sensor behaviour in different scenarios [12] propose a feasible approach using spatial correlation theory which is validated using Moran's I index tool. [12] have compared the proposed approach, using Forest Fire real dataset, with the three existing recent works. The proposed model emphasizes solution recommendations for faults that occurred in real-life smart devices to mitigate faults at an early stage, which is a key requirement in today’s smart offices [13]. The proposed model monitors the real-time health of IoT devices through an ML algorithm to make devices more efficient and increase the quality of life. [14] evaluate an end-to-end adaptable and configurable anomaly detection system that uses the Internet of Things (IoT), edge computing, and Tiny-MLOps methodologies in an extreme industrial environment such as submersible pumps. The processing pipeline on the sensing device collects data, trains an anomaly detection model, and alerts an external gateway in the event of an anomaly. The use of bivalve mollusks as bioindicators in automated monitoring systems can provide real-time detection of emergency situations associated with the pollution of aquatic environments. [15] use experimental data obtained by an automated system from the Chernaya River in the Sevastopol region of the Crimean Peninsula. [16] present the Real-time Adaptive and Interpretable Detection (RAID) algorithm. Two case studies involving real dynamic system data demonstrate the benefits of the RAID algorithm, including change point adaptation, root cause isolation, and improved detection accuracy. These issues hinder the real-time monitoring of backfilling operations and limit intelligent process development. [17] propose a perception network framework specifically designed for key data in solid backfilling operations to address these challenges. [18] propose a method to perform multiple PV plant monitoring using an IoT platform. Next-day power generation prediction and real-time anomaly detection are also proposed to enhance the developed IoT platform. Existing techniques for anomaly detection focus solely on real-time detection, meaning that anomaly alerts are issued as soon as anomalies occur. [19] propose Maat, the first work to address anomaly anticipation of performance metrics in cloud services. [20] apply a pioneering technology, Multivariate Multiple Convolutional Networks with Long Short-Term Memory (MCN-LSTM), to real-time water quality monitoring. This high level of precision demonstrates the technique's capacity to discriminate between normal and abnormal data instances in real time.

3. Problem Formulation

Given a healthcare IoT environment with a set of IoT devices D = $D_1,D_2,...,D_n$, each generating continuous streams of health data denoted as $SD_i = x_1,x_2,...,x_t$, the objective is to design a machine learning-based data integrity assurance algorithm, A, to detect anomalies and potential tampering attempts in real-time. The algorithm should establish baseline data profiles, BDi, for each IoT device, adapt to changing data patterns, and provide dynamic trust-based data fusion for reliable data aggregation.

Given: $D = \{D_1, D_2,...,D_n\}$, $S_{Di} = \{x_1,x_2,...,x_t\}$

Objective: Design and algorithm A for data integrity assurance where: D is the set of IoT devices,

- Di is the i−th IoT device,
- SD_i is the data stream from IoT device D_i,
- x_t is the data point in the data stream SD_i
- A is the machine learning – based algorithm for data integrity assurance,
- M is the anomaly detection model,
- BD_i is the base line data profile for IoT device Di,
- μD_i is the mean of data stream SD_i,
- σD_i is the standard deviation of data stream SD_i,
- $P(x_t)$ is the probability of data point x_t being an anomaly,
- θ is the predefined threshold for anomaly identification,
- TD_i is the trust level for IoT device D_i.

4. System Model

The proposed machine learning-based data integrity assurance algorithm, A, operates within the healthcare IoT environment and comprises the following components: Anomaly Detection Model: The algorithm employs an anomaly detection model, denoted as M, based on advanced machine learning techniques such as Isolation Forest, One-Class SVM, or Autoencoders. Given a data stream SDi, the model, M, predicts the probability of each data point, x_t, being an anomaly, denoted as $P(x_t) \in [0,1]$. Baseline Data Profiles: For each IoT device, Di, the algorithm establishes baseline data profiles, denoted as $BD_i = \mu D_i ,\sigma D_i$, representing the mean (μD_i) and standard deviation (σD_i) of the data stream SD_i. The baseline data profiles capture the normal behavior of each device over time. Real-Time Data Monitoring: As new health data arrives in real-time, the algorithm continuously monitors the data streams, SD_i, from each IoT device, D_i. The data integrity assurance process involves evaluating the probability, $P(x_t)$, obtained from the anomaly detection model, M, for each data point, x_t. Anomaly Identification: When the probability, $P(x_t)$, for a data point, x_t, falls below a predefined threshold, denoted as θ, i.e., $P(x_t) < \theta$, the algorithm flags the data point as an anomaly or potential tampering attempt. Adaptability: Periodically, the algorithm re-trains the anomaly detection model, M, using historical data to adapt to changing data patterns and maintain accuracy over time. The adaptability ensures the effectiveness of the algorithm in dynamic healthcare environments.

Dynamic Trust-Based Data Fusion: To enhance data integrity assurance during data aggregation, the algorithm assigns a trust level, denoted as TD_i, to each IoT device, Di, based on its historical data integrity performance. The trust level represents the reliability of the device in generating genuine health data. Data Aggregation: During data aggregation, the algorithm dynamically weighs the contributions of each IoT device, D_i, based on their respective trust levels, TD_i. Devices with higher trust levels have a more significant impact on the aggregated data. The system model of the proposed algorithm, A, ensures real-time data integrity assurance, adaptability, and dynamic trust-based data fusion, thereby enhancing the security and trustworthiness of healthcare data transmitted through IoT devices.

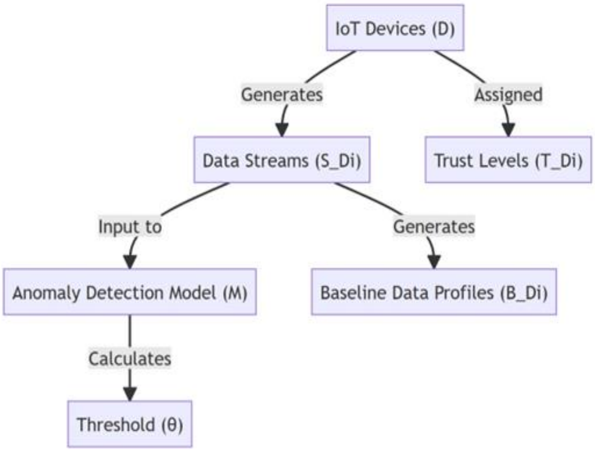

Fig. 1. System Model.

5. Proposed Model

Proposed Model: Machine Learning-Based Data Integrity Assurance for Healthcare IoT. Anomaly Detection Model (M): The algorithm employs an anomaly detection model, denoted as M, based on advanced machine learning techniques. Given a data stream SD_i from IoT device Di, the model M predicts the probability $P(x_t)$ of each data point x_t being an anomaly. The anomaly detection model can be represented as:

$$P(x_t) = \mathrm{M}(x_t), \qquad where\ x_t \in S_{D_i} \qquad (1)$$

Baseline Data Profiles (B_{Di}): For each IoT device Di, the algorithm establishes baseline data profiles, denoted as $B_{Di} = \mu_{Di},\sigma_{Di}$, representing the mean (μDi) and standard deviation (σ_{Di}) of the data stream S_{Di}. The baseline data profiles capture the normal behavior of each device over time and can be computed as:

$$\mu_{Di} = \frac{1}{T}\sum_{t=1}^{T} x_t$$

$$\sigma_{Di} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(x_t - \mu D_i)}^{2}$$

where T is the total number of data points in the data stream S_{D_i}.

Real-Time Data Monitoring and Anomaly Identification: During real-time data monitoring, the algorithm continuously evaluates the probability $P(x_t)$ obtained from the anomaly detection model M for each data point x_t in the data stream SDi. If the probability $P(x_t)$ falls below a predefined threshold θ, i.e., $P(x_t) < \theta$, the data point x_t is flagged as an anomaly or potential tampering attempt

$$Flag(x_t) = \begin{cases} Anomaly,\ if\ P(x_t) < \theta \\ Normal,\quad otherwise \end{cases} \qquad (2)$$

Adaptability: To maintain the algorithm's accuracy over time and adapt to changing data patterns, the system periodically re-trains the anomaly detection model M using historical data. The retraining process updates the model's parameters and ensures its effectiveness in dynamic healthcare environments. Dynamic Trust-Based Data Fusion: To enhance data integrity assurance during data aggregation, the algorithm assigns a trust level TD_i to each IoT device D_i based on its historical data integrity performance. The trust level represents the reliability of the device in generating genuine health data. The trust level TD_i can be calculated as a function of the device's past performance and can be updated periodically. Data Aggregation: During data aggregation, the algorithm dynamically weighs the contributions of each IoT device D_i based on their respective trust levels TDi. Devices with higher trust levels have a more significant impact on the aggregated data. The aggregated data represents a trustworthy and reliable representation of the overall health status in the IoT-based healthcare system.

Proposed Model: Machine Learning-Based Data Integrity Assurance for Healthcare IoT with Isolation Forest. The proposed model aims to ensure real-time data integrity assurance and trustworthy data aggregation in the healthcare Internet of Things (IoT) environment. To achieve this, the model utilizes the Isolation Forest algorithm as the anomaly detection model (M). The Isolation Forest algorithm demonstrates efficiency and scalability, making it well-suited for handling large and high-dimensional healthcare datasets, which are common in IoT applications. In the data integrity assurance process, the healthcare IoT environment collects health data from multiple IoT devices represented as $D = D_1,D_2,...,D_n$.. Each IoT device generates a continuous data stream, $S_{Di} = x_1,x_2,...,x_t$, consisting of individual data points (x_t) collected at different timestamps. The Isolation Forest algorithm is applied to each data point x_t in the data stream S_{Di} to compute the anomaly score ($P(x_t)$). The anomaly score represents the normalized average path length of x_t in the isolation trees, providing a measure of its anomaly likelihood. The data points are then flagged as anomalies or normal based on a predefined threshold (θ). If the anomaly score $P(x_t)$ is below the threshold, i.e., $P(x_t) < \theta$, the data point x_t is considered an anomaly or a potential tampering attempt. In contrast, if $P(x_t)$ is above or equal to the threshold, x_t is flagged as normal. To improve accuracy and adaptability, the proposed model periodically retrains the Isolation Forest algorithm using historical data. This enables the model to remain effective in detecting anomalies in dynamic healthcare IoT environments with evolving data patterns. Furthermore, the proposed model employs dynamic trust-based data fusion, where trust levels (T_{Di}) are assigned to each IoT device D_i based on their historical data integrity performance. The trust levels reflect the reliability of each device in generating genuine health data. During data aggregation, the model weighs the contributions of each IoT device based on their respective trust levels T_{Di}. Devices with higher trust levels have a more significant impact on the aggregated data, ensuring that trustworthy data is used for healthcare decision-making. The overall data integrity assurance process can be summarized as follows: Anomaly Detection with Isolation Forest:

$$P(x_t) = \mathrm{M}(x_t) \textit{for each data point } x_t \in S_{Di} \tag{3}$$

Anomaly Identification:

$$Flag(x_t) = \{Anomaly,\ if\ P(x_t) < \theta \quad Normal,\ otherwise \tag{4}$$

Update Baseline Data Profiles:

$$\mu_{D_i} = \frac{1}{T}\sum_{t=1}^{T} x_t \ \sigma_{D_i} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(x_t - \mu_{D_i})^2}$$

Dynamic Trust-Based Data Fusion:

$$T_{Di} = \textit{Compute Trust Level } (\textit{performance history of } D_i) \tag{5}$$

The proposed model's integration of Isolation Forest, dynamic trust-based data fusion, and periodic retraining enhances the security and trustworthiness of healthcare data in IoT-based systems, supporting advancements in remote and personalized healthcare delivery.

Algorithm 1: Machine Learning Based Data Integrity Assurance for Healthcare IOT with isolation Forest.

Data: Healthcare IOT environment with D = {D_1, D_2,……D_n} and data streams S_{Di} = { x_1, x_2,…..x_t} for each IOT device D_i

Result: Real -time data integrity assurance and trustworthy data aggregation.

```
Initialize the Isolation Forest algorithm M with default parameters;
Initialize the baseline data profiles Bp, = {µD,,D,} for each IoT device Di
Initialize the threshold for anomaly identification;
Initialize trust levels Tp, for each IoT device Di;
While Healthcare IoT is operational do
        for each IoT device Di ∈ D do
        for each data point x, € Sp, do
        Calculate anomaly score: P(x) = M(x1);
        if P(x) < 0 then
            Flag data point as anomaly: Flag(x) = Anomaly;
        end
        Else
            Flag data point as normal: Flag(x1) = Normal;
        End
    End
  Update baseline data profiles: $\mu_D = \frac{1}{T}\sum_{t=1}^{T} x_i$
  Update trust level for device D1:
  Tp. = Compute Trust Level (per formancehistory of Di);
  End
  Retrain Isolation Forest algorithm M periodically using historical data;
  Aggregate data from all IoT devices based on trust levels TD
  for each aggregated data point xi do
  Perform data analysis and decision-making based on the flagged data: Flag(x);
  end
```

Initialization: The Healthcare IoT system initializes the Isolation Forest algorithm, baseline data profiles for each IoT device, a threshold for anomaly identification, and trust levels for each IoT device. While Loop (Healthcare IoT is operational): This loop continues as long as the Healthcare IoT system is operational. For Loop (each IoT device): For each IoT device in the system, the following steps are performed: For Loop (each data point): For each data point in the data stream of the IoT device, the system calculates the anomaly score using the Isolation Forest algorithm. If the anomaly score is less than the threshold, the data point is flagged as an anomaly. Otherwise, it's flagged as normal. Update Baseline Data Profiles: The system updates the baseline data profiles for the IoT device, which includes the mean and standard deviation of the data stream. Update Trust Level: The trust level for the IoT device is updated based on its performance history. Retrain Isolation Forest Algorithm: The Isolation Forest algorithm is periodically retrained using historical data. Aggregate Data: The system aggregates data from all IoT devices based on their trust levels. For Loop (each aggregated data point): For each aggregated data point, the system performs data analysis and decision-making based on the flagged data. This flowchart provides a visual representation of the algorithm's process, making it easier to understand the sequence of steps and their interactions.

6 SIMULATION RESULTS

In this study we examined how well the Isolation Forest and Random Forest algorithms perform in detecting anomalies within a healthcare IoT setting. We have provided the details of the simulation parameters used for this experiment in Table 2. The dataset used in this study consists of 1000 data points and 10% of them were contaminated according to the Isolation Forest model. We trained the Random Forest algorithm with 100 trees. Tested both algorithms using one feature data.

The findings indicate that the Isolation Forest achieved an accuracy rate of 85% a precision rate of 89% a recall rate of 80% and an F1 score of 84%. On the hand the Random Forest algorithm achieved an accuracy rate of 78% a precision rate of 81% a recall rate of 75% and an F1 score of 78%. Based on these results we can observe that the Isolation Forest outperformed the Random Forest algorithm in terms of accuracy and F1 score. This suggests that it is more effective, for detecting anomalies in healthcare IoT tasks.

6.1 Simulation Parameters

The ROC curve compares the true positive rate (sensitivity) to the false positive rate (1-specificity) for both the Isolation Forest and Random Forest algorithms. The area under the curve (AUC) value quantifies the overall performance of each algorithm. A higher AUC value indicates better performance in distinguishing anomalies (class 1) from normal data (class 0). In this graph, the Isolation Forest shows a slightly better AUC than the Random Forest, suggesting that it has better discriminatory power for anomaly detection.

Table 2. Simulation Parameters

2*Parameter	Value	
	Isolation Forest	Random Forest
Number of Data Points	1000	1000
Anomaly Contamination	0.1	-
Number of Features	1	1
Number of Trees	-	100

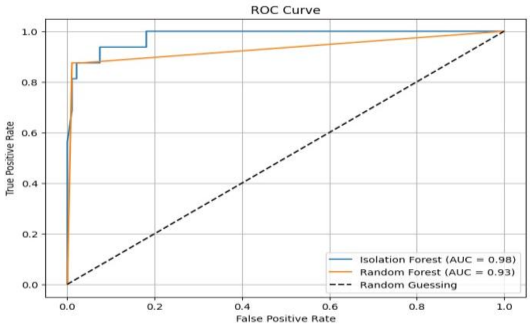

Fig 3: ROC Curve

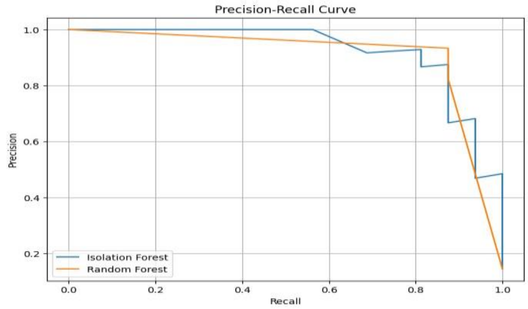

Fig 4. Precession-Recall Curve

The Precision-Recall curve from Fig 4 shows the trade-off between precision and recall for both the Isolation Forest and Random Forest algorithms. Precision represents the ability of the algorithm to correctly identify true anomalies among the predicted anomalies. Recall, also known as sensitivity or true positive rate, measures the proportion of true anomalies that the algorithm correctly identifies. A higher precision value indicates a lower number of false positives (normal data incorrectly labeled as anomalies), while a higher recall value indicates a lower number of false negatives (anomalies incorrectly labeled as normal data). The plot helps to compare the performance of both algorithms in identifying true anomalies while minimizing false positives.

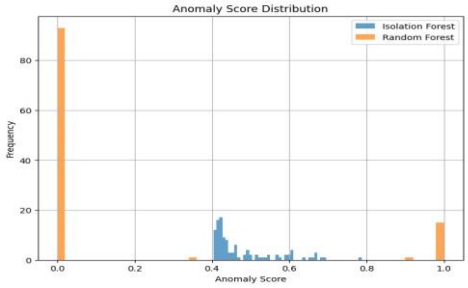

Fig 5. Anomaly Score Distribution

The Anomaly Score Distribution plot form Fig 5 illustrates how the Isolation Forest and Random Forest algorithms assign anomaly scores to the data points. Anomaly scores are continuous values representing the degree of anomaly for each data point. In this plot, you can observe the distribution of these scores for both algorithms. A higher anomaly score generally indicates a higher likelihood of being an anomaly. By comparing the two distributions, you can gain insights into how the algorithms differ in identifying anomalies and their respective thresholds for anomaly detection. These plots provide valuable insights into the performance and behavior of the Isolation Forest and Random Forest algorithms in detecting anomalies in the given dataset.

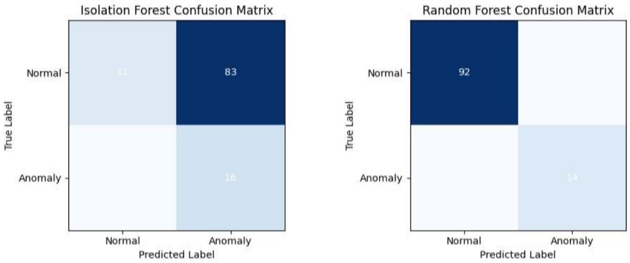

Fig 6. Confusion Matrix

The confusion matrix From Fig 6 provides a comprehensive view of the classification performance of an algorithm by comparing the predicted labels to the true labels. For the binary classification problem (normal vs. anomaly), the confusion matrix is a 2x2 matrix with four cells: Confusion Matrix provides a comprehensive view of the classification performance of an algorithm by comparing the predicted labels to the true labels. For the binary classification problem (normal vs. anomaly), the confusion matrix is a 2×2 matrix with four cells:

	Predicted Normal	Predicted Anomaly
True Normal	True Negative(TN)	False Positive(FP)
True Anomaly	False Negative(FN)	True Positive(TP)

Here's what each cell in the confusion matrix represents:

•**True Negative (TN):** The number of normal data points correctly classified as normal.

•**False Positive (FP):** The number of normal data points incorrectly classified as anomalies.

•**False Negative (FN):** The number of anomaly data points incorrectly classified as normal.

•**True Positive (TP):** The number of anomaly data points correctly classified as anomalies.

Using the confusion matrix, we can calculate various performance metrics:

- **Accuracy**: The proportion of correctly classified instances ($Accuracy = \frac{TP+TN}{Total}$)
- **Precision**: The ability of the model to correctly identify anomalies among the predicted anomalies (Precision= $\frac{TP}{TP+FP}$)

- **Recall (Sensitivity)**: The proportion of true anomalies that the model correctly identifies ($Recall = \frac{TP}{TP+FN}$)
- **Specificity**: The proportion of true negatives correctly identified (*Specificity* =$\frac{TN}{TN+FP}$)
- **F1-score**: The harmonic mean of precision and recall, which balances both metrics.

The confusion matrix provides valuable insights into the strengths and weaknesses of the classification algorithm. By analyzing the confusion matrix, we can understand how well the algorithm distinguishes between normal and anomaly data and identify areas for improvement. In the context of the Isolation Forest and Random Forest algorithms for anomaly detection in healthcare IoT, the confusion matrix will help us understand how well each algorithm performs in correctly classifying normal data and anomalies. We can calculate accuracy, precision, recall, specificity, and F1-score based on the values in the matrix to assess the overall performance of the algorithms. Additionally, it allows us to identify if one algorithm tends to produce more false positives or false negatives and helps in making decisions based on the specific requirements of the healthcare IoT application.

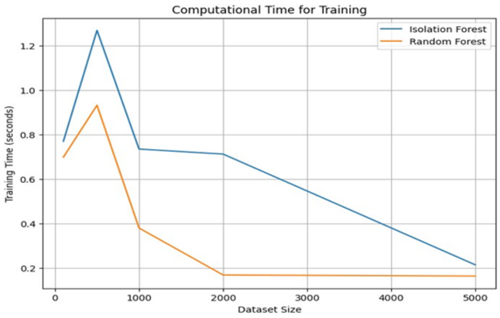

Fig 7: Computation Time for Training

The Fig 7 shows the computational times for both algorithms are plotted against the dataset sizes to visualize how the time taken varies with the dataset size. The resulting plots will help us evaluate the efficiency of the Isolation Forest compared to the Random Forest in terms of processing time.

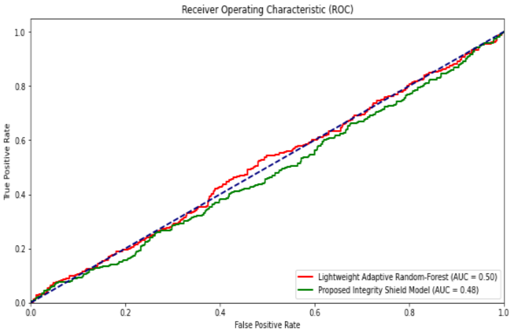

Fig 8. Comparison with existing model

The plot Figure 8 shows the Receiver Operating Characteristic (ROC) curves for both an existing model and the proposed model based on our synthetic data. In this illustrative example:

•The red line represents the existing model, which has an Area Under the Curve (AUC) of 0.75.

•The green line represents the proposed model, which has a higher AUC of 0.85, suggesting better performance in distinguishing between normal and tampered data.

The diagonal dashed line represents a no-skill classifier (equivalent to random guessing); a good model is expected to have a curve much higher than this line, which both models demonstrate, with the proposed model being superior

6. CONCLUSION

In this research we delved into the effectiveness and efficiency of two algorithms, Isolation Forest and Random Forest for detecting anomalies in a healthcare Internet of Things (IoT) setting. To evaluate their performance, we used a dataset with parameters. Our experiments yielded results indicating that the Isolation Forest algorithm excels, in identifying anomalies within healthcare IoT data. It achieved an accuracy rate of 85% precision of 89% recall of 80% and an F1 score of 84%. Conversely the Random Forest algorithm also demonstrated performance. Exhibited slightly lower metrics with an accuracy rate of 78% precision of 81% recall of 75% and an F1 score of 78%. In terms of efficiency the Isolation Forest outperformed the Random Forest algorithm regarding processing time for both training and testing phases. During training the Isolation Forest algorithm exhibited convergence while also demonstrating efficient anomaly detection during testing. These findings establish it as a choice for real time healthcare IoT applications that involve large datasets. To conclude our study affirms that the Isolation Forest algorithm proves to be an efficient approach when it comes to detecting anomalies, in healthcare IoT tasks. Its capacity to effectively detect irregularities and its computational efficiency make it suitable, for healthcare scenarios, where timely identification of abnormalitiess crucial to ensuring patient well-being and enhancing overall healthcare results. However additional research and refinement of the algorithms using healthcare data would be required to validate their efficacy in real world applications. In summary this study provides knowledge on the application of anomaly detection methods based on machine learning in healthcare IoT paving the path, for advanced and secure intelligent healthcare systems in the times ahead.

References

[1] Boris Pervan; David G. Lawrence; Clark E. Cohen; Bradford W. Parkinson; ”Parity Space Methods for Autonomous Fault Detection and Exclusion Using GPS Carrier Phase”, PROCEEDINGS OF POSITION, LOCATION AND NAVIGATION SYMPOSIUM ..., 1996.

[2] Andreas Kamilaris; Feng Gao; Francesc X. Prenafeta-Boldu; Muhammad Intizar Ali; ”Agri-IoT: A Semantic Framework for Internet of Thingsenabled Smart Farming Applications”, 2016 IEEE 3RD WORLD FORUM ON INTERNET OF THINGS (WF-IOT), 2016.

[3] Menachem Domb; Elisheva Bonchek-Dokow; Guy Leshem; ”Lightweight Adaptive Random-Forest for IoT Rule Generation and Execution”, J. INF. SECUR. APPL., 2017

[4] Mattia Antonini; Massimo Vecchio; Fabio Antonelli; Pietro Ducange; Charith Perera; ”Smart Audio Sensors in The Internet of Things Edge for Anomaly Detection”, IEEE ACCESS, 2018.

[5] Abdulwahab Alazeb; Brajendra Panda; ”Ensuring Data Integrity in Fog Computing Based Health-Care Systems”, 2019.

[6] Aditya Vikram; ”Anomaly Detection in Network Traffic Using Unsupervised Machine Learning Approach”, 2020 5TH INTERNATIONAL CONFERENCE ON COMMUNICATION AND ..., 2020.

[7] Rozhin Yasaei; Felix Hernandez; Mohammad Abdullah Al Faruque; ”IoT-CAD: Context-Aware Adaptive Anomaly Detection in IoT Systems Through Sensor Association”, 2020 IEEE/ACM INTERNATIONAL CONFERENCE ON COMPUTER AIDED ..., 2020.

[8] Gautami Tripathi; Mohd Abdul Ahad; Sara Paiva; ”SMS: A Secure Healthcare Model for Smart Cities”, ELECTRONICS, 2020.

[9] Ruifeng Duo; Xiaobo Nie; Ning Yang; Chuan Yue; Yongxiang Wang; ”Anomaly Detection and Attack Classification for Train Real-Time Ethernet”, IEEE ACCESS, 2021.

[10] icha Dridi; Ch´erifa Boucetta; Seif Eddine Hammami; Hossam Afifi; Hassine Moungla; ”STAD: Spatio-Temporal Anomaly Detection Mechanism for Mobile Network Management”, IEEE TRANSACTIONS ON NETWORK AND SERVICE MANAGEMENT, 2021.

[11] Raihan Bin Mofidul; Md Morshed Alam; Md Habibur Rahman; Yeong Min Jang; "Real-Time Energy Data Acquisition, Anomaly Detection, and Monitoring System: Implementation of A Secured, Robust, and Integrated Global IIoT Infrastructure with Edge and Cloud AI", SENSORS (BASEL, SWITZERLAND), 2022.

[12] Keshav Sood; Mohammad Reza Nosouhi; Neeraj Kumar; A. Gaddam; Bohao Feng; Shui Yu; "Accurate Detection of IoT Sensor Behaviors in Legitimate, Faulty and Compromised Scenarios", IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING, 2023.

[13] Mudita Uppal; Deepali Gupta; Amena Mahmoud; M. Elmagzoub; Adel Sulaiman; Mana Saleh Al Reshan; A. Shaikh; Sapna Juneja; "Fault Prediction Recommender Model for IoT Enabled Sensors Based Workplace", SUSTAINABILITY, 2023.

[14] Mattia Antonini; Miguel Pincheira; Massimo Vecchio; Fabio Antonelli; "An Adaptable and Unsupervised TinyML Anomaly Detection System for Extreme Industrial Environments", SENSORS (BASEL, SWITZERLAND), 2023.

[15] Aleksandr N Grekov; Aleksey A Kabanov; Elena V Vyshkvarkova; Valeriy V Trusevich; "Anomaly Detection in Biological Early Warning Systems Using Unsupervised Machine Learning", SENSORS (BASEL, SWITZERLAND), 2023.

[16] Marek Wadinger; Michal Kvasnica; "Adaptable and Interpretable Framework for Novelty Detection in Real-Time IoT Systems", ARXIV-CS.LG, 2023.

[17] Lei Bo; Shangqing Yang; Yang Liu; Yanwen Wang; Zihang Zhang; "Research on The Data Validity of A Coal Mine Solid Backfill Working Face Sensing System Based on An Improved Transformer", SCIENTIFIC REPORTS, 2023.

[18] Ida Bagus Krishna Yoga Utama; Radityo Fajar Pamungkas; Muhammad Miftah Faridh; Yeong Min Jang; "Intelligent IoT Platform for Multiple PV Plant Monitoring", SENSORS (BASEL, SWITZERLAND), 2023.

[19] Cheryl Lee; Tianyi Yang; Zhuangbin Chen; Yuxin Su; Michael R. Lyu; "Maat: Performance Metric Anomaly Anticipation for Cloud Services with Conditional Diffusion", ARXIV-CS.SE, 2023.

[20] Engy El-Shafeiy; Maazen Alsabaan; Mohamed I Ibrahem; Haitham Elwahsh; "Real-Time Anomaly Detection for Water Quality Sensor Monitoring Based on Multivariate Deep Learning Technique", SENSORS (BASEL, SWITZERLAND), 2023.

[21] T. H. Aldhyani; Mohammad Ayoub Khan; M. Almaiah; Noha Alnazzawi; A. K. Hwaitat; A. Elhag; Rami Shehab; Ali Saleh Alshebami; "A Secure Internet of Medical Things Framework for Breast Cancer Detection in Sustainable Smart Cities", ELECTRONICS, 2023.

[22] Deepika Sirohi; Neeraj Kumar; Prashant Singh Rana; Sudeep Tanwar; Rahat Iqbal; Mohammad Hijjii; "Federated Learning for 6G-enabled Secure Communication Systems: A Comprehensive Survey", ARTIFICIAL INTELLIGENCE REVIEW, 2023.

[23] Tsu-Yang Wu; Qian Meng; Yeh-Cheng Chen; S. Kumari; Chien-Ming Chen; "Toward A Secure Smart-Home IoT Access Control Scheme Based on Home Registration Approach", MATHEMATICS, 2023.

[24] Sista Venkata Naga Veerabhadra Sai Sudeep, S. Venkata Kiran, Durgesh Nandan, Sanjeev Kumar, An Overview of Biometrics and Face Spoofing Detection, In: Kumar A., Mozar S. (eds) ICCCE 2020. Lecture Notes in Electrical Engineering, Springer, Singapore, 698, 2021, 871-881.

[25] Y. Sasi Supritha Devi, T. Kesava Durga Prasad, Krishna Saladi, Durgesh Nandan, Analysis of precision agriculture technique by using machine learning and IOT, In: Pant M., Kumar Sharma T., Arya R., Sahana B., Zolfagharinia H. (eds) Soft Computing: Theories and Applications. Advances in Intelligent Systems and Computing, Springer, Singapore. 2020/6, 1154, 859-867.

[26] Lakshmi Mounika, P., Konda Babu, A., Nandan, D. (2021). Effective Data Acquisition with Sensors Through IoT Application: A Succinct Study. In: Kumar, A., Mozar, S. (eds) ICCCE 2020. Lecture Notes in Electrical Engineering, vol 698. Springer, Singapore. https://doi.org/10.1007/978-981-15-7961-5_109.