Classifying ECGs with the PTB-XL ECG Dataset

Joseph Killian

Abstract - *Electrocardiograms (ECGs) are an increasingly common medical procedure that is done to help determine the health of a person's heart. They are a noninvasive and relatively easy procedure whose frequency has been increasing, and this increases the time required to read and accurately interpret the results. Some deep learning techniques have already been developed to try and solve this problem utilizing this dataset. The goal of this report is to examine how deep learning for classification for a subset of ECGs could be conducted using the PTB-XL. The methods used in this paper utilized python and jupyter notebooks to demonstrate the steps taken to classify the ECG waveforms. The accuracy achieved on the subset of 5 of the classes from the dataset was 81%. This shows that even with a simple model and light training, reasonable results can be achieved on this subset of classes from the data.*

1. INTRODUCTION

With Cardiovascular diseases being the leading cause of death worldwide, the ability to accurately detect, diagnose, and treat them is crucial. ECGs (Electrocardiograms) are a routine medical exam carried out to detect and diagnose cardiovascular diseases. This technology has been developing steadily, and its use has been widespread due to it being a noninvasive and relatively easy test to carry out. It is common for most appointments at a cardiologists office to include an ECG, or at a hospital when a patient is having symptoms consistent with cardiovascular diseases. The problem lies in that reading these ECGs accurately can be more time consuming and generally requires a cardiologist to read them.

Having systems trained on to solve this problem could make getting results cheaper, and faster, and allow doctors to spend less time reading results, and only need to confirm the findings of a decision system. The general lack of publicly available datasets has been problematic, as it makes research on this topic more difficult [1] There is also the issue of not so well defined metrics for evaluating the algorithms and methods used [1]. This is not just a problem with ECGs, but with many other problems in the medical field, where obtaining data requires a lot of caution with HIPPA laws. The data used for this project: The PTB-XL ECG Dataset, is a step in the right direction. It is a publicly available dataset that will hopefully allow further developments to be made towards solving the problem of classifying ECGs.

The dataset consists of 21837 records from 18885 patients, each 10 seconds in length [1]. This data was annotated by two cardiologists, along with some assistance from machines, with 5 overarching superclasses, as well as 24 subclasses [1]. The data itself was collected from 1989 -1996, and covers a wide variety of ages [1]. There are many statements regarding each of the records commenting on form, and rhythm, as well as other technical details regarding ECGs [1], but the main things for this project are regarding the diagnostic class and the waveforms themselves. Each waveform comes in two different sampling rates, 100, and 500. There are similar datasets out there, but the ones of similar size are only for private use. This dataset is the largest public one available for this area [1]. The following sections give some background on ECGs and their components.

WHAT IS AN ECG/HOW DOES IT WORK?

This section is meant to give a brief overview of how an ECG works, and what can be learned from the results. To start, nerve and muscle cells in our bodies communicate with both electrical and chemical signals, with regular electrical signals being sent to control our heartbeat [3]. In the right atrium of the heart, there is a group of cells called the sinoatrial (SA node) which send these electrical signals, which then spread through the heart muscle [3]. This leads to the atria, and then the ventricles of the heart to contract [3]. What the ECG does is measure the changes in voltage of these signals at the surface of the skin, and produces a graph. Measuring these signals at the surface of the skin also leads to noise which can make reading the graphs more challenging.

A normal heart that is beating steadily will produce what is known as a QRS complex. This QRS complex will be preceded by a P-wave, and then succeeded by a T-wave [3]. The diagram below gives a general idea of what this QRS complex looks like.

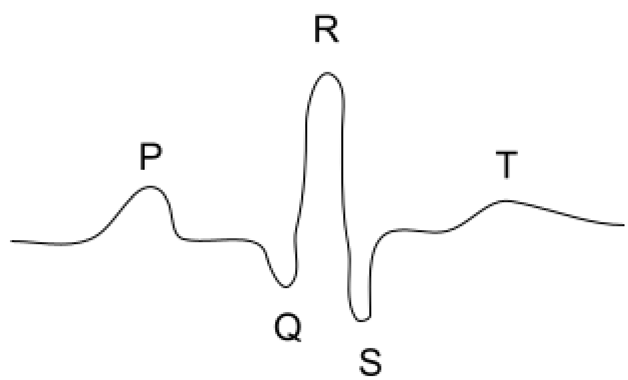

Figure 1: QRS Complex

Each "lead" of the ECG gives information about the voltage of the electrical signals of the heart. The leads of the standard 12-Lead ECG that is used in the dataset, will be explained in the following section. From these graphs produced, cardiologists can read them and determine things such as the narrowing of coronary arteries, a heart attack, irregular heartbeats, and other irregularities or diseases of the heart [3]. There are a few different types of ECGs: Resting (involves lying flat on your back with little movement), Exercise (measurements are made during physical activity, typically on an exercise bike or treadmill), and a Holter Monitor which measures electrical activity over a period of 24 hours.

WHAT ARE THE LEADS OF A 12-LEAD ECG?

This section explains what each of the leads mean in an ECG, and what their purpose is. The 12-lead ECG is the standard method for taking measurements. Each lead gives a tracing from 12 different electrical positions of the heart [4]. An ECG lead consists of two electrodes of opposite polarity where one is positive, and one is negative, or where one is positive and one is at a reference point [4]. Opposite polarity leads are called bipolar leads, and lead consisting of a positive electrode and a reference point are called unipolar leads [4]. These leads are calculated as combinations of multiple electrodes. There are only 10 electrodes which are used to calculate each of the 12 leads. Leads 1-6 are called the limb leads, and consist of three bipolar limb leads, and three unipolar limb leads. Leads 7-12 are chest leads, and all of these six are unipolar [4]. The reasoning behind the 12 different leads is that they are all taken at different angles with respect to the heart, and this allows for pinpointing exactly where in the heart irregularities are occurring [4]

GOAL FOR THIS PROJECT

The goal of this project was to try and do some classification of these waveforms into their diagnostic superclasses using deep learning. The method that was examined to try and do this was to sort the data into five superclasses are: NORM (Normal ECG), CD (Conduction Disturbance ECG), STTC (ST/T Change ECG), MI (Myocardial Infarction ECG), HYP (Hypertrophy ECG). All twelve leads for each data point were used for training and testing. Only those waveforms with 100% confidence were used in this implementation.

2. SOFTWARE CHALLENGES

The software related problems that needed to be completed for this task can be divided into: loading the raw data, sorting the data based on class, transforming the data, defining a model for classification, and then making predictions based off of the model. Much of the issues that were run into occurred early on when trying to load everything correctly in the right format. Once that had been completed, the training and testing went relatively smoothly. The format that the dataset came in originally consisted of header and dat files, so some extra work needed to be done in order to obtain all the wav files for each of the ECG records.

There are two csv files that contain information about the dataset, and those are linked to the folders that actually hold all the ECG data. The ptbxl_dataset.csv file contained all the information about the records, the metadata associated with each one, as well the path to each file in their respective folders. The records with a sampling rate of 500 were chosen for classification here, as it seemed to give the most accurate picture of what these waveforms look like.

Most of the issues that were run into were problems with making sure all the data was in the right folders, and in the right directory. The data folders were structures similar to the data in assignment 4 and the code snippets from class7 [5]. This made it easier to adapt the code that was used for classifying speech into classifying the waveforms for the ECGs. Software wasn’t a huge problem for this project, but there is more work and variations that could have been accomplished with more time available to work with.

3. WORKS COMPLETED

The first thing that was completed for this project was some initial poking around with the PTB-XL ECG dataset to understand best how to use it. This dataset is very detailed, and contains many attributes for each of the records available, each record comes in sampling rates of 100 and 500, along with all the other associated information. These additional attributes for each record were inconsistent as every ECG was not taken under identical circumstances. These are things like whether the report needed a second opinion, what noise was present, and other more technical descriptors for this procedure such as: infarction stadium, baseline drift, heart axis, etc. The dataset provides other metadata and identifiers, but for this project the scp_codes(Standard Communications Protocol Codes), and the waveform are of most importance.

The scp_codes are associated with the diagnostic classes, which are further broken up into subclasses and superclasses. There are 71 different types of scp_statements, 45 of which can be linked to a subclass or superclass, with the other 36 being descriptors of the waveform or associated with other conditions that were not prevalent or distinctive enough to be put into their own classification. For this project only the five main superclasses will be used for classification, any possibly further breakdown into some subclasses if time permits. These five superclasses are: NORM (Normal ECG), CD (Conduction Disturbance ECG), STTC (ST/T Change ECG), MI (Myocardial Infarction ECG), HYP (Hypertrophy ECG).

To filter the records into these classes, each record had to go through a few steps to determine where it should be placed. The first step is to determine the likelihood of the diagnosis. The likelihoods as given by the paper describing the dataset are: 15% (cannot rule out), 35% (likely), 50% (probably), 80% (very likely), 100% (Diagnose) [1]. Then, if the classification is already the superclass, then sort it into that category, if a subclass is provided, then it needs to be mapped from its subclass to its superclass. The method I used to sort these records is by taking only the records with 100% likelihood of being in a particular class, and also are unique only to that class. This excludes 8515 records from the dataset, and the numbers of each record are as follows:

Superclass:	# of Records:
NORM	7036
CD	2142
STTC	2331
MI	1001
HYP	181

Another step that was taken was to display some examples of each of the classes, and what the waveform looks like. This just gives an idea as to what the waveforms that will be used in this project look like. One record was looked at per class, as well as the waveforms for each of the leads. Below an example of each of waveforms for each class are displayed, each lead is represented by a different color. All of these examples were sampled at 500 samples/second, for 10 seconds. The notebook attached contains more graphs of each of the individual leads for each of these examples.

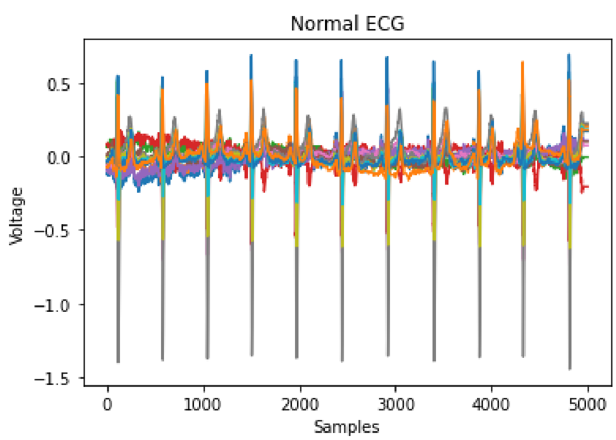

Figure 2: Normal ECG

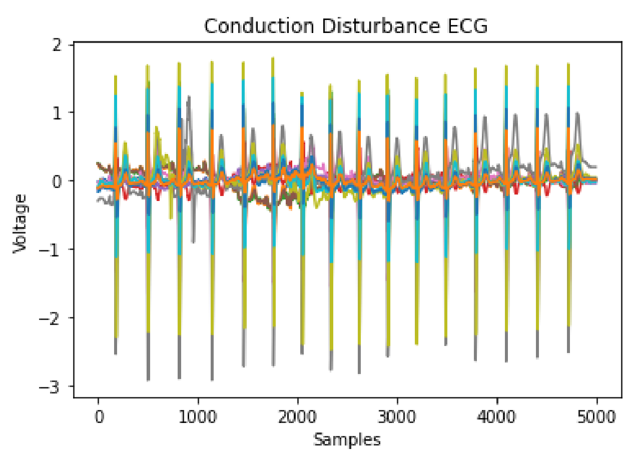

Figure 3: Conduction Disturbance

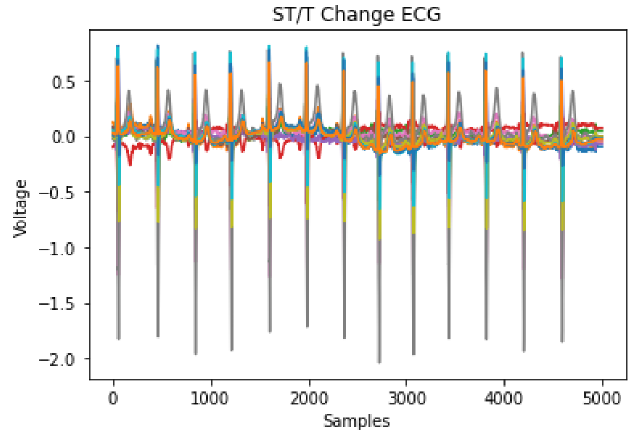

Figure 4: ST/T Change

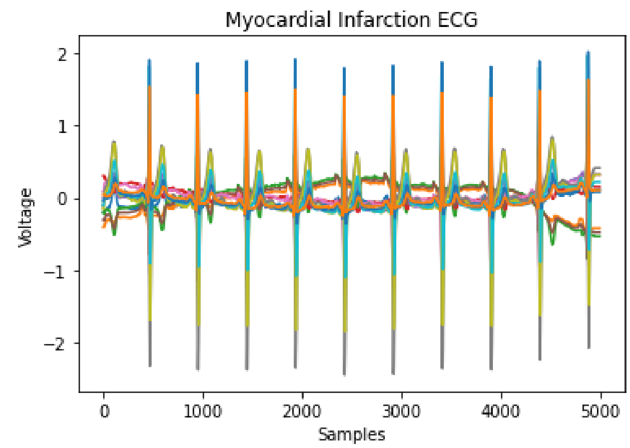

Figure 5: Myocardial Infarction

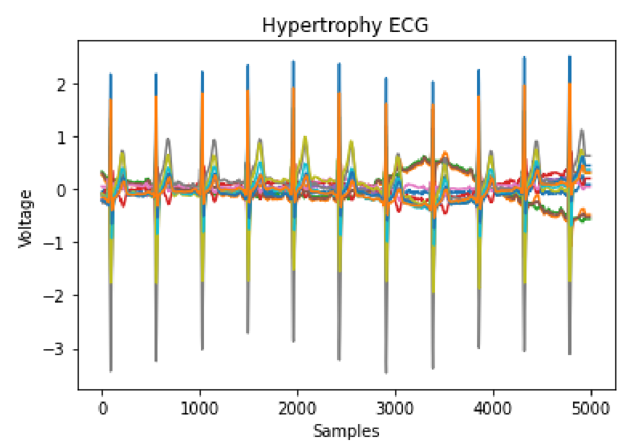

Figure 6: Hypertrophy

The next task to accomplish was to load all the .wav files into their respective folders. A waveforms directory was created, and in that folder, a folder for each class of ECG was created. The ECGs were then written into those folders as .wav files so that they could be more easily worked with as opposed to the .dat and header files that came with the original dataset. Once all the files were written to the folders, then needed to be loaded into a form that could be input into the deep learning model. Each of the twelve channels for each waveform was converted into a tensor along with the sampling rate and label associated with it. The sampling rate was the same for all the files at 500, and there was no reason to down sample as the training is relatively quick with only a few thousand records that fit the criteria that were chosen for evaluation.

Once all the files for each of the respective classes were loaded in, they were all appended to the same tensor and shuffled. This full dataset was then split into training and test sets, with a 90/10 split. The model used and all the functions defined regarding the deep learning model are all very similar to the functions and code from the PyTorch documentation, as well as the code snippets from class [5]. Only some slight modifications were necessary, the output classes had to be changed to 5, and the number of channels also had to be changed to twelve corresponding to the twelve leads of each of the ECGs. Once all that was completed, all that was left was to train the model and test it.

After around 100 epochs the test set accuracy got up to 81%. The model very quickly gets to around 75% and then soon to 80% accuracy, and plateaus out at around 80% with only marginal improvement after more that around 20 epochs of training. After doing the training, some examples were given of some of the predictions that were done on the testset to give an idea of how well it does. The next section goes over some possible improvements and other additional work that could have been accomplished with more time.

4. SELF EVALUATION

Overall there are a lot of areas for improvement for this particular project and the notebooks associated with it. There is a lot more experimentation that could be done to show some different methodologies to compare and contrast how they would all perform. The only method that was used here was a more or less already well defined network from the documentation with only slight modification, and using a subset of the data that was available (only about 60% of the total data).

Some other things that could have been attempted were examining each lead individually. Instead of using all the leads for the model, focusing on just some of the individual ones could be interesting to see how each lead affects a particular diagnosis with a cardiovascular issue. Other model architectures could have also been chosen to see if better results could have been obtained than the around 80% accuracy that was achieved. The subset of the data chosen was also a limiting factor as it only focused on the mutually exclusive records for each class that had 100% confidence for diagnosis. In reality there are varying levels of confidence for most ECGs so using all the data with more labels and confidence levels of diagnosis. Attempting to denoise the signal, is also another avenue that could be taken in order to obtain some better results

Given the time allotted for the project, there was some good background that was learned about how an ECG works more in depth, and what each of the leads mean. The waveforms of an ECG and the methods used for classification, are not different from those that are used for classifying audio samples. The lower sampling rates that ECG machines have is a potential issue that increases the difficulty of classification.

5. CONCLUSION

ECGs are one of the most common and noninvasive procedures that can be done that provide a window into the function and health of the heart. Correctly identifying problems that present in the heart is an important issue, and as more and more are done, the load on cardiologists to read all the test results increases. This project gave some background on the topic of ECGs, and an example implementation of how classification with the PTB-XL dataset could be conducted with deep learning. The accuracy achieved at classifying the five classes of: Normal , Conduction Disturbance, ST/T Change), Myocardial Infarction, Hypertrophy, was around 81%. More work can be done on this problem looking at more classes of cardiovascular diseases, as well as examining each of the individual leads of an ECG.

REFERENCES

[1] Wagner P, Strodthoff N, Bousseljot RD, Kreiseler D, Lunze FI, Samek W, Schaeffter T. PTB-XL, a large publicly available electrocardiography dataset. Sci Data. 2020 May 25;7(1):154. doi: 10.1038/s41597-020-0495-6. PMID: 32451379; PMCID: PMC7248071.

[2] Śmigiel S, Pałczyński K, Ledziński D. ECG Signal Classification Using Deep Learning Techniques Based on the PTB-XL Dataset. Entropy (Basel). 2021 Aug 28;23(9):1121. doi: 10.3390/e23091121. PMID: 34573746; PMCID: PMC8469424.

[3] "What is an echocardiogram?," *InformedHealth.org [Internet]*., 31-Jan-2019. [Online]. Available at: https://www.ncbi.nlm.nih.gov/books/NBK395556/. [Accessed: 20-Mar-2022].

[4] Nurseslearning.com. 2022. *EKG Interpretation*. [Online] Available at: <https://www.nurseslearning.com/courses/nrp/NRP1619/Section%205/index.htm#:~:text=The%2012-lead%20ECG%20gives%20a%20tracing%20from%2012,to%20see%20the%20heart%20from%20many%20different%20angles.> [Accessed 27 March 2022].

[5] Andriamanalimanana, B 2022, code snippets class7, Speech and Audio Processing CS 548, SUNY Polytechnic Institute, delivered Spring 2022