Anomaly Detection in Biosensor Waveforms

Data 599 Capstone Proposal

Justine Filion, Sara Hall, Saisree GR, Neethu Gopalakrishna

May 2022

1. Introduction

1.1 Problem, Motivation, and Purpose

As the healthcare sector evolves to incorporate more technology for faster and more accurate diagnostics, there is an increasing demand for commercially produced biomedical devices. For these instruments to help improve patient care, they must be accurate and reliable and thus undergo many tests for quality control. With the development of machine learning techniques in recent years, we are uniquely positioned to automate certain aspects of the quality control process.

One area of development is for systems that perform blood analysis, as it is one of the most common tests run by health professionals and is involved in the diagnosis of many different conditions [1]. Being able to get fast and accurate live test results has the potential to improve patient outcomes by enabling treatment to be modified based on results [2]. This is especially pertinent in high-pressure emergency situations like on ambulance decks where the first few minutes of care are often critical [2]. The epoc system fulfills these needs as a handheld wireless system that enables comprehensive blood analysis testing at the patient's side within minutes on a single room temperature test card [2].

The epoc system has three components: a test card which contains the components necessary to perform the testing for various different analytes, a reader which has a card slot for the test card and records the test results, and a host that interfaces with the reader to provide information to the healthcare professional [3] As a device that is used in patient care, the test cards must undergo comprehensive quality control testing before being sent out for use. This testing involves making sure a low number of cards yield unsuccessful readings. A reading may be unsuccessful for different reasons such as inadequate pin contact between the test card and the reader or air bubbles in the sample. Operators spend a significant amount of time trying to diagnose these problems with unsuccessful tests by analyzing their waveforms, but manual identification is cumbersome and not always accurate because of:

1. Human bias in the identification of causes.
2. Overlap between the characteristics of different problems that underlie unsuccessful readings.

As a result, the goal of this project is to use unsupervised machine learning techniques on the data from the readings for a particular analyte to see if they can be automatically clustered into different types of errors. Of particular interest is whether unsupervised methods will yield a cluster that corresponds well to a lack of pin contact between the test card and the reader.

1.2 Background

While not a lot of previous work has looked specifically at blood analysis systems, there is a large body of literature describing time series analysis and anomaly detection in various domains including electrocardiography, electroencephalography, and structural health monitoring, among many other domains [4]–[7]. In general, there are a couple of steps that need to be taken when

performing machine learning with waveforms. First, some sort of noise reduction like filtering must be applied followed by some feature engineering techniques like Fourier transforms or discrete wavelet transforms to produce better features for the models [6]. Once these steps are complete, clustering methods can be used to find groups in the anomalies.

Filtering of a noisy waveform is crucial in the identification of the actual signal. In the medical domain where stakes are extremely sensitive, noise cancellation is critical. Anam Mahmud et.al [4] proposed a signal processing method for structural health monitoring where noises were removed by using a simple Butterworth filter, while other work has used various other techniques [4], [8].

After the cleaned signals are obtained, it is important to generate better information from the time series data. This will ensure that the models have better features to train on. Nawaz, Menaa, et al. proposed the usage of many different methods to extract features from waveforms [6]. The methods include those that encompass frequency-based information like discrete Fourier transforms and discrete wavelet transforms, as well as those that pull out the information that contains the most variability like principal and independent component analyses [6]. Finally, dynamic time warping can also be used to determine the similarity (distance) between different waveforms [5].

These features generated from the previous step are fed to supervised, unsupervised or semi-supervised models depending on the problem statement for prediction or exploration. Previous work has utilized Deep Belief Nets, a type of multi-layer generative neural network to perform semi-supervised anomaly detection in electroencephalography waveforms [7]. Other work includes performing unsupervised clustering methods for detecting anomalies in vital signals like heart rate and respiration rate [9].

2. Aims and Objectives

This project aims to compare various machine learning pipelines that perform clustering to identify pin contact errors from biosensor readings.

Key objectives include :

1. Statistically show that the successful readings are similar enough to consider only a subset of them for our training dataset. This will alleviate issues resulting from having an unbalanced dataset.
2. Resample our data to obtain multiple training sets, each having more balanced classes.
3. Leverage the aggregate predictors of the readings to find possible clusters.
4. Remove the wet-up period from the waveforms as it provides no information on whether or not the reading is successful.
5. Split the waveforms into different time windows such as calibration window, sample window and post window.
6. Apply noise filters to the waveforms and feed them to machine learning algorithms for clustering.
7. Compare whether or not the clusters obtained by the raw data and the waveform data match.

Key research questions :

1. How accurately can a machine learning algorithm that uses sensor waveforms and aggregate predictors identify pin contact errors as compared to manual classification done by domain knowledge experts?
2. Which machine learning pipelines are effective, and which are not when it comes to identifying anomalies in biosensor readings?

3. Dataset Description and Preparation

The dataset we are working with consists of several CSV files which describe sensor readings for a particular analyte from many different tests. The data generally consist of three broad categories - successful readings, unsuccessful readings, and readings lacking pin contact, which have been labelled manually by teams of people and separated out into different files. Overall, there are 411076 successful readings, 9855 unsuccessful readings, and 84 readings lacking pin contact. This means that the data are heavily unbalanced, and we will have to account for this in our machine learning pipelines.

Each reading has two records associated with it: a raw signal time series, and a series of predictors which have been calculated from the raw data. These records can be linked together using their unique test identifiers. Within the predictors derived from the raw waveforms are metrics like slope, mean signal, and noise from several different windows including the calibration and sample periods. The time series for different tests are different lengths so that is another thing we will need to account for in our analyses (Figure 1).

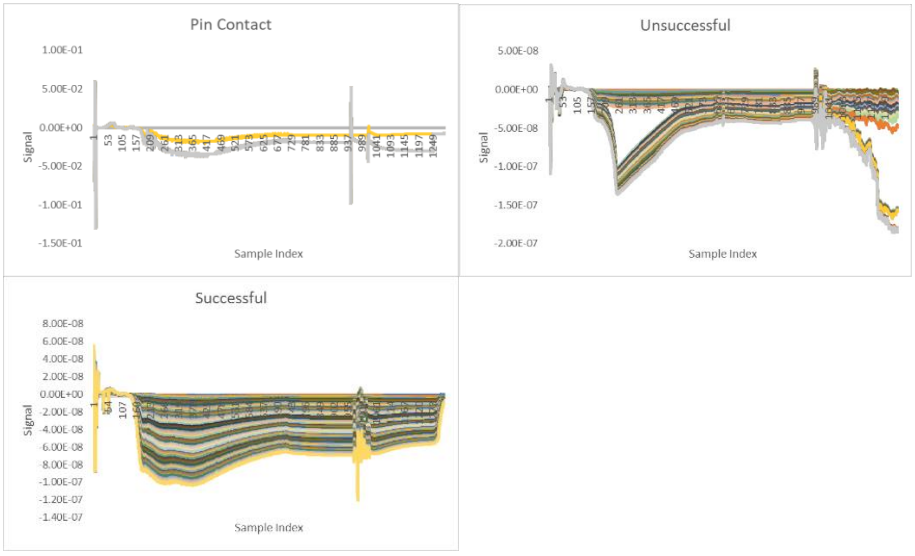

Figure 1. Sample waveforms for 82 tests of each of the three categories.

Both the predictor and time-series files will need some amount of preprocessing. With the time series data, we will need to perform some sort of noise filtering. We will then need to plot and look at the waveforms to split them into windows that make sense for clustering readings into different groups. We may also want to look into ways to normalize the signals. For the predictors, we will need to make sure we are not missing too many missing values or outliers prior to proceeding with our analyses. Before modelling, it may also be necessary to perform different techniques for feature extraction.

Finally, it seems there might be some duplicated test records for the predictors as there are more records stored as summary statistics than as raw waveforms. As a result, we will need to filter out the tests that are present in both forms and get rid of any duplicates so that we are able to use them together should we choose to do so.

4. Deliverables, Schedule, and Responsibilities

4.1 Deliverables

We will produce well commented modular Python code for various machine learning pipelines. Siemens Healthineers does not expect our code to handle exceptions or to include docstrings. A final report documenting our analysis and the results we have obtained will also be delivered.

4.2 Schedule

We will officially start working on the project on May 9th and complete it by June 22nd. During this period, we have two set deadlines, one being the mid-project presentation on May 31st and the other being the final presentation on June 22nd. We will meet with Siemen's data science team twice a week on Mondays and Fridays, as well as present to Siemen's advisory committee once every two weeks. In order to meet our client's expectations, we have come up with a preliminary project timeline.

Table 1. Preliminary project timeline.

Week	Task	Description
May 2 - May 8	Project proposal	• Establish a solid understanding of the project expectations. • Research relevant literature. • Start exploring the data. • Establish research questions. • Develop a general timeline that we can follow for the next 7 weeks.

May 9 - May 15	Data preprocessing and research on time series analysis	• Figure out ways to perform noise reduction. • Look into ways to deal with unbalanced data (determine how different the successful readings are from each other to see if we can justify using fewer samples). • Use visualizations to see how we can divide the time series into different windows. • Research methods for time series clustering. • Clean and wrangle the data. • Figure out how to build our training/testing sets.
May 16 - June 5	Modelling	• Build various unsupervised machine learning pipelines to figure out what works and what doesn't in terms of clustering different types of readings and identifying pin contact errors.
June 6 - June 12	Tuning	• Focus on improving the most promising model(s). • If we have time, look into data augmentation methods.
June 13 - June 19	Documenting	• Write the final report. • Generate slides and practice for the final presentation.
June 20 - June 24	Presenting	• Submit the final report. • Present the final results.

4.3 Team member's role and responsibilities

Having access to both time series data and aggregate predictors data, we have decided to work in teams of two, where each team focuses on one type of data. Our daily meetings will ensure that all four members understand the progress that is being made by the other team. Furthermore, as the project evolves, we will potentially have to combine our work, in which case we will delegate duties accordingly.

We have agreed to rotate who will be recording the minutes during the meetings. Also, we expect every team member to participate equally in each weekly presentation.

References

[1] "Blood Tests - Blood Tests | NHLBI, NIH." https://www.nhlbi.nih.gov/health/blood-tests (accessed May 04, 2022).

[2] "epoc® Blood Analysis System." https://www.siemens-healthineers.com/en-ca/blood-gas/blood-gas-systems/epoc-blood-analysis-system (accessed May 04, 2022).

[3] "epoc® Blood Analysis System Resource Guide".

[4] M. A. Mahmud, A. Abdelgawad, K. Yelamarthi, and Y. A. Ismail, "Signal processing techniques for IoT-based structural health monitoring," *Proceedings of the International Conference on Microelectronics, ICM*, vol. 2017-December, pp. 1–5, Jan. 2018, doi: 10.1109/ICM.2017.8268825.

[5] D. Azariadi, V. Tsoutsouras, S. Xydis, and D. Soudris, "ECG signal analysis and arrhythmia detection on IoT wearable medical devices," *2016 5th International Conference on Modern Circuits and Systems Technologies, MOCAST 2016*, Jun. 2016, doi: 10.1109/MOCAST.2016.7495143.

[6] M. Nawaz, J. Ahmed, G. Abbas, and M. Ur Rehman, "Signal Analysis and Anomaly Detection of IoT-Based Healthcare Framework," *2020 Global Conference on Wireless and Optical Technologies, GCWOT 2020*, Oct. 2020, doi: 10.1109/GCWOT49901.2020.9391621.

[7] D. Wulsin, J. Blanco, R. Mani, and B. Litt, "Semi-supervised anomaly detection for EEG waveforms using deep belief nets," *Proceedings - 9th International Conference on Machine Learning and Applications, ICMLA 2010*, pp. 436–441, 2010, doi: 10.1109/ICMLA.2010.71.

[8] T. Köhler and D. Lorenz, "A comparison of denoising methods for one dimensional time series".

[9] A. Mahmoudzadeh, I. Azimi, A. M. Rahmani, and P. Liljeberg, "Lightweight photoplethysmography quality assessment for real-time IoT-based health monitoring using unsupervised anomaly detection," *Procedia Computer Science*, vol. 184, pp. 140–147, 2021, doi: 10.1016/J.PROCS.2021.03.025.