Can we Teach a Machine to be a Cardiologist?

Motivation and Background

Cardiovascular disease (CVD) is a prevalent issue in society and a leading cause of death. In Australia, CVD accounted for a quarter of all deaths in 2019. CVD can often be prevented, including by improving diet and exercise, so it is important to diagnose before an event (such as stroke) occurs.

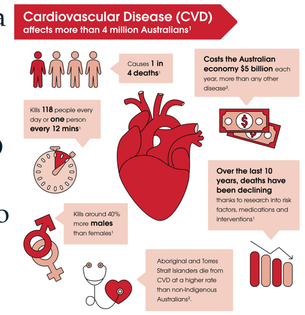

Figure 1: Statistics on CVD, from [1]

The Electrocardiogram (ECG)

An ECG is a graph of the electrical signals produced by the heart, and is obtained by placing the electrodes on the skin. Physicians use a number of waves and points on the ECG to diagnose heart disease, as shown in Figure 2.

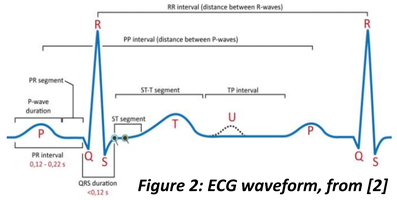

Figure 2: ECG waveform, from [2]

Machine Learning

Computers can be programmed to classify data. In our case, this was to distinguish between different heart diseases using ECG signals. Although each algorithm operates differently, they follow the same general process:

1. Pre-processing: filtering, noise removal, transforms;
2. Feature extraction: small set extracted from raw data;
3. Classification: of signal based on feature set.

Aim of the Project

The aim of this project was to investigate whether machine learning can be used to teach a computer to accurately distinguish between normal and abnormal heart patterns, and even between different heart diseases.

References:
[1] Heart Foundation, "Key Statistics: Cardiovascular Disease" (2020), [Online], Available: https://www.heartfoundation.org.au/activities-finding-or-opinion/key-stats-cardiovascular-disease
[2] *Clinical ECG Interpretation*, [Online], Available: https://ecgwaves.com/course/the-ecg-book/
[3] M. Dorraki, A. Fouladzadeh, A. Allison, B.R. Davis, and D. Abbott. "On Moment of Velocity for Signal Analysis" in *Royal Society Open Science*, vol. 6, iss. 3, (2019), DOI: https://doi.org/10.1098/rsos.182001

Methodology

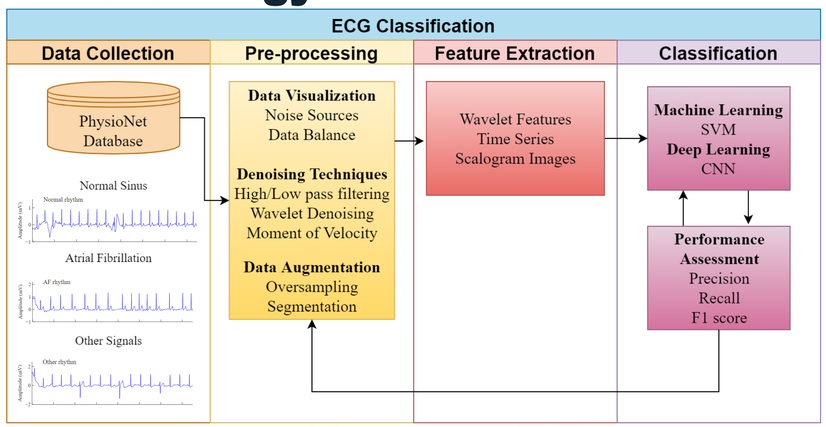

Figure 3: Methodology Diagram of ECG Classification

Pre-processing

Collected ECG data contains noise and movement artifacts. Filtering techniques, wavelet denoising, and moment of velocity [3] were used to clean up and/or transform the ECGs prior to classification.

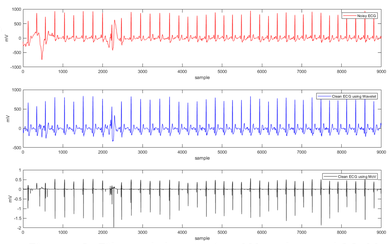

Figure 4: Filtered data using Wavelet and MoV denoising.

Features Extraction

Feature extraction involves isolating a handful of useful information from the raw ECG data. It is needed to reduce the size of data input to classifiers, and hence run time and classification accuracy. Some useful features from time domain and frequency domain extracted including, wavelet decomposition (Figure 5), beat-to-beat variability, and time-frequency images (Figure 6).

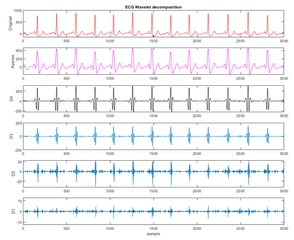

Figure 5: Wavelet Decomposition.

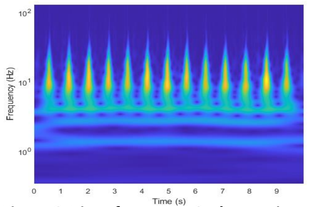

Figure 6: Time-frequency Scalogram image

Machine Learning

A range of traditional machine learning and modern deep learning algorithms were used to classify heart conditions.

Results

Three machine learning models have been tested, each with three pre-processing and feature extraction techniques.

Our results show:

- The SVM classifier is not as effective as the CNN classifier or the LSTM classifier;
- Classification results are greater when pre-processing is used before feature extraction; and,
- Wavelet denoising coupled with a CNN model produced the greatest results of the methods tested.

For comparison, the LSTM classifier was used without any pre-processing or feature extraction, i.e. the classifier used only the raw ECG data to classify the heart condition. Not only did this take significantly longer than the methods shown in Figure 7, but it produced a low F1-score of 0.507. This demonstrates the importance of pre-processing the data and using a suitable set of features.

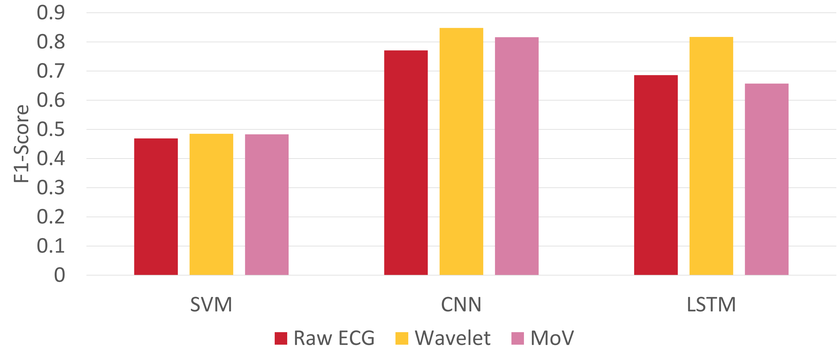

Figure 7: Performance of various Machine Learning techniques

Conclusion & Future Work

So, can we teach a machine to be a cardiologist? The answer is yes, it is possible to enable the computer to accurately distinguish between different heart problems.

Future development could include improving these results by modifying the classifier, pre-processing and/or feature extraction processes; or the model could be adapted to classify a greater range of cardiovascular conditions.

Sonia Kleinig and Hien Long Nguyen

Supervisors: Derek Abbott, Mohsen Dorraki | Electrical and Electronic Engineering | 13434 | HS21| 2021