CIS 520 Spring 2019 Project Proposal

Fatal or Non-fatal: Comparative Study of Classification Algorithms for Cardiac Arrhythmias Discrimination

Team Members:

Alëna Rodionova	PennKey: nellro	Email: nellro@seas.upenn.edu
Haochen Han	PennKey: hanhc	Email: hanhc@seas.upenn.edu
Haotian Zhu	PennKey: htzhu	Email: htzhu@seas.upenn.edu
Po-Yuan Wang	PennKey: poyuanw	Email: poyuanw@seas.upenn.edu

Instructor:
Prof. Shivani Agarwal

Project Mentor:
???

1 Introduction

In this project we will compare classification methods for electrogram (ECG) arrhythmia discrimination. The set of methods we analyze include Deep Neural Network (DNN), Support Vector Machine (SVM), k-Nearest Neighbours (k-NN) and Decision Tree.

The following two datasets will be used:

- The MIT-BIH Arrhythmia Database [1]. It contains 48 half-hour two-channel ECG recordings, obtained from 47 subjects. The recordings are digitized at 360 samples per second. There are approximately $110,000$ ECG beats.
- EGM Database [2]. It consists of 1920 EGM signals, equally split into 960 VTs and 960 SVTs. The EGMs were generated by the heart model that has been validated for realism by cardiologists [2].

2 Problem formulation

According to the World Health Organization (WHO), heart disease is the leading cause of death around the world. More than 1000 Americans died each day in 2016 from heart attacks. Although a single arrhythmia heartbeat will not have a serious impact on life, continuous arrhythmia can result in fatal circumstances. For instance, *Ventricular Tachycardia* (VT) is a potentially fatal arrhythmia. On the other hand, *Supraventricular Tachycardia* (SVT) is not considered to be dangerous for most of people.

The problem we decided to work on is cardiac arrhythmias discrimination. We view this problem as a binary classification task that discriminates whether a patient has VT (fatal tachycardia) or SVT (non-fatal tachycardia). The will analyze, implement and compare four classification algorithms: DNN, SVM (with and without PCA), decision tree and k-NN.

We will use the following performance measures:

- Precision, TPR (# correctly detected VT/# true VTs), TNT (# correctly detected SVT/# true SVTs). Those are very important performance measures in case of tachycardia discrimination. Though FP (misclassifying SVT as VT) might lead to a process of further examination or treatment episodes which are painful for the patient, FN (misclassifying VT as SVT) might put the patient in the risk of death.
- For DNN and SVM with PCA performance we will also measure training and test time.
- Since PCA is used to reduce calculation costs, we will specifically measure how much training time was reduced after using PCA while test error is still above threshold.
- Since k-NN is a memory-based algorithm we will report its test times.

3 Related work

Due to high importance of the problem, hundreds of methods for arrhythmia classification have been presented in the literature, many of them use machine learning techniques. For instance, in [3] authors perform multi-class classification of ECG signals using convolutional neural network (CNN) to discriminate between different tachycardia types. Authors in [5] present a 95.16% accurate k-NN classifier. In 2012, other group of researchers reported a 99.34% accurate multi-class arrhythmia classification algorithm using *ensemble decision trees*. In [4] authors show that under specific conditions decision trees perform better than SVM classifiers.

4 Solution methods

The set of methods we analyze implement and compare include:

- Deep Neural Network (DNN) (Alëna),
- Decision Tree (Haochen),
- Support Vector Machine (SVM) (Haotian),
- k-Nearest Neighbours (k-NN)(Po-Yuan).

Related work review shows that decision trees can achieve high accuracy results in arrhythmia classification based on ECG signals. As the database for our project is also in ECG form, decision tree is a valid candidate for distinguishing fatal/non-fatal tachycardias. Compared with neural networks, decision trees have an advantage that they are easy to interpret. Unlike neural network, the transparency of decision tree gives us a better understanding of the intermediate processes.

k-NN algorithm is known for its implementation simplicity and fast training time. It is robust to noisy data, but sensitive to irrelevant features. The biggest disadvantage of k-NN is its expensive computations in the testing phase. Moreover, in case of high-dimensional data, the distance between the two data points becomes less meaningful and the accuracy may decrease [6].

After we complete the main part of our project, we also plan to show beneficial results of applying Principal Component Analysis (PCA) for speeding up the performance. Though PCA is usually used in unsupervised learning, we will show how it can be applied to supervised learning algorithms such as DNN and SVM. We will also focus on possible modifications of the algorithms such as cost-sensitive decision tree, combining k-NN, and AdaBoost.

5 Plan of work

The project time-line up to the milestone meeting is presented in a Table 1.

	Algo	**Week 1**	**Week 2**	**Week 3 & buffer time**
Alëna	Neural Network	Data pre-processing, data cleaning	NN implementation, hyper-parameters identification	Performance improvements, algorithms comparison
Haochen	Decision tree	ECG/EGM signals feature extraction	Modeling, literature review	Implementation. Cost-sensitive decision tree
Haotian	SVM	SVM and PCA theory/tutorials, literature review	SVM algorithm implementation, analysis of possible improvements without PCA	SVM with PCA. Parameter identification. Performance comparison
Po-Yuan	k-NN	k-NN literature review, problem analysis	Implementation, results analysis, performance improvements	Resolving implementation issues, improving computation-time, complexity-reduction

Table 1: Project time-line: from the moment of the project proposal submission up to the milestone meeting.

Acknowledgments

The authors would like to thank Kuk Jin Jang for the valuable discussion on the existing arrhythmia monitoring algorithms and publicly available ECG databases.

References

[1] G. B. Moody and R. G. Mark, "The impact of the mit-bih arrhythmia database," *IEEE Engineering in Medicine and Biology Magazine*, vol. 20, no. 3, pp. 45–50, 2001.

[2] Z. Jiang, H. Abbas, K. J. Jang, M. Beccani, J. Liang, S. Dixit, and R. Mangharam, "In-silico pre-clinical trials for implantable cardioverter defibrillators," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2016, pp. 169–172.

[3] T. J. Jun, H. M. Nguyen, D. Kang, D. Kim, D. Kim, and Y.-H. Kim, "Ecg arrhythmia classification using a 2-d convolutional neural network," *arXiv preprint arXiv:1804.06812*, 2018.

[4] M. MOHANTY, P. BISWAL, and S. SABUT, "Ventricular tachycardia and fibrillation detection using dwt and decision tree classifier," *Journal of Mechanics in Medicine and Biology*, p. 1950008, 2018.

[5] I. Assadi, A. Charef, T. Bensouici, and N. Belgacem, "Arrhythmias discrimination based on fractional order system and knn classifier," 2015.

[6] K. Beyer, J. Goldstein, R. Ramakrishnan, and U. Shaft, "When is nearest neighbor meaningful?" in *International conference on database theory*. Springer, 1999, pp. 217–235.