DATA SCIENCE CHALLENGE 2023

CARDIAC ELECTROPHYSIOLOGY

Zhexing Li (UCR), Achyuth Kolluru (UCM), Arial Tolentino (UCR), and Andrii Dovhaniuk (UCR)

UC RIVERSIDE UC MERCED Lawrence Livermore National Laboratory

Harnessing electrocardiogram (ECG) technology, our project attempts to use advanced machine learning (ML) techniques and deep learning (DL) methodologies for better cardiac diagnostics. Through the implementation of algorithms such as K-Nearest Neighbors (KNN), Random Forest, Decision Trees, XGBoost, Convolutional Neural Networks (CNN) etc., we have developed models that can pinpoint subtle cardiac anomalies often overlooked by conventional analyses. Our project aims to improve cardiac diagnostics, combining computation and medicine for faster and more accurate detection of heart abnormalities, potentially revolutionizing therapeutic interventions.

HEARTBEAT CLASSIFICATION

Healthy Heartbeat vs Irregular Heartbeat

- Dataset divided into 60% training, 20% testing, and 20% validation (10506 normal, 4046 abnormal cases)
- Used two models to perform binary classification
  - Logistic Regression:
    - Accuracy: 81%, Precision: 85%, Recall: 91%, F1: 88%
  - Gradient Boosting Classifier:
    - Accuracy: 97%, Precision: 97%, Recall: 99%, F1: 98%
    - This achieved the best accuracy among all models and we show the classification result in Figure 1 in the form of a confusion matrix.

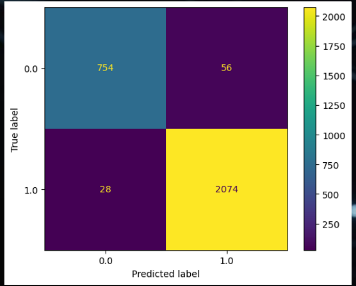

Figure 1: Confusion Matrix of Gradient Boosting Classifier.

LinkedIn

AT ZL AK AD

MULTI-CLASS HEARTBEAT CLASSIFICATION

Diagnosing Irregular Heartbeats

- Five classes: 0: normal (18118 cases), 1: Atrial Premature (556 cases), 2: Premature Ventricular Contraction (1448 cases), 3: Fusion of Ventricular and Normal (162 cases), 4: Fusion of Paced and Normal (1608 cases)
- Dataset divided into 80% training, 20% testing

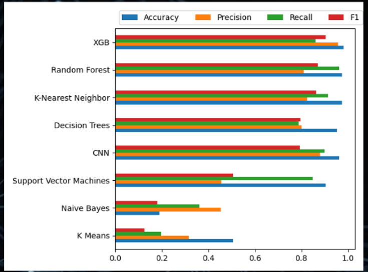

Figure 2: Classification scores of different classifiers sorted in descending order of F1 score based only on the based model.

We select some of the models that performed the best on the base model and further tune it with grid search, where the training subset further divided for 10-fold cross validation.

Final model comparison:

Classifier	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)
KNN	97	89	82	85
CNN	97	90	80	84
Decision Tree	96	81	82	81

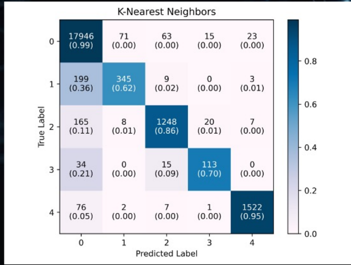

Figure 3: Confusion matrix for multiclass heartbeat classification with K-Nearest Neighbors, the overall best model.

ACTIVATION MAP RECONSTRUCTION

Sequence-to-vector prediction

- Reconstruct a cardiac activation map (75,1) containing activation times of 75 points within the myocardium from the 12-lead ECG signal (12, 500).
- Network architecture inspired by the AlexNet, achieve a final R^2 score of 95%.

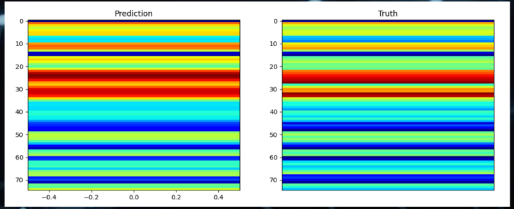

Figure 4: Activation times of the 75 points within the myocardium. Bluer and redder color represent shorter and longer activation time, respectively.

TRANSMEMBRANE POTENTIAL RECONSTRUCTION

Sequence-to-sequence prediction

- Reconstruct a cardiac activation map (75, 500) containing transmembrane voltages of 75 points within the myocardium from the 12-lead ECG signal (12, 500).
- Utilized a similar architecture as above and achieved an R^2 score of 94%.

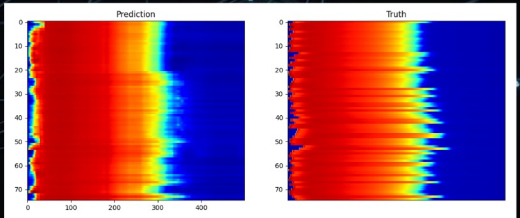

Figure 5: Activation map of transmembrane voltages within the myocardium. X axis is time (ms) and Y axis is voltages (mV). Bluer and redder color represent lower and higher voltage levels, respectively.

TAKEAWAYS

ML and DL are transforming the landscape of cardiac diagnostics, providing us with tools to interpret complex patterns in heart rhythm data with unprecedented accuracy. Algorithms we implemented allow us to analyze ECG readings to detect subtle signs of heart disease that might be overlooked in traditional analyzes. thereby opening doors to timely and effective treatments.

** Some of the text on this poster are generated by GPT-4.