Classification of cardiomyocytes based on their action potentials

Anil Palepu, Arman Koul, Jenn Heiko, Maggie Wodicka, and Samiksha Ramesh

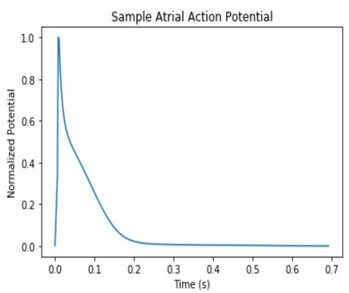
EN.553.436 | Johns Hopkins University | Spring 2019

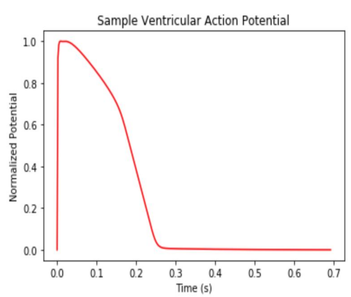
Introduction

The location of cardiac cells, or cardiomyocytes, can be identified as either atrial or ventricular by the shape of the action potentials they produce. Having such a classification of location happening in real time can be useful in a clinical setting to monitor a patient's heart health, particularly to identify errant heartbeats. Therefore, we sought to classify a set of action potentials as either atrial or ventricular, testing several different methods to see how accurately and quickly we could do so to be useful in hospitals settings.

Data Exploration and Preparation

- We obtained a dataset from our BME professor Dr. Benjamín Bejár containing samples of action potential data scaled to the same time window
- Upon initial screening of our dataset, we observed that we have 2000 different action potentials, with an even class split
- The output variable is a binary label: atrial or ventricular
- We divided the dataset into 400 training samples and 1600 testing samples.
- Sample atrial and ventricular action potentials are shown below.

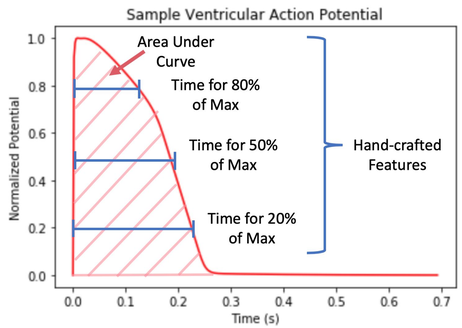


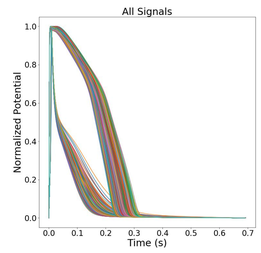


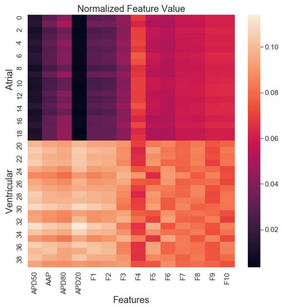
Materials and Models

We decided to test several different classifiers in order to find the best for both classification accuracy and time to complete classifying. We used the following classification models: **Logistic Regression, Support Vector Classification, K-NN, Naïve Bayes, Decision Tree, and Random Forest.** We compared the time elapsed to classify with the frequency of heart beat of a typical patient with heart problems to quantify the lag during real-time classification.

Feature Creation







We created 4 hand-crafted features and also looked at 10 Fourier features. We felt these would be most useful to classify the two classes of action potentials based on their differences in shape. The hand-crafted features were the area under the curve of the action potential as well as the times to decay to 80%, 20%, and 50% of the maximum value of the action potential. Features are normalized for centering and proper scaling.A comparison of the 14 features for the two classes is shown in the heatmap

Initial Results

Classifier	Accuracy	Time (sec)
Logistic Regression	1.0	0.0648
SVC	0.9989	0.0519
KNN	1.0	0.0189

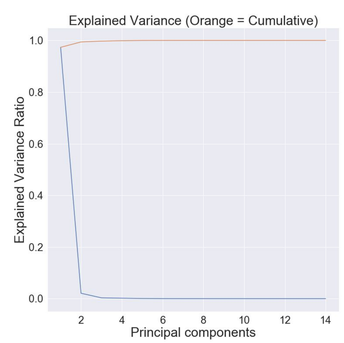
Classifier	Accuracy	Time (sec)
Naïve Bayes	1.0	0.0080
Decision Tree	1.0	0.0090
Random Forest	1.0	0.0330

To determine the optimal classification algorithm for our dataset, we ran six models and measured each of their accuracies in prediction.We found that they were all very accurate but that the Decision Tree was the fastest classifier.

PCA

The orange line demonstrates the aggregate score of the dataset based on each principal component. The score shoots up with a single component, and then levels off afterwards, demonstrating that using a single principal component can explain a large majority of the variance

As a result, we chose to modify our training and testing data to use just the first principal component.



After PCA

After running PCA, accuracy didn't significantly change. However, times for classification decreased for all 6 models.

Classifier	Accuracy	Time (sec)
Logistic Regression	1.0	0.0249
SVC	0.9989	0.0109
KNN	1.0	0.005

Classifier	Accuracy	Time (sec)
Naïve Bayes	1.0	0.003
Decision Tree	1.0	0.003
Random Forest	1.0	0.0289

Conclusion

Summary:
Among the 6 methods of classification that we performed on our dataset, we determined that all except SVC performed with perfect accuracy. These classification methods can be used to accurately classify an action potential as either atrial or ventricular. However, the Naive Bayes and Decision Tree methods were the quickest classifier for both the 14 original and 1 PCA features, and are therefore the best options.

Future Work:
We observed that while we originally began with 14 features, we only needed 1 PCA feature to characterize the action potentials. In future work, we would like to see how we can use this data to help detect changes in these signals which are predictive of cardiac arrest/failure. Using this, we hope that we can improve the administration of preventive treatment in clinical settings using ECG data which is already collected in the hospital.

References

[1] A. Nygren, C. Fiset, L. Firek, J. Clark, D. S Lindblad, R. Clark, and W. R Giles. Mathematical model of an adult human atrial cell : The role of k+ currents in repolarization. Circulation research, 82:63–81, 01 1998.

[2] T. O'Hara, L. Vir´ag, A. Varr´o, and Y. Rudy. Simulation of the undiseased human cardiac ventricular action potential: Model formulation and experimental validation. PLOS Computational Biology, 7(5):1–29, 05 2011.

Acknowledgements

We would like to thank Dr. Tamás Budavári as well as all the TAs for their guidance throughout this semester. We would also like to thank Drs. Rene Vidal and Benjamín Bejár for giving us data and background into data science as used in cardiology.