Machine Learning Methods for the Detection of Heart Disease

Albert Ho[1], Diego Zubieta[2], Jessy Garcia[3], Abylay Amanbayev*,[1]

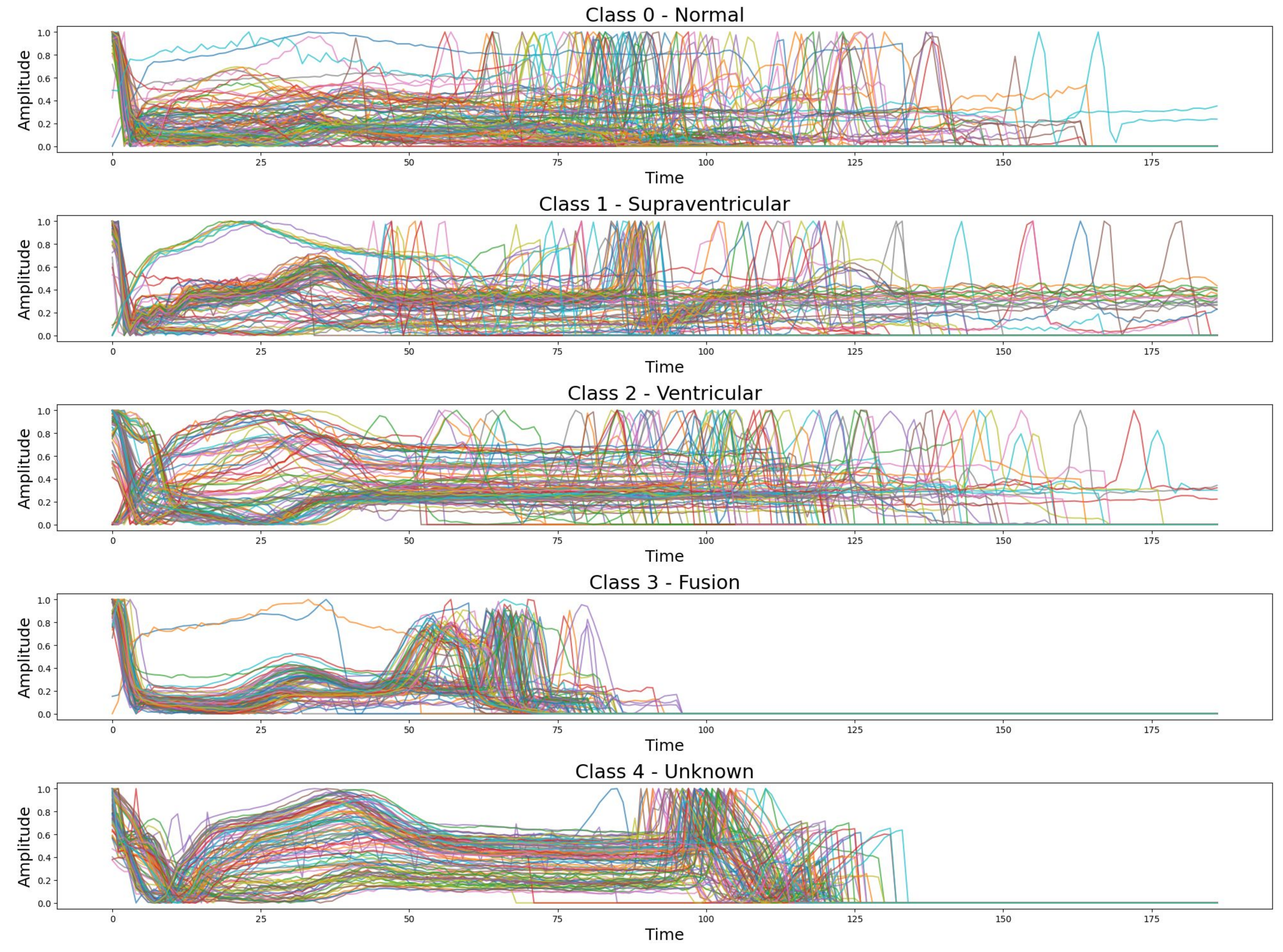
[1] University of California, Merced; [2] Case Western Reserve University; [3] University of California, Riverside; * Team Lead

ABSTRACT

Heart disease is the leading cause of death in the US, and early detection using accessible ECG data is crucial. This study employs classical machine learning techniques to classify heart conditions and 1D SqueezeNet neural network architectures to map the activation potential of the heart. By leveraging advanced machine learning and data science techniques, our model effectively predicts heart disease, providing a framework for future research and clinical applications. The results demonstrate the potential for improving diagnostic accuracy and patient outcomes using automated ECG analysis.

MOTIVATION

- Heart disease is the leading cause of death in the US
- Decrease costs and minimizing time for detection
- Minimize incorrect (false negative) detections, possibly fatal



Distribution of heartbeat sample data used in classification modeling.
Forty samples for each group

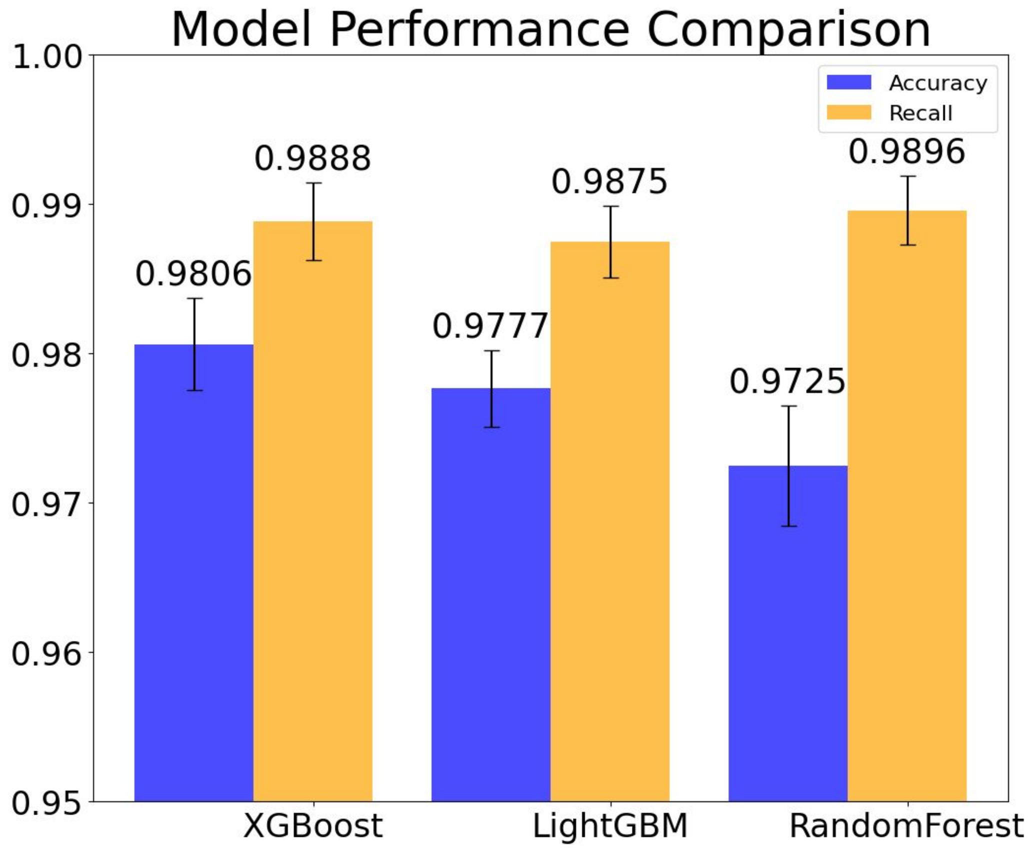
APPROACH

Models: XGBoost, LightGBM, Random Forest, CNNs (1D Squeeze Net)
Sampling: Oversampling (SMOTE), Undersampling
Optimizing: Grid Search (adjusting hyperparameters)

RESULTS

Confusion Matrix for Best Model: XGBoost

Actual \ Predicted	1	0
1	2083	15
0	28	785

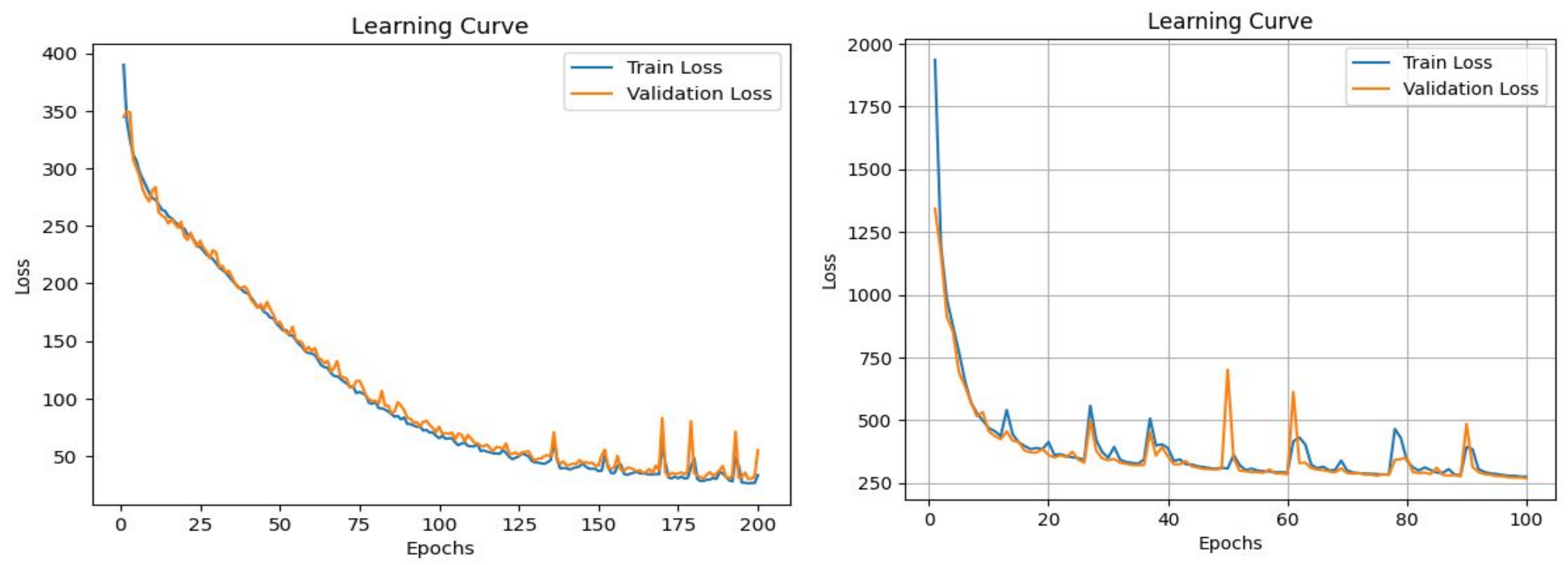


Task 1: Classification of normal and abnormal heartbeats. Optimized XGBoost Confusion Matrix after Grid Search (left) and comparison of optimized models measuring accuracy and recall results (right).

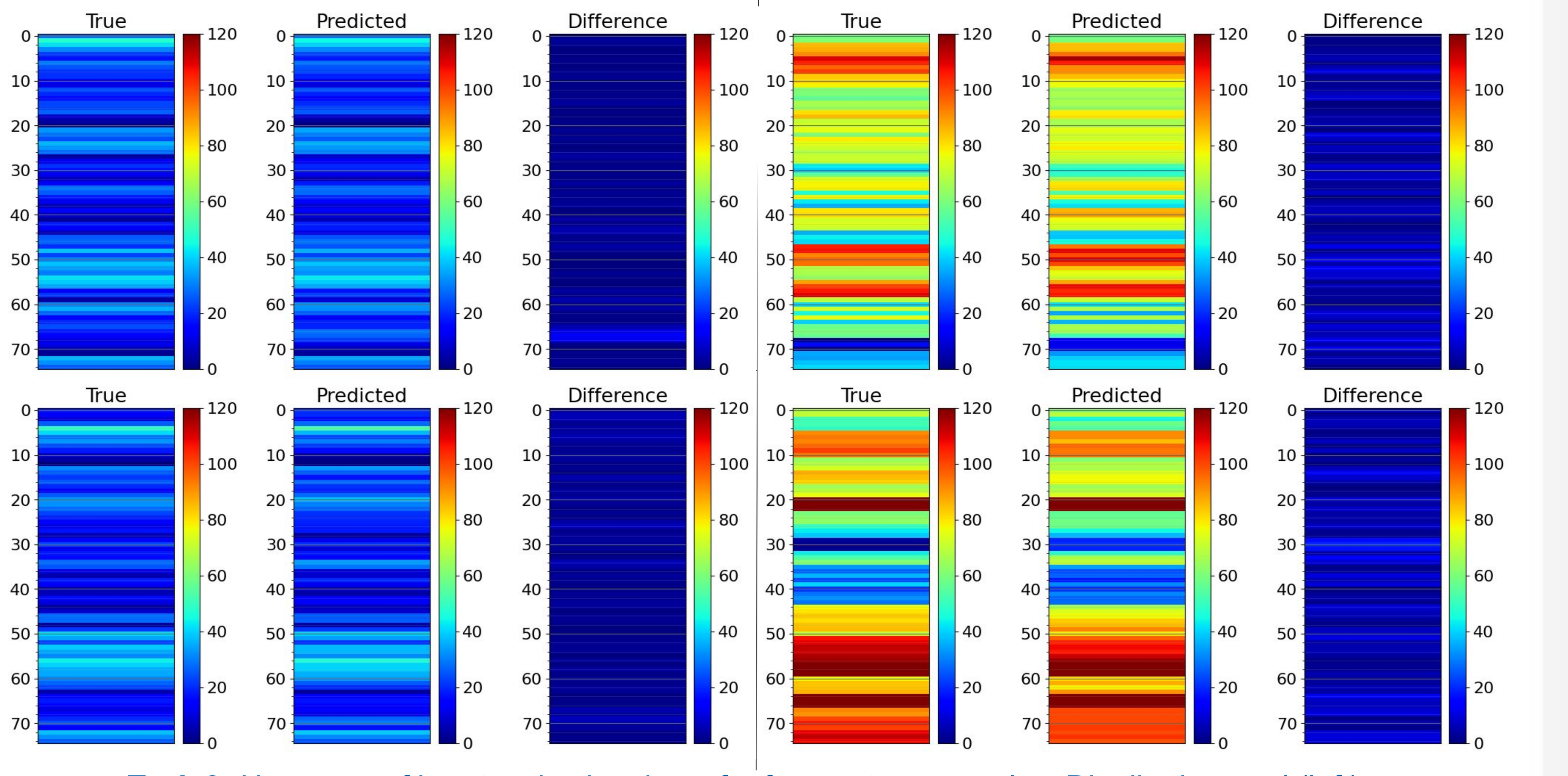
Confusion Matrix (Percentage)

Actual \ Predicted	N	S	V	F	U
N	99.27	0.37	0.20	0.09	0.07
S	20.86	77.34	1.26	0.00	0.54
V	3.11	0.21	95.30	1.04	0.35
F	11.73	0.00	7.41	80.86	0.00
U	1.68	0.12	0.19	0.00	98.01

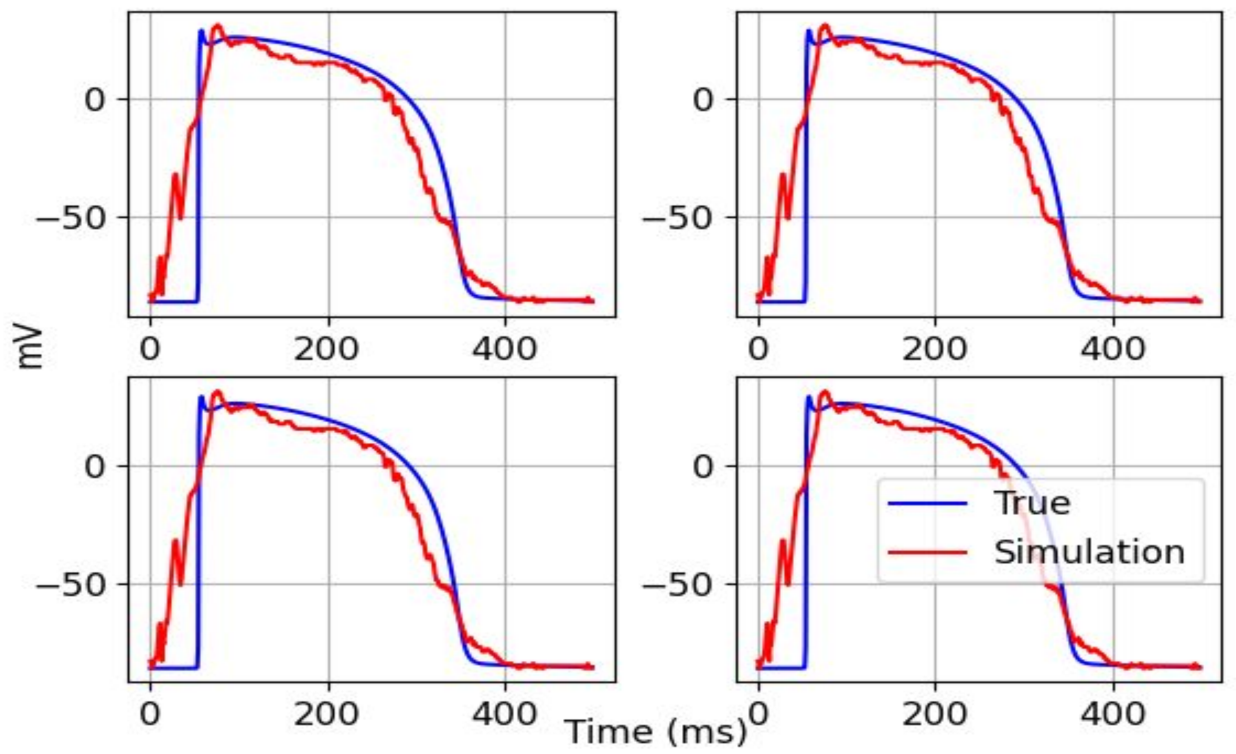
Task 2: Confusion matrix for multiclass classification of heartbeats impacted by arrhythmias and myocardial infarction using an XGBoost model.
Legend - N: Normal, S: Supraventricular, V: Ventricular, F: Fusion, U: Unknown



Learning curves of SqueezeNet for Activation Map Reconstruction (left), and Transmembrane Potential Reconstruction (right)



Task 3: Heat map of heart activation times for four test set samples. Distribution: real (left), predicted with SqueezeNet (middle), and absolute difference (right)



Task 4: Transmembrane Potentials (blue) and their Reconstructions (red) using SqueezeNet

Developing Machine Learning Classification and Neural Network Architecture can Lead the Future in Disease Detection.

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.