CSC 522 – Spring 2020 Automated Learning and Data Analysis

Arrhythmia Prediction and Diagnosis using Data Analysis

Presented By:
Project Group: P21
Mangalnathan Vijayagopal – mvijaya2
Nishcal Badarinath Kashyap – nkashya
Pawandeep Mendiratta – psmendir
Shreyas Chikkballapur Muralidhara - schikkb

Index:

- Introduction and Related Work
- Dataset
- Pre-Processing
- Feature Selection
- Support Vector Machine
- K – Nearest Neighbors
- 1D - Convoluted Neural Network
- Multi Layer Perceptron
- Results
- Conclusion
- References

Introduction

- To detect and predict the type of arrhythmia based on Electrocardiogram (ECG) tool using machine learning models and algorithms.

Related Work

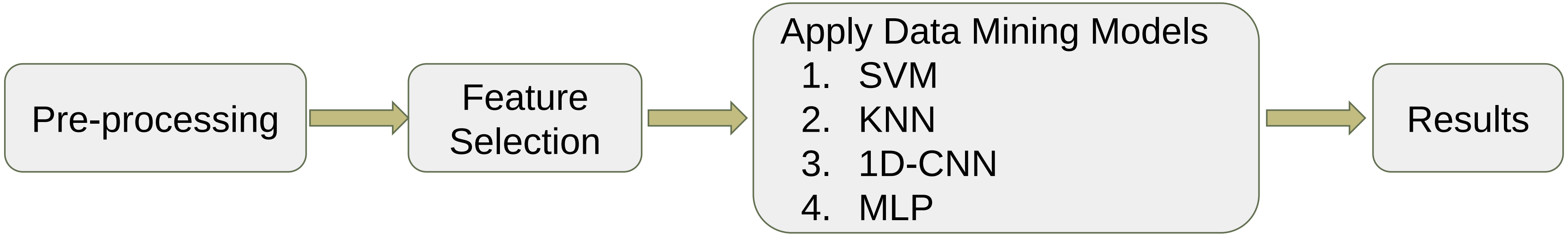
- Automated Screening of Arrhythmia Using Wavelet Based Machine Learning Techniques
- Machine Intelligent Diagnosis of ECG for Arrhythmia Classification Using DWT, ICA and SVM techniques
- An integrated ECG feature extraction scheme using PCA and wavelet transform
- Heart rate dynamics distinguish among atrial fibrillation, normal sinus rhythm and sinus rhythm with frequent ectopy

Data Set

http://archive.ics.uci.edu/ml/datasets/Arrhythmia

- Arrhythmia Data Set – Referenced from UCI Machine Learning Repository
- The data set contains 452 instances with 279 attributes
- The data set represents tabular form of ECG data of 452 patients along with patient information
- Multivariate dataset with Categorical, Integer and Real Attributes
- Class Label values range from 1 to 16 (Types of Arrhythmia)
- A train-test split in ratio 70:30 into training and testing data containing 316 and 116 records respectively.

Process Flow

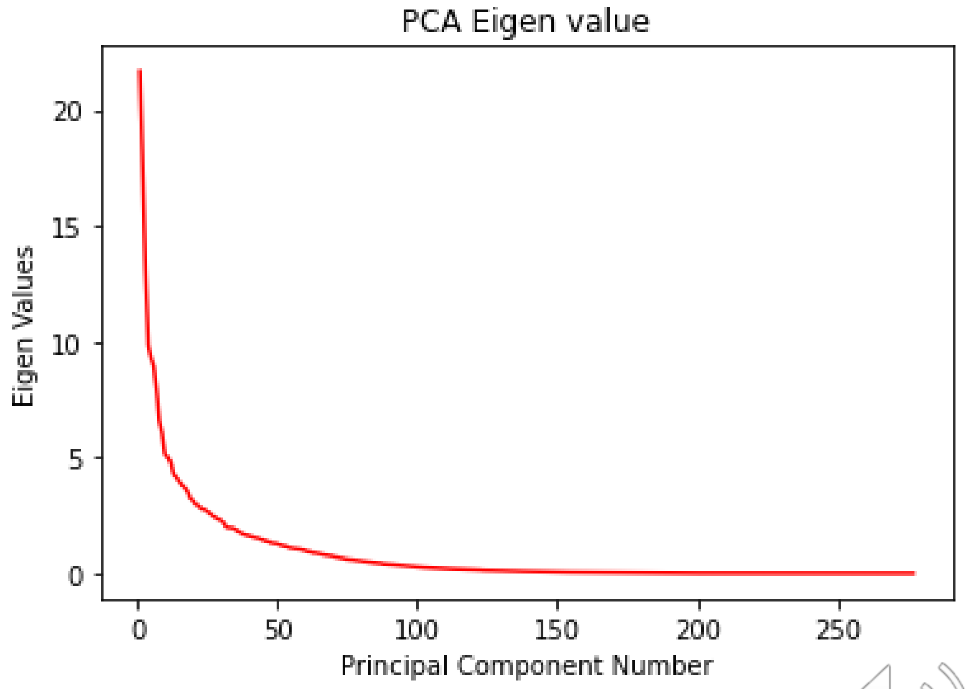


Pre-Processing

- Remove Unwanted Columns
- Replace Missing Values
- Attribute Scaling – Normalize values to the range of 0 to 1

Feature Selection

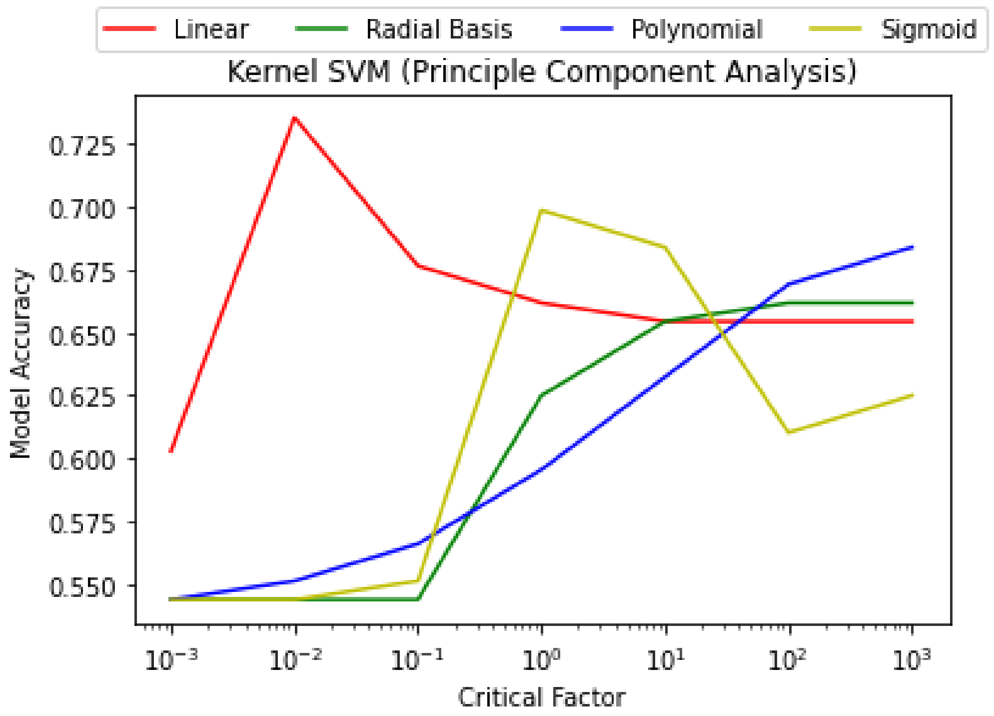
- Principal Component Analysis
  Top 88 features were chosen based on PCA output

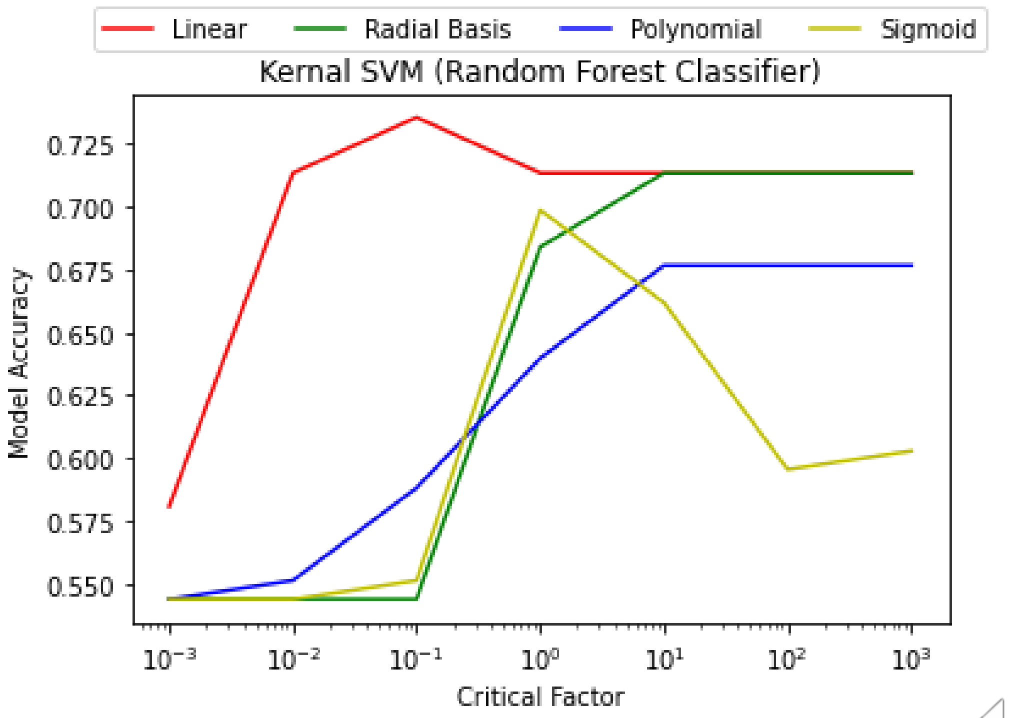


- Random Forest
  Top 99 features were chosen based on Random Forest output

Support Vector Machines

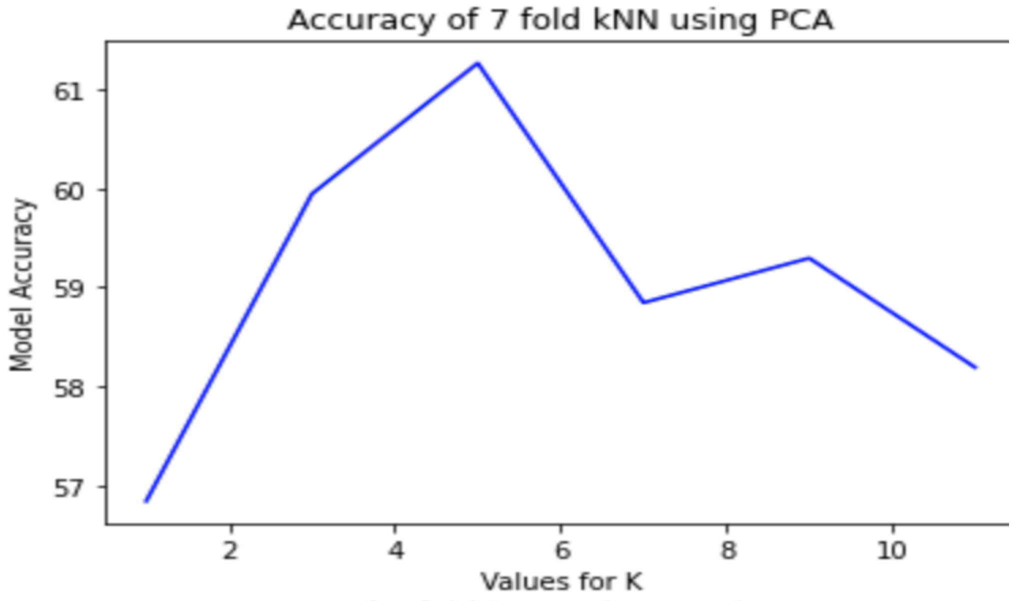
- For SVM , we trained the model with different kernel functions: Linear, Polynomial, Gaussian RBF and Sigmoid kernel function.
- Linear Kernel function gave us the best performance of 73.53 %

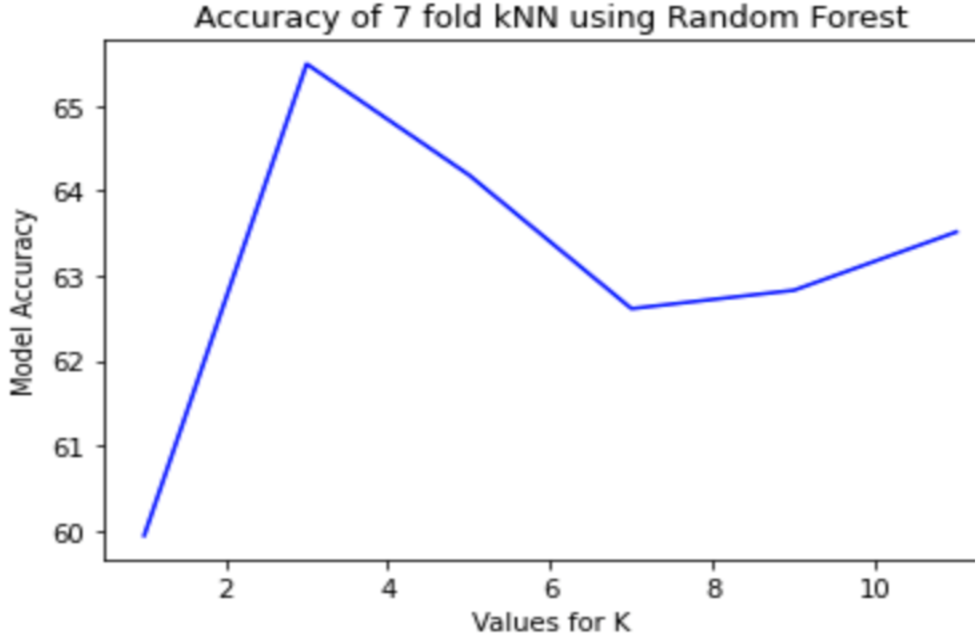




K Fold K Nearest Neighbor

Principal Component Analysis	
Hyperparameters	
Folds	7
Accuracy(5 NN)	61.26
Random Forest Classifier	
Hyperparameters	
Folds	7
Accuracy(3 NN)	65.51



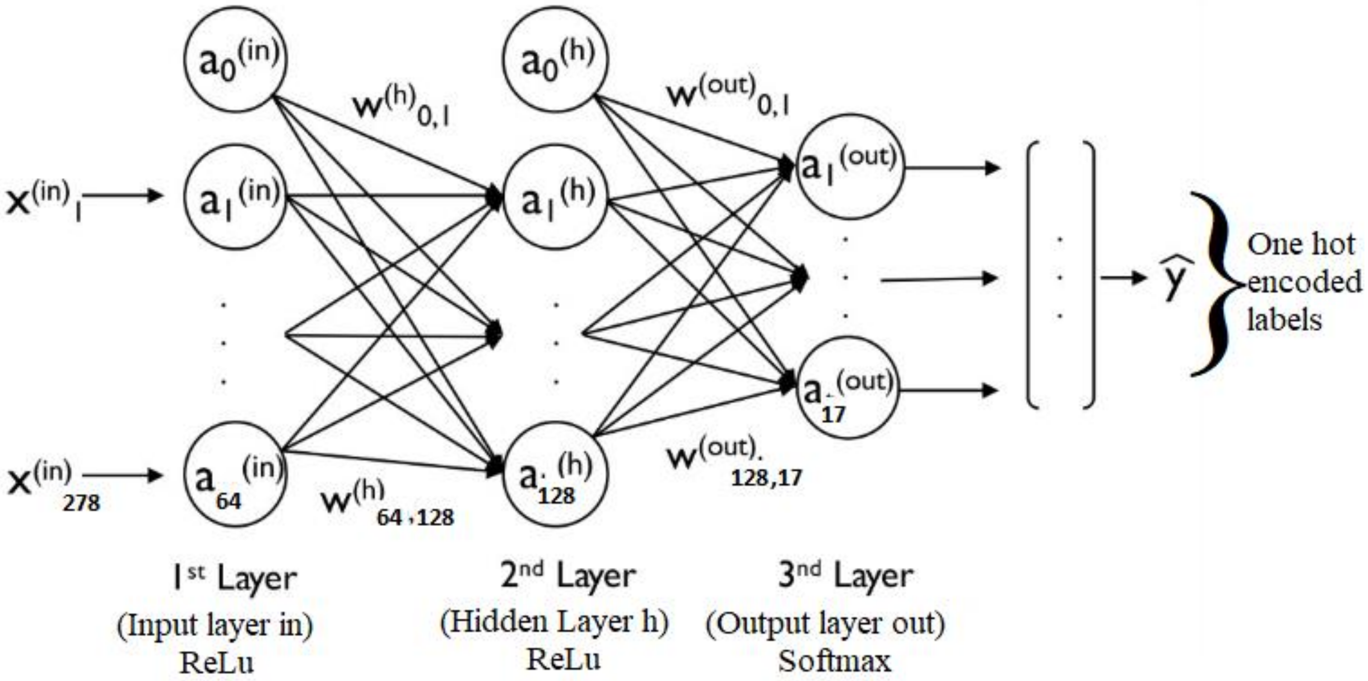


Multi Layer Perceptron - Method

- Feed forward Neural network with 2 fully connected layers with ReLU activation. Predictions generated from Final Softmax layer.

- Weighted Loss Function for compensating imbalanced class.

- Hyper-parameters:
  - Optimizer – Adam
  - Learning Rate – 0.01
  - Batch size – 8
  - Epoch - 30
  - Neurons per layer – [64, 128, 17]

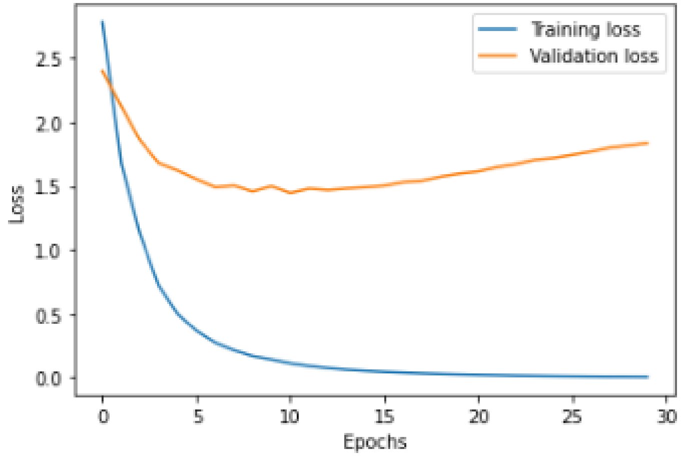


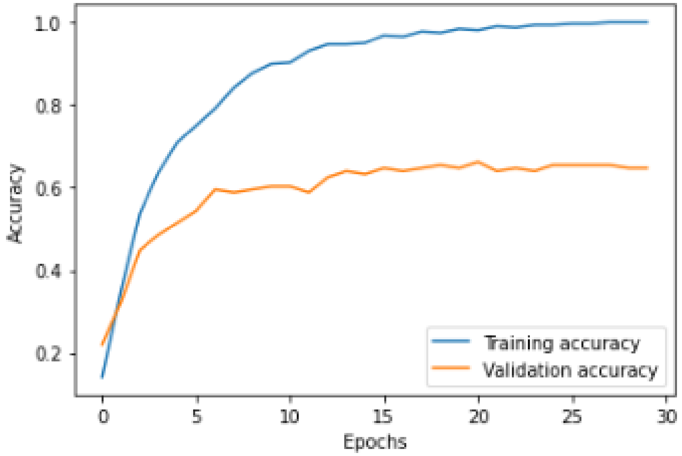
Multi Layer Perceptron - Result

- Test Accuracy for model is 66.91%, while the validation accuracy is 68.63%

Classification Report for model

	precision	recall	f1-score	support
1	0.74	0.85	0.79	74
2	0.40	0.46	0.43	13
3	0.80	1.00	0.89	4
4	0.75	0.75	0.75	4
5	0.67	0.50	0.57	4
6	0.43	0.38	0.40	8
7	0.00	0.00	0.00	1
8	0.00	0.00	0.00	1
9	1.00	0.67	0.80	3
10	0.75	0.40	0.52	15
14	1.00	1.00	1.00	1
15	0.50	1.00	0.67	1
16	0.00	0.00	0.00	7
accuracy			0.67	136
macro avg	0.54	0.54	0.52	136
weighted avg	0.65	0.67	0.65	136



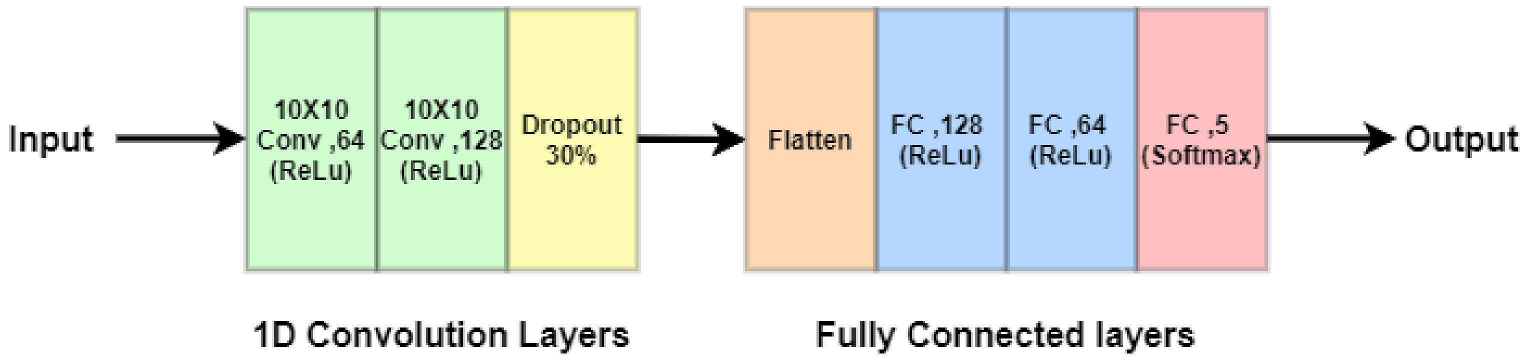


1D – Convolutional Neural Network - Method

- Model is divided into 2 blocks
  - Convolution block – 1D Convolution layers and Dropout.
  - Fully Connected block – Extract features required for class prediction.

- Compiling model with Weighted Loss Function to penalize misclassification of imbalanced classes.

- Hyper-parameters:
  - Optimizer – Adam
  - Learning Rate – 0.01
  - Batch size – 16
  - Epoch - 20



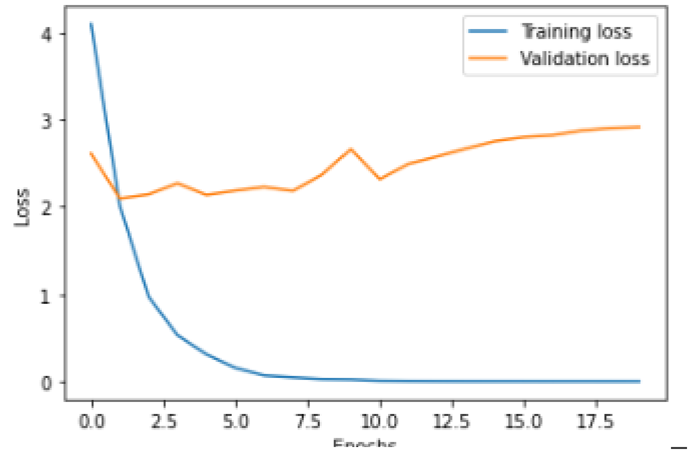
1D – Convolutional Neural Network - Result

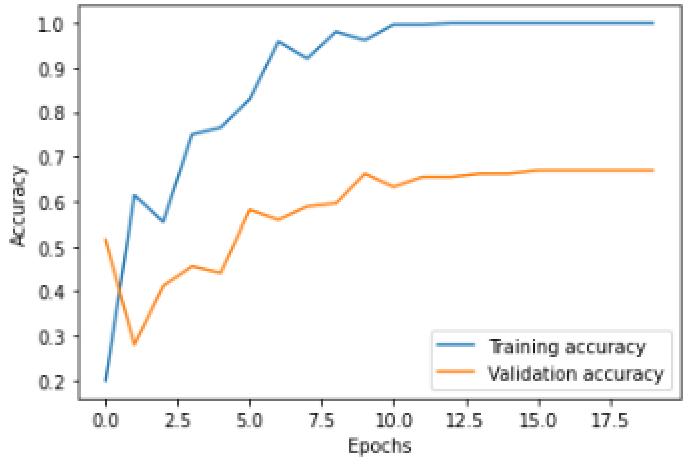
- Test Accuracy for model is 69.85%, while the validation accuracy is 70.63%

```
Classification Report for model
              precision    recall  f1-score   support

           1       0.70      0.92      0.80        74
           2       0.54      0.54      0.54        13
           3       0.80      1.00      0.89         4
           4       1.00      0.75      0.86         4
           5       0.50      0.25      0.33         4
           6       0.67      0.25      0.36         8
           7       0.00      0.00      0.00         1
           8       0.00      0.00      0.00         1
           9       1.00      0.67      0.80         3
          10       0.80      0.53      0.64        15
          14       0.00      0.00      0.00         1
          15       0.00      0.00      0.00         1
          16       0.00      0.00      0.00         7

    accuracy                           0.70       136
   macro avg       0.46      0.38      0.40       136
weighted avg       0.65      0.70      0.66       136
```





Discussion

- For SVM, the data is linearly separated. From our work, we conclude that linear kernel works best for the data set by comparing accuracy scores with other multi-dimensional kernels.
- For KNN, this was a novel approach which was not implemented in prior related work.
- Random Forest is equally good as PCA for feature selection.
- None of the related work referenced has implemented 1D-CNN and MLP.

References

[1] A. Das, F. Catthoor and S. Schaafsma, "Heartbeat Classification in Wearables Using Multi-layer Perceptron and Time-Frequency Joint Distribution of ECG," 2018 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE), Washington, DC, USA, 2018, pp. 69-74.\newline

[2] N. Kalkstein, Y. Kinar, M. Na'aman, N. Neumark and P. Akiva, "Using machine learning to detect problems in ECG data collection," 2011 Computing in Cardiology, Hangzhou, 2011, pp. 437-440. \newline

[3]Martis, R.J., Krishnan, M.M.R., Chakraborty, C. et al. Automated Screening of Arrhythmia Using Wavelet Based Machine Learning Techniques. J Med Syst 36, 677–688 (2012). https://doi-org.prox.lib.ncsu.edu/10.1007/s10916-010-9535-7

[4] U. Desai, R. J. Martis, C. G. Nayak, Sarika K. and G. Seshikala, "Machine intelligent diagnosis of ECG for arrhythmia classification using DWT, ICA and SVM techniques," 2015 Annual IEEE India Conference (INDICON), New Delhi, 2015

[5]R. J. Martis, C. Chakraborty and A. K. Ray, "An Integrated ECG Feature Extraction Scheme Using PCA and Wavelet Transform," 2009 Annual IEEE India Conference, Gujarat, 2009, pp. 1-4.

THANK YOU!