EEP 596: AI and Health Care || Mini-project 1

Dr. Karthik Mohan

Univ. of Washington, Seattle

Apr 18, 2022

Mini-Project Overview

Arrhythmia or "irregular heart beats" is a very common heart rate problem and often goes un-diagnosed. This mini-project looks into super fine-grained data on heart beats and characteristics of normal and abnormal heartbeats. Having ML algorithms that can automate detection of possible Arrhythmia is super impactful in helping doctors and hospitals be more efficient and effective in diagnosis and treatment of heart rate issues and also avert medical emergencies, prevent deaths. Since this is a mini-project, it is much more involved than a regular assignment - Plan a good amount of time for going through all the deliverables. To make it easy for you, we give you two deadline dates for the mini-project - An early deadline to help you get your feet wet and submit code on early models. And a final deadline to submit all your work. Mini-project has twice the grade weightage as a regular assignment and so you also get two weeks to do it. Enjoy!

Submission Guidelines

- You get to work in teams of 2 for the Kaggle and modeling piece!! Please make sure each person of the team gets to work on all aspects of the mini-project and mention at the top of your report the contributions from each person.
- The submission is in 3 parts
- **Code:** Please submit a Jupyter/IPython notebook file, report and Kaggle predictions as part of your submission. You can start with the template notebook provided and add in your solutions to it.
- **Report:** The report should be in a pdf format and have plots, correlation matrices and tables added in as mentioned in the **Heart Rate Deliverables below**. Feel free to use either latex or word for creating it. Include answers to conceptual questions, and your insights as well. Ideally you should NOT use comments in ipynb to answer any conceptual question.
- **Kaggle Contest:** There is a Kaggle competition as well, where you submit predictions on a "held out" data set.

Dataset Description

We have the following classes of heartbeats present in the dataset :

- N: Normal beat
- L: Left bundle branch block beat
- R: Right bundle branch block beat
- A: Atrial premature beat
- V: Premature ventricular contraction
- U: All other types of beats should be classified as this (this would require relabelling of the data)

Dataset description

The database contains 44 half-hour excerpts of two-channel ambulatory ECG recordings, obtained from 43 subjects studied by the BIH Arrhythmia Laboratory between 1975 and 1979. The recordings were digitized at 360 samples per second per channel, and were labelled manually by cardiologists. You can only use MLII information to train the model for all the part except the last one, this is because we have maximum availability of this feature. The txt file contains time, sample number and type of the heartbeat.

Required preprocessing

The first objective is to split MIT-BIH record at the R-peaks into individual heartbeat records. This can be done by creating a file which shall have the required information from the csv and txt files. The txt file contains time, sample number and type of the heartbeat. For each row of txt files, take 180 samples before and 179 samples after this sample number to create a time series from the corresponding csv file with the corresponding type as the label. Hence the final file you shall create shall have 360 features, along with it's label.
Feel free to try out any other method for pre-processing, and clearly explain the steps taken for it in the report. (There are no additional points for this)

Heartbeat Prediction Deliverables

- Do the preprocessing defined in the previous page. Then plot 3 heartbeats which are classified N and 3 which are classified as some other class. Is there a visible difference between these? How many heartbeats do you have in total? (30 points)
- Data normalization(can normalize to range [0,1]). Feel free to add any other pre-processing you deem useful. (5 points) (5 bonus points for any other pre-processing added)
- Class imbalance handling
  - Show the class imbalance present in the database with the help of plots. Use an autoencoder to augment data for classes with lesser data(especially for the A class). (20 points)
  - Show some plots of true anomalies and generated anomalies - And compare them side by side visually. (5 points)
  - Use a Variational autoencoder(VAE) architecture for the same. Discuss it's performance. (10 bonus points)

Heartbeat Prediction Deliverables

- Data denoising
  - Apply any noise reduction method(like Fourier transform, wavelet transform etc). Then plot the heartbeat with and without this filtering, and discuss the differences. Briefly describe how is your method useful. (HINT: Find a method to make the frequency component of noise zero) (10 points)
- Run at least one non-deep learning (ML) model on the processed dataset. Do hyperparameter tuning for the same. Show the confusion matrix, f1 score and accuracy score. Specifically mention the metrics for 'A' class as well.(5 points)
- Apply a feed-forward neural network and discuss it's performance w.r.t. the machine learning model used(on metrics defined in previous question). (10 points)

Heartbeat Prediction Deliverables

- Implement the neural network architecture from any recent paper on the MIT BIH Arrhythmia Database (check references for some papers). It is expected that the implementation should be your own and briefly describe the approach taken. How was the performance of this model? Were you able to get similar scores to the reference paper? Submit your implementation as well. (40 points)
- Plot the curves of training, validation and test sets losses and accuracy scores with number of epochs on the x-axis. Show a table with performances of different models. (10 points)
- Run the last neural network model with usage of one more feature in addition to MLII for which there is enough data. You can choose any one of V1, V2, V3, V4, V5. Compare it's performance with the usage of only MLII feature on the accuracy score. (15 points)

Heartbeat Prediction Deliverables

- Interpretability - Print/plot examples or time-series snippets of mis-classified arrhythmia (False positives) and also false negatives. Why do you think the model might have done a mis-classification here? (10 points)

References

- MIT-BIH Arrhythmia Database
- Noise Reduction in ECG Signals Using Fully Convolutional Denoising Autoencoders
- ECG arrhythmia classification by using a recurrence plot and convolutional neural network
- ECG Heartbeat Classification Using Convolutional Neural Networks
- Generalization of Convolutional Neural Networks for ECG Classification Using Generative Adversarial Networks
- AMSOM: artificial metaplasticity in SOM neural networks—application to MIT-BIH arrhythmias database