Computer Engineering Department

Platform for automated analysis using clinical time series data

Project Advisor: Vijay Eranti

Jacky, Chow (MS Software Engineering)
Jumana, Nadir (MS Computer Engineering)
Jerry, Huang (MS Software Engineering)

Introduction

We aim to create a low-cost and effective platform which will give deep insights to a patient's data for various diseases which are evaluated based on the time-series data. Our web application is user-friendly and delivers easy insights not just to healthcare professionals but also to the patients. For training our model efficiently, we require a large amount of open-source data that supports signal processing and patient-specific knowledge and understanding. Resources like PhysioNet are an essential part of our findings[4]. Physionet and other healthcare organizations support a large number of datasets for machine learning research. We are gathering large and relevant datasets for our model.

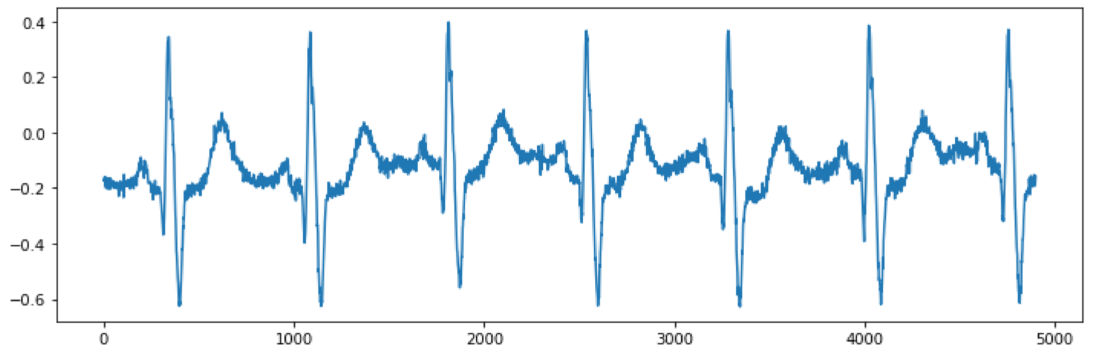
Wave-like features of the electrical impulses allow doctors to determine whether a person's heart is beating normally. Advancement in Artificial Intelligence is making it possible to analyze large data and give results in real-time.

In this project, we propose a Deep Learning Platform for the classification of such time-series ECG data in a real ECG test. Reducing the testing time, which otherwise takes up to 3 days in the US to over a week in other countries. Data collected is in the form of raw electrical signals, with reference electrodes on the right arm. This waveform data is used to distinguish between a normal and abnormal wave. We compare and propose deep learning models which can effectively classify these signals.

Methodology

Noise Cancellation

Noise contamination sources for the time series data is usually due to power line interference, electrode contact noise, motion artifacts, muscle contraction, baseline wandering and due to random noise. We need to remove this noise before building our model. We performed various experiments to remove noise from our data, below are the results of denoising. The graph below is the raw data with the noise.

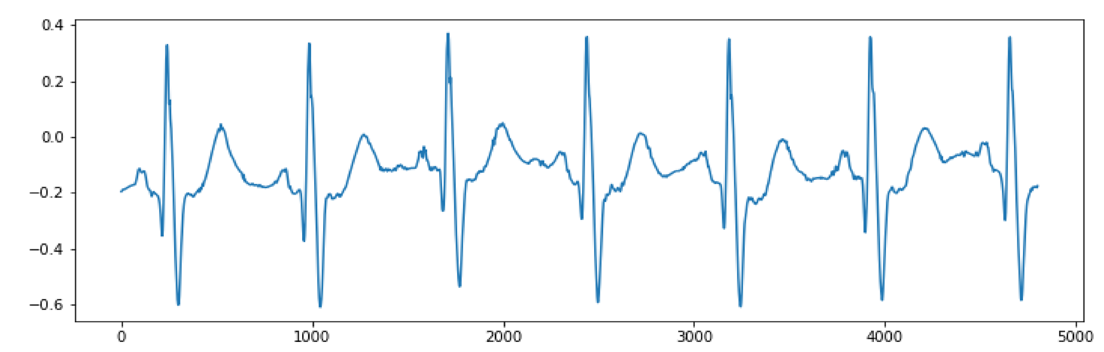


$$RMS = \sqrt{} 1\ N \sum [ypred(n) - yclean(n)]\ N\ 2\ n{=}0$$

$$SNR = 10 log 10 \sum [yclean(n)]$$

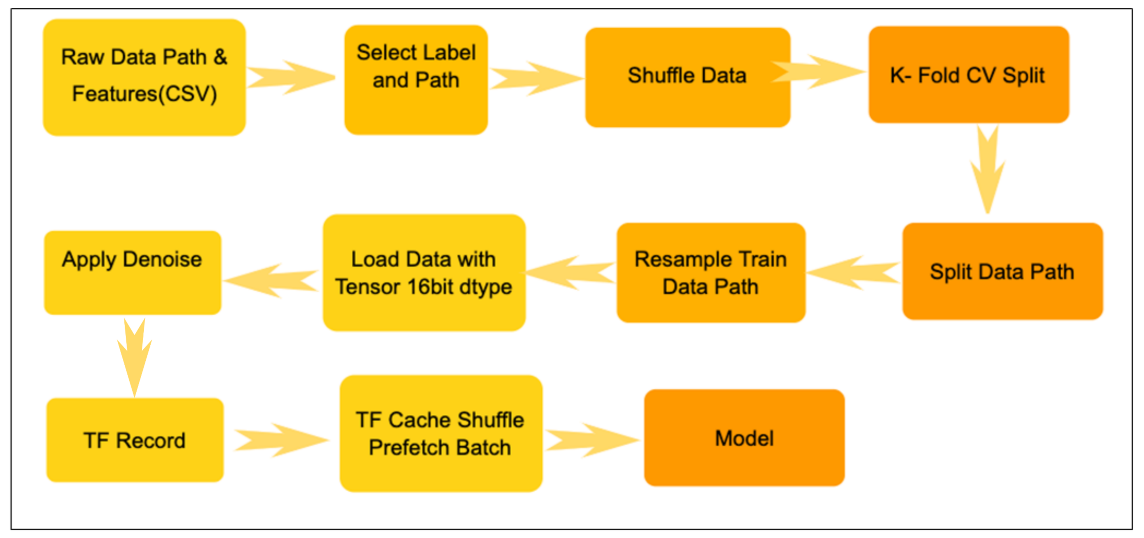
$$N\ 2\ n{=}1 \sum [ypred(n){-}yclean(n)]\ N\ 2\ n{=}1$$

Where yclean is the clean ECG signal, ypred is the predicted ECG signal, n is an index, N is the total number of sample points forming an ECG signal, which can have a duration of 1 second, 10 seconds, or any other time duration chosen by us. The engineering literature is abundant about discussions on what is a good or a bad RMS value or a threshold value for SNR. Statistical significance has been usually taken at 5% or even lower depending on the domain of study. In this work, it is proposed that the RMS limit value to be taken 5% of the difference between the maximum and the minimum values of the ECG signal being investigated. And below is the graph of the data after noise cancellation.



These maximum and minimum values for a clean ECG signal usually do not go above 3 mV and below 40μV: however, because of the different ECG apparatus/machines used to record the ECG signal, then the clean ECG signal may vary between 3 mV and -3 mV, which give a difference of 6 mV. This interval (i.e. [-3 3]) is also characteristic for the clean ECG signals used in this work. It results in an RMS limit value of 0.3 mV (i.e. 5% of 6mV), which will be used through the remaining part of this investigation. With regard to an SNR threshold, values higher than 10 dB are usually considered acceptable, while others are reporting SNR higher than 6 dB also worth to be taken into consideration. This depends also on the type of signal to be investigated that is for example signals from DSL cables or wireless signals and so on. In this work it is of interest the denoising of real ECG signals with SNR much below 0 (e.g., SNR= -6 dB, SNR = -8 dB) and especially the identification of the spikes corresponding to the R points in the P–Q-R-S-T wave. Therefore, it will be considered an SNR= 8 dB the threshold points above which it is considered that the denoised ECG signal may still carry some valuable information[3]. We used the wavelet library to do the denoise function and we set the level of dB to 9. Which is a high level, but it is the best one for our data[5].

Data pipeline



We decide to make the dataset into the TF records first and then load the TF records and use the batch function directly and send the batch into the models. At the same time, we used the build function cache, shuffle, prefetch on the tf records data before we train the model. we find in this way the evaluation will increase around 5% compared to if we do not use the tf record and the tf build in function Dataset.

Analysis and Results

ECG2004-PTB

This database contains 549 records from 290 subjects, both men and women. Each subject is represented by one to five records. There are no subjects numbered 124, 132, 134, or 161. Each record includes 15 simultaneously measured signals: the conventional 12 leads (i, ii, iii, avr, avl, avf, v1, v2, v3, v4, v5, v6) together with the 3 Frank lead ECGs (vx, vy, vz). Each signal is digitized at 1000 samples per second, with 16-bit resolution over a range of ± 16.384 mV[1]. The diagnostic classes of the 268 subjects are summarized below

Diagnostic class	Number of subjects
Myocardial infarction	148
Cardiomyopathy/Heart failure	18
Bundle branch block	15
Dysrhythmia	14
Myocardial hypertrophy	7
Valvular heart disease	6
Myocarditis	4
Miscellaneous	4
Healthy controls	52

We first used dataset ECG 2004 from PhysoNet. This raw dataset has 13 labels, but many of labels were lacking of data. Therefore, we selected Myocardia Infraction and the normal labels as our binary class and we did the training on three different models and did the hyperparameter tuning on each of them. The result shown on table below. [reference]

Model	Normal support	MI Support	Normal precision, Recall, F1(%	MI precision, recall, F1 (%)	accuracy(%)
CNN	52	52	61, 67, 64	61, 52, 56	61
LSTM	5705	5705	88, 69, 77	77, 90, 82	83
CNN+LSTM	5783	5712	61, 90, 72	80, 41, 54	66

ECG2020-PTBXL

The PTB-XL ECG dataset is a large dataset of 21837 clinical 12-lead, Each record includes 12 simultaneously measured signals: the conventional 12 leads (i, ii, iii, avr, avl, avf, v1, v2, v3, v4, v5, v6). ECGs from 18885 patients of 10 second length. The raw waveform data was annotated by up to two cardiologists, who assigned potentially multiple ECG statements to each record. The in total 71 different ECG statements conform to the SCP-ECG standard and cover diagnostic, form, and rhythm statements[2]. The diagnostic classes of the subjects are summarized below

#Records	Superclass	Description
9528	NORM	Normal ECG
5486	MI	Myocardial Infarction
5250	STTC	ST/T Change
4907	CD	Conduction Disturbance
2655	HYP	Hypertrophy

We are using the PTB-XL as our raw database. We firstly set the training dataset with binary labels. We set normal as 0, and we set the rest of the labels which are MI, CD, STTC, HYP all to 1. And then we train our models and do the hyperparameter tuning and results are shown as below. [reference]

Binary class result

model	accuracy	recall	precision	auc
ANN	0.8308	0.83	0.8308	0.9139
CNN	0.8359	0.831	0.8301	0.9205
LeNet-5	0.7959	0.781	0.796	0.853
VGG16	0.8689	0.8629	0.8629	0.9328
VGG19	0.8618	0.862	0.8616	0.9364
Inception	0.8424	0.841	0.8434	0.9177
LSTM	0.6322	0.6362	0.6339	0.6825
LSTM-CNN	0.573	0.6322	0.6343	0.7015

We also train our model using the single disease labels. We set our training dataset also to binary class. For each training set only contain the normal feature as 0 and the target label to 1. And we used the best model VGG19 as our model to train the different training sets.

Feature	loss	accuracy	recall	precision	auc
STTC	0.1969	0.9251	0.9251	0.9251	0.9789
MI	0.2343	0.8632	0.8632	0.8638	0.9124
CD	0.2803	0.8982	0.899	0.8975	0.9634
HYP	0.1922	0.9272	0.9281	0.9264	0.925

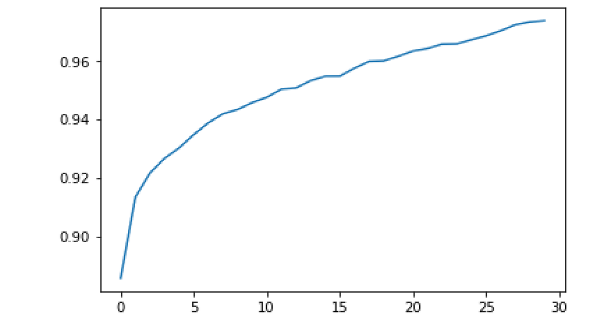
Multiple class result

We trained our model using the multiple classes and we also use the VGG19 as our model to do the hyperparameter tuning. In 5 labels multiple class, we set normal as 0 and the rest as label as single labels. In 4 labels multiple class, we set normal as 0 and we only use STTC, MI, CD as other labels. Cause the lack data of HYP in our raw dataset, we just remove the HYP feature here. And get 1% increase compare to 4 labels with AUC evaluation.

model	accuracy	recall	precision	auc
4 labels	0.7199	0.6426	0.7579	0.89
5 labels	0.7119	0.6769	0.7627	0.8839

Evaluation metrics

The area under the receiver operating characteristic curve (AUC) is frequently used as a performance measure for medical tests. It is a threshold-free measure that is independent of the disease prevalence rate. We evaluate the utility of the AUC against an alternate measure called the average positive predictive value (AP), in the setting of many medical screening programs where the disease has a low prevalence rate[6].



Summary/Conclusions

Time series data is one of the most common diagnostic data Procedures performed in hospitals and doctors' offices. We imagine the great potential of automatic timeseries data interpretation Algorithms in different medical applications, but we see the current progress in the field due to lack of Appropriate benchmark data sets and well-defined assessments program. We provide the flow of building the processes to classified on the ECG data. Also, we provide the efficiency data pipeline for those researchers how to handle a large dataset in local machine. We found the VGG19 will provide the best evaluation with our ECG data. We think as the development of the AI and medical technology, AI will help people getting a better and healthy life in the future. We provide a basic pipeline flow for medical timeseries data, which we think will be helped for future researchers keep developing in medical AI. For the future work, researchers can try to use the AUTO-ML[13] to increase the evaluation score. And researchers can try to use feature selection to select the best channel base on the different channel's importance to different disease label. Meanwhile, collecting more and more data on the other diseases labels and provide a more multiple classification is also possible.

Key References

[1] Ralf-Dieter Bousseljot, PTB Diagnostic ECG Database(2004).

[2] Patrick Wagner, Nils Strodthoff, Ralf-Dieter Bousseljot, Wojcich Samek, Tobias Schaeffter, PTB-XL, a large publicly available electrocardiography dataset(2020).

[3] Shoaib Ahmed Siddiqui. *et al.* TSViz: Demystification of Deep Learning Models for the Time-series Analysis. *IEEE Access.* Volume PP (99) (2019).

[4] Eduardo Jose da S. *et al*. ECG-based heartbeat classification for arrhythmia detection: A survey. *Computer Methods and Programs in Biomedicine*, Volume 127 (2016).

[5] Muhammad Ryan, Muhammad RyanWhat is Wavelet and How We Use It for Data Science(2019)

[6] Karimollah Hajian-Tilaki, PhD, Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation

Acknowledgements

We are deeply indebted to our Professor Vijay Eranti . He gave us lots of suggestions on our project especially on the data and the modeling. Ashish Jagtiani who is the co-founder and CTO at Chronus Health, and Sreenath Reddy is software expert also from Chronus Health . Thanks for their invaluable assistance and help in the preparation of this project. We would like to extend special thanks to Anand Parikh who is the CEO at Chronus Health.

We are extremely grateful to Prof. Dan Harkey for helping us throughout this journey. Special thanks to our writing instructor Bonnie and Maire.