

Data Collection and Preprocessing Phase

Date	6/20/2025
Project Title	Classification of Arrhythmia by Using Deep Learning with 2-D ECG Spectral Image Representation.
Maximum Marks	6 Marks

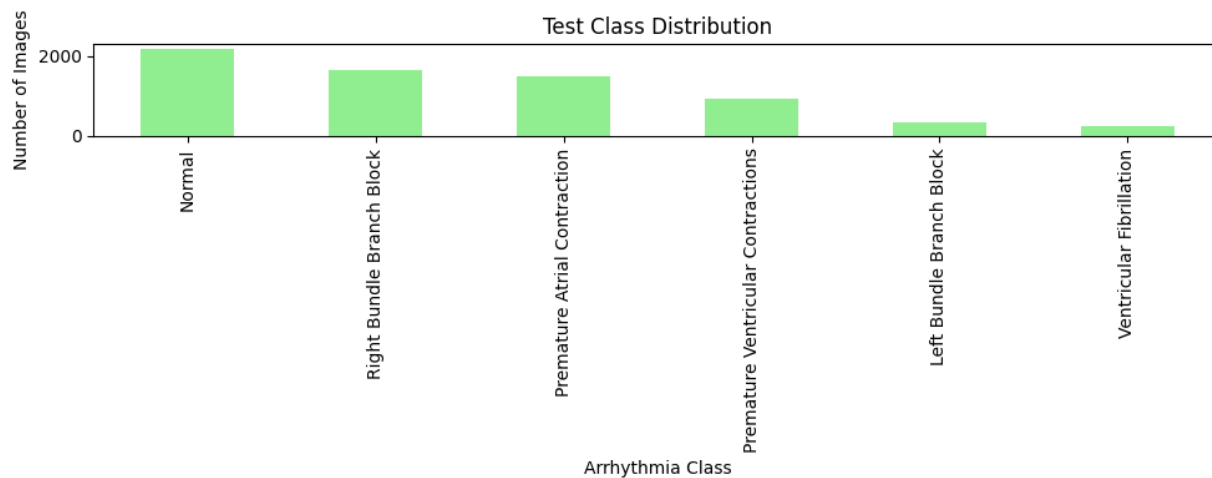
Data Exploration and Preprocessing Report

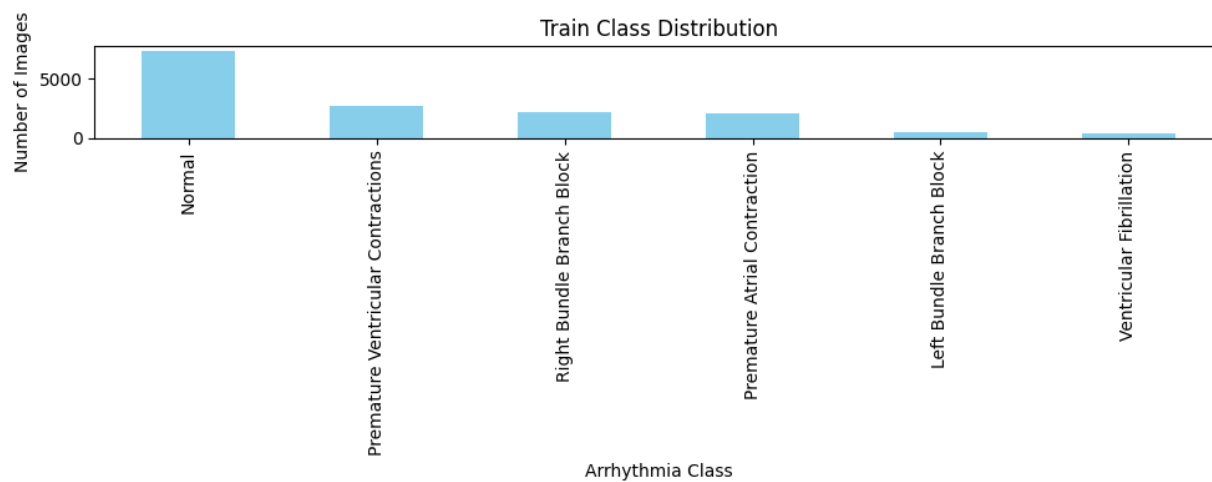
Dataset variables will be statistically analyzed to identify patterns and outliers, with Python employed for preprocessing tasks like normalization and feature engineering. Data cleaning will address missing values and outliers, ensuring quality for subsequent analysis and modeling, and forming a strong foundation for insights and predictions.

Section	Description
Data Overview	<p>Dimension:</p> <ul style="list-style-type: none"> - Training set: 15,742 images across 6 classes - Test set: 6,765 images across 6 classes <p>Class Names:</p> <ul style="list-style-type: none"> - Left Bundle Branch Block - Normal - Premature Atrial Contraction - Premature Ventricular Contractions - Right Bundle Branch Block - Ventricular Fibrillation <p>Train Class Distribution:</p> <ul style="list-style-type: none"> - Normal: 7,346 - Premature Ventricular Contractions: 2,759 - Right Bundle Branch Block: 2,239 - Premature Atrial Contraction: 2,054 - Left Bundle Branch Block: 504 - Ventricular Fibrillation: 439 <p>Test Class Distribution:</p> <ul style="list-style-type: none"> - Normal: 2,179

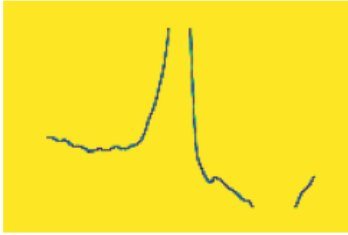
Section	Description
	<ul style="list-style-type: none"> - Right Bundle Branch Block: 1,645 - Premature Atrial Contraction: 1,503 - Premature Ventricular Contractions: 915 - Left Bundle Branch Block: 341 - Ventricular Fibrillation: 242
Descriptive Statistics / Univariate Analysis	<ul style="list-style-type: none"> - The dataset is imbalanced, with the "Normal" class being the most represented and "Ventricular Fibrillation" the least. - No missing or corrupted images were found. - Bar plots of class distribution (see attached figures) clearly show the imbalance. - Example images from each class were visually checked and are representative of their labels.
Bivariate/Multivariate Analysis	<ul style="list-style-type: none"> - No additional features present beyond class labels. - Visual inspection of sample images from each class confirms distinct ECG patterns for each arrhythmia type. - No multicollinearity or feature correlation analysis required for image data. - Both train and test sets contain all six classes, supporting robust evaluation.
Outliers and Anomalies	<ul style="list-style-type: none"> - No corrupted or unreadable images detected. - All images are consistently formatted (PNG/JPG).

Section	Description
	<ul style="list-style-type: none"> - Class imbalance is present, with "Ventricular Fibrillation" and "Left Bundle Branch Block" underrepresented. - No extreme outliers in image sizes or file formats.
Data Preprocessing Code Screenshots	<ul style="list-style-type: none"> - Loaded data using Python (Pandas and glob). - Checked for corrupted images using PIL and found none. - Images will be resized to a consistent input size (e.g., 128×128 or 224×224 pixels) for model training. - Pixel values will be normalized to 1. - Data augmentation (rotation, shift, zoom, flip) will be applied to improve generalization, especially for minority classes. - Training data will be split into training and validation sets as needed. - <i>(Screenshots of code and plots are attached as per template.)</i>
Save Processed Data	<ul style="list-style-type: none"> - Cleaned and preprocessed images are ready for use in model training and validation. - Data is organized in class-wise folders under "train" and "test" directories. - No further cleaning required.

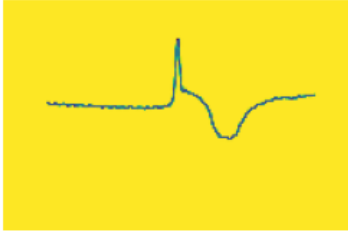




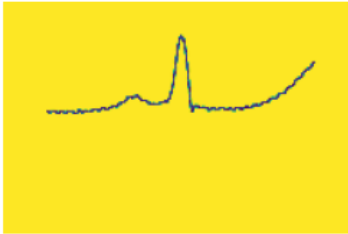
Class: Premature Ventricular Contractions



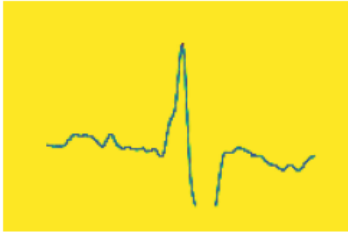
Class: Premature Atrial Contraction



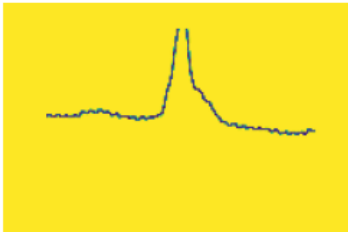
Class: Normal



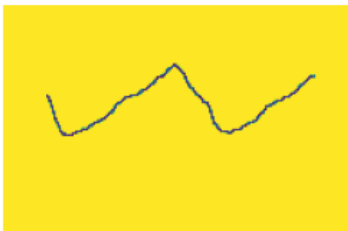
Class: Right Bundle Branch Block



Class: Left Bundle Branch Block



Class: Ventricular Fibrillation



```
Corrupted images (0): []
label Left Bundle Branch Block Normal Premature Atrial Contraction \
split
test 341 2179 1503
train 504 7346 2054

label Premature Ventricular Contractions Right Bundle Branch Block \
split
test 915 1645
train 2759 2239

label Ventricular Fibrillation
split
test 242
train 439
```