A Project Report

On

Bio-CPS Device Privacy and Security

BY

Patel Teerth Vasant

2021A7PS2090H

Under the supervision of

PROF. CHITTARANJAN HOTA

SUBMITTED IN PARTIAL FULLFILLMENT OF THE REQUIREMENTS OF

CS F376: DESIGN PROJECT

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE PILANI (RAJASTHAN)

HYDERABAD CAMPUS

(MARCH 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: 12 March 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

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.

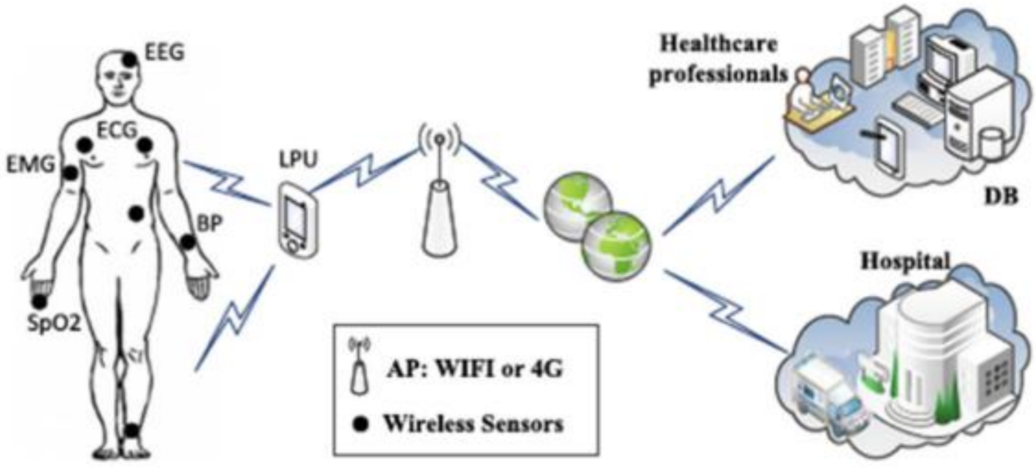

Figure 1: Wireless Body Area Networks

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.

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.

As far as we are aware, not much research has been done on using the isolation principle to identify anomalies in the data streams collected by Wireless Body Area Networks (WBANs). Therefore, this research uses the MIMIC-1 dataset to apply the isolation forest (iForest) algorithm, together with the support vector machine and local outlier factor for comparison in the setting of WBANs.

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.

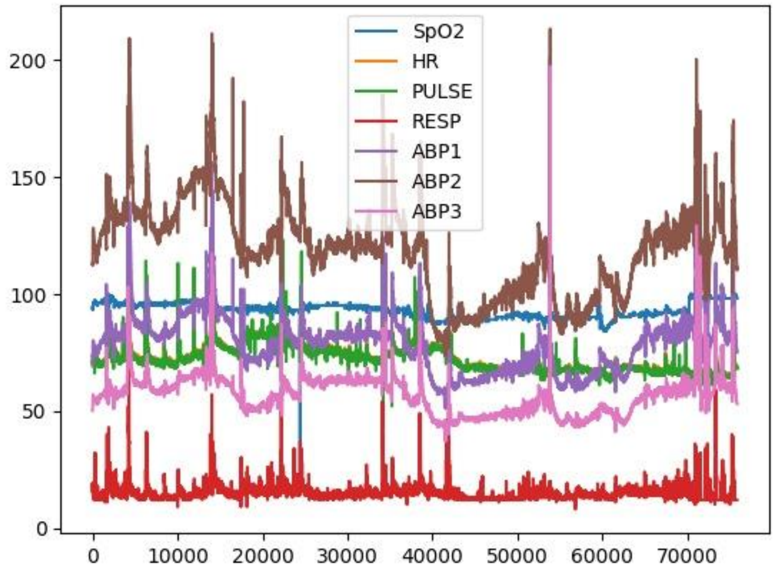

Figure 2: Sensor readings

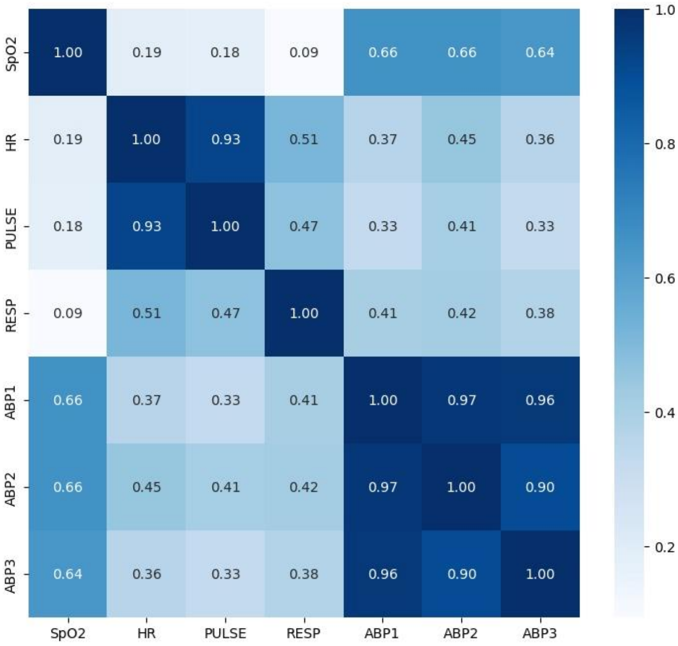

Figure 3: Heat map for correlation visualization

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.

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.

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.

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.

RESULTS

The empirical evaluation of our proposed model entailed a comparative analysis with two established unsupervised anomaly detection methodologies outlined in prior literature, denoted as LOF and OCSVM, utilizing patient records 401 and 442 for assessment. Our findings reveal that our proposed model exhibits a notable performance, achieving an Area Under the Curve (AUC) metric of 61.01%, juxtaposed against 64.90% for OCSVM and 59.10% for LOF. Noteworthy is the ranking of the isolation forest, securing the second position in performance, whereas LOF ranks third. Additionally, it's imperative to highlight that the utilization of the isolation concept mitigates the computational overhead typically associated with traditional machine and deep learning models, as it obviates the necessity for distance measure calculations.

CONCLUSION

The project is still under further research and development. Hence, we don't have any conclusions yet.

REFERENCES

[1] S. K. a. M. J. Nagdeo, "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] M. U. a. S. F. a. N. I. U. a. S. A. a. T. N. Harun Al Rasyid, "Anomalous Data Detection in WBAN Measurements," in *2018 International Electronics Symposium on Knowledge Creation and Intelligent Computing (IES-KCIC)*, 2018, pp. 303-309.

[3] O. a. A. K. a. M. A. a. B. R. Salem, "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] A. a. R. M. A. Albattah, "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.