A Project Report

On

Bio-CPS Device Privacy and Security

BY

Patel Teerth Vasant

2021A7PS2090H

Under the supervision of

PROF. CHITTARANJAN HOTA

SUBMITTED IN FULLFILLMENT OF THE REQUIREMENTS OF

CS F376: DESIGN PROJECT

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE PILANI (RAJASTHAN)

HYDERABAD CAMPUS

(MAY 2024)

ACKNOWLEDGMENTS

I would like to extend my sincere appreciation to all those who have contributed to the successful culmination of this project report on Bio-CPS Device Privacy and Security, guided by Prof. Chittaranjan Hota at Birla Institute of Technology and Science (BITS) Pilani, Hyderabad Campus.

Foremost, I wish to express my deep gratitude to Prof. Chittaranjan Hota for his invaluable guidance, consistent support, and mentorship throughout this project. His expertise, insightful recommendations, and motivation have significantly influenced this report's development and enriched my comprehension of WBAN systems.

I am grateful to my peers and colleagues for their collaborative efforts, constructive input, and support, all of which have greatly enhanced the quality of this project report.

Lastly, I want to convey my thanks to my family and friends for their steadfast support, understanding, and encouragement during the duration of this project. Their collective contributions have undeniably played a crucial role in the successful completion of this endeavor.

Birla Institute of Technology and Science-Pilani,

Hyderabad Campus

Certificate

This is to certify that the project report entitled "**Bio-CPS Device Privacy and Security**" submitted by Mr. Patel Teerth Vasant (ID No. 2021A7PS2090H) in partial fulfillment of the requirements of the course CS F376, Design Project Course, embodies the work done by him under my supervision and guidance.

Date: 4 May 2024

(Prof. Chittaranjan Hota)

BITS- Pilani, Hyderabad Campus

ABSTRACT

Wireless Body Area Networks (WBANs) play a vital role in healthcare by providing continuous monitoring capabilities, allowing patients to seamlessly integrate monitoring into their daily routines. WBANs utilize non-invasive sensors placed on the skin to monitor various physiological attributes. However, data transmission within WBANs faces challenges such as interference, sensor faults, inaccuracies in measurements, and potential malicious attacks aimed at data tampering.

To address these challenges, this paper introduces a novel approach to anomaly detection in WBANs called Isolation Forest-based anomaly detection for WBANs (iForestBAN-AD). Unlike traditional methods that rely on distance measures or density functions, the iForest method takes a fully unsupervised approach by using isolation to detect anomalies in the data.

To evaluate the effectiveness of this approach, extensive experiments were conducted using real-world physiological network records from Physionet. The results demonstrate the robustness and effectiveness of the iForestBAN-AD model, achieving an accuracy of approximately 61%. This research contributes to enhancing the security and reliability of WBANs, thereby promoting the widespread use of continuous monitoring in healthcare environments.

TABLE OF CONTENT

1) INTRODUCTION

The need for continuous monitoring of vital signs has become increasingly apparent in societies with longer lifespans and growing elderly populations, notably observed in regions like Europe. This demographic shift strains healthcare systems, necessitating the development of monitoring systems capable of efficiently overseeing large numbers of patients. Moreover, the increasing demand for ICU admissions underscores the necessity for automated monitoring systems to support healthcare professionals in making timely decisions.

The Internet of Medical Things (IoMT) involves collecting, analyzing, and storing health-related data using miniature sensors forming body area sensor networks. This data encompasses vital signs such as blood pressure (BP), oxygen saturation (SPO2), and pulse rate. Figure 1 illustrates various sensors on the human body for monitoring vital signs both at home and in ICUs.

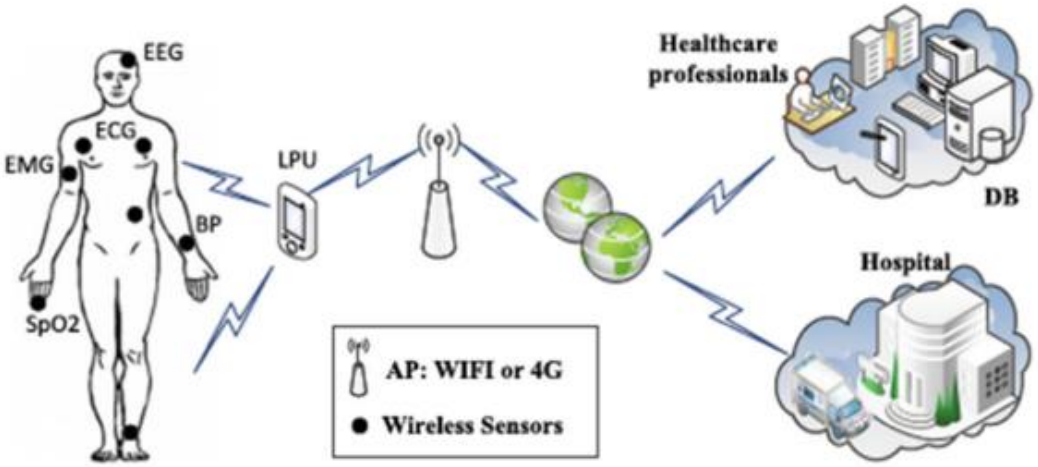


Figure 1: Wireless Body Area Networks

Ensuring data quality in Wireless Body Area Networks (WBANs) for healthcare monitoring is a significant research focus, often addressed through anomaly detection methods to identify abnormal observations. While various approaches exist, many rely on complex techniques, potentially limiting their use in time-sensitive healthcare monitoring. Additionally, some methods focus solely on individual signs, overlooking simultaneous monitoring of multiple parameters.

In light of these considerations, this paper addresses the detection of anomalous observations in multivariate healthcare data by leveraging the concept of isolation. To this end, we employ the Isolation Forest (iForest) algorithm, considering six vital signs recorded in ICU settings collectively to construct an efficient detection model. The isolation concept, as elucidated and applied in prior works, offers the advantage of low linear time complexity and minimal memory requirements by eschewing distance measure calculations.

The contributions of this paper include proposing a novel anomaly detection model for WBANs based on the iForest technique and conducting a comparative analysis against existing models. Subsequent sections review related literature, introduce the proposed model and the background on isolation, present experimental results, and conclude with insights and future directions.

2) RELATED WORK

[1] demonstrates the use of machine learning to identify abnormal data and sensor malfunctions in Wireless Body Area Networks (WBANs) used for remote healthcare monitoring. Using an Artificial Neural Network (ANN) to classify physiological parameters as normal or abnormal is the first stage in the two-step process. Then, Ensemble Linear Regression (LinReg) is used to forecast aberrant parameter values and identify anomalies based on comparisons with felt values. The approach's efficacy is demonstrated through performance evaluation with actual patient data, where it outperforms other approaches including J48 decision trees, Support Vector Machines (SVM), and Linear Regression in terms of accuracy, false positive detection rates, and error rates.

[2] suggests an innovative method for anomaly identification in healthcare applications, such as wireless body area network (WBAN) readings for remote patient monitoring. It compares predicted sensor values with actual measurements using prediction methods on historical data, dynamically adjusting the threshold based on data variability. A majority voting system uses several physiological markers to separate anomalies from real medical disorders. High detection rates, low false positives, and quick processing times are all revealed by evaluation on real datasets. Furthermore, the Gaussian process prediction approach performs better when compared to SMO regression.

[3] presents a Markov model-based method for detecting anomalies in Wireless Body Area Networks (WBANs) used for health surveillance. The strategy seeks to reduce transmission mistakes and energy usage by utilising forecasting tools. Based on actual physiological data, the results show a 99.98% detection accuracy with a low false alarm rate of 5.2%. Notably, by taking advantage of spatiotemporal dependencies, the method successfully separates errors from medical emergencies.

[4] presents a unique model that uses a hybrid Convolutional Long Short-Term Memory (ConvLSTM) technique to detect anomalies in Wireless Body Area Networks (WBANs). Through utilising correlations found in physiological data, the model is able to identify anomalies that are both contextual and point-related. An average F1-measure of 98% and accuracy of 99% are shown in the performance test on the MIMIC dataset, outperforming standalone CNN and LSTM approaches. This development has the potential to improve healthcare services by effectively detecting sensor errors and harmful data patterns.

A novel anomaly detection approach for WBAN called iForestBAN-AD was presented in [5]. Because it makes use of the Isolation Forest technique, an unsupervised learning strategy, this model is especially well-suited to situations in which there is a dearth of labelled data. The model

was created to guarantee the accuracy of data in WBANs used for monitoring medical care, particularly in intensive care units (ICUs). Due to its lack of distance measure calculations, the iForestBAN-AD model showed computational efficiency. The model outperformed many other unsupervised algorithms when tested on real-world physiological network information from Physionet, achieving an excellent Area Under the Curve (AUC) of almost 95%. Future study, according to the researchers, could look at the idea of data drifting in combination with isolation to take patient health context into account almost instantly.

[6] describes BSiForest, an anomaly detection technique for Wireless Sensor Networks (WSNs) that enhances the conventional Isolation Forest algorithm1. In order to filter sub-datasets and choose high-accuracy isolation trees to create a base forest anomaly detector, it uses a box plot to handle problems with randomness, generalisation performance, and stability. The Breast Wisconsin dataset and datasets from a university data centre were used to test the approach, and the results showed improved performance with higher AUC values. The strategy is important for IoT systems to guarantee security and lower financial losses brought on by security threats in WSNs.

3) APPROACH

1. Dataset

The physionet website's MIMIC-1 dataset includes information from 121 patients and includes vital physiological parameters like heart rate (HR), pulse, and oxygen saturation (SpO2). It also includes metrics for arterial blood pressure (ABP), such as systolic (ABP sys), diastolic (ABP dias), and mean arterial blood pressure (ABP mean). The collection also includes alarm signals, giving a complete picture of the monitoring status and patient health. For the purpose of developing and validating algorithms and models for a range of clinical applications, such as anomaly detection, predictive modelling, and decision support systems, this large and varied dataset is a great resource for healthcare analytics research.

2. Preprocessing

Initially, the data.txt files were converted into Excel format during the preprocessing stage of the data collected from patients 401 and 442 to facilitate modification and analysis. We then carried out data cleaning techniques to polish the dataset. This involved eliminating any unnecessary columns that weren't thought to be significant to the analysis. To further guarantee data consistency and integrity, rows with null values or zeros were removed, with the exception of the alarm column. All remaining columns were changed to the float64 data format in order to make additional analysis easier. In

order to ensure the quality and dependability of the dataset for significant insights and interpretation in healthcare research and analytics, these preprocessing activities are crucial for preparing the data for further stages of analysis.

Next, we modified the preprocessing strategy and determined a basic normal range of values for each attribute in the dataset. Then, for every attribute in isolation, we created alarm indications by utilising these defined ranges. After then, a composite alarm indication was created by combining these distinct alarm signs. Nonetheless, we maintained the attributes' connections within this composite indicator. For example, heart rate and pulse have a link, and different forms of blood pressure also show intercorrelation.

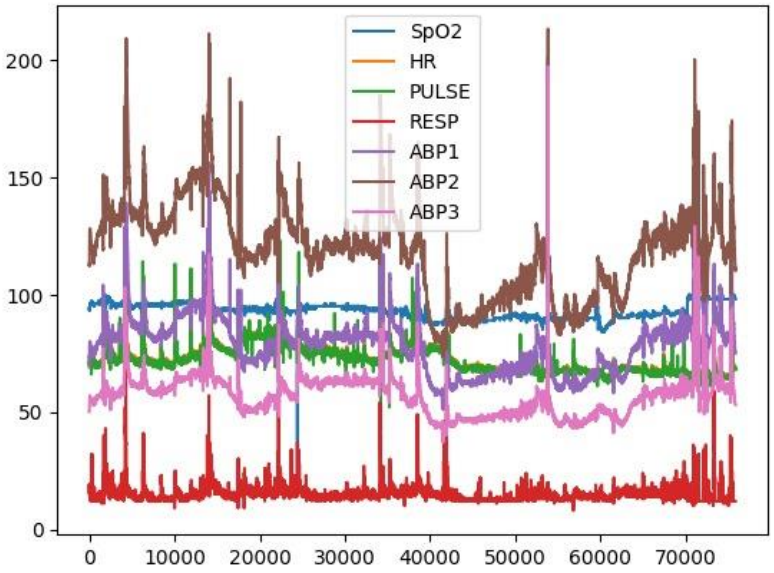


Figure 2: Sensor readings

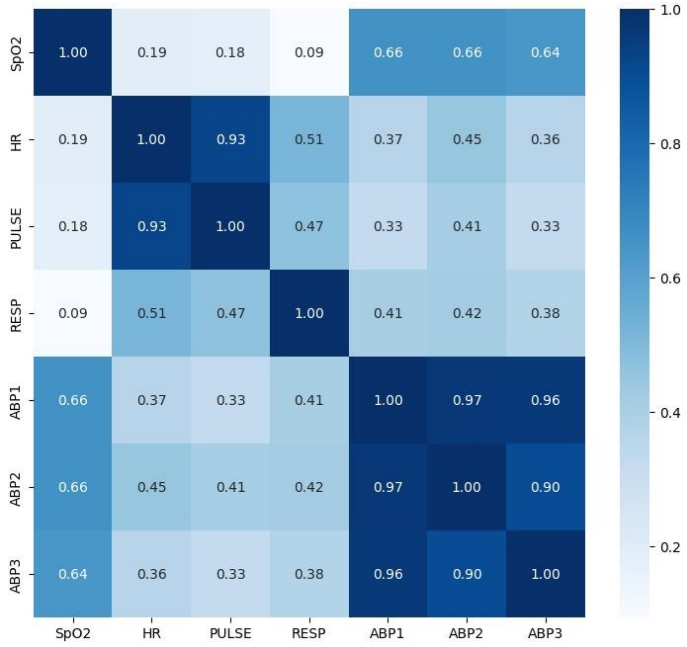


Figure 3: Heat map for correlation visualization

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

r = correlation coefficient
x_i = values of the x-variable in a sample
$\bar{x}$ = mean of the values of the x-variable
y_i = values of the y-variable in a sample
$\bar{y}$ = mean of the values of the y-variable

3. Model Implementation

The implementation of the isolation forest, local outlier factor, and support vector machines (SVM) models involved distinct methodologies tailored to the specific characteristics of each algorithm.

We used the scikit-learn Python module, which offers a stable and effective implementation of the technique, for the isolation forest model. By randomly dividing feature space, the isolation forest algorithm isolates instances by building an ensemble of decision trees. By employing iterative partitioning, anomalies are identified as those cases that necessitate fewer partitions for isolation, taking advantage of anomalies inclination to be more isolated and less common within the dataset. Through cross-validation, hyperparameters including the number of trees, maximum features, maximum samples, and maximum tree depth were adjusted to maximise performance.

$$Anomaly\ Score\ (S) = 2^{\frac{-E(h(k,m,N)}{c(n)}}$$

$$,where\ c(n) = 2(\ln(n-1) + 0.5772156649) - 2\left(\frac{n-1}{n}\right)$$

$$,where\ n\ is\ a\ number\ of\ data\ points\ in\ a\ chosen\ sample$$

$$,where\ E(h(k,m,N)) = \frac{\sum_{i=1}^{N} \begin{cases} if\ k == 1, \sum_{j=1}^{M} 1 \\ else,\ \sum_{j=1}^{M} 1 + c(k) \end{cases}}{N}$$

$$,where\ N\ is\ a\ total\ number\ of\ trees$$

$$,where\ M\ is\ a\ total\ number\ of\ binary\ splits$$

$$,where\ k\ is\ a\ total\ number\ of\ data\ points\ in\ the\ final\ node\ (exit\ node)$$

On the other hand, the scikit-learn library was also utilised in the implementation of the local outlier factor (LOF) algorithm. LOF is an outlier detection technique based on density that evaluates the local deviation of a data point's density in relation to its neighbours. Each data point's LOF score is calculated by this procedure; higher scores denote a higher probability of an outlier. Parameters like the number of neighbours taken into account and the distance metric used to compute local densities had to be adjusted during implementation.

$$\mathrm{LOF}(\mathbf{x}_i) = \frac{1}{k} * \frac{\sum_{j:j\in N_k(\mathbf{x}_i)} d(\mathbf{x}_i, \mathbf{x}_j)}{\sum_{j:j\in \mathcal{N}_k(\mathbf{x}_i)} \sum_{l:l\in N_k(\mathbf{x}_j)} d(\mathbf{x}_j, \mathbf{x}_l)}$$

We also used the scikit-learn library for support vector machines (SVM), which provides effective SVM implementations for applications like anomaly detection and classification. SVMs look for the hyperplane that divides anomalies or examples of various classes from typical instances as much as possible. Choosing the right kernel function (linear, polynomial, or radial basis function, for example) and adjusting hyperparameters like the regularisation parameter (C) and kernel-specific parameters (e.g., gamma, for radial basis function kernel) were necessary for implementation.

Maximize (in α_i)

$$\tilde{L}(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j k(\mathbf{x}_i, \mathbf{x}_j)$$

subject to (for any $i = 1, \ldots, n$)

$$\alpha_i \geq 0,$$

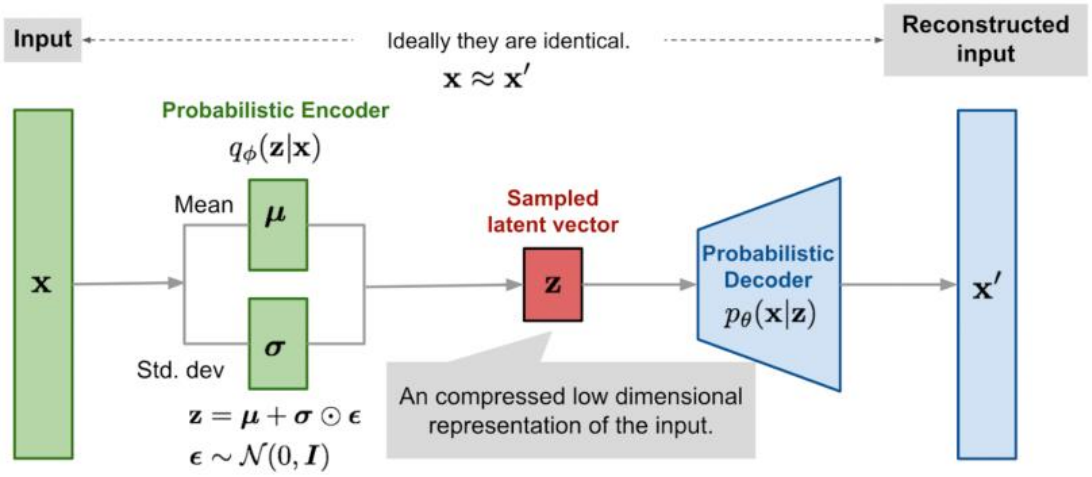
and to the constraint from the minimization in b

$$\sum_{i=1}^{n} \alpha_i y_i = 0.$$

Our Wireless Body Area Network (WBAN) system's anomaly detection capabilities have been greatly improved by the addition of autoencoders. We have developed a strong mechanism to detect anomalies in the physiological data that the WBAN sensors have acquired by utilising autoencoders.

Unsupervised learning neural network design in the form of autoencoders enables us to efficiently rebuild input data while discovering its underlying patterns. We train autoencoders on typical physiological data in our WBAN system so they can recognise the inherent pattern of healthy measures.

The autoencoder reconstructs the input when it receives fresh data while it is in use. A notable departure from the original data and the rebuilt data indicates abnormal behaviour. This strategy provides a proactive way to identify irregularities in real-time, enabling prompt responses and safeguarding WBAN users' general health and wellbeing.



We have greatly improved our anomaly detection capabilities in our Wireless Body Area Network (WBAN) system by adding a voting classifier on top of the LOF, OCSVM, and Isolation Forest algorithms.

Through the combination of predictions from various algorithms, our voting classifier provides a complete method for detecting abnormalities in physiological data obtained from WBAN sensors. The voting classifier uses the distinct advantages of each base algorithm and their combined strength to improve anomaly detection's overall robustness and accuracy.

Our WBAN system can efficiently capture a broad range of aberrant patterns thanks to its ensemble method, giving medical practitioners insightful information about possible health issues. Furthermore, the voting classifier's implementation
enhances the scalability and flexibility of our anomaly detection platform, guaranteeing its effectiveness in actual healthcare environments.

Overall, the implementation of these models required careful consideration of algorithm-specific parameters and optimization techniques to ensure accurate and reliable anomaly detection in the context of wireless body area networks (WBANs) and healthcare monitoring applications.

4) RESULTS

Our anomaly detection models' efficacy was evaluated by the application of the area under the curve (AUC) measure. In particular, we assessed the effectiveness of five different techniques: the voting classifier, the isolation forest, the autoencoder, the local outlier factor (LOF), and one-class support vector machine (OCSVM). The patient records 401 and 442 were the subjects of the evaluation.

• **LOF:** The LOF approach performed rather well in identifying abnormalities in the patient data, with an AUC of 59.10% (without additional preprocessing) and 75.49% (with new preprocessing).

• **OCSVM:** The OCSVM strategy demonstrated marginally better performance than LOF without the new processing, with an AUC of 64.90% (without new preprocessing) and 63.56% (with new preprocessing). However, we observe a reduction in performance when we apply new preprocessing.

• **Isolation Forest:** The Isolation Forest method exhibited an AUC of 61.01% (without new preprocessing) and 77.02% (with new preprocessing), positioning it between LOF and OCSVM in terms of anomaly detection effectiveness without new preprocessing but outperforms all the models when used along with new preprocessing.

• **Autoencoder:** By utilising autoencoder architecture, our model outperformed the other techniques examined in anomaly detection, as evidenced by its much higher AUC of 78.24%. This outcome is exclusive to the updated preprocessing.

• **Voting Classifier:** Using LOF, OCSVM, and Isolation Forest algorithms, the voting classifier ensemble approach produced an AUC of 74.53%, demonstrating its ability to combine the predictions of several anomaly detection algorithms. This result is with new preprocessing, same as autoencoder.

Overall, the results underscore the efficacy of the autoencoder-based approach in detecting anomalies within patient records, outperforming traditional methods such as LOF, OCSVM, Isolation Forest, and even the ensemble method represented by the voting classifier.

5) CONCLUSION

Our analysis of anomaly detection techniques on patient records 401 and 442 showed that different approaches performed differently. Preprocessing had a significant impact on the performance of conventional techniques like Local Outlier Factor (LOF) and One-Class [5]Support Vector Machine (OCSVM), despite their moderate effectiveness. After new preprocessing, Isolation Forest surpassed LOF and OCSVM, originally placing it between them in terms of performance. The autoencoder-based strategy, however, outperformed all other

techniques and demonstrated amazing anomaly detection capabilities. This highlights the promise of deep learning methods, especially in identifying intricate patterns in physiological data, opening up exciting new directions for the field of medicine. The voting classifier demonstrated the possibility of integrating many anomaly detection techniques for improved performance, even though it did not outperform the autoencoder. Overall, our findings provide valuable insights into the efficacy of different anomaly detection approaches, with the autoencoder-based method showing particular promise for advancing anomaly detection in healthcare settings.

6) REFERENCES

[1] Nagdeo, Sumit Kumar and Mahapatro, Judhistir, "Wireless Body Area Network Sensor Faults and Anomalous Data Detection and Classification using Machine Learning," in *2019 IEEE Bombay Section Signature Conference (IBSSC)*, 2019, pp. 1-6.

[2] Harun Al Rasyid, M. Udin and Setiawan, Fajar and Nadhori, Isbat Uzzin and Sudarsonc, Amang and Tamami, Niam, "Anomalous Data Detection in WBAN Measurements," in *2018 International Electronics Symposium on Knowledge Creation and Intelligent Computing (IES-KCIC)*, 2018, pp. 303-309.

[3] Salem, Osman and Alsubhi, Khalid and Mehaoua, Ahmed and Boutaba, Raouf, "Markov Models for Anomaly Detection in Wireless Body Area Networks for Secure Health Monitoring," *IEEE Journal on Selected Areas in Communications,* vol. 39, pp. 526-540, 2021.

[4] Albattah, Albatul and Rassam, Murad A., "A Correlation-Based Anomaly Detection Model for Wireless Body Area Networks Using Convolutional Long Short-Term Memory Neural Network," *Sensors,* vol. 22, no. 5, 2022.

[5] Rassam, Murad A., "Isolation Forest Based Anomaly Detection Approach for Wireless Body Area Networks," Cham, 2023.

[6] Chen, Junxiang and Zhang, Jilin and Qian, Ruixiang and Yuan, Junfeng and Ren, Yongjian, "An Anomaly Detection Method for Wireless Sensor Networks Based on the Improved Isolation Forest," *Applied Sciences,* vol. 13, p. 702, 01 2023.