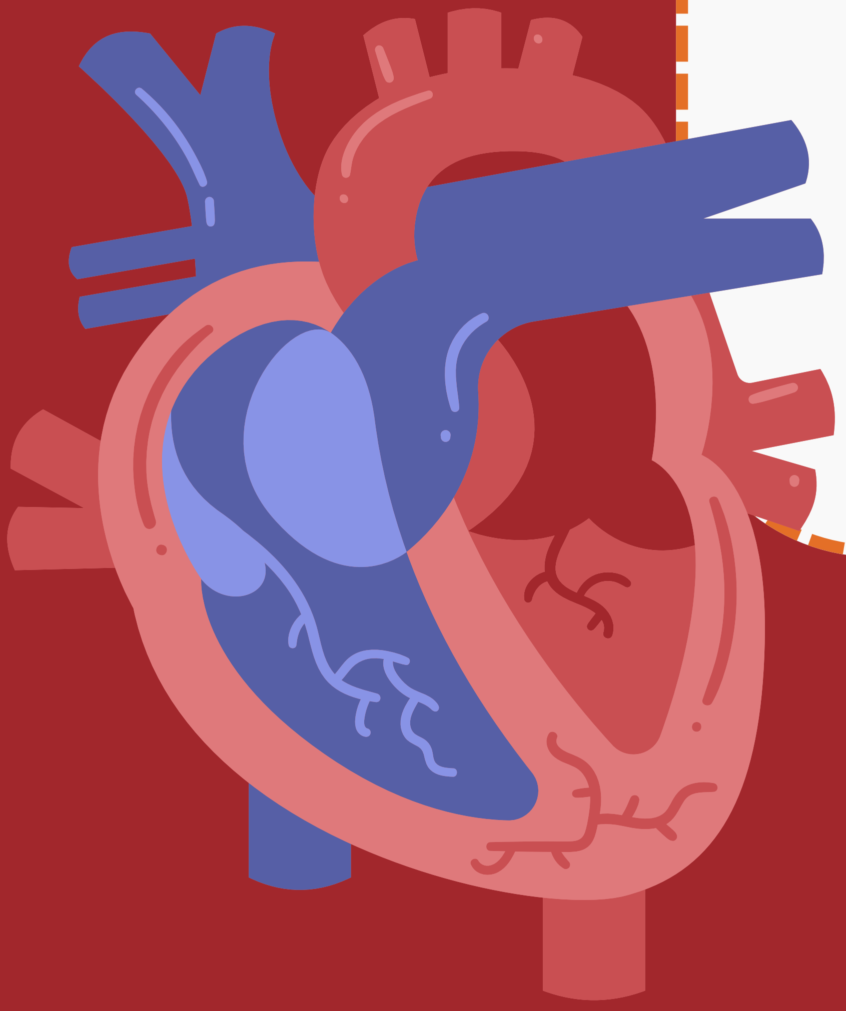
Detection of Heart Abnormalities and Disease

Presented by Group-16

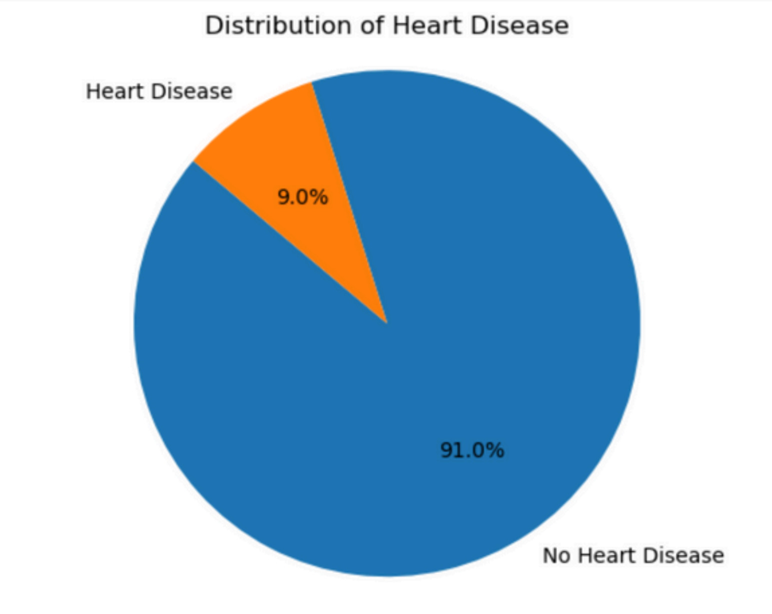
Ilham Ghori, Juhanah Madhiyyah, Kanana Muchiri, Raiqah Shameer, Sherif Fares



Feature selection and EDA

Dataset Overview

- Heart Disease dataset with 301,717 samples and 18 features.
- Balanced dataset using Random Undersampling (27,261 instances per class).



Feature Selection

Encoding:

- One-Hot Encoding applied to categorical features.
- K-1 Dummy Encoding used for binary features to reduce dimensionality.

Correlation Analysis:

- Point-Biserial Correlation for continuous features (e.g., BMI, PhysicalHealth).
- Cramer's V for categorical features (e.g., Smoking, Diabetic).
- Features ranked by correlation with the target variable (HeartDisease).

Feature Sets:

- Created datasets with Top 10, Top 20, and All Features for experimentation.
- We saved separate datasets for original and under sampled data.

Key Findings

Top Correlated Features with Target Class:

- GenHealth (0.38), DiffWalking (0.28), Diabetic (0.25), PhysicalHealth (0.23), Stroke (0.23).

Image Pre-Processing

Image Dataset Overview

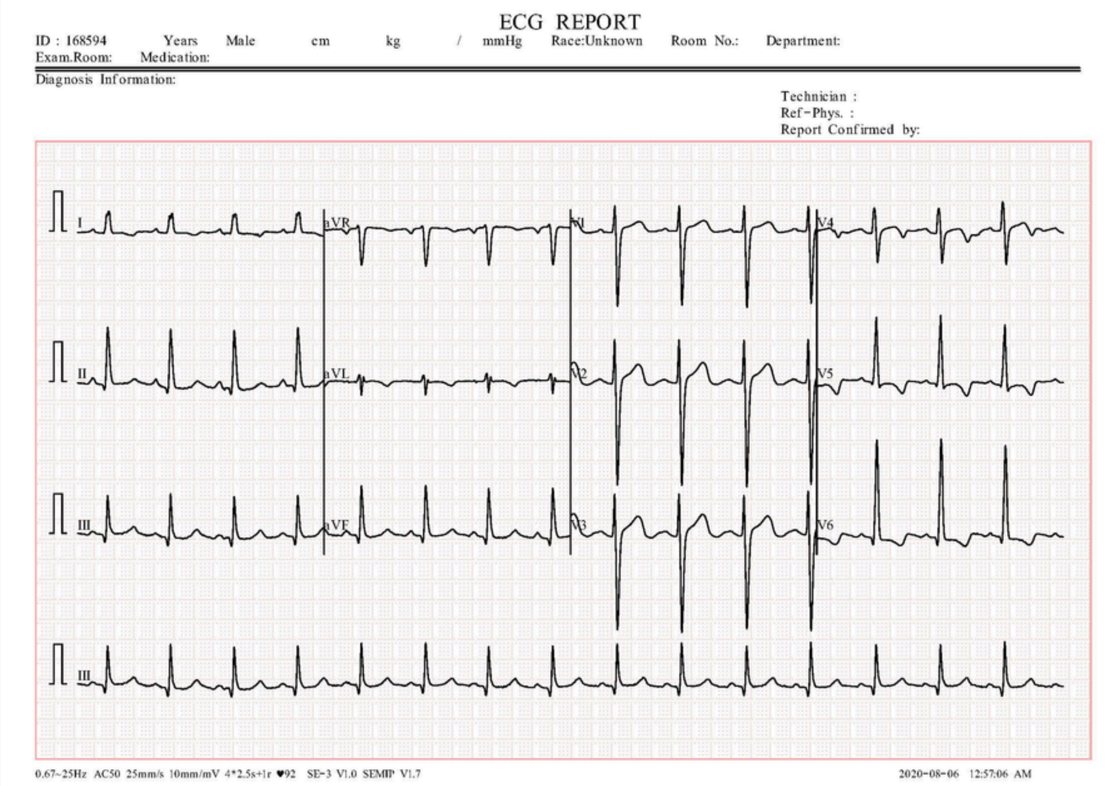
- ·Abnormal Heartbeat (HB): 240
- ·Myocardial Infraction (MI): 233
- ·Normal Patient: 172
- ·Patients with a history of Myocardial Infraction (PMI): 284

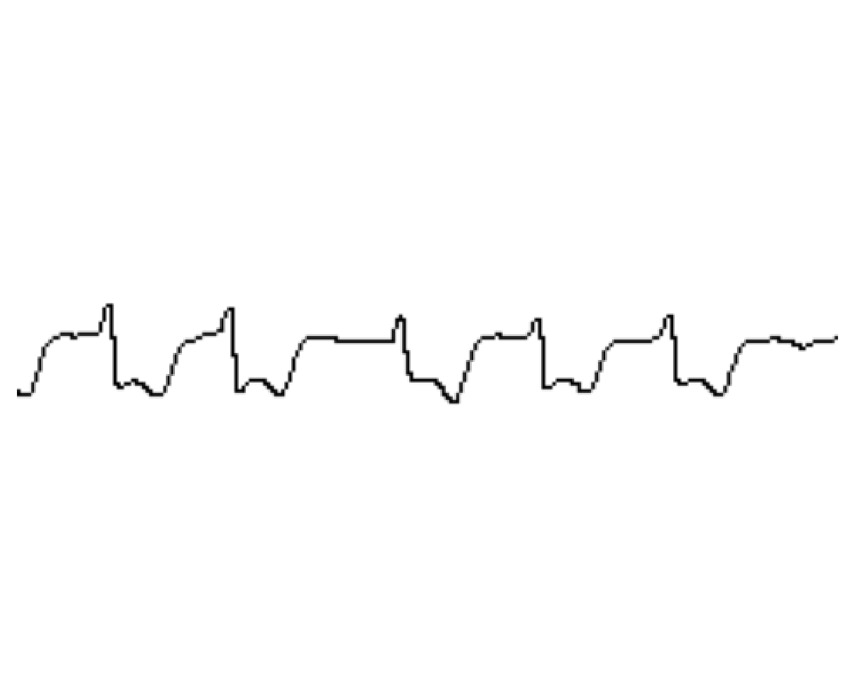
Image dataset after processing

- ·Abnormal Heartbeat HB: 2796
- ·Myocardial Infraction MI: 2868
- ·Normal Patient: 3408
- ·Patients with a history of Myocardial Infraction: 2064

Image dataset after preprocessing

Entire image consists of 12 signals (leads) for a single patient. The leads were extracted, the dataset became each instance is a lead.
The data was processed through performing the following actions: Cropping, Grayscale, Conversion, Smoothing, Thresholding and resizing





CLASSIFICATION

Objective

- Address class imbalance and assess the effect of feature selection and undersampling.

Our Approach:

Primary classifier: **Decision Trees**

Key Metrics Before Undersampling:

- High Accuracy: 91%
- Poor Recall: Only 7-25% of heart disease cases detected
- F1-Score: 0.12
- AUC: 0.74 (with 10 features)

After Undersampling:

- Balanced Accuracy: 70%
- Improved Recall: 66% heart disease detection
- F1-Score: 0.69
- More reliable predictions overall

GridSearch Results:

- Parameter Testing:
  - max_depth: [None, 10, 20, 30, 40, 50]
  - min_samples_split: [2, 5, 10, 20]
  - min_samples_leaf: [1, 2, 5, 10]
  - criterion: ['gini', 'entropy']
- Key Finding: Simpler trees (depth=10) with gini criterion performed best, achieving 0.91 cross-validation accuracy while avoiding overfitting

Additional Classifiers

Random Forest:

- Accuracy: 0.90 -> 0.73
- F1-Score: 0.18 -> 0.74
- AUC: 0.78 -> 0.80
- Precision: 0.38 -> 0.72
- Recall: 0.12 -> 0.77

Logistic Regression:

- Accuracy: 0.91 -> 0.76
- F1-Score: 0.17 -> 0.76
- AUC: 0.83 -> 0.83
- Precision: 0.54 -> 0.75
- Recall: 0.10 -> 0.78

KNN:

- Accuracy: 0.90 -> 0.71
- F1-Score: 0.19 -> 0.72
- AUC: 0.70 -> 0.77
- Precision: 0.36 -> 0.71
- Recall: 0.13 -> 0.74

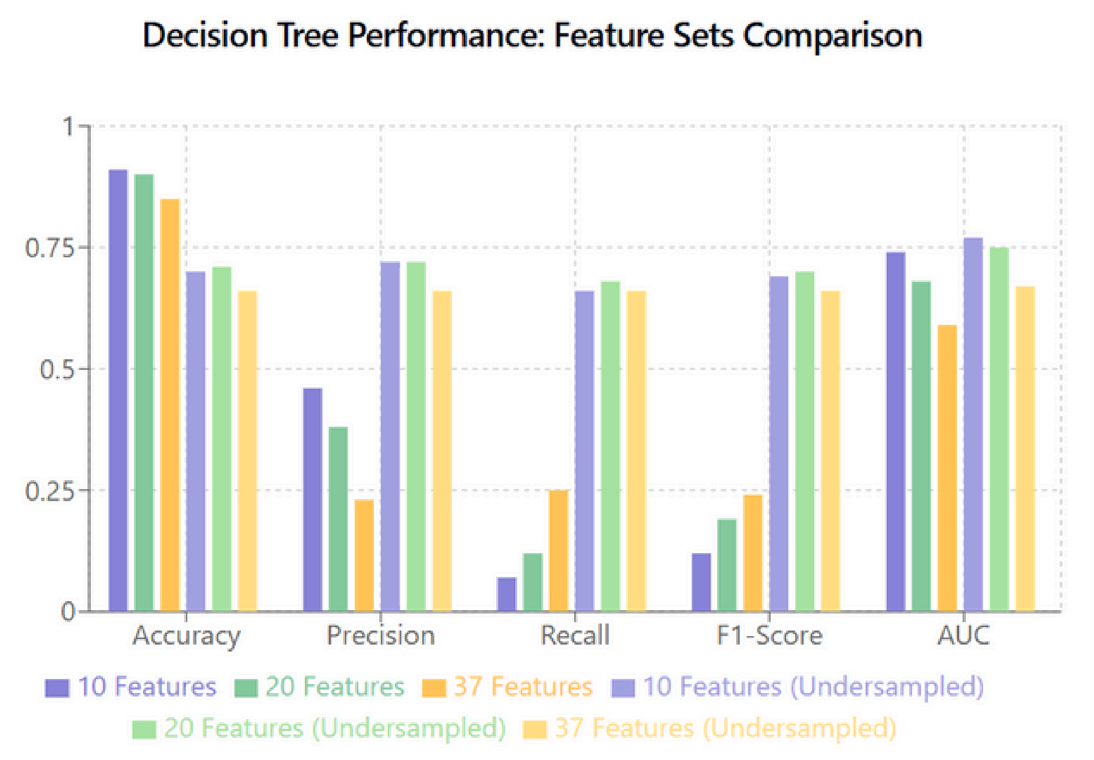
Key Cross-Model Findings:

1. **Feature Selection Impact**

- 10 Features: Often insufficient information
- 20 Features: Optimal for most models
- 37 Features: Diminishing returns/overfitting

2. **Undersampling Effects**

- All models showed significant improvement
- Recall improved by 50-60% across models
- F1-scores improved from ~0.15 to ~0.70



CLUSTERING

Objective

- The objective of clustering was to group patients or data points into two distinct clusters—those with heart disease ("Yes Heart Disease") and those without heart disease ("No Heart Disease")—based on shared characteristics within the dataset.

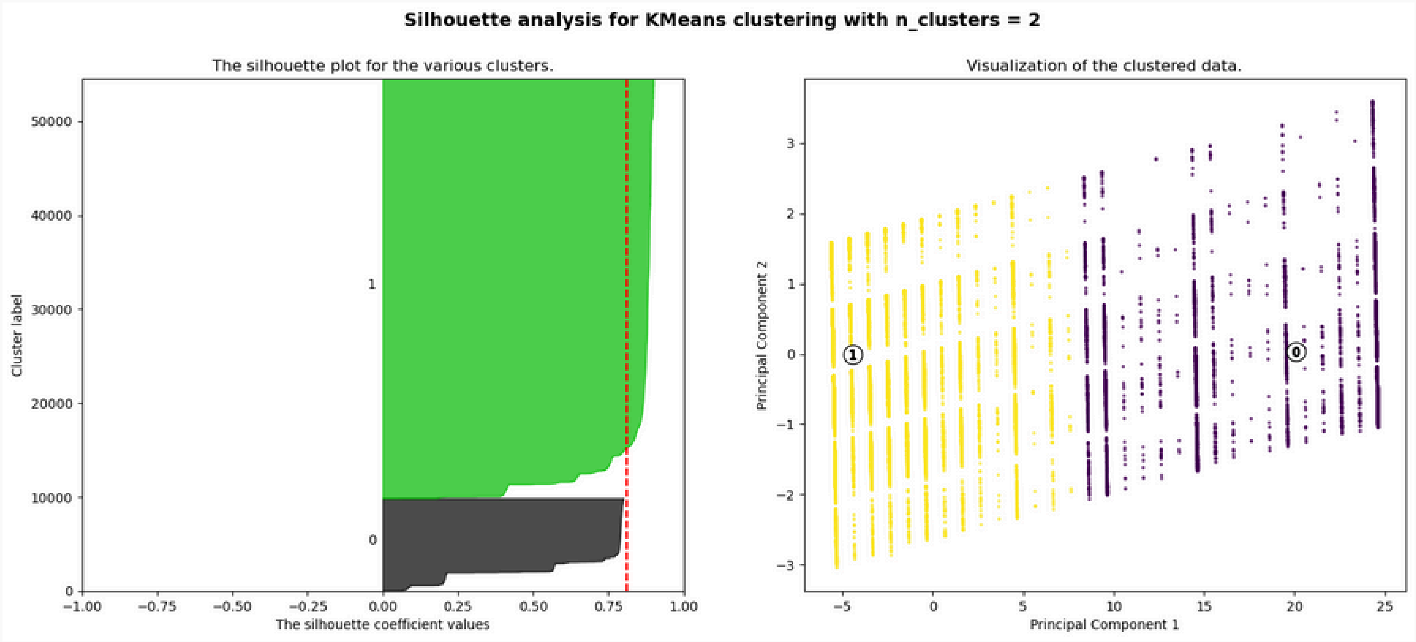
Our Approach:

Primary algorithm: **K-Means**

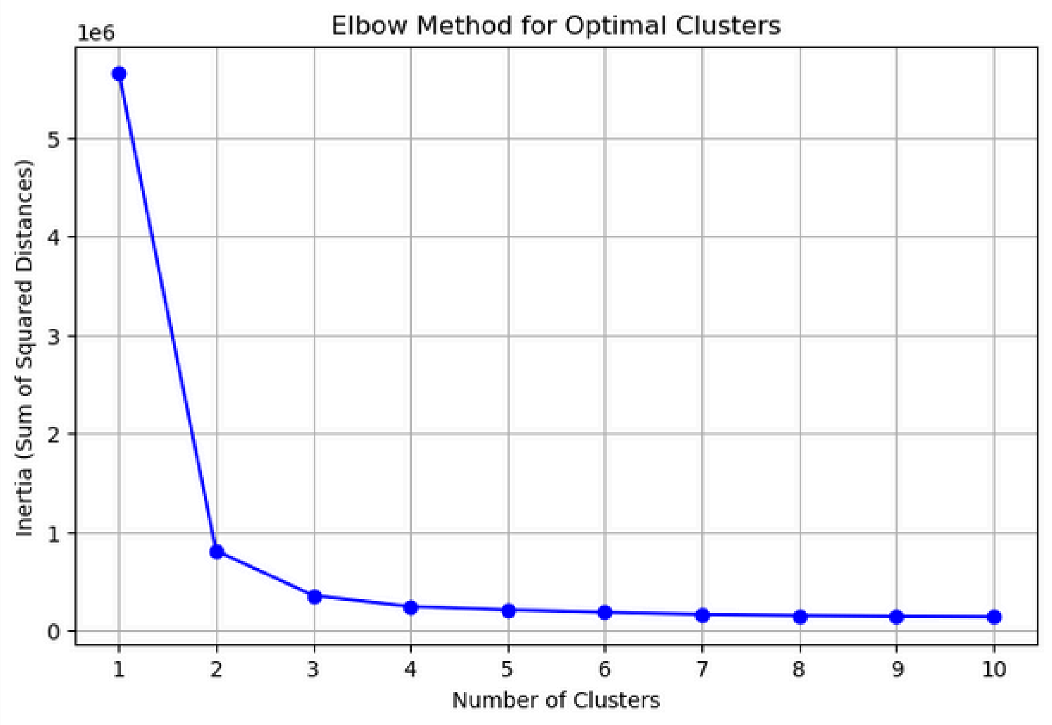
Dimensionality reduction with PCA - to simplify visualization and reduce computational complexity for clustering
Evaluation Metrics: Silhouette Analysis, Elbow Method

Results

Best Silhouette Score: 0.811 for k=2.
Cluster A (No Heart Disease): 90% purity.
Cluster B (Yes Heart Disease): 26% purity (due to overlap).
Overall Clustering Accuracy: 82%.



Silhouette Plot: Optimal k=2 clusters



Elbow Method Plot for the optimal no. of clusters

Additional Algorithms

Gaussian Mixture Models (GMM):

- Explored soft clustering but increased overlap.
- Using BIC (Bayesian Information Criterion) and Silhouette Scores, the optimal number of clusters was determined to be 2.

MLP

MLP configuration:

Input Layer: Flattening 2D array
Hidden Layers: 200, 100, 32
Dropout Rate: 0.7, 0.6, 0.6
Output Layer: 4

Observation:

Detecting an Abnormal Heartbeat (HB)

- Had a high rate of false positives with PMI and Normal and misses 33% of cases with abnormal heartbeat

Detecting Myocardial Infarction (MI)

- Very low false positive rate and correctly identifies all instances of MI

Detecting Patients with a History of Myocardial Infarction (PMI)

- High rate of false positives for HB, a few for HB and misses 29% of PMI

Detecting Patients with Normal Heart Health (Normal)

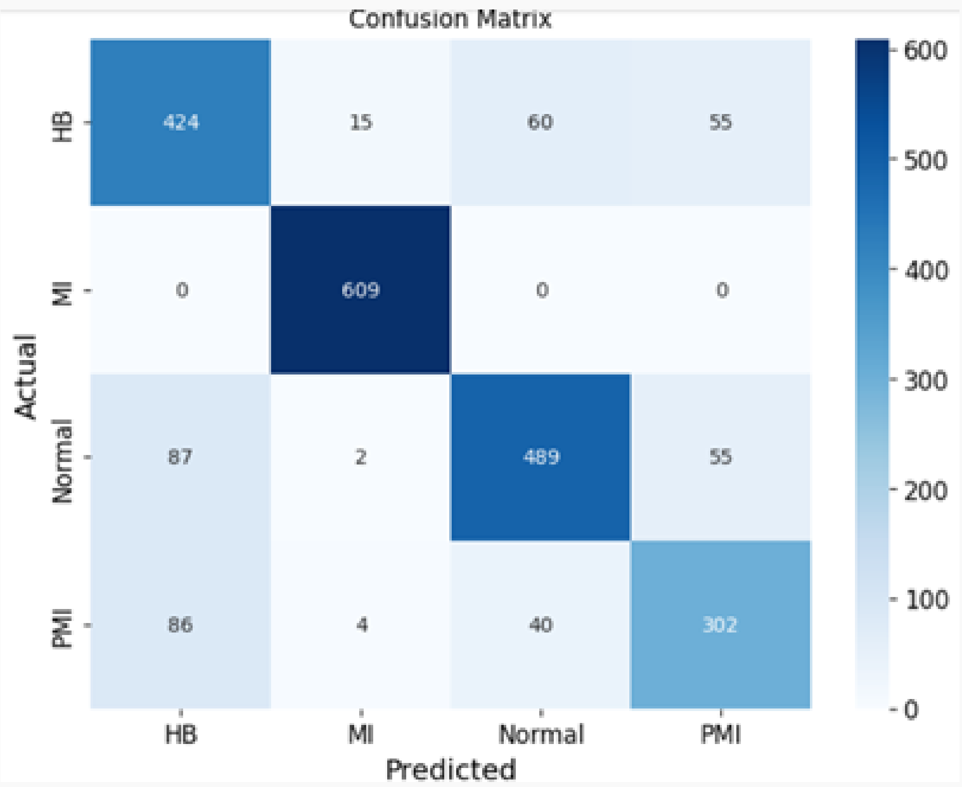
- High rate of false positives, misses fewer cases of normal cases

Performance:

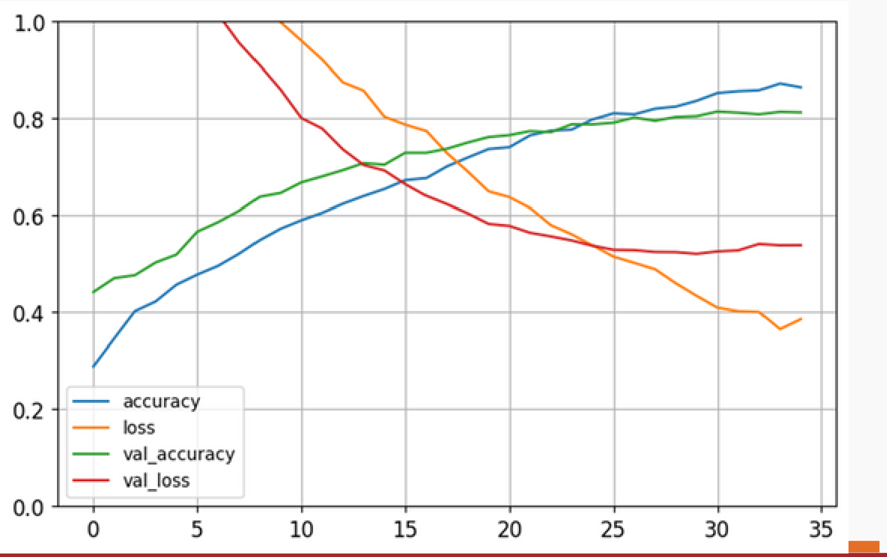
The model struggled with overfitting, as can be seen with the results between the training, validation and test accuracy.
Training Accuracy: 96.3%
·Validation Accuracy: 87.2%
·Test Accuracy: 85.4%

Conclusion:

The model excels at identifying cases of Myocardial Infarction; however, it misses quite a high number of Abnormal Heartbeats, and Patients with a history of MI. There is confusion between PMI and HB, but also a high rate of false positives for normal ECGs. Therefore, the model is good at identifying heart attacks but, has an issue identifying patients with heart problems.



Class	Precision	Recall	F1-score	Support
HB	0.78	0.67	0.72	554
MI	0.97	1	0.99	609
Normal	0.76	0.84	0.79	633
PMI	0.71	0.71	0.71	432
Accuracy			0.81	2228
macro avg	0.81	0.8	0.8	2228
weighted avg	0.81	0.81	0.81	2228



CNN

CNN configuration:

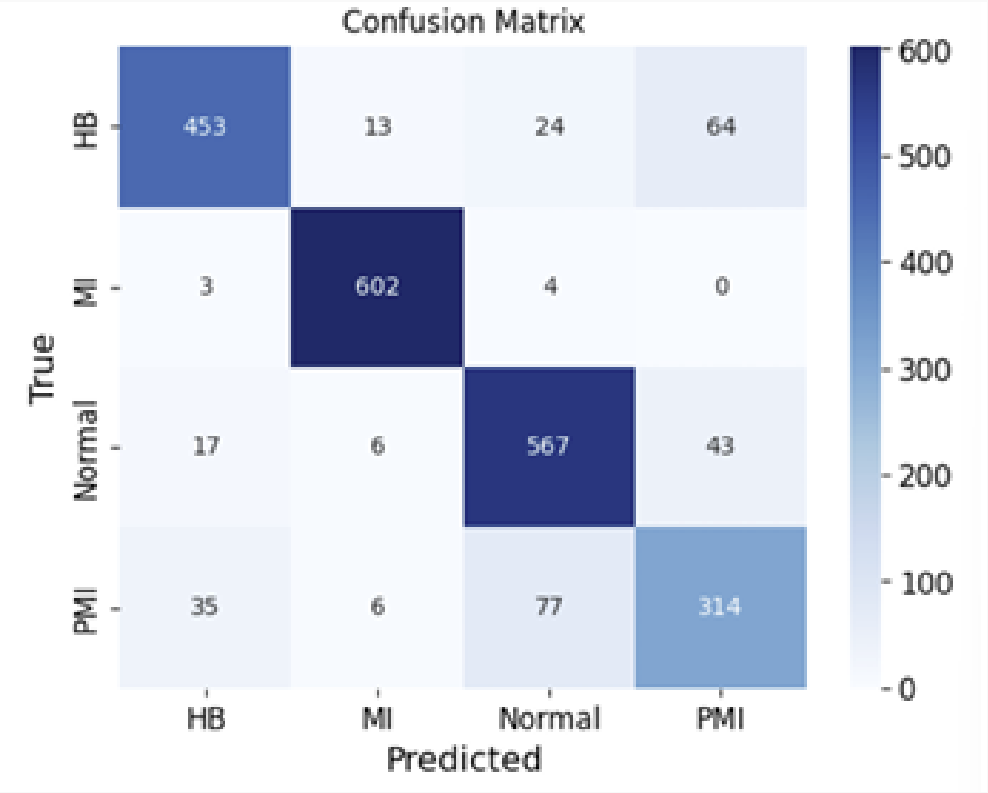
Convolutional layers: **3**
Filter sizes: **(32, 64, 128)**
Dense layer: **128 neurons**
Activation function: **ReLU**
Dropout rate: **0.5**
Output layer: **Softmax**

Train and Validation split: **80% Training and 20% validation**

Optimizations made:

- **Increased dropout rate:** 0.3 and 0.4 for Convolutional layers , 0.5 for dense layer
- **Data Augmentation:** Rotation, Zoom, Horizontal flipping.
- **Reduced learning rate:** 1×10^-4.

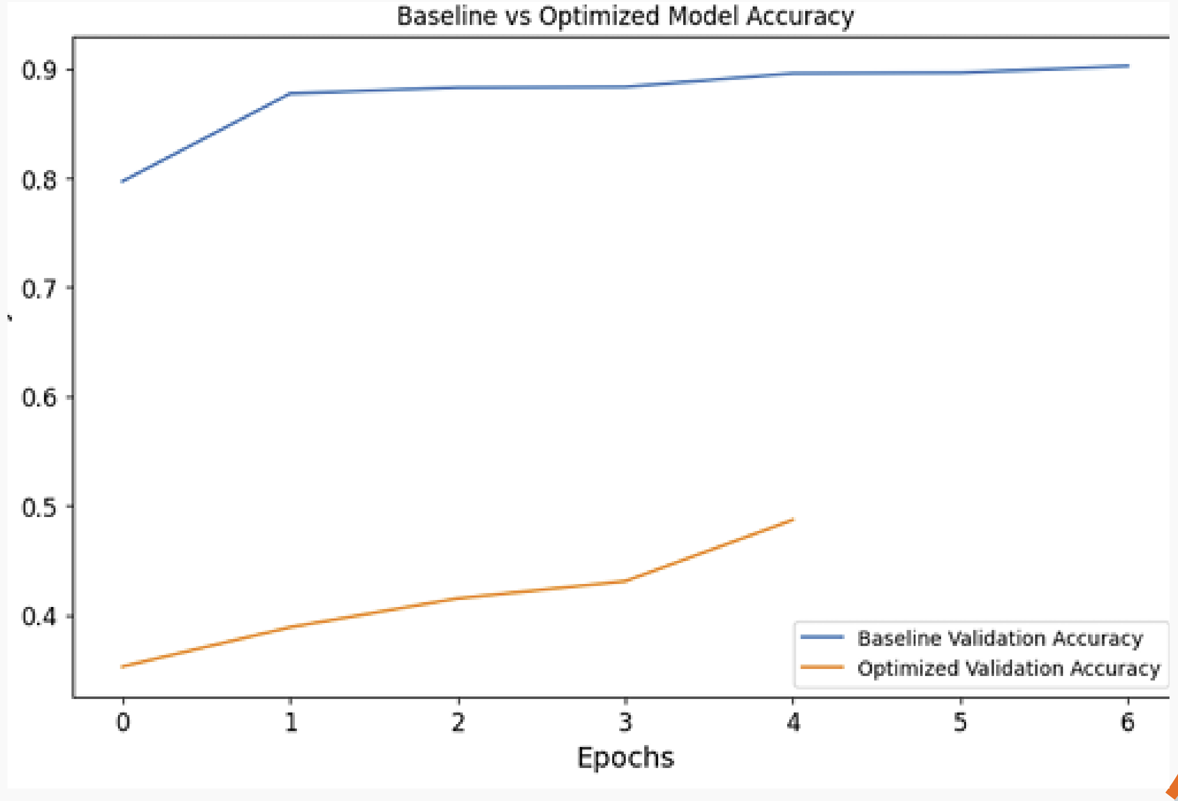
Metric	MLP Baseline	CNN Baseline	Optimized CNN
Training Accuracy	96.3%	99.9%	99.9%
Validation Accuracy	87.2%	90.2%	87.6%
Test Accuracy	85.4%	87.0%	TBD




```
              precision    recall  f1-score   support

          HB       0.89      0.82      0.85       554
          MI       0.96      0.99      0.97       609
      Normal       0.84      0.90      0.87       633
         PMI       0.75      0.73      0.74       432

    accuracy                           0.87      2228
   macro avg       0.86      0.86      0.86      2228
weighted avg       0.87      0.87      0.87      2228
```



Thank you