



LLNL Data Science Challenge 2023: Hearts with Machine Learning

J. Pastore¹, I. Romero³, H. Rath⁴, A. Kaattari-Lim²

¹Dept. of Psychology, ²Dept. of Engineering UC Riverside,

³Dept. of Applied Math, ⁴Dept. of Engineering UC Merced



Task 1 – Binary Classification of healthy and irregular ECG Data

Description: ECG Heartbeat Categorization by performing binary classification for healthy heartbeat vs. irregular heartbeat

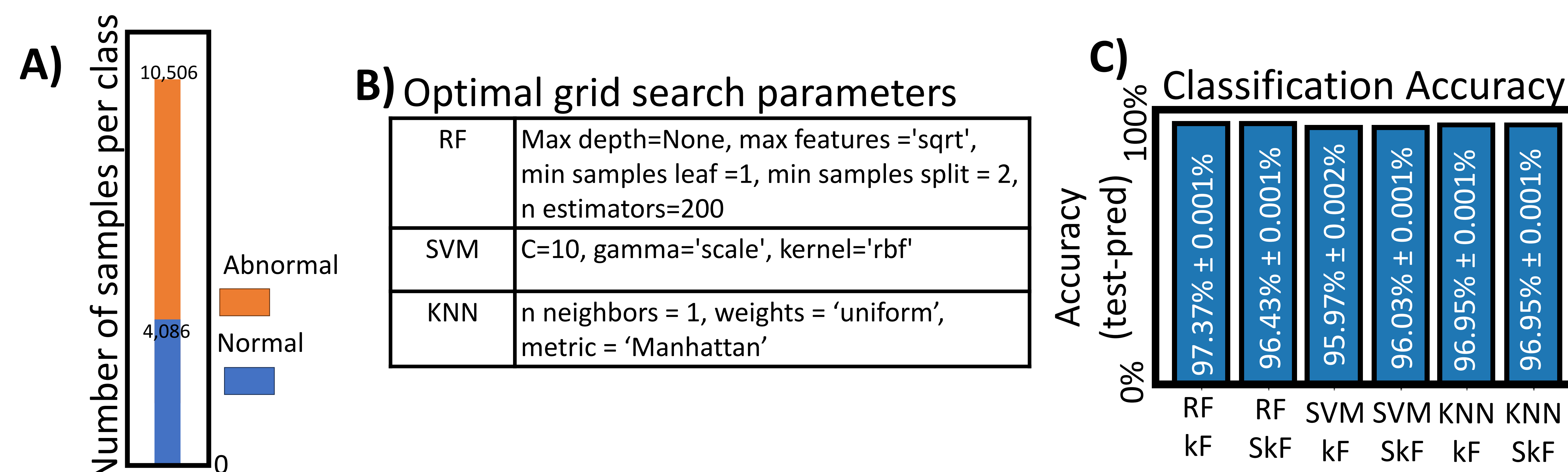


Figure 1. **A)** Number of samples per class, abnormal ECG (orange) vs normal ECG (blue). **B)** Parameters chosen from a grid search of random forest (RF), Support Vector Machines (SVM), and K-Nearest Neighbors (KNN). **C)** Classification accuracy for RF, SVM, and KNN with k-fold cross validation (kf) and Stratified k-fold cross-validation (SkF). Each test was run with 10 folds. Error bars represent SEM.

Task 2 – Multi-Class Classification of healthy and irregular ECG Data

Description: By using the ECG Heartbeat Categorization Dataset, we performed multiclass classification to diagnose the various ECG recordings.

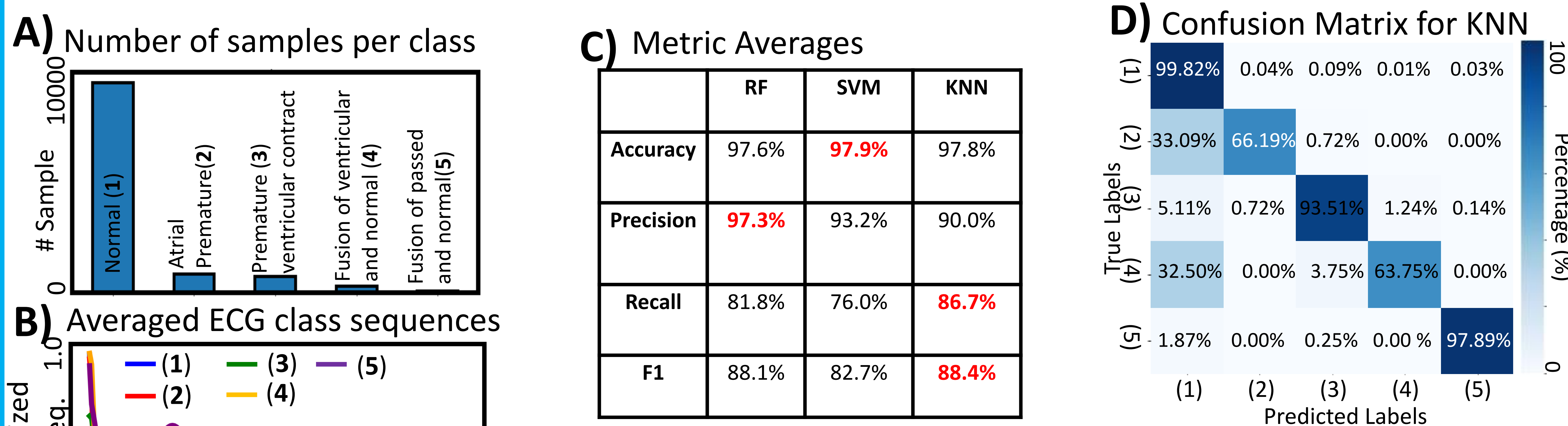


Figure2. **A)** Number of samples per class. **B)** Average ECG signal for each class across patients. **C)** Classification accuracy for Random Forest (RF), Support Vector Machine (SVM), and k-Nearest Neighbors (KNN) with stratified k-fold cross validation (Skf, k= 10). Precision: tp / (tp + fp), Recall: tp / (tp + fn), F1: weighted harmonic mean of the precision and recall. **D)** Confusion matrix for KNN with an average stratified k-fold cross-validation accuracy of 97.8%. Error bars represent SEM.

Conclusion / Take away

- Machine learning classification of binary and multiclass datasets is highly valuable as it enables automated decision-making processes. These models can accurately categorize data into different classes, aiding in tasks such as medical diagnosis of various heart conditions.
- The utilization of 1D convolutional neural networks (CNNs) for classification and reconstruction of training and testing data proves highly effective in tasks involving time-series data. With their ability to capture temporal patterns, the 1Ds CNNs allowed for precise reconstruction of the activation sequences for 75 virtual locations within the heart given ECG recordings from 12-leads.

Task 3– Reconstruction activation map given ECG leads

Description: Sequence-to-vector prediction of voltage recordings and ECG Signals to perform activation map reconstruction (i.e. transform 12x500 standard leads recordings to a 75x1 activation map using a neural network)

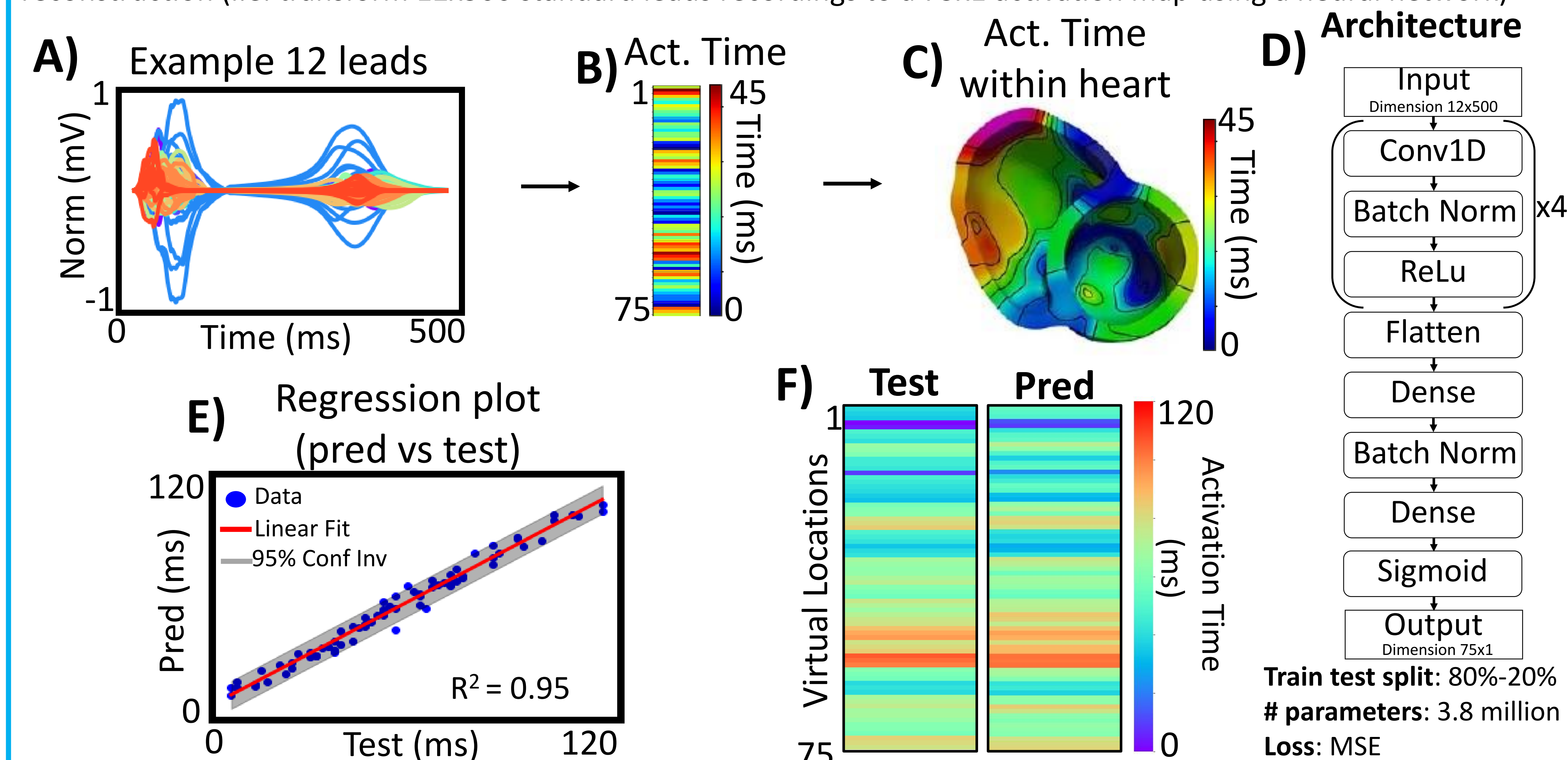


Figure 3. **A-C)** Example with 1 patient. **A)** Example of 12 ECG leads in mV. **B)** Example activation time of all 75 virtual locations within the heart. **C)** Reconstruction of virtual locations within the heart and corresponding activation times. **D)** Network architecture. **E)** Linear regression fit of predicted vs test data activation times. **F)** Test and predicted activation times in milliseconds (ms).

Task 4 – Reconstruction of transmembrane potential given ECG leads

Description: Sequence-to-sequence prediction using voltage recordings and ECG Signals to perform transmembrane potential (TP) reconstruction (i.e transform 12x500 to 75x500 using a neural network)

*Note only trained and tested on 5000 of 16000 patients due to hardware constraints

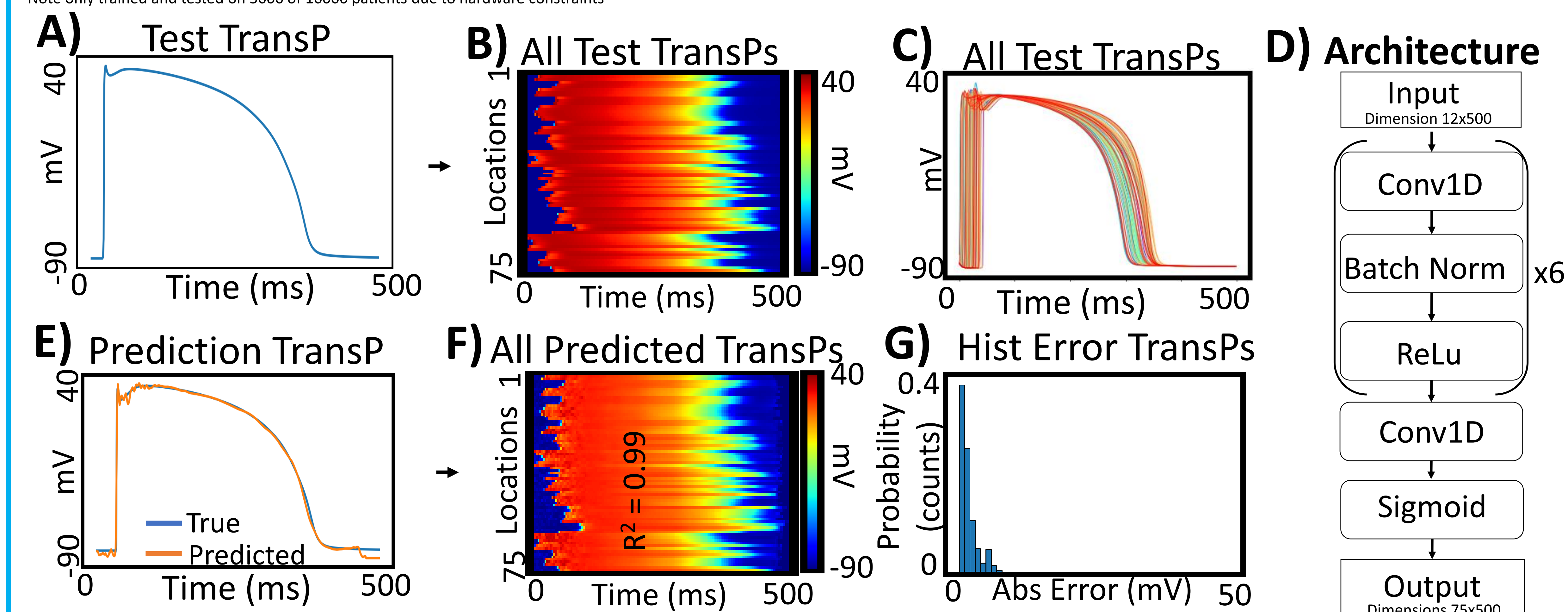


Figure 4. **A)** Example transmembrane potential (TransP). **B)** Heatmap of all TPs over all test 75 virtual locations. **C)** All TPs for all 75 virtual locations. **D)** Neural network architecture. **E)** True vs predicted reconstruction of TPs. **F)** Reconstruction for all 75 virtual locations for 1 patient, with an average $R^2 = 0.985$ **G)** Histogram of absolute error (test – prediction) for all 75 reconstruction virtual locations (mV).