A Project Report
in partial fulfillment of the degree

Bachelor of Technology
in
Computer Science & Artificial Intelligence

By:
2203A52102
Swetha Masadi

To:
Ramesh Dadi

School of Computer Science & Artificial Intelligence
SR University, Ananthsagar, Hasanparthy (M), Warangal, Telangana 506371, India

2024-25

I. Data Sets:

1. Dataset Description

- **Dataset Name:**
  Heart Failure Prediction
- **Source:**
  Kaggle (Author: Andrew Mvd)
- **Type:**
  CSV
- **Description:**
  This dataset includes clinical records for 299 heart failure patients. There are 12 medical features for each patient: Age, Anaemia, Creatinine Phoshokinase, Diabetes, Ejection Fraction, High Blood Pressure, Platelets, Serum Creatinine, Serum Sodium, Sex, and Smoking. The target variable is DEATH_EVENT, which refers to whether or not the participant died during the service period.
- **Dataset Size:**
  - **Total Samples:** 299
  - **Features:** 12 clinical features
  - **Target Variable:** DEATH_EVENT (Binary: 0 = No death event, 1 = Death event)
- **Data Split:**
  This dataset does not have predefined training, validation, or test splits. Data partitioning should be performed based on the user's requirements.

2. Dataset Description

- **Dataset Name:**
  CAD Cardiac MRI Dataset
- **Source:**
  Kaggle (Author: Danial Sharifrazi)
- **Type:**
  Image
- **Description:**
  This dataset contains cardiac MRI images that are suitable for coronary artery disease (CAD) diagnosis. It is especially well-suited for developing and training deep learning models for medical image classification tasks.
- **Dataset Size:**
  - **Image Format:** JPG
  - **Number of Samples:** 63.4k

- **Classes:** 2 classes (Normal and Sick)

- **Data Split:**
  The dataset does not include predefined training, validation, or test splits. Users should partition the data according to their project requirements.

3. Dataset Description

- **Dataset Name:**
  Heartbeat Sounds
- **Source:**
  Kaggle (Author: kinguistics)
- **Type:**
  Audio
- **Description:**
  The dataset consists of heart beats sound recordings obtained with a stethoscope. It has been designed for the research and machine learning development for heart sound classification tasks to class normal and abnormal heartbeats. The dataset consists of audio samples and corresponding metadata in form of CSV files, including information such as timing and classifications.
- **Dataset Size:**
  - Total Samples: 835 files
  - Audio Files: 832 .wav files
  - Metadata Files: 3 .csv files
  - CSV Columns: 12 attributes per row (set_a.csv, set_b.csv, set_a_timing.csv)
- **Data Split:**
  The dataset is divided into two primary sets:
  - Set A
  - Set B

II. Implementation Overview

a) Dataset: Heart Failure Prediction (CSV)

1. Data Pre-processing

- The dataset was imported using the pandas library for continued analysis.
- Missing values checks were completed to confirm a clean dataset.
- The features (independent variables) and the target (DEATH_EVENT) were separated for model training.

- The dataset was then split into training and testing sets using an 80:20 ratio by using the train_test_split() function from Scikit-learn.
- There was no need for feature scaling to be applied, since the scale of the dataset's numeric features were compatible for the models used.

2. Models Used

- **Logistic Regression**
  Regression is a type of statistical model that is often used for binary classification problems. Logistic Regression fits data to a logistic function to estimate the probability of an event occurring. For this project, a logistic regression model was applied for predicting heart failure outcomes. Logistic regression is particularly useful when the dependent variable is categorical, most often binary, such as predicting whether a patient will survive or not. It calculates the log-odds of the outcome as a linear combination of the input features. In our context, features such as age, ejection fraction, and serum creatinine were fed into the model to determine the likelihood of heart failure. Due to its simplicity and interpretability, logistic regression is often considered a strong baseline model in medical data analysis.
- **Decision Tree Classifier**
  Decision Trees are a supervised learning method that split the data with respect to feature thresholds resulting in a tree-like structure. The internal nodes are a decision with respect to a feature, and the leaves represent an output label. This model was used for predicting heart failure outcomes from the clinical features of the dataset. The structure of a decision tree is highly intuitive, as it mimics human decision-making with a flowchart-like approach. At each node, the algorithm selects the feature and corresponding threshold that results in the best split based on a chosen criterion like Gini impurity or entropy. This hierarchical structure makes it easy to visualize which features contribute most to the final prediction. However, decision trees are prone to overfitting, especially when they grow deep, which is why pruning or setting depth constraints is essential for optimal performance.
- **Random Forest Classifier**
  Random Forest is an ensemble version of the Decision Tree, as an ensemble learns from the predictions of several (ideally diverse) decision trees; thus, making a better and more accurate prediction. Each tree makes a classification, and the random forest returns a classification based on the majority of votes from the trees. This model was used to increase predictive performance while decreasing the possibility of overfitting from a decision tree. Random Forest adds an element of randomness during training by selecting random subsets of data and features for each individual tree. This randomness helps ensure that the trees are de-correlated, resulting in a more generalized model. Additionally, the ensemble approach improves robustness to noise and outliers in the data. In this project, Random Forest was

particularly beneficial because it maintained high accuracy while reducing the risk of overfitting, which is critical when working with limited or imbalanced clinical datasets.

3. Model Evaluation

- Each model was evaluated using the following metrics:
    - **Accuracy Score**: This represents the proportion of correctly predicted samples.
    - **Classification Report**: This includes precision, recall, and F1-score for each class (0: No heart failure event, 1: heart failure event).

The three models were then compared to assess their performance in order to select the most reliable model for heart failure event prediction.

b) Dataset: CAD Cardiac MRI Dataset (Images)

1. Data Preprocessing

- The raw dataset was divided into two classes, Normal and Sick, with subfolders that contained the MRI images.
- The images underwent preprocessing with OpenCV and PyTorch transforms
    - The images were loaded using grayscale images to minimize complexity of interpretation while still containing the features of medical images.
    - The images were resized to a determined size so that all input images retained dimensionality.
    - The dataset underwent training, validation, and testing splits.
    - The data augmentation was applied and accomplished using PyTorch's transforms of resize, normalize, and convert to tensor.

2. Models Used

- **Convolutional Neural Network (CNN)**
  A custom CNN architecture was used for the binary classification of the MRI images. The architecture had:
    - Convolutional layers with ReLU activations,
    - Max Pooling to reduce image dimensionality, and
    - Fully connected (Dense) layers used for the classification.

This model was trained using the torch.nn module on GPU-compatible devices.

3. Model Training and Evaluation

- The model was trained with the torch.optim optimizer, using the CrossEntropyLoss function.
- The training included:
    - Forward propagation, loss calculation, and backpropagation.
    - Validation took place at every epoch to assess the model's generalization performance.
- The final evaluation occurred on the unseen test set.
- Performance was assessed using:
    - Accuracy.
    - Confusion Matrix.
    - ROC Curve (Receiver Operating Characteristic) to evaluate the model's ability to discriminate Normal from Sick cases.
    - ANOVA statistical analysis to confirm/substantiate stability of accuracy over time across epochs.

c) Dataset: Heartbeat Sounds Dataset (Audio)

1. Data Preprocessing

- The dataset included .wav audio recordings of heart sounds that were organized into set_a and set_b folders.
- The Preprocessing steps applied were:
    - Audio files were loaded using Librosa at a sample rate of 22,050 Hz.
    - The audio clips were trimmed or padded to a consistent 5 seconds duration.
    - Mel-frequency cepstral coefficients (MFCCs) were extracted from each audio file to use as additional feature representations.
    - Features were normalized and saved into NumPy arrays for faster computation.
    - Labels were encoded for supervised training.
    - The final dataset was split into training and testing sets using a stratified approach.

2. Models Used

- **Convolutional Neural Network (CNN)**
  A deep Convolutional Neural Network (CNN) was built to predict the beats of the heart sounds into classes.

- The input layer accepted MFCC features as input with the added channel dimension.
- Several convolutions with max pooling and batch normalization.
- The dense layers with dropout to prevent overfitting.
- Softmax activation in the output layer for multiclass classification.

3. Model Training and Evaluation

- The model was compiled with:
  - Adam as the adaptive learning rate optimizer.
  - Categorical crossentropy as the loss function.
- The training was performed in the following manner:
  - Model was trained on MFCC-based feature inputs.
  - Validation was done after each epoch to monitor learning.
- Performance metrics included:
  - Accuracy of the models on both the training and test datasets.
  - Confusion Matrix and classification report.
  - Visualization of training history: Accuracy vs Epochs and Loss vs Epochs graph

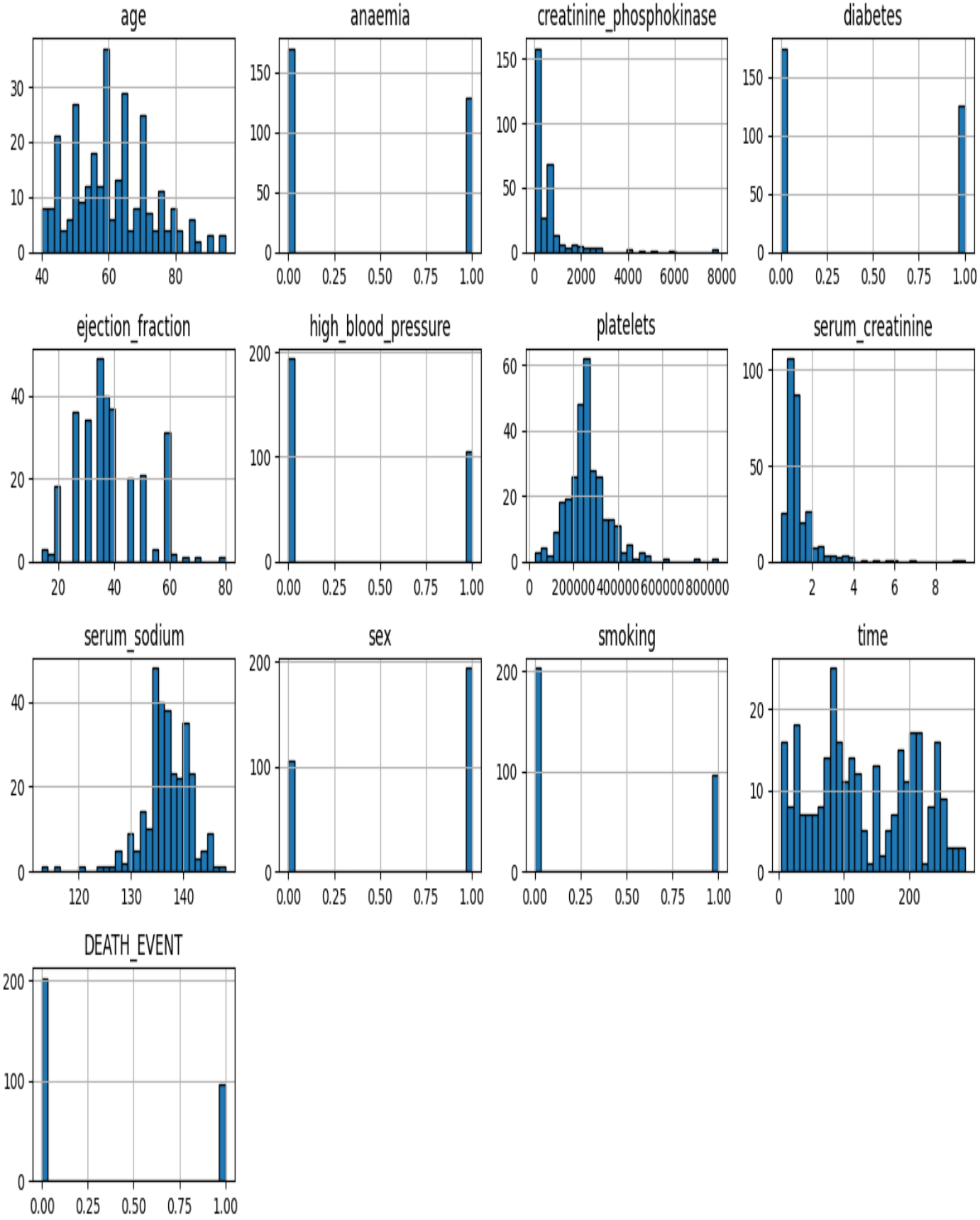
III. Conclusion

Conclusion

In this project, we investigated three separate types of biomedical datasets - structured clinical data, medical images, and recorded heart sounds - to build models for predicting and classifying heart diseases. The project examined end-to-end workflows including data preprocessing, feature extraction, model selection, training, and assessment.

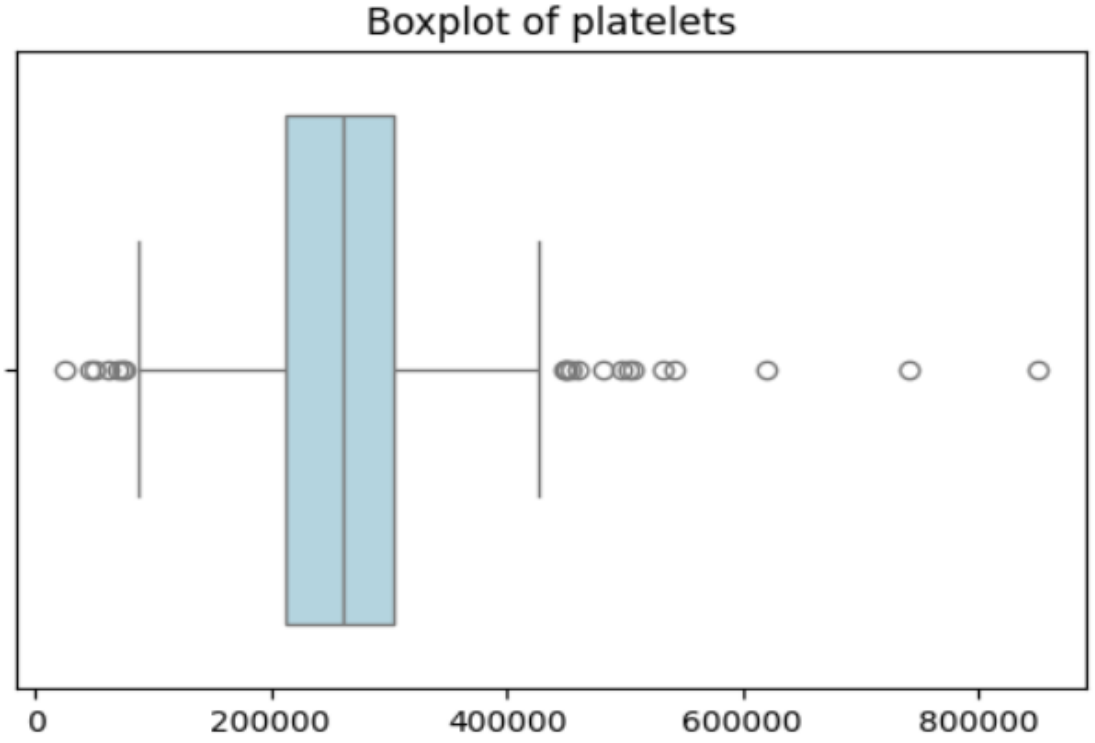
Outcomes

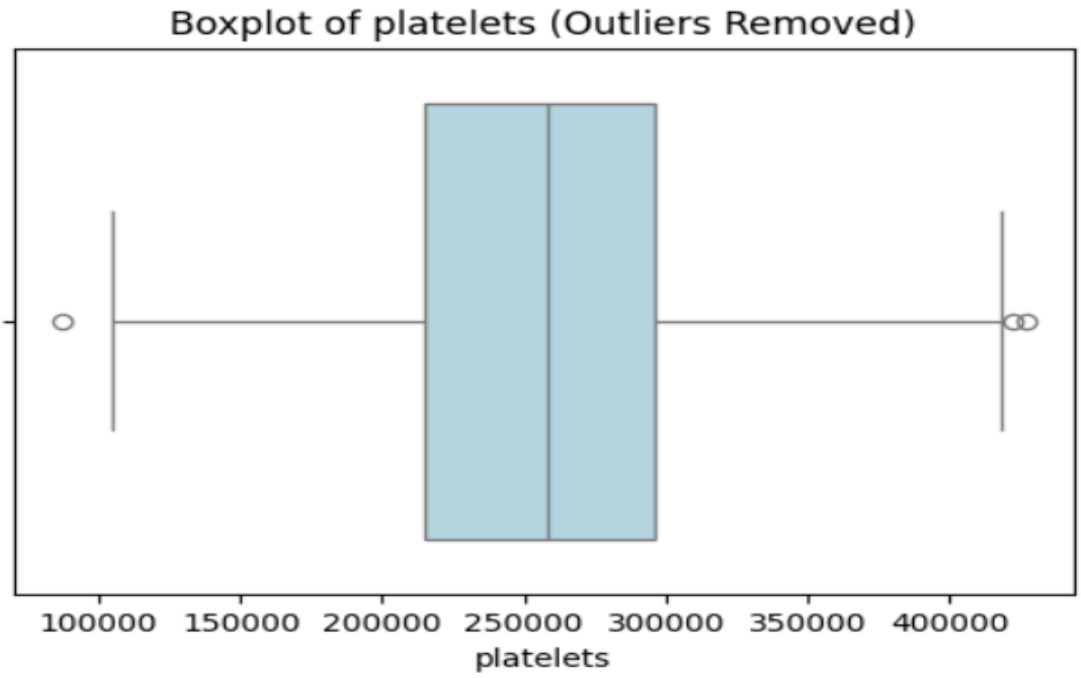
a) **For CSV Dataset:**


age
anaemia
creatinine_phosphokinase
diabetes
ejection_fraction
high_blood_pressure
platelets
serum_creatinine
serum_sodium
sex
smoking
time
DEATH_EVENT

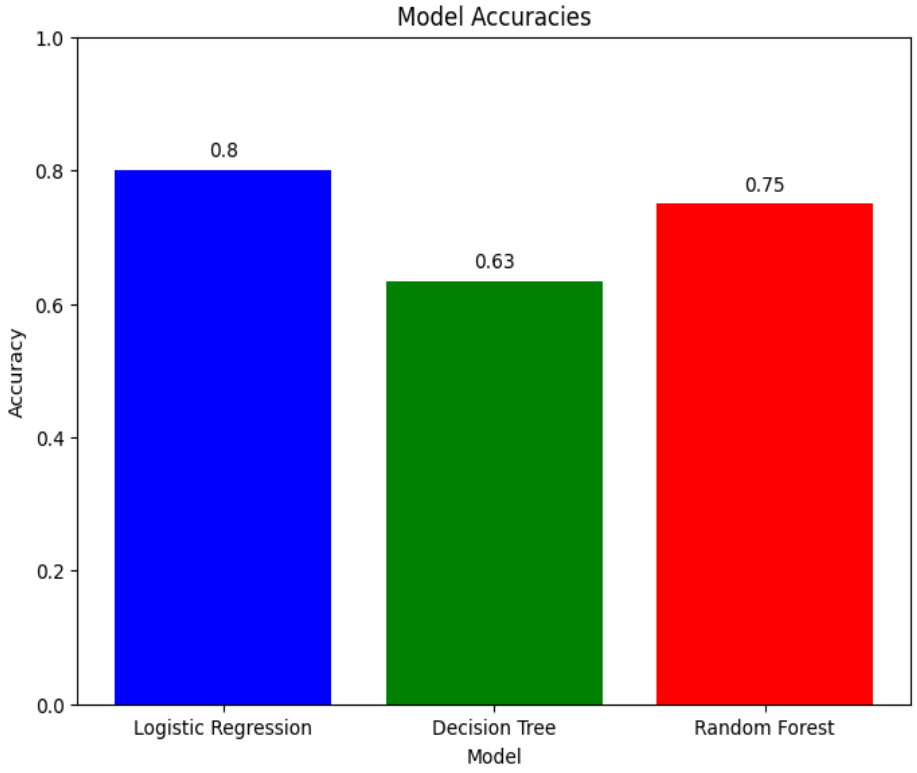

age
anaemia
creatinine_phosphokinase
diabetes
ejection_fraction
high_blood_pressure
platelets
serum_creatinine
serum_sodium
sex
smoking
time
DEATH_EVENT

```
platelets:
  IQR: 91000.0
  Lower Bound: 76000.0
  Upper Bound: 440000.0
  Outliers: 21 records
```



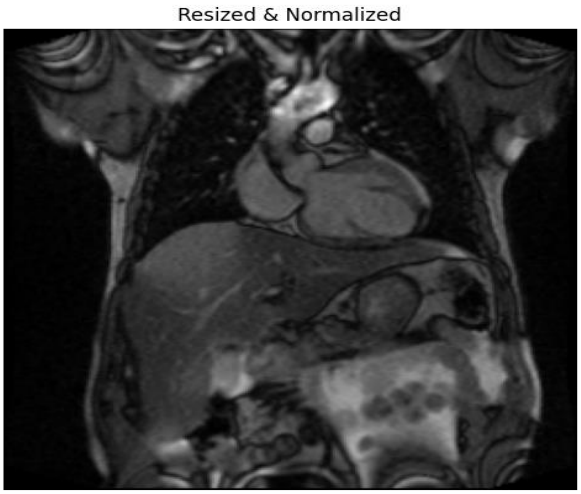


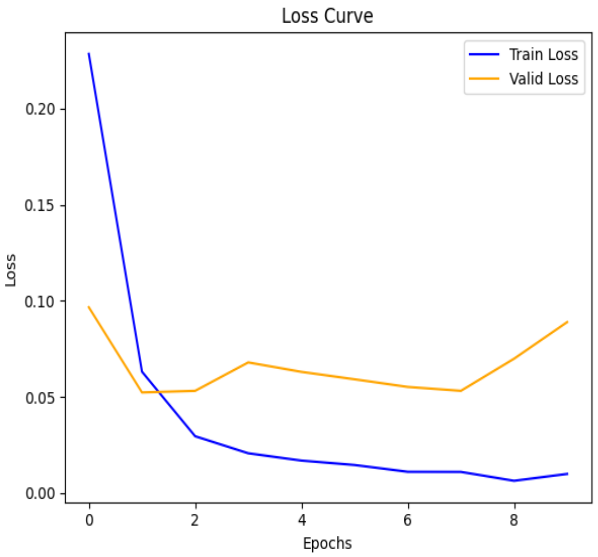
Metric	Logistic Regression	Decision Tree	Random Forest
Accuracy	0.80	0.63	0.75
Precision (0)	0.77	0.66	0.72
Recall (0)	0.94	0.77	0.94
F1-score (0)	0.85	0.71	0.81
Precision (1)	0.88	0.58	0.86
Recall (1)	0.60	0.44	0.48
F1-score (1)	0.71	0.50	0.62
Macro Avg F1	0.78	0.61	0.72
Weighted Avg F1	0.79	0.62	0.73

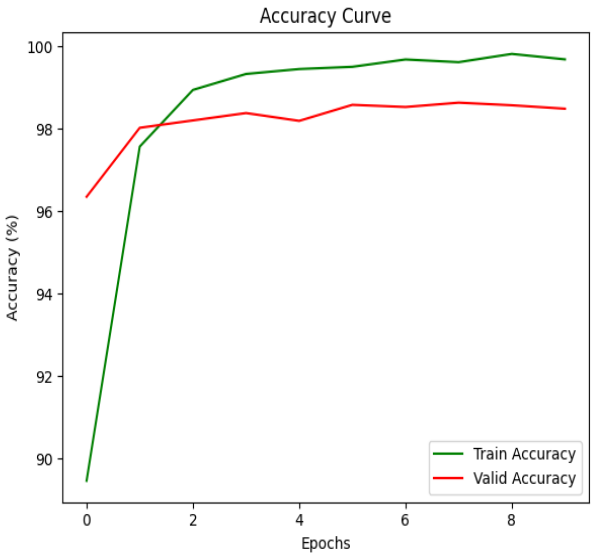


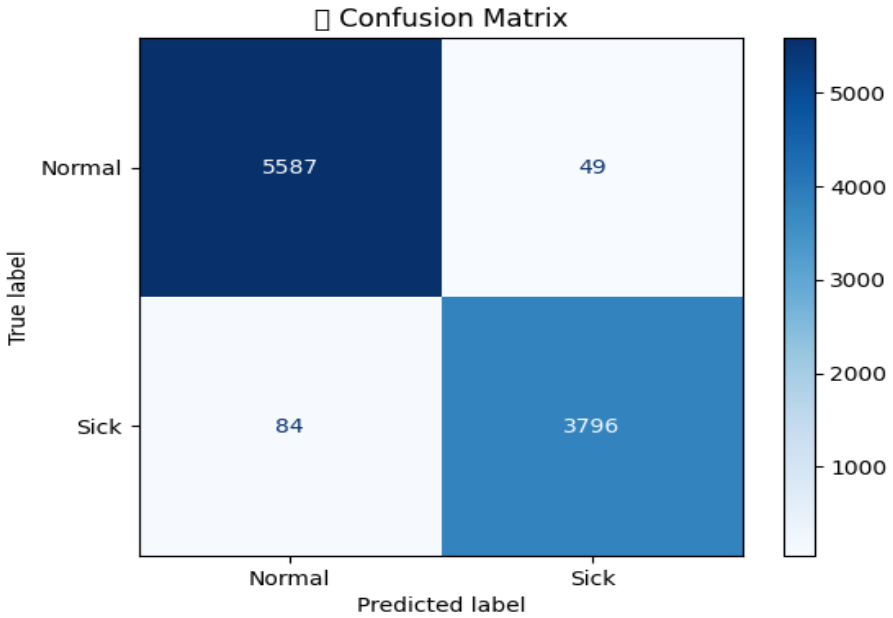
For the Heart Failure Clinical Dataset (CSV) used classical machine learning models such as Logistic Regression, Decision Trees, and Random Forest Models. The Random Forest model provided the most consistent results, indicating that there were complexities and interactions of features in clinical datasets the ensemble methods were able to extract in useful ways.

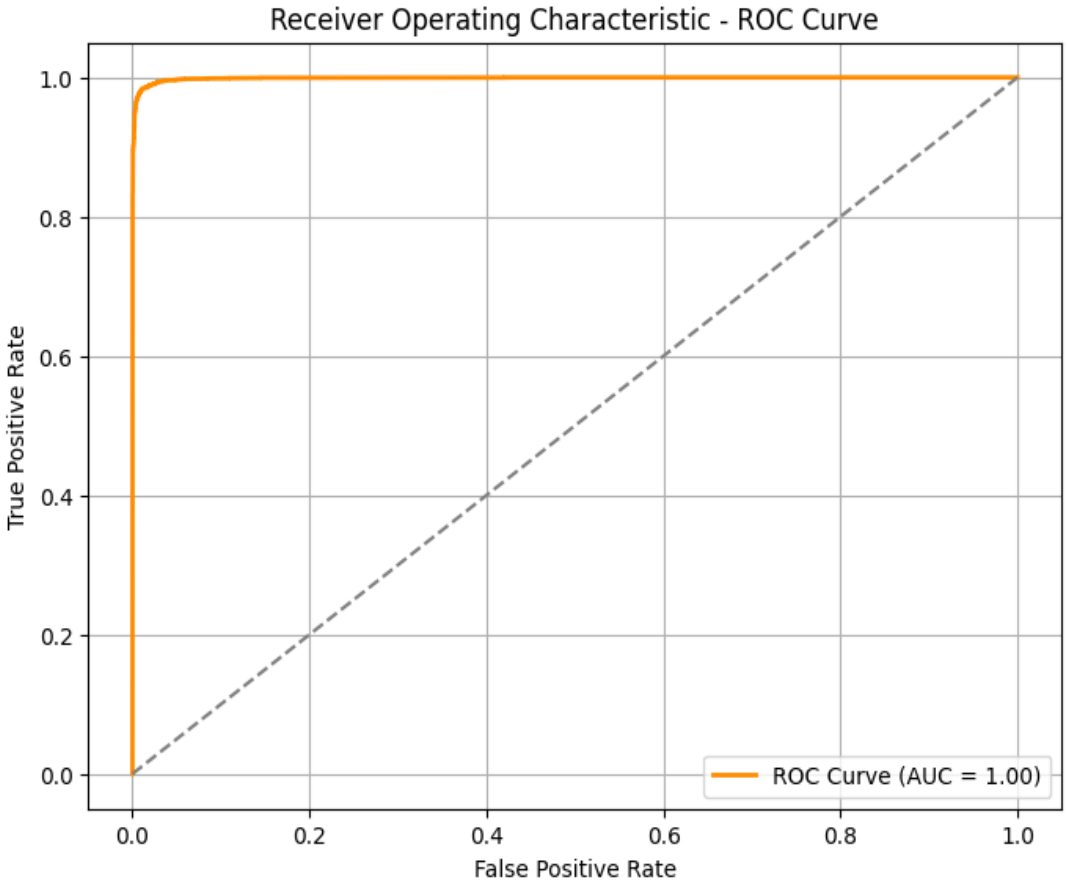
b) For Image Dataset:







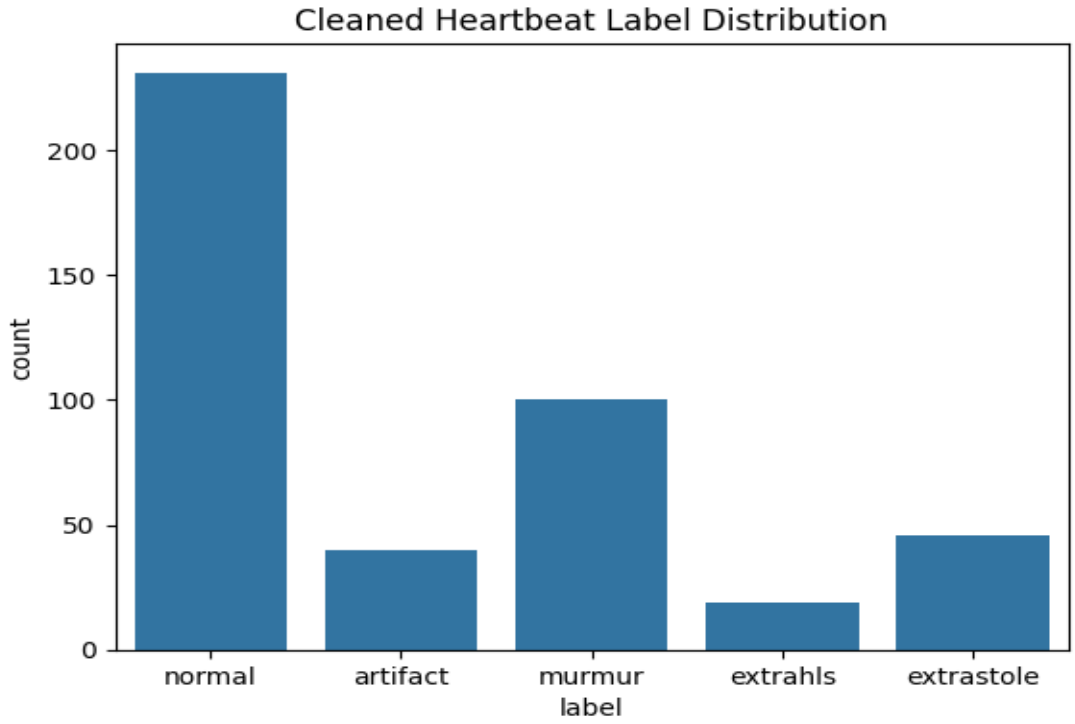




Test Type	Statistic Value	p-value	Interpretation
Z-Test	Z = 0.1069	0.9149	⚖️ No significant difference between training and validation accuracies.
T-Test	T = 0.1069	0.9170	⚖️ No significant difference between training and validation accuracies.
ANOVA	F = 0.1265	0.8817	⚖️ No significant difference found between training, validation, and test accuracies.

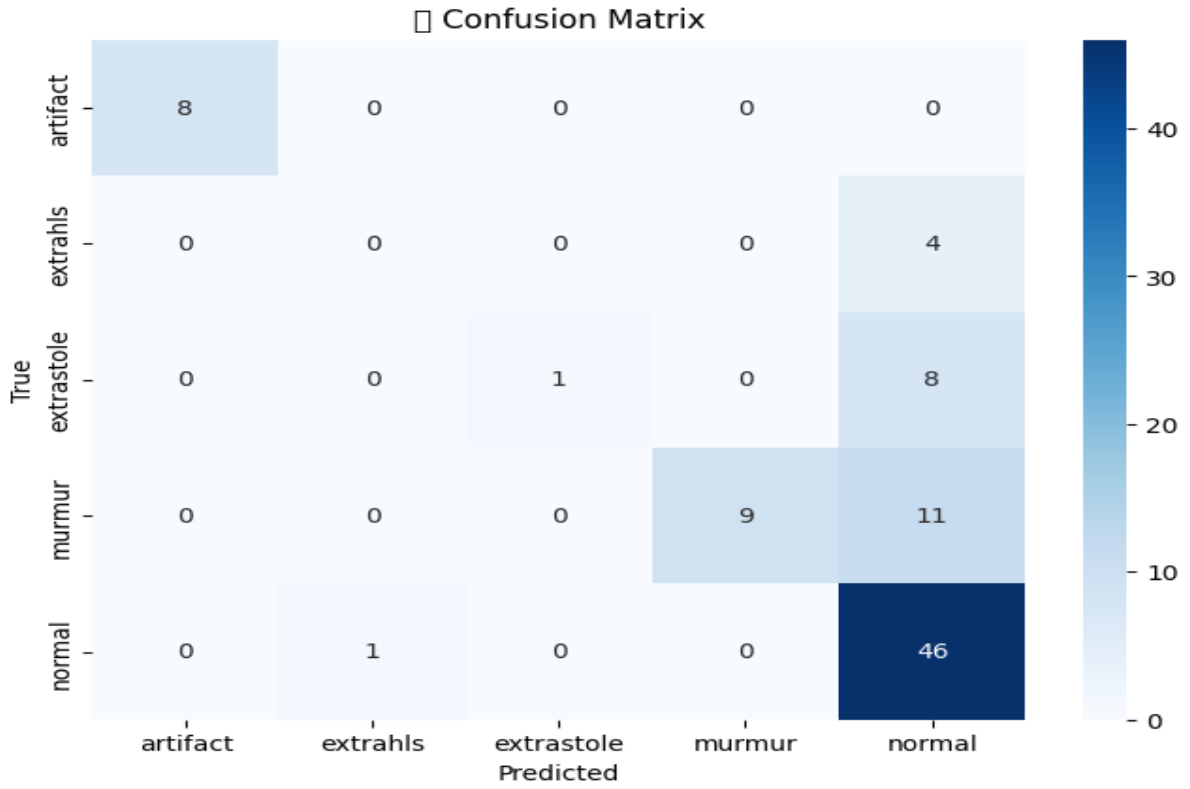
In the case of the Cardiac MRI Dataset (Images) deep learning models were developed by using Convolutional Neural Networks (CNNs). The model learned distinguish between normal and sick heart MRI images, providing evidence that visual features in medical images can be extracted and classified with deep learning methods.

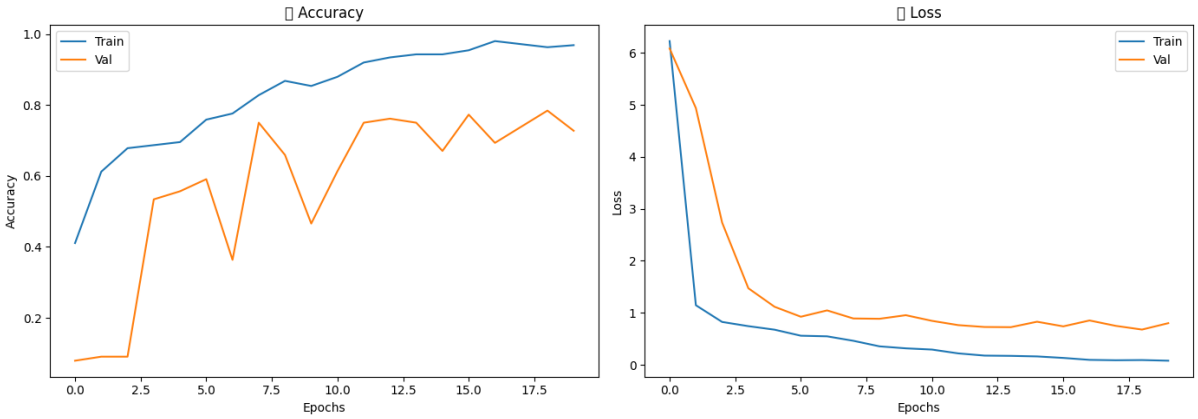
C) For Audio Dataset:

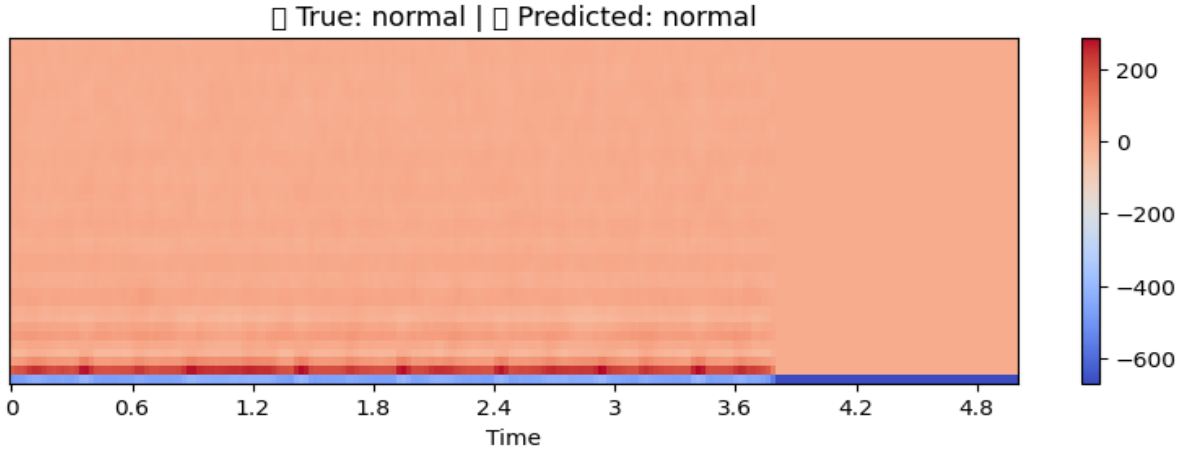


Class	Precision	Recall	F1-score	Support
artifact	1.00	1.00	1.00	8
extrahls	0.00	0.00	0.00	4
extrastole	1.00	0.11	0.20	9
murmur	1.00	0.45	0.62	20
normal	0.67	0.98	0.79	47

Metric	Value
Accuracy	0.73
Macro Avg F1	0.52
Weighted Avg F1	0.68


Confusion Matrix
True
Predicted
artifact
extrahls
extrastole
murmur
normal
8 0 0 0 0
0 0 0 0 4
0 0 1 0 8
0 0 0 9 11
0 1 0 0 46
40
30
20
10
0


Accuracy
Train
Val
Accuracy
Epochs
Loss
Train
Val
Loss
Epochs


True: normal | Predicted: normal
Time
0 0.6 1.2 1.8 2.4 3 3.6 4.2 4.8
200
0
−200
−400
−600

For the Heartbeat Sounds Dataset (Audio), the audio was preprocessed using the MFCC feature extraction and the features were used to train a CNN-based classifier. The model was able to classify the heart sounds into health conditions accurately, demonstrating the possibilities of sound-based diagnostic tools.