COMP6208: Advanced Machine Learning

Brent De Hauwere[1]**, Matthew De Vries**[2]**, Dimitris Mallios**[3]**, Ilias Kazantzidis**[4]

[1]`bdh1g19`, [2]`mrdv1n19`, [3]`dm1n19`, [4]`ik3n19`

1 Introduction

Electrocardiogram (ECG) analysis has been established at the core of cardiovascular pathology diagnosis since its development in the twentieth century [2]. The main problem with manual analysis of ECG signals, similar to many other time-series data, lies in difficulty of detecting and categorizing different waveforms and morphologies in the signal. For a human, this task is both extensively time-consuming and prone to errors [1].

2 Data

This dataset is composed of heartbeat signals derived from the famous dataset in heartbeat classification, the MIT-BIH Arrhythmia Dataset. The number of samples is large enough for training a deep neural network as well as using other advanced methods in machine learning. The dataset is described:

- Number of Samples: 109446
- Number of Categories: 5
- Sampling Frequency: 125Hz
- Data Source: Physionet's MIT-BIH Arrhythmia Dataset
- Classes: ['N': 0, 'S': 1, 'V': 2, 'F': 3, 'Q': 4]

This dataset has been used in exploring heartbeat classification using deep neural network architectures, and observing some of the capabilities of transfer learning on it. The signals correspond to electrocardiogram (ECG) shapes of heartbeats for the normal case and the cases affected by different arrhythmias and myocardial infarction.
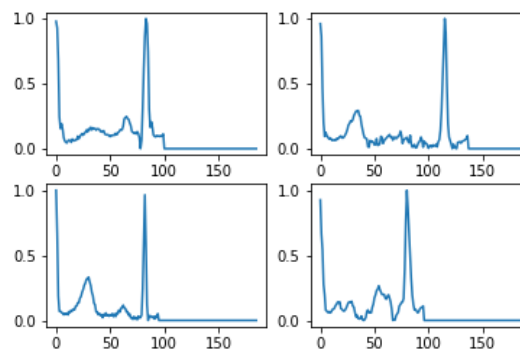

Figure 1: Examples of the data

3 Reasons

We chose this dataset due to the intriguing, challenging and useful aspect of medical topics and especially the ones concerned with the heart. A slight improvement in the real-time recognition of an abnormal heartbeat by an implanted device to a patient's heart could be crucial for them. We are not used to working with time series or signal data and this presents a challenge. Moreover, the specific dataset will give us the chance to apply both baseline traditional techniques and state-of-the-art deep learning approaches (LSTM, CNN, etc) in order to potentially demonstrate the superiority of the later in large datasets like this. We also hope that transfer learning and other advanced techniques will prove beneficial. We expect many challenges with preprocessing but none that we will shy from.

References

[1] Mohammad Kachuee, Shayan Fazeli, and Majid Sarrafzadeh. Ecg heartbeat classification: A deep transferable representation. *2018 IEEE International Conference on Healthcare Informatics (ICHI)*, Jun 2018.

[2] Alfaras Miquel, Soriano Miguel C., and Ortín Silvia. A fast machine learning model for ecg-based heartbeat classification and arrhythmia detection. *Frontiers in Physics*, 7:103, 2019.