Disease classification from 2-lead ECGs

Yue Zhang, ID: 10647902
Supervisor: David Wong

1 Introduction

Heart disease is currently one of the most fatal diseases in the world. The electrocardiogram (ECG) is the most commonly used tool for the detection of heart abnormalities. [24] It clearly shows the movement of the depolarization wave in each heartbeat. The depolarization wave is a wave of positive charge, which depends on the electrode that people use. This electrode is called the lead. The lead of the electrocardiogram is the placement position of the electrode on the human body and the connection between the electrode and the electrocardiograph galvanometer. When the myocardium is excited, the weak current generated is conducted to all parts of the human body. The lead can capture this current and reflect it on our electrocardiogram. The more leads are put, the more information we get about the heart. 12-lead ECG is a conventional method used to have a comprehensive understanding of heart activity. The standard 12-lead include six chest leads(V1, V2, ..., V6) and six limb leads (I, II, III, aVR, aVL, aVF). 6-lead (I, II, III, aVR, aVL, aVF), 3-lead(I, II, V2) and 2-lead(II, V5) ECG also contain much disease information. [22]

Compared with cardiologist diagnosis, computer diagnosis is cheaper, faster and can be used anywhere, anytime. Therefore, many scholars and companies try to use these advantages to make a device that can detect heart abnormalities. Companies like Alivecor have developed cardiac abnormality detectors using single-lead and six-leads. [15, 23] Although the single-lead instrument is easy to operate, the results are inaccurate. [23] The result of the six-lead instrument is accurate, but it is difficult for users to get stable ECG signals at one time, and it often takes many attempts to get the result. Can we make good use of heart information on 2-lead ECGs to achieving high accuracy and also convenience for users to access the ECG signal?

The aim of this project is to build a model to distinguish and classify different types of heart disease by using 2-lead ECG signals. My objectives are as follows:

- Identify potential useful algorithms for ECG classification from existing literature.
- Compare classification results using existing models on 2-lead and 12-lead.
- Consider how to extract ECG features manually and how to combine deep learning model with traditional machine learning models (such as decision tree) to create an ensemble model.

2 Background

In the 2020 year's PhysioNet/CinC Challenge, the classifier of 12-lead ECG signals can be divided into three types: traditional machine learning methods, deep learning methods, and ensemble models combining machine learning and deep learning models.

Some competitors try to use traditional models to classify myocardial diseases. Żyliński and Cybulski use bagged decision trees to solve this problem. They select Global Electrical Heterogeneity(GEH), AF features, The ratio of PVC beat, ECG periods and ECG morphology parameters. [1] Smisek et al. detect QRS waves and cluster and average these waves. Then get the onsets and offsets of P waves, T waves and QRS waves. They divide heart disease into two categories according to the different selection threshold criteria. For the first category, they use Standard clinical diagnosis to define the threshold, while for the second category, they use the decision tree to get the threshold. They made the eleventh in the competition. [2] Rajpal et al. build an XGBoost based classification model with Butterworth bandpass filter and "db4" discrete wavelet transform to deal with the ECG signals, and these signals are divided into four categories. [3] Perkins et al. combine PCA and traditional R-peak detection algorithm with a random forest model. [4] García-Isla et al. also combine PCA and Pan Tompkins' algorithm with decision trees but tried both bagging trees and boosted trees. [5] However, results on machine learning models are not as good as deep learning models.

ResNet performed very well in last year's competition. Jia et al. use a standard SE-ResNet34 to classify heart disease. Their test score is 0.359. [6] Chen et al. use two paths to get the ECG features and get 0.411 on the test set. [7] Based on SE-ResNet, Zhu et al. use an external public dataset and add a rule-based model on their network. They get 0.514 on the test set and place 3. [8] Zhao et al. They create an Adaptive lead-weighted ResNet. Their network has eight residual blocks with size 15 convolutional kernels, and each block has a SE layer. They also add gender features in their fully connection layers and add drop out layers. By doing this, they get 0.52 on the test set and won second place in the competition. [9] In addition to using ResNet to build a network for processing features, Jiang et al. also used Graph Convolutional Network for classification. From the result point of view, the effect is not as good as that of using ResNet. They only placed 34 in the competition, and their score on the test set is -0.012. [18] Apart from this, some participants also combine the time sequence model with CNN to a better result. Hasani et al. build a VGG-like branched CNN model with LSTM. Their test score is 0.437. [10] Fayyazifar et al. combine Bi-Directional LSTM with Convolutional network. They do not filter data manually. Their ranking in the competition is ninth, and their score in the test set is 0.382. [17] Lin et al. add global average pooling to the convolutional neural network. Their score on the validation set is 0.58(Fail to run on the test set). [19] EfficientNet also performed very well in this challenge. Nonaka and Seita use EfficientNet with Augmented data. They get 0.456 on the test set. [11] The top-10 models are all based on Neural Networks.

Combining traditional models with Neural Networks are also promising. Cai et al. build two models, one based on ResNet and another based on CNN and GRU. Then, train deep features in the XGBoost classifier. The score of the test set is 0.109. [12] Natarajan and his partners divide ECG features into "wide" feature and "deep" feature. They use the random forest to get the weight of each "wide" feature and use CNN to get the representation of the

"deep" features. Then, use a fully connection layer to combine these two types of features. Their model is the best last year, and their score on the test set is 0.533. [13] As the number of leads decreases, the effect of using traditional machine learning methods will become more and more significant.

Many scholars are also committed to using 2-lead ECG signals for classification. In 2015, Yan et al. used the restricted Boltzmann machine on the MIT-BIH data set to classify arrhythmia and achieved 98.829% accuracy. [20] In 2019, Ramirez, Melin and Prado-Arechiga using neural networks and fuzzy logic to classify the arrhythmia on data in the 2-lead signals. They put the lead II and lead I (or lead III or lead 5) signals into different units and then combined the two units. Using fuzzy KNN algorithm, MLP-GDM, and MLP-SCG training models, respectively, the best accuracy obtained by comparing the models applied on the MIT-BIH dataset is 94.20%. [21] Ochiai et al. combine CNN with Denoising Autoencoders(DAE) on MIT-BIH Arrhythmia Database to identify arrhythmia. [25] In 2019, Yang et al. used DL-CCANet and TL-CCANet to detect arrhythmia on two datasets. They achieve 95.2% accuracy on the MIT-BIH dataset by using DL-CCANet and 95.52% on IN CART data set by using TL-CCANet. [26] Ma et al. compare the effect of the iterative models(SVM and LS-SVM) and non-Iterative models(Random Forest) for 2-lead ECG. The result shows that the effect of random forests is the best. [27] Their experimental results have given us some inspiration for choosing models, but different models may have different effects on different data sets. This article also has its limitations. Most classifiers are used to detect arrhythmia but not all classes, so we need to consider migrating the 12-lead ECG classification method to 2-lead ECG.

3 Research Methodology

This project aims to build a classifier to do cardiac abnormality detection by using 2-lead ECG signals. Although we have achieved good predictions with deep neural networks on 12-lead ECG signals, this does not mean that neural networks can also achieve good predictions on 2-lead ECG signals. Because 2-lead ECG signals miss a lot of information than 12-lead ECG signals, only using the deep learning model does not guarantee the classification results. Therefore, we need to consider adding traditional machine learning models to deep learning models. For traditional machine learning models, the most important thing is to consider how to process features. For Deep Neural Networks, the problem of overfitting needs to be considered. The work plan is outlined below:

3.1 In the prepare phase:

Determine the project's data set:
The source of the data is the PhysioNet/Computing in Cardiology Challenge 2021
Theoretical background study (from journals and the Internet) of:

1. Clinical diagnosis of electrocardiogram
2. The processing method of electrocardiogram signals
3. The architectures of machine learning and deep learning models

Also, analyze the statistics of the data set and compare the results of 2-lead and 12-lead on an existing model, and compare the differences between the results is very vital.

By studying background knowledge, I should be able to identify which clinical diagnoses of the low-scoring category were not classified well, learn how to deal with waveforms and select machine learning and deep learning methods that offer inspiring ideas.

By analyzing the statistics of the data set and comparing the results, I can analyze the causes of the results and select a few categories that have the most opportunity for improvement.

3.2 In the data preprocessing phase:

The ECG signal is susceptible to noise. It has three common noises: power-line interference, EMG interference, and baseline wander.

Power-line interference refers to the interference caused by the signals around the equipment responsible for collecting ECG signals. Its waveform is very similar to a sinusoidal signal, which often overwhelms useless signals and also affects the detection of P and T waves.

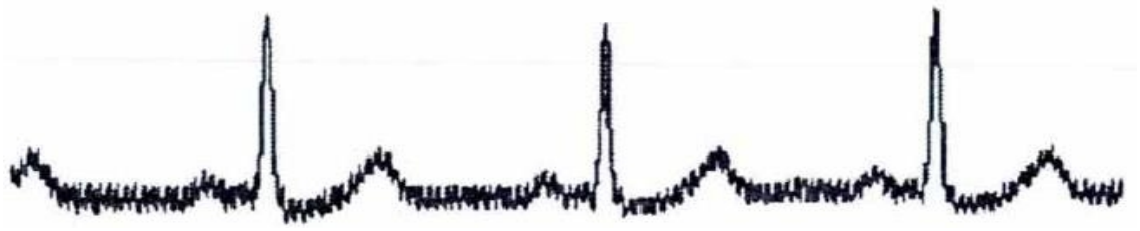

Figure 1: power-line interference

EMG interference is caused by the tremor of human muscles, there is no regularity, the waveform will change rapidly, so useful signals are ignored.

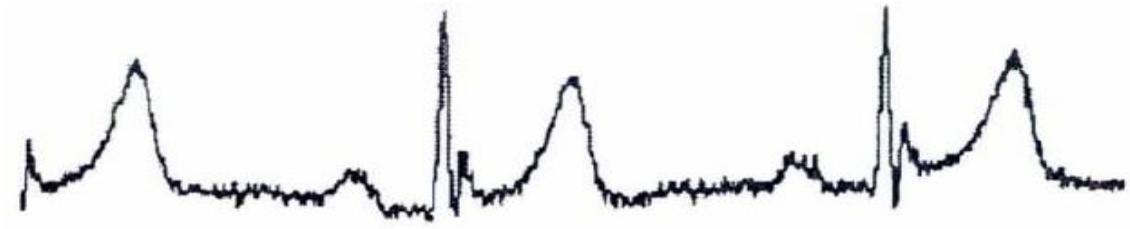

Figure 2: EMG interference

The baseline wander is caused by the sliding of the electrode position or the breathing of the human body, and the amplitude and frequency will change all the time. The PR band and ST band in the ECG signal are very susceptible.

Therefore, data processing is a significant part of the entire project, which determines the machine learning model's performance.

I plan to preprocess data as following:

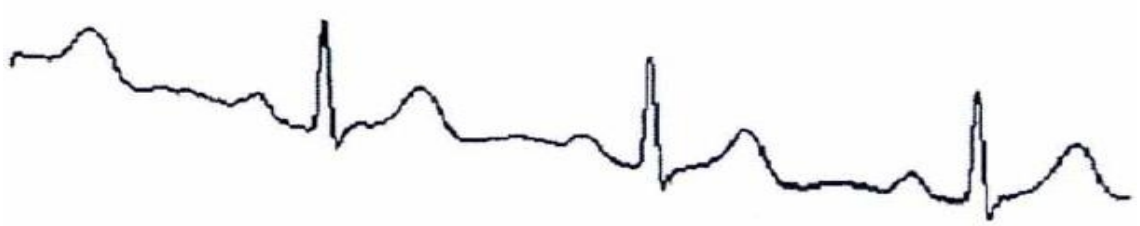

Figure 3: baseline wander

1. Data selecting and filtering
2. Zero padding and truncate data
3. Add some others techniques such as shifting and downsample to deal with data.
4. Use method like wavelet transform to deal with the problems that I have method above.
5. Especially focus on those classes with low scores and manually analyze the waveforms of these classes to extract essential features for classification.

3.3 In the training phase:

In this phase, I will implement a model based on the best model I found when I was doing a background study with PyTorch. The training will be on 2-lead dataset.

3.4 In the validation phase:

I will use cross-validation to verify the model's results, and uses loss entropy (or a loss function defined by myself) as my loss function.

3.5 In the evaluation phase:

I will choose AUROC, AUPRC and F_measure for each class I classified as my standard to evaluate my model.

The AUROC is the area under the ROC curve. AUROC can reflect the model's ability to distinguish between positive and negative examples. For practical problems like the ECG problem, the AUROC of 0.8 indicates that the model has a good distinguishing ability. [14]

The AUPRC is the area under the Precision-Recall curve. It is handy for evaluating imbalanced data that people want to find more positive cases. The data set used by the project has significantly fewer positive cases than negative cases, so it is significant to choose AUPRC.

As the harmonic mean of recall and precision, F-measure is a comprehensive consideration of the classification results of positive and negative examples.

I will take the average value of these three methods as my final criteria.

4 Ethics and Professional Considerations

The aim of ethics is to protect mainly data and human. The project's data is authorized for public use on The PhysioNet/Computing website [22], so this project does not have specific ethical issues to consider—no personal information except age and sex in this dataset. People cannot locate a specific person through the information in the dataset.

According to the ethics decision tool [16] in UoM, my project does not require NHS or HRA permission.

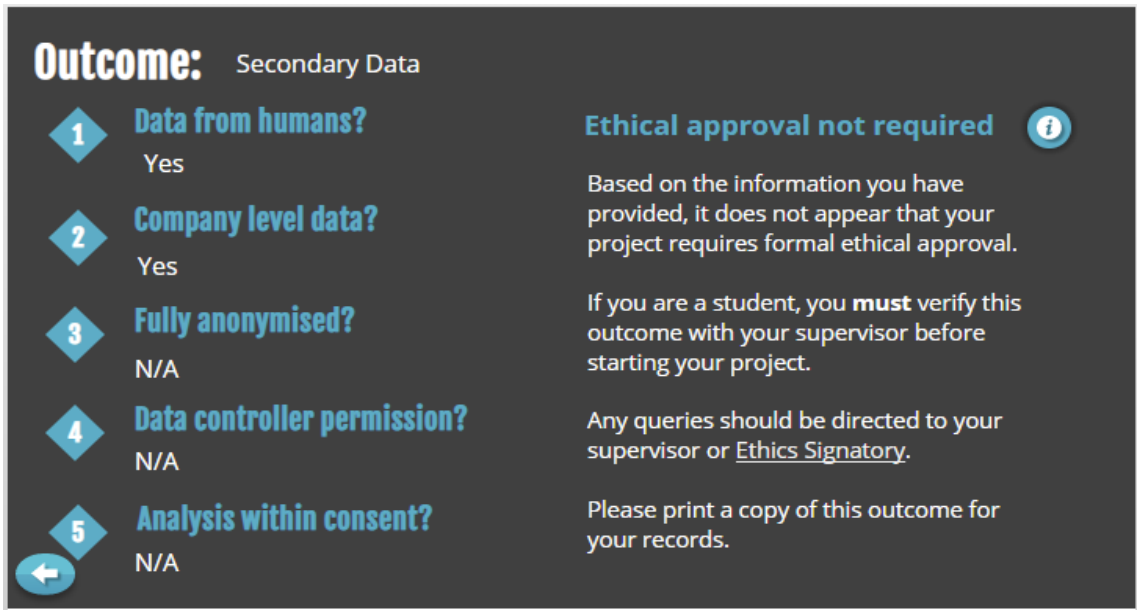

Figure 4: The result of UoM ethics decision tool

As shown in Figure 4, although the data is from humans, I do not need ethical approval in my project.

If the accuracy of the model is not high, ethical problems will also occur. The datasets of the project have a large amount of data, a wide range of patient ages, and different nationalities. However, deviations may still occur. Therefore, the results predicted by this project cannot completely replace doctors, but should be used as a reference for patients with possible cardiac abnormalities.

5 Risk Consideration

The common risks that researchers might face are shown in Table 1:

Risks type	Definition of risks	Risks in my project
Social risks	Disclosures that may infect the participants' life in the community, their family and their work.	No
Legal risks	Activities that may lead to the commission of a crime by the participant, researchers and/or the University; Activities can lead to criminal activities disclosed to scholars. Scholars require reporting to the authorities; Activities that may lead to a civil claim for damages.	No
Economic harm	Cause financial losses to participators, academics or universities through disclosure or other events.	No
Reputational risk	In the views of sponsors, research groups or the public, damage the public's perception of the university or the reputation of the university/researcher.	No
Safeguarding risks	Risk to adolescents, vulnerable adults or researchers due to improper behavior, abuse or exploitation. Researchers may face complex situations and may raise allegations of illegal activities.	No
Health and safety risk	Risks of harm to the health of participants or researchers, both physical and mental harm are included.	No

Table 1: The common risks of research

This project does not have these common risks. Apart from that, additional project risks are as followed:

- **Problem**: The risk of the project's final result is very poor
  Solution: Start coding early and need to read more papers in order to improve the model when the scores are low. Try to use different ways to deal with signals, use a different model and evaluate the model according to the standards of part 3.
- **Problem:** Computer damage or computer data loss
  Solution: Use Github to store the project's code
- **Problem:** Physical or psychological problems
  Solution: Exercise, and seek medical attention if there is a problem. If it is really serious, report to the school in time.
- **Problem:** Unfamiliarity with python's library causes coding difficulties
  Solution: Start as early as I can. Make a plan about what I should learn and execute it. Use online public resources like Coursera to learn those libraries.
- **Problem:** Spend too much time on the training model
  Solution: use Google Colab and the university's high-performance computers to ensure the training time is not too long. Coding early can also avoid this problem. If this happens, the plan must be adjusted in time for the project to proceed smoothly.

6 Project Evaluation

I will evaluate the success of the project using the project's objectives as the guidelines. My evaluation standards are as followed:

1. Complete a literature review of ECG models.
2. Select and implement the model of ECG classification(generate ensemble model). The selected method is declared in the paper and provide evidence of high model accuracy. Complete the code of the selected model.
3. Identify classes commonly misclassified. Declare in the paper which three categories are misclassified.
4. Develop a new model to improve on these classes. Successfully write the code of the new model and show the result of the new model.
5. Create an ensemble model and compare the specific classes' score mentioned in part 3 with the old model's (model that is not modified) score. Successfully write the code of ensemble model, and declare the score of the ensemble model in the paper.

For papers, the paper should have a clear structure, refined language, and meet the requirements of academic writing.

Item one to four would be a minimum standard, items one to five would be a medium standard, and the highest level would be items one to five, with a better score than the original one.

7 Planning

A reasonable plan can ensure the smooth progress of the project. I use a Gantt chart to visually express my plan. The plan is shown in Figure 5:

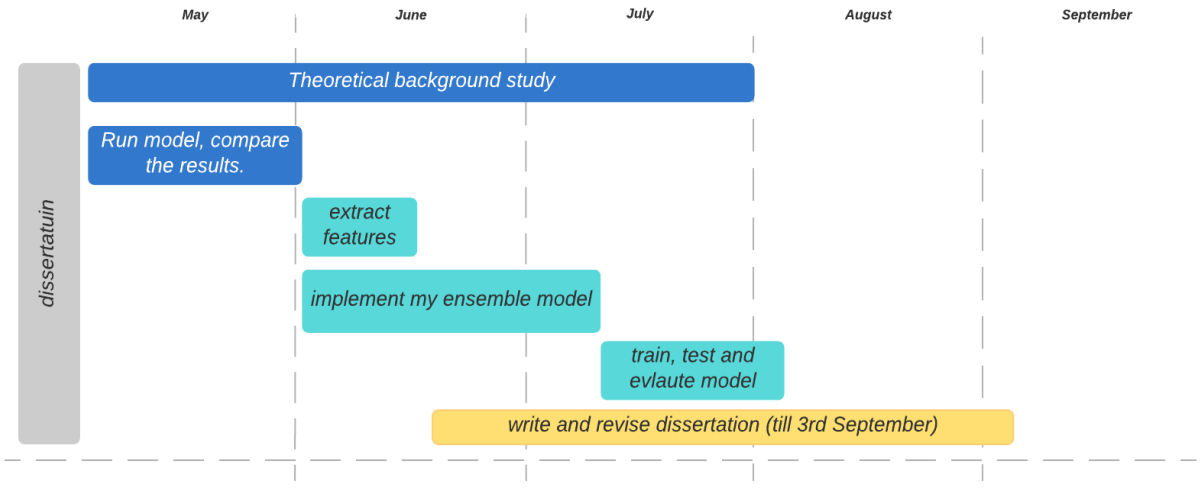

Figure 5: plan

I already have done a literature review, Ethics and professional considerations and risk assessment. I will not show these in the plan description and Gantt chart. According to time, I divided my plan into three stages :

From May to July, I will:

1. Doing background study
2. Run the existing model on the computer and compare the results.
3. try to extract features by myself (preprocess data)
4. implement my ensemble model
5. train and test my model

In August, I will:

1. Evaluate my model and improve my model.
2. do Write-up work.

In 1st September to 3st September, I will:

1. Revise my report.
2. Check and submit my dissertation.

At the first stage, I need to do most of the modelling work, considering that the course module is not over from May to early June, so most of the work may be done from mid-June to July. At this stage, I need to do much reading work to ensure that a high-precision model can be built. At the same time, it is also necessary to record details during modelling to prevent part of the content from being forgotten when writing the dissertation.

At the second stage, my main job is evaluating the model and writing a dissertation.

At the last stage, I should revise my dissertation, make sure it is academic, and submit it.

The plan cannot be set in stone. If there is any change in the plan, it needs to be adjusted in time. The dissertation should be written at least one week in advance to deal with unforeseen problems.

References

[1] M. Żyliński and G. Cybulski, "Selected Features for Classification of 12-lead ECGs," *2020 Computing in Cardiology*, 2020, pp. 1-4, doi: 10.22489/CinC.2020.061.

[2] R. Smisek, A. Nemcova, L. Marsanova, L. Smital, M. Vitek and J. Kozumplik, "Cardiac Pathologies Detection and Classification in 12-lead ECG," *2020 Computing in Cardiology*, 2020, pp. 1-4, doi: 10.22489/CinC.2020.171.

[3] H. Raipal, M. Sas, C. Lockwood, R. Joakim, N. S. Peters and M. Falkenberg, "Interpretable XGBoost Based Classification of 12-lead ECGs Applying Information Theory Measures From Neuroscience," *2020 Computing in Cardiology*, 2020, pp. 1-4, doi: 10.22489/CinC.2020.185.

[4] G. Perkins, C. McGlinn, M. Rizwan and B. M. Whitaker, "Detecting Cardiac Abnormalities From 12-lead ECG Signals Using Feature Selection, Feature Extraction, and Machine Learning Classification," *2020 Computing in Cardiology*, 2020, pp. 1-4, doi: 10.22489/CinC.2020.217.

[5] G. García-Isla, R. Laureanti, V. D. Corino and L. T. Mainardi, "ECG Morphological Decomposition for Automatic Rhythm Identification," *2020 Computing in Cardiology*, 2020, pp. 1-4, doi: 10.22489/CinC.2020.435.

[6] W. Jia, X. Xu, X. Xu, Y. Sun and X. Liu, "Automatic Detection and Classification of 12-lead ECGs Using a Deep Neural Network," *2020 Computing in Cardiology*, 2020, pp. 1-4, doi: 10.22489/CinC.2020.035.

[7] J. Chen et al., "SE-ECGNet: Multi-scale SE-Net for Multi-lead ECG Data," *2020 Computing in Cardiology*, 2020, pp. 1-4, doi: 10.22489/CinC.2020.085.

[8] Z. Zhu et al., "Classification of Cardiac Abnormalities From ECG Signals Using SE-ResNet," *2020 Computing in Cardiology*, 2020, pp. 1-4, doi: 10.22489/CinC.2020.281.

[9] Z. Zhao et al., "Adaptive Lead Weighted ResNet Trained With Different Duration Signals for Classifying 12-lead ECGs," *2020 Computing in Cardiology*, 2020, pp. 1-4, doi: 10.22489/CinC.2020.112.

[10] H. Hasani, A. Bitarafan and M. S. Baghshah, "Classification of 12-lead ECG Signals With Adversarial Multi-Source Domain Generalization," *2020 Computing in Cardiology*, 2020, pp. 1-4, doi: 10.22489/CinC.2020.445.

[11] N. Nonaka and J. Seita, "Electrocardiogram Classification by Modified EfficientNet With Data Augmentation," *2020 Computing in Cardiology*, 2020, pp. 1-4, doi: 10.22489/CinC.2020.063.

[12] W. Cai, S. Hu, J. Yang and J. Cao, "Automatic 12-lead ECG Classification Using Deep Neural Networks," *2020 Computing in Cardiology*, 2020, pp. 1-4, doi: 10.22489/CinC.2020.039.

[13] A. Natarajan et al., "A Wide and Deep Transformer Neural Network for 12-Lead ECG Classification," *2020 Computing in Cardiology*, 2020, pp. 1-4, doi: 10.22489/CinC.2020.107.

[14] D'Agostino Sr, Ralph B., et al. "Cardiovascular disease risk assessment: insights from Framingham." *Global heart*, 8.1 (2013): 11-23.

[15] KardiaMobile6L https://store.alivecor.co.uk/products/kardiamobile6l

[16] amended_interactive-decision-tree-amended-jan-2021-draft https://www.training.itservices.manchester.ac.uk/uom/ERM/ethics_decision_tool/story_html5.html

[17] N. Fayyazifar, S. Ahderom, D. Suter, A. Maiorana and G. Dwivedi, "Impact of Neural Architecture Design on Cardiac Abnormality Classification Using 12-lead ECG Signals," *Physiological measurement*, pp. 1-4, doi: 10.22489/CinC.2020.161.

[18] Jiang, Z., Almeida, T. P., Schlindwein, F. S., Ng, G. A., Zhou, H., & Li, X. (2020). Diagnostic of multiple cardiac disorders from 12-lead ECGs using Graph Convolutional Network based multi-label classification.

[19] Y. -C. Lin, Y. -C. Lee, W. -C. Tsai, W. -K. Beh and A. -Y. A. Wu, "Explainable Deep Neural Network for Identifying Cardiac Abnormalities Using Class Activation Map," *Physiological measurement*, pp. 1-4, doi: 10.22489/CinC.2020.072.

[20] Y. Yan, X. Qin, Y. Wu, N. Zhang, J. Fan and L. Wang, "A restricted Boltzmann machine based two-lead electrocardiography classification," *2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, 2015, pp. 1-9, doi: 10.1109/BSN.2015.7299399.

[21] Ramirez, E., Melin, P., & Prado-Arechiga, G. (2019). Hybrid model based on neural networks, type-1 and type-2 fuzzy systems for 2-lead cardiac arrhythmia classification. *Expert Systems with Applications*, 126, 295-307.

[22] Will Two Do? Varying Dimensions in Electrocardiography: The PhysioNet/Computing in Cardiology Challenge 2021 https://physionetchallenges.org/2021/

[23] KardiaMobile https://store.alivecor.co.uk/products/kardiamobile

[24] Kaur, A., & Arora, J. (2018). HEART DISEASE PREDICTION USING DATA MINING TECHNIQUES: A SURVEY. *International Journal of Advanced Research in Computer Science*, 9(2).

[25] Ochiai, K., Takahashi, S., & Fukazawa, Y. (2018). Arrhythmia detection from 2-lead ECG using convolutional denoising autoencoders. *Proc. KDD*, pp. 1-7. 2018.

[26] Yang, W., Si, Y., Wang, D., & Zhang, G. (2019). A novel approach for multi-lead ECG classification using DL-CCANet and TL-CCANet. *Sensors*, 19(14), 3214.

[27] KMa, Z., Chen, B., Xu, Y., Li, Y., & Zhang, Y. (2019, October). Comparing Performance of Iterative and Non-Iterative Classifiers for 2-Lead ECGs on Multi-Feature Schemes. *2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, pp. 1-6. IEEE, 2019.