Machine Learning Diagnosis of Irregular Heartbeats

A cost-effective solution for arrhythmia detection

Jeffrey Tucker (UCM), Avash Adhikari (UCM), Xiyuan Wu (UCR), Nicole Collart (CWRU)

Explore More!

Lawrence Livermore National Laboratory

Abstract

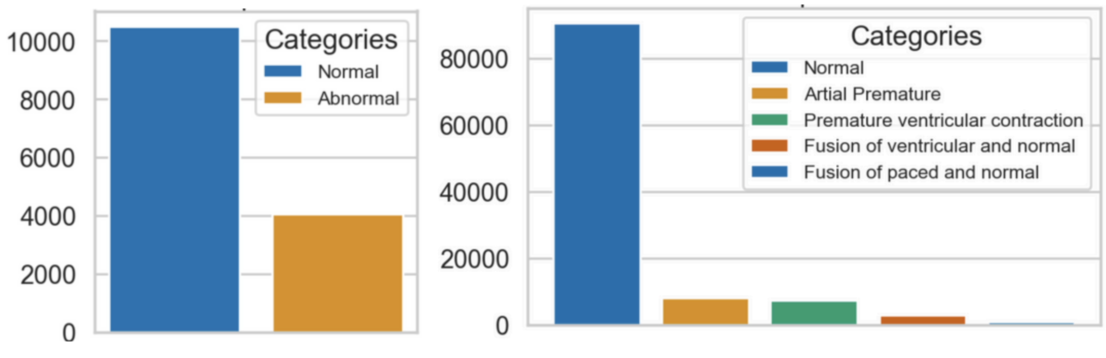
Arrhythmia is a medical condition characterized by irregular heartbeats, which, if left untreated, can have life-threatening effects. This project leverages machine learning to detect heart abnormalities using inexpensive electrocardiogram (ECG) data. Our model can classify ECG data as usual or one of four abnormal heartbeats with a false negative rate of only 0.87%. Additionally, our model's predictions for myocardial activation times are, on average, only 2 ms off. These results suggest that machine learning is a promising, cost-effective solution for arrhythmia detection.

Background

Heart disease is the leading cause of death in the United States. Arrhythmias are a result of underlying heart problems and classifying them can provide life-saving care. A common method of getting heartbeat data is a standard 12 lead ECG, where electric signals from the heart are measured. We aim to use these signals to predict when specific parts of the heart activate.

Data Exploration

Task 1&2: A class imbalance existed in the datasets. We decided to try using SMOTE and undersampling of the majority class to reduce possible bias.



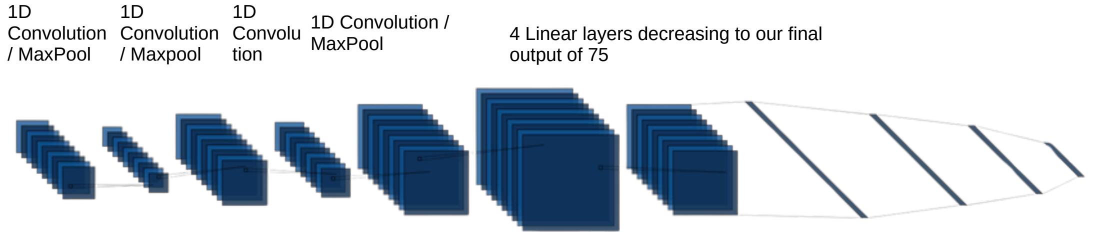
Class Imbalance for Task 1 (left) and Task 2 (right)

Task 3: A different dataset was used here, which contained the 10-lead ECG signals paired with transmembrane voltage data for each patient. The 10-lead ECG signal can be used to calculate the 12-lead ECG, and transmembrane voltage data is used to get the activation map.

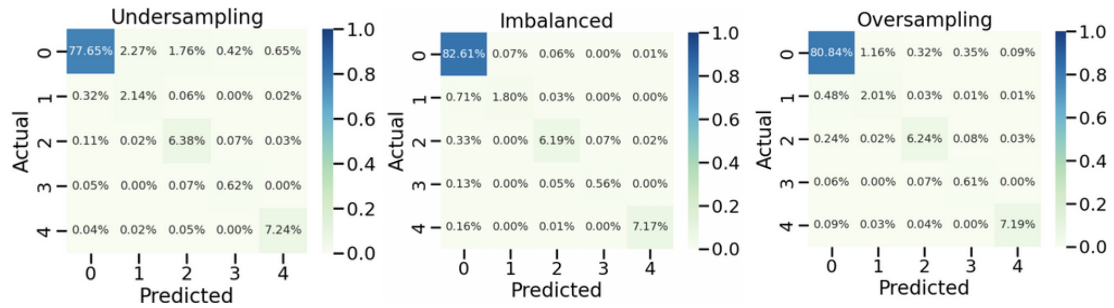
Methods

Task 1&2: We compared the results of various machine learning models (including decision tree, random forest, boosting, etc) and made confusion matrix heatmaps to visualize the accuracy of each model.

Task 3: here is the structure of our 1D Convolutional Neural Network
✓Kernel 5 / Stride 1 / Padding 2 for all Convolutional Layers
✓Stride / Kernel 2 for 1st two MaxPool | Stride/Kernel 4 for last Maxpool
✓Batch Normalization after every Convolutional Layers
✓40 Percent Dropout before our Linear layers
✓80 / 10 /10 Training/Testing/Validation Split
✓64 Batch Size ✓100 Epochs ✓Mean Squared Error Loss Function



Results



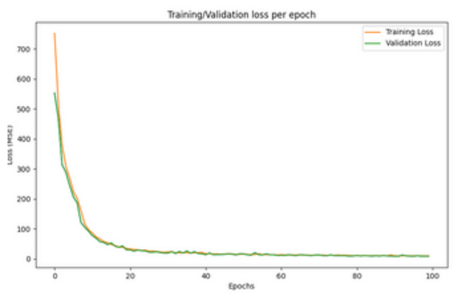
Confusion matrix for XGB classifier with undersampling (left), the original dataset (middle), and with SMOTE (right). Percentages are based on the total number of data points (18,853, 87,554, and 362,355 respectively)

Diagnosis	Imbalanced	Oversampling	Undersampling
Normal	0.9911	0.9831	0.9650
Atrial Premature	0.8137	0.6989	0.6110
Premature Ventricular Contraction	0.9549	0.9369	0.8550
Fusion of Ventricular and Normal	0.8133	0.6819	0.6683
Fusion of Paced and Normal	0.9859	0.9804	0.9471
Accuracy	**Imbalanced**	**Oversampling**	**Undersampling**
Overall Accuracy	0.9833	0.9689	0.9403

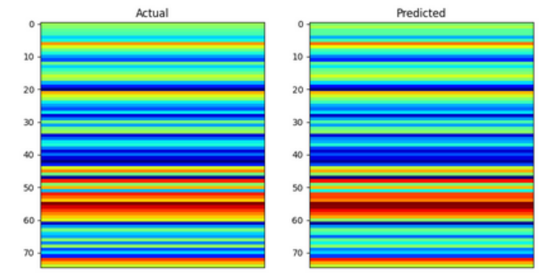
F1 scores and accuracy of the model trained on the original imbalance dataset, the oversampled dataset and undersampled dataset

For Task 1&2, The model with highest accuracy, precision, and recall was XGB classifier. Both oversampling and undersampling resulted in less false negatives. The accuracy with oversampling was 96.89%.

For Task 3, we had an MAE (average mean absolute error) of 2.3 ms as well as an MSE (average mean squared error) of 10.4 on our testing set.



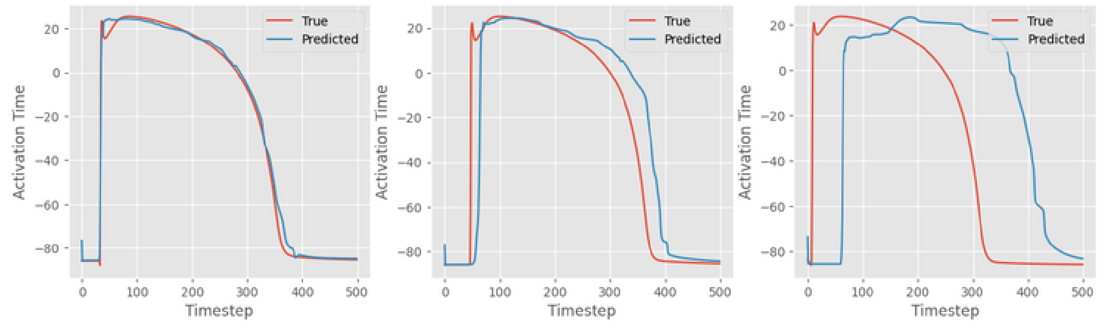
Training and validation loss as model progresses through epochs



An example of a true activation map versus an activation map our model predicted

Future Work

Task 4: Due to time constraints, we have yet to delve deeply into the LSTM model—this model with 2 layers: LSTM Layer and Linear Layer.
The image below displays 3 sample curves comparing actual versus predicted values. While some predictions align closely with the true curves, others do not perform as well. The model's best performance shows a loss of 7, indicating the predictions are on average 7 ms off from the true curve.



Three sample curves true vs. predicted

WHY LSTM? LSTMs capture long-term dependencies in time series data, which is crucial for accurately predicting entire sequences.

Conclusion

We utilized models from the sklearn and XGBoost libraries to achieve a 0.98 F1 score in classifying heartbeats as healthy or unhealthy. For subsequent tasks, we developed deep learning models to predict cardiac activation times, achieving a mean squared error of 10 and a mean absolute error of 2 milliseconds. However, further validation by medical professionals is essential to ascertain the clinical reliability of our predictions.

Jeffrey Tucker (jtucker7@ucmerced.edu), Avash Adhikari (aadhikari4@ucmerced.edu), Xiyuan Wu (xwu136@gmail.com), Nicole Collart (nlc62@case.edu)