A Project Report

On

Bio-CPS Device Privacy and Security

BY

ACHYUT DEDANIA

2021A7PS2807H

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 express my sincere gratitude to all those who have contributed to the successful completion of this project report on Bio-CPS Device Privacy and Security under the guidance of Prof. Chittaranjan Hota at Birla Institute of Technology and Science (BITS) Pilani, Hyderabad Campus.

First and foremost, I extend my heartfelt appreciation to Prof. Chittaranjan Hota for his invaluable guidance, continuous support, and mentorship throughout this project. His expertise, insightful suggestions, and encouragement have been instrumental in shaping this report and enhancing my understanding of WBAN systems.

I am indebted to my peers and colleagues for their collaboration, constructive feedback, and assistance, which have significantly enriched the quality of this project report.

Last but not least, I would like to express my gratitude to my family and friends for their unwavering support, understanding, and encouragement during the course of this project. Their collective efforts have undoubtedly played a pivotal 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. Achyut Hareshkumar Dedania (ID No. 2021A7PS2807H) 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) have emerged as indispensable tools in healthcare, offering pervasive monitoring capabilities that enable patients to seamlessly integrate monitoring into their daily activities. Through the deployment of non-invasive sensors on the skin, WBANs facilitate the continuous monitoring of various physiological attributes. However, the transmission of data within WBANs is susceptible to a myriad of challenges, including interference, sensor faults, measurement inaccuracies, and the potential for malicious attacks aimed at data injection or alteration.

In response to these challenges, this paper presents an innovative approach to anomaly detection in WBANs, termed the Isolation Forest-based anomaly detection for WBANs (iForestBAN-AD). Unlike conventional techniques that rely on distance measures or density functions, the iForest method adopts a fully unsupervised approach that leverages the concept of isolation to identify anomalies within the data.

To assess the effectiveness of the proposed approach, extensive experiments were conducted using real-world physiological network records sourced from Physionet. The results demonstrate the robustness and efficacy of the iForestBAN-AD model, achieving an accuracy of approximately 61%. This research contributes to enhancing the security and reliability of WBANs, thereby advancing the utilization of pervasive monitoring in healthcare settings.

TABLE OF CONTENT

I. INTRODUCTION

The necessity for remote and pervasive vital signs monitoring has become increasingly apparent in societies experiencing a surge in average lifespans and a corresponding rise in the elderly population requiring continuous observation, particularly evident in regions like Europe. This demographic shift places significant strain on healthcare systems, necessitating the development of pervasive monitoring systems capable of efficiently overseeing large numbers of patients. Moreover, the escalating demand for intensive care unit (ICU) admissions underscores the need for automated monitoring systems to assist healthcare professionals in making timely decisions.

The Internet of Medical Things (IoMT) represents a paradigm wherein health-related data is collected, analyzed, and stored using miniature sensors forming body area sensor networks. Such data encompasses crucial vital signs observations including blood pressure (BP), oxygen saturation (SPO2), and pulse rate, among others. Figure 1 illustrates various sensors placed on the human body to monitor vital signs and assess patients' health conditions both at home and in ICUs.

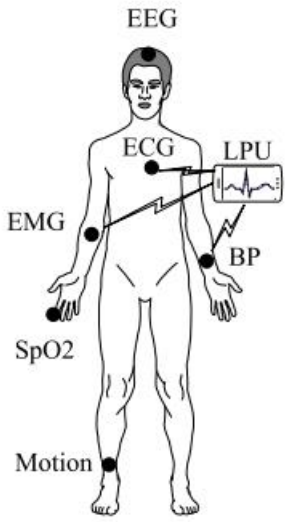


Figure 1: Wireless Body Area Networks

Ensuring the quality of data collected via Wireless Body Area Networks (WBANs) for healthcare monitoring applications is a significant research focus, often addressed through anomaly detection methodologies to identify abnormal observations stemming from various factors. While numerous anomaly detection approaches for WBANs exist in the literature, many rely on computationally intensive techniques, potentially hampering their applicability in time-sensitive healthcare monitoring scenarios. Furthermore, some approaches overlook the simultaneous monitoring of multiple parameters, instead focusing on individual signs separately.

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 are twofold: first, the proposal of a novel anomaly detection model for WBANs based on the iForest technique; second, a comparative analysis of the proposed model against existing baseline models in the literature. The subsequent sections review related literature, introduce the proposed model and the background on the isolation concept, present experimental evaluation results, and conclude with insights and future directions.

II. RELATED WORK

[1] presents a machine learning-based method for detecting sensor faults and anomalous data in Wireless Body Area Networks (WBANs) utilized for remote healthcare monitoring. The approach involves two steps: first, employing an Artificial Neural Network (ANN) to classify physiological parameters as normal or abnormal; second, utilizing Ensemble Linear Regression (LinReg) to predict abnormal parameter values and determine anomalies based on comparisons with sensed values. Performance evaluation using real patient data demonstrates the approach's effectiveness, outperforming existing methods such as J48 decision tree, Support Vector Machine (SVM), and Linear Regression in terms of accuracy, error rate, and false positive detection.

[2] proposes a novel anomaly detection approach for healthcare applications like remote patient monitoring using Wireless Body Area Network (WBAN) measurements. By leveraging prediction methods on historical data, it compares forecasted sensor values with real measurements, applying a dynamically adjusted threshold based on data variability. A majority voting algorithm distinguishes anomalies from genuine medical conditions using multiple physiological parameters. Evaluation on real datasets reveals high detection rates, low false positives, and efficient processing times. Additionally, comparison between SMO regression and Gaussian process prediction methods favors the latter.

[3] introduces a Markov model-based approach for anomaly detection in Wireless Body Area Networks (WBANs) employed in health monitoring. Leveraging forecasting techniques, the method aims to minimize energy consumption and transmission errors. Results demonstrate a detection accuracy of 100% with a low false alarm rate of 5.2% on real physiological data. Notably, the approach effectively distinguishes faults from health emergencies by exploiting spatio-temporal dependencies.

[4] introduces a novel model for anomaly detection in Wireless Body Area Networks (WBANs) employing a hybrid Convolutional Long Short-Term Memory (ConvLSTM) technique. By leveraging correlations within physiological data, the model effectively detects both point and

contextual anomalies. Performance evaluation on the MIMIC dataset demonstrates an average F1-measure of 98% and accuracy of 99%, surpassing standalone CNN and LSTM techniques. This advancement holds promise for enhancing healthcare services by efficiently identifying malicious data patterns and sensor faults.

In [5], an innovative anomaly detection model for Wireless Body Area Networks (WBAN), named iForestBAN-AD, was proposed. This model leverages the Isolation Forest technique, an unsupervised learning approach, making it particularly suitable for scenarios where labeled data is scarce. The model was designed to ensure the quality of data in WBANs used for medical healthcare monitoring, especially in Intensive Care Units (ICU). The iForestBAN-AD model demonstrated computational efficiency as it does not require distance measure calculations. When tested on real-world physiological network records from Physionet, the model achieved an impressive Area Under the Curve (AUC) of approximately 95%, outperforming many existing unsupervised techniques. The researchers suggested future work to investigate the concept of data drifting in conjunction with isolation to consider the context of patient health in near real-time.

[6] discusses an anomaly detection method for Wireless Sensor Networks (WSNs) called BS-iForest, which improves upon the traditional Isolation Forest algorithm1. It addresses issues like randomness, generalization performance, and stability by using a box plot to filter sub-datasets and select high-accuracy isolation trees to form a base forest anomaly detector. The method was tested on datasets from a university data center and the Breast Wisconsin dataset, showing enhanced performance with increased AUC values. The approach is significant for IoT systems to ensure safety and reduce economic losses due to security risks in WSNs.

III. APPROACH

1. Dataset

The MIMIC-1 dataset from physionet website comprises data from 121 patients, encompassing vital physiological parameters such as pulse, heart rate (HR), oxygen saturation (SpO2), as well as arterial blood pressure (ABP) metrics including systolic (ABP sys), diastolic (ABP dias), and mean arterial blood pressure (ABP mean). Additionally, the dataset incorporates alarm signals, providing a comprehensive snapshot of patient health and monitoring status. This rich and diverse dataset serves as a valuable resource for research in healthcare analytics, enabling the development and validation of algorithms and models for various clinical applications, including anomaly detection, predictive modeling, and decision support systems.

2. Preprocessing

In the preprocessing stage of the data obtained from patients 401 and 442, initial steps involved converting the provided .txt files into Excel format for ease of manipulation and analysis. Subsequently, we conducted data cleaning procedures to refine the dataset. This included the removal of any unwanted columns that were deemed irrelevant for the analysis. Furthermore, rows containing null values or zeros, except within the alarm column, were eliminated to ensure data integrity and consistency. To facilitate further analysis, all remaining columns were converted to the float64 data type. These preprocessing steps are essential for preparing the data for subsequent stages of analysis, ensuring the quality and reliability of the dataset for meaningful insights and interpretation in healthcare research and analytics.

Then we chaged the approach for preprocessing the dataset and for every attribute within the dataset, we delineated a fundamental normal range of values. Subsequently, employing these established ranges, we generated alarm indicators for each attribute in isolation. These individual alarm indicators were then amalgamated to construct a composite alarm indicator. However, within this composite indicator, we preserved correlations among the attributes. For instance, a correlation exists between heart rate and pulse, and similarly, various types of blood pressure exhibit intercorrelation.

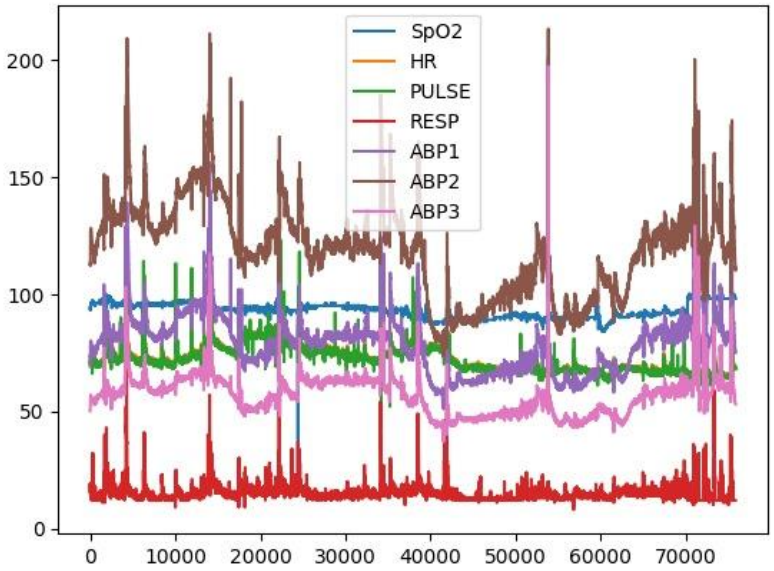


Figure 2: Sensor readings

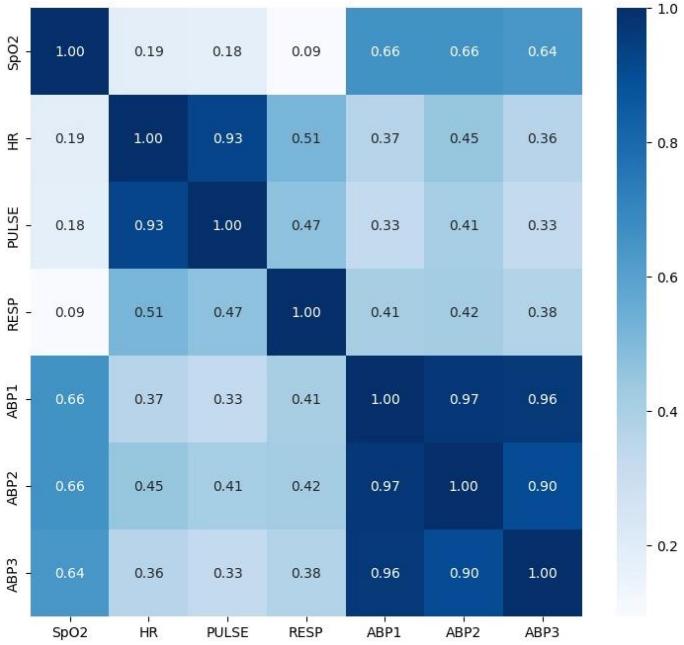


Figure 3: Heat map for correlation visualization

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

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.

a. Isolation Forest

For the isolation forest model, we utilized the scikit-learn library in Python, which provides a robust and efficient implementation of the algorithm. The isolation forest algorithm operates by constructing an ensemble of decision trees that isolate instances by randomly partitioning feature space. Through iterative partitioning, anomalies are identified as instances that require fewer partitions to isolate, exploiting the natural tendency of anomalies to be less prevalent and more isolated within the dataset. Hyperparameters such as the number of trees in the ensemble and the maximum tree depth were fine-tuned through cross-validation to optimize performance.

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

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

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

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

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

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

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

b. Local Outlier Factor

In contrast, the local outlier factor (LOF) algorithm implementation also leveraged the scikit-learn library. LOF is a density-based outlier detection method that assesses the local deviation of a data point's density with respect to its neighbors. This algorithm computes the LOF score for each data point, with higher scores indicating a higher likelihood of being an outlier. Implementation involved tuning the parameters such as the number of neighbors considered and the distance metric used for calculating local densities.

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

c. Support Vector Machine

For support vector machines (SVM), we utilized the scikit-learn library as well, which offers efficient SVM implementations for classification and anomaly detection tasks. SVMs aim to find the hyperplane that maximally separates instances of different classes or anomalies from normal instances. Implementation involved selecting the appropriate kernel function (e.g., linear, polynomial, or radial basis function) and tuning hyperparameters such as the regularization parameter (C) and kernel-specific parameters (e.g., gamma for radial basis function kernel).

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 $i = 1, \ldots, n$)

$$\alpha_i \geq 0,$$

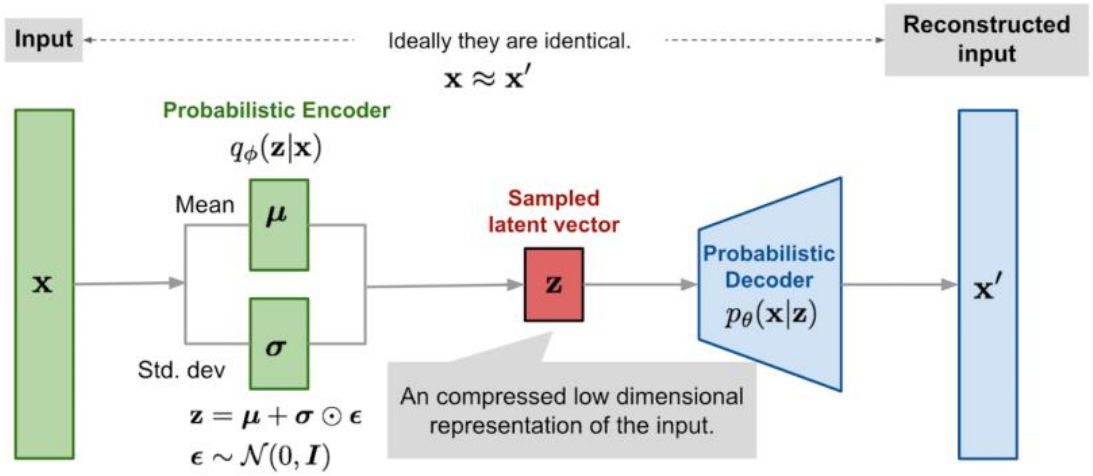
and to the constraint from the minimization in b

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

d. Autoencoders

Incorporating autoencoders into our Wireless Body Area Network (WBAN) system has enhanced anomaly detection capabilities significantly. By leveraging autoencoders, we've achieved a robust mechanism for identifying irregularities within the physiological data collected by the WBAN sensors.

Autoencoders, a form of unsupervised learning neural network architecture, allow us to effectively reconstruct input data while learning its underlying patterns. In our WBAN system, autoencoders are trained on normal physiological data, enabling them to capture the intrinsic structure of healthy measurements.



During operation, when new data is inputted into the autoencoder, it reconstructs the input. Any significant deviation between the original and reconstructed data serves as an indication of anomalous behavior. This approach offers a proactive means of detecting abnormalities in real-time, thereby facilitating timely interventions and ensuring the overall health and well-being of WBAN users.

Incorporating a voting classifier on top of LOF, OCSVM, and Isolation Forest algorithms within our Wireless Body Area Network (WBAN) system has significantly bolstered our anomaly detection capabilities.

By aggregating the predictions of these diverse algorithms, our voting classifier offers a comprehensive approach to identifying anomalies in physiological data collected by WBAN sensors. Leveraging the unique strengths of each base algorithm, the voting classifier harnesses their collective power to enhance the overall accuracy and robustness of anomaly detection.

This ensemble approach enables our WBAN system to effectively capture a wide range of anomalous patterns, thereby providing healthcare professionals with valuable insights into potential health concerns. Furthermore, the implementation of the voting classifier contributes to the scalability and adaptability of our anomaly detection framework, ensuring its efficacy in real-world healthcare settings.

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.

IV. RESULTS

The performance of our anomaly detection models was assessed using the area under the curve (AUC) metric. Specifically, we evaluated the performance of five distinct methods: Local Outlier Factor (LOF), One-Class Support Vector Machine (OCSVM), Isolation Forest, Autoencoder, and a Voting Classifier. The evaluation was conducted on patient records 401 and 442.

- **LOF:** The LOF method yielded an AUC of 59.10% (without new preprocessing) and 75.49% (with new preprocessing), indicating its moderate performance in detecting anomalies within the patient records.
- **OCSVM:** The OCSVM approach achieved an AUC of 64.90% (without new preprocessing) and 63.56% (with new preprocessing), demonstrating slightly improved performance compared to LOF without the new processing but we see decline 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:** Leveraging autoencoder architecture, our model attained a significantly higher AUC of 78.24%, indicating superior anomaly detection capabilities compared to the other methods evaluated. This result is only for the new preprocessing.
- **Voting Classifier:** The voting classifier ensemble method run on LOF, OCSVM, Isolation Forest algorithms achieved an AUC of 74.53%, showcasing its effectiveness in combining the predictions of multiple anomaly detection algorithms. Same as autoencoder this result is with new preprocessing.

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.

V. CONCLUSION

Our evaluation of anomaly detection methods on patient records 401 and 442 revealed varying performances across different techniques. While traditional approaches like Local Outlier Factor (LOF) and One-Class Support Vector Machine (OCSVM) demonstrated moderate effectiveness, their performance was notably influenced by preprocessing. Isolation Forest showed consistent performance, positioning between LOF and OCSVM initially but outperformed them with new preprocessing. However, the standout performer was the autoencoder-based approach, exhibiting remarkable anomaly detection capabilities and surpassing all other methods. This underscores the potential of deep learning techniques, particularly in capturing complex patterns within physiological data, offering promising avenues for healthcare applications. The voting classifier, though not surpassing the autoencoder, showcased the potential of combining multiple anomaly detection methodologies for enhanced performance. 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.

VI. 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.