ECG Classification Project

Robert Posch

Purpose:

The purpose of this project was to be able to reliably detect which leads are placed where on the body so that a completely wireless 12-lead ECG system (imagine a series of 2-inch diameter circular ‘sensing’ nodes) can be placed on a patient, and a receiver hub can wirelessly receive the signal data, classify which lead is located where on the body, and use that information to output the filtered signals to into a standard medical grade ECG machine, and perform appropriate calculations for the augmented leads (the ECG machine that normally receives the signals from wired lead).

Data Recording:  
Three different lead placements were chosen for close proximity (V1, V2, V3) to replicate the worst-case signal classification that could be determined when recording an ECG. Theoretically, successfully classifying the closest proximity leads would mean that the rest of the lead placements could also be classified. The three different leads were placed approximately 2 inches apart on the chest and fed into a small-signal filter to reduce noise and for amplification. Originally a differential amplifier was used, but the signal was extremely noisy, so an Adafruit “heartbeat” sensor was used. Each lead was measured time-independent of the others for approximately 10 minutes each while at rest and consciously controlling breathing rate. An Arduino’s ADC was used to sample the filtered signals. This data was recorded using Python to read the Arduino serial data and write to text files.

Classification:

For creating the model, Python was used with popular open-source data science libraries such as Tensorflow, Numpy, Matplotlib, and Keras. Tensorflow is Google’s API for many different machine learning algorithms, and Keras is a machine learning API layer that uses Tensorflow’s neural network functionality for rapid-prototyping and experimentation. Numpy and Matplotlib were used for data manipulation and graphing.

Using a convolutional neural network, the ECG signal data were used to train the network using different filter window sizes, number of layers, activation functions, pooling sizes, and loss minimizing functions. The ECG signals were chunked into various lengths, like having a series of images (the only difference is that this is a 1-dimensional time series, in comparison to 2-dimensional for black and white images, or 3-dimensional for color images).

For this project’s final result, a convolutional neural network with eight layers was trained. Each 10-minute long ECG signal was chunked into segments of 200 voltage samples. This is approximately corresponding to the length of each heartbeat. The model successfully classifies the different leads with about 98% accuracy, sometimes with 100% accuracy. The final accuracy of the trained neural network depends on the initial (randomized) starting weights of the convolution filter windows and neural network weight matrices. With bad initial conditions, the gradient descent does not converge, and the accuracy sits at about 33% (essentially doing nothing because there are only 3 possible classifications). When this happens, retraining the model will likely the result in the ~98% accuracy over the course of about 20 epochs (iterations over the data).

Future Improvements:

Simultaneous measurement of the leads would be valuable because this could result in more similar signals. For example, sitting in a certain position, having a different heart rate, or breathing differently could cause the three lead signals to be more different than they would be if they were all measured simultaneously. Additionally, the use of 10 leads for the full system model as it would be implemented, would likely affect the final accuracy of the neural network.

For real world use of this classification system, there would need to be much more data taken from many kinds of cardiac functioning; arrythmias and other cardiac problems, resting heart rates, BMI, etc. There will be problems with overfitting the model without enough data. Each person would need to be measured while performing different activities as well, such as sprinting, jogging, walking, sitting, laying down. Multiple neural networks trained for each different activity would greatly improve the general accuracy of the model.