DSP LAB PROJECT REPORT

QRS DETECTION AND HEARTBEAT CLASSIFICATION ON SEGMENTED ECG SIGNAL

Submitted By:

Vedant Baldwa	23UCC611
Vedant Wadhwa	23UCC612

Course: Digital Signal Processing Lab

Submission Date: 19-11-2025

Objective:

"To Implement a compact DSP pipeline that localizes the R-peak inside each heartbeat snippet from the Kaggle MIT-BIH dataset and to evaluate localization quality and basic heartbeat classification performance across beat classes".

QRS (R-Peak) Localization and Classification on Heartbeat Snippets

---

Abstract

This project implements a compact digital-signal-processing pipeline in MATLAB to localize R-peak inside segmented ECG Heartbeats (Kaggle MIT-BIH snippet dataset). Each snippet (187 samples) is band-pass filtered (5-25 Hz) to emphasize QRS energy, absolute value of the filtered signal is scanned for locally prominent peaks and a canter-preference rule selects the R-index. The detector returned an index for every snippet (100% detection rate). Using a 20-smaple central window, localization accuracy was 60.01% on the training set and 59.99% on test set, per-class accuracies are reported in the Result section. We additionally extract simple DSP features (R amplitude, QRS width, RMS, spectral centroid, etc.) and train a linear SVM classifier, which achieves 89.91% test accuracy. Per-class metrices and template visualizations are provided and interpreted.

---

1. Introduction

Electrocardiography (ECG) is a fundamental tool for monitoring cardiac activity. The QRS complex particularly the R-peak marks ventricular depolarization and is used to compute heart rate, RR intervals, and heart rate variability (HRV). Robust automatic localization of R peaks is therefore a core preprocessing step in many cardiac signal processing pipelines.

In this project we implemented a DSP-based R-peak localization pipeline in MATLAB and evaluate its performance on publicly available segmented heartbeat data derived from the MIT-BIH Arrhythmia Database (Kaggle). The aim is to demonstrate classical DSP building blocks (filtering and peak detection), produce per class localization statistics, and analyse failure modes. We demonstrated how DSP features can feed a lightweight classifier to separate common beat morphologies.

Because the Kaggle dataset consists of isolated beats without continuous time stamps, this work focuses on the within-snippet localization quality and per-beat morphology analysis rather than heart-rate or HRV estimation.

2. Dataset

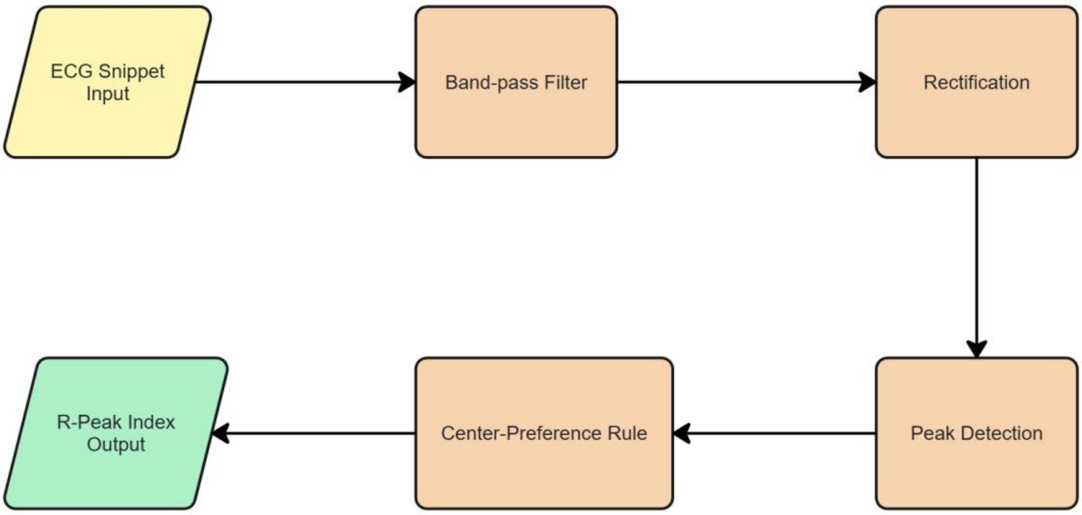
Source: Kaggle "ECG Heartbeat Categorization" (derived from MIT-BIH).

Dataset used:

- mitbih_train.csv and mitbih_test.csv from *Kaggle ECG Heartbeat Categorization* collection (derived from MIT-BIH).
- Each CSV row is a single heartbeat snippet (~187 samples) followed by a numeric class label (0-4). Labels correspond to canonical MIT-BIH beat types: Normal, Supraventricular, Ventricular, Fusion and Unknown.
- Since, snippets are isolated, inter-beat timing information Is not available, therefore heart-rate/HRV analysis is outside this projects scope.

Notes: snippets are independent (no continuous order or timestamps). Because of this, clinical metrices based on inter-beat intervals (HR,HRV) cannot be computed from CSVs alone. The project therefore reports localization and morphology/classification metrices appropriate to segmented data.

3. Method / Algorithm



1. **Band-pass filter:** 2^{nd} order Butterworth band-pass with 5-25 Hz (applied using zero-phase filtering `filtfilt` when possible). This removes baseline wander and high frequency noise while preserving QRS energy.
2. **Rectification:** Absolute value of the band-passed waveform (**`abs(x)`**) so both positive and negative QRS deflections produce strong peaks.
3. **Peak detection:**
   - **`findpeaks`** is applied to the absolute band passed signal. The key parameter is **MinPeakProminence**, set adaptively per snippet as `prom_scale × dynamic_range` (where `dynamic_range = max(abs(xf))-min(abs(xf))`) this provides local thresholding that adapts to beat amplitude.
   - A minimum peak distance (`min_dist_ms`) prevents multiple detections inside a single QRS.
4. **Canter-preference Rule:** Most dataset snippets are approximately centered on the R peak. When multiple candidate peaks are present, the detector prefers peaks within

±`center_window_ms` of the snippet centre, otherwise the most prominent candidate is chosen.

5. **Fallback:** If no peak meets threshold, the global maximum of the band-passed signal is used (ensures an index is always returned).

Key Parameters Used:

- Sampling Frequency `fs = 360 Hz.`
- Bandpass `: 5-25 Hz, order = 2.`
- `Prom_scale = 0.25` (per-snippet prominence fraction).
- `min_dist_ms = 120ms.`
- `center_window_ms = 80ms.`
- Central Window for evaluation: `W = +-20 samples` (~55ms at 360 Hz).

4. Results

4.1. Localization metrices(detector):

Training set (mitbih_train.csv) – central window = +-20 samples

- Detection rate: **100%**
- Overall central-window accuracy: **60.01%**
- Per-label central accuracy:
  - Label 0: 58.71%
  - Label 1: 46.42%
  - Label 2: 47.44%
  - Label 3: 8.11%
  - Label 4: 95.85%

Figure 1: Distribution of detected R sample index across all snippets

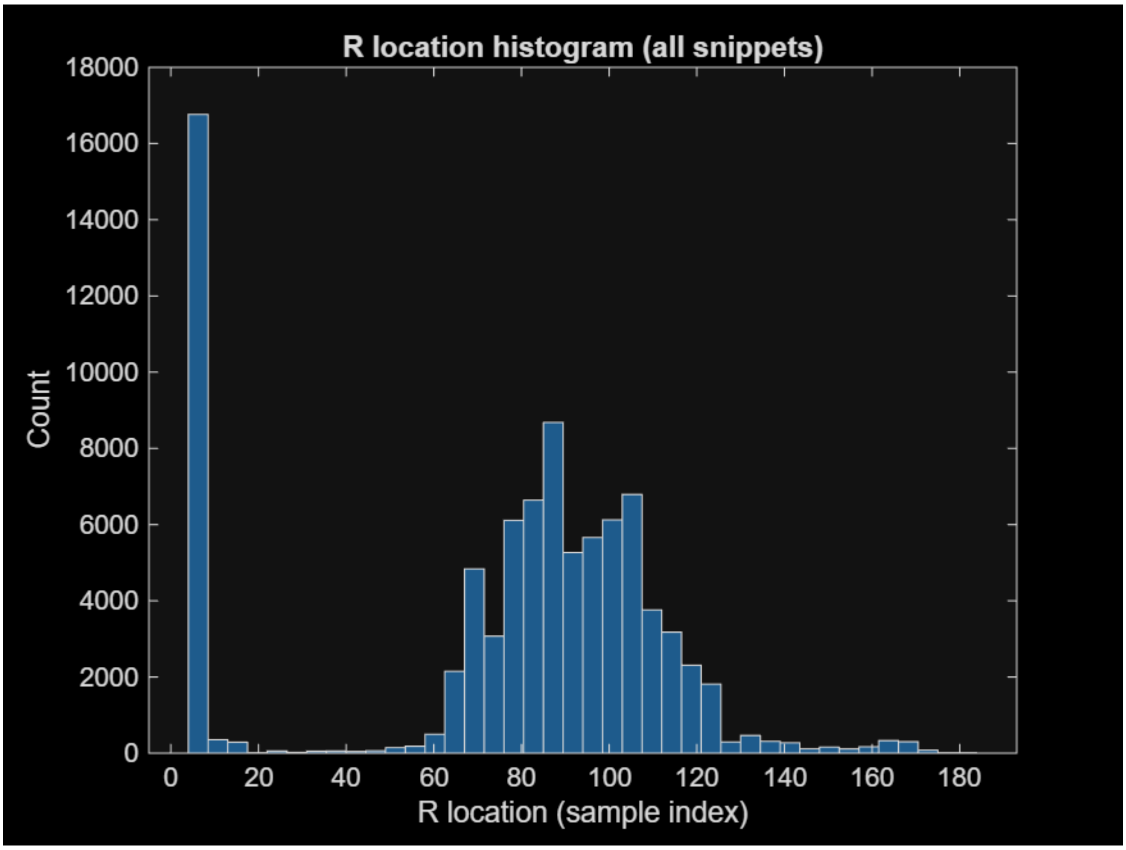


Figure 2: R-location histogram by class, shows class-specific centring and spread.

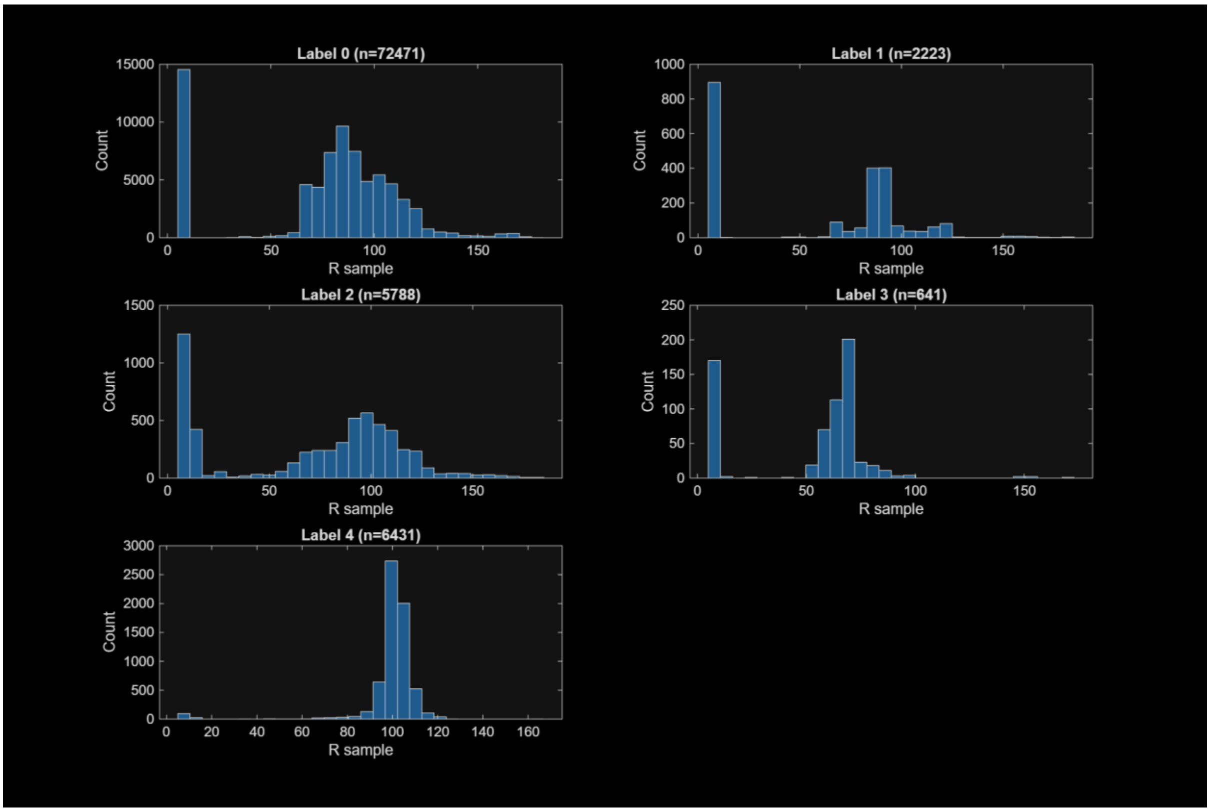


Figure 3: Aligned mean waveform per class after shifting detected R to center (±1 std shaded).

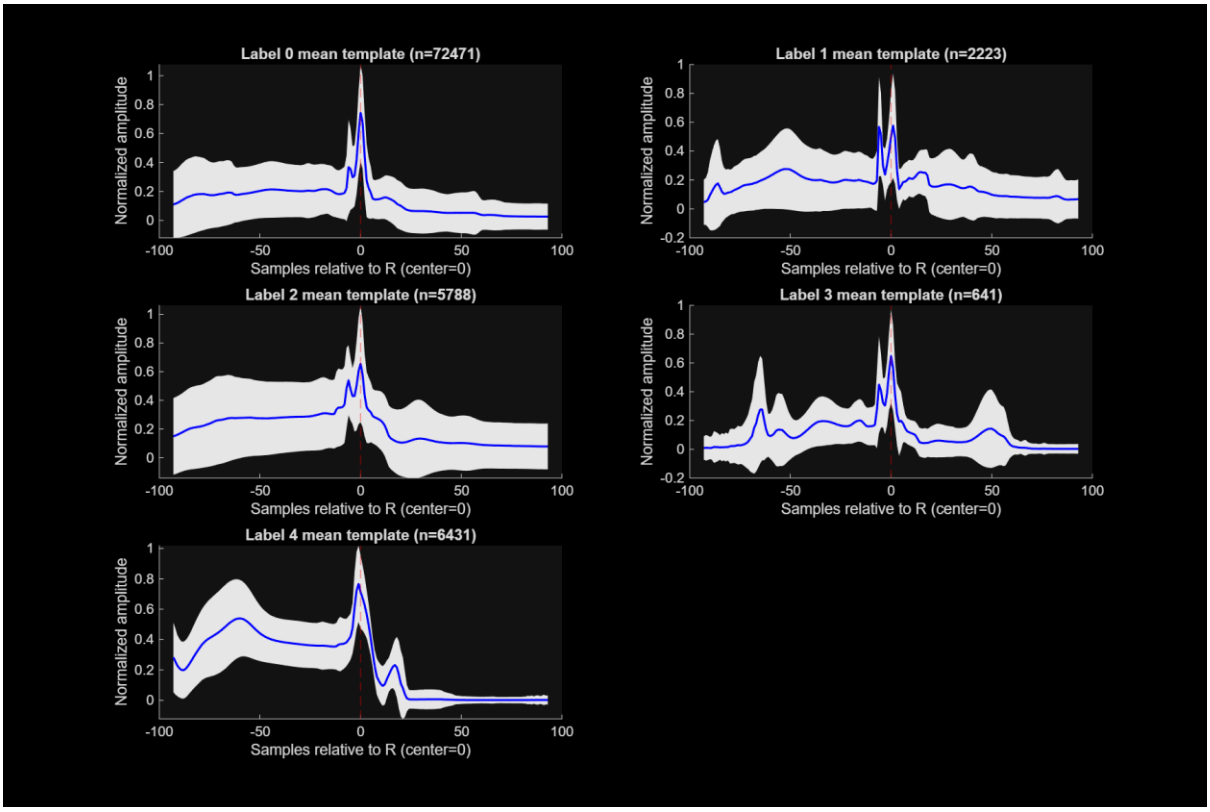
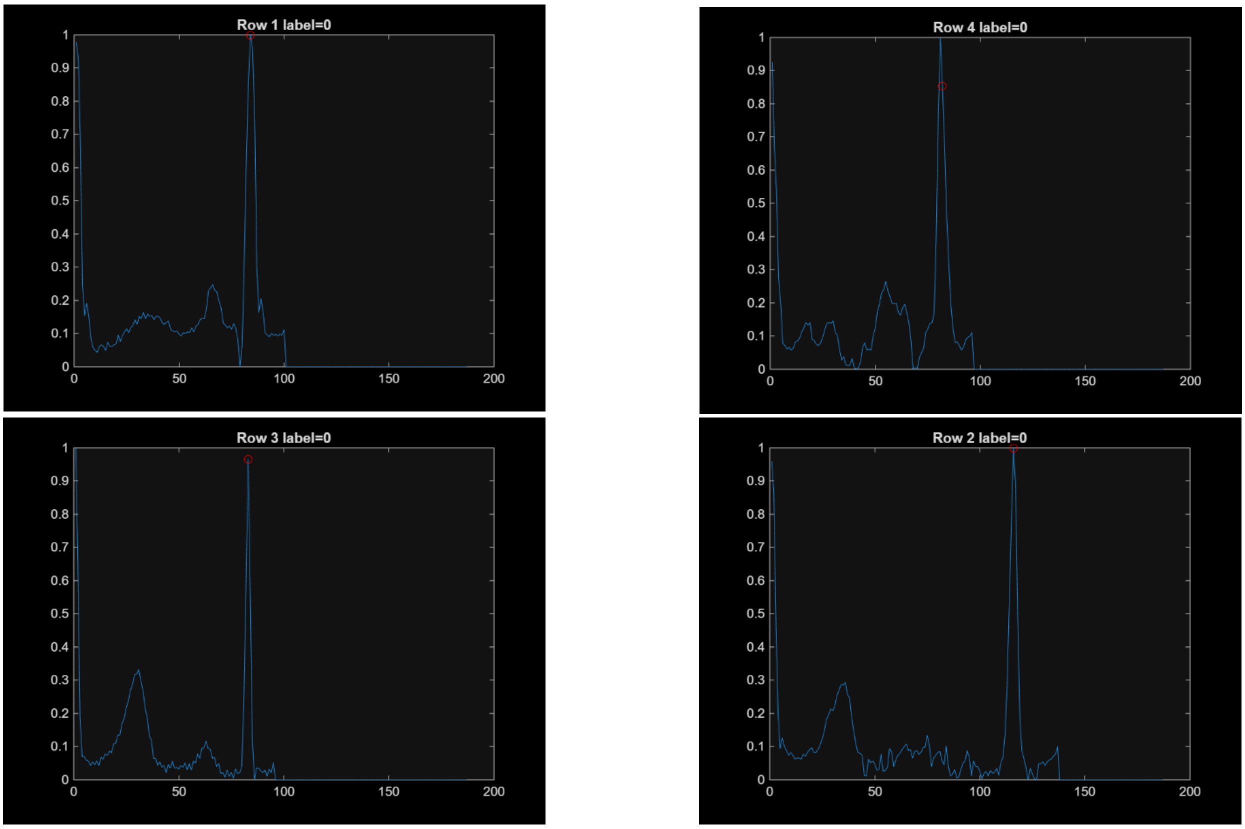


Figure 4: Detected signal Examples



4.2. Heartbeat Feature Analysis & Classification

Features Extracted: R amplitude, QRS width (ms), peak-to-peak, QRS area, RMS energy, skewness, kurtosis, spectral centroid (Hz), spectral Bandwidth (Hz).

Classifier: multi-class linear SVM (MATLAB `fitcecoc` with linear SVM learners), trained on the training set features and evaluated on the test features.

Classification Performance:

- **Accuracy:** 89.91%
- **Per-label precision / recall / F1:**

Classification accuracy on test set: 0.8991

Label	Precision	Recall	F1
0	0.91374	0.98984	0.95027
1	0	0	0
2	0.69476	0.42127	0.52451
3	0	0	0
4	0.8212	0.70833	0.7606

Figure 1: Classification Confusion matrix

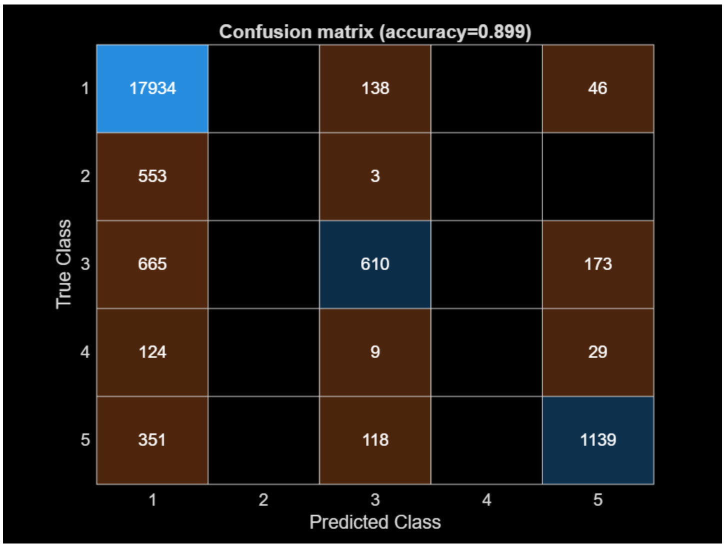


Figure 2: Boxplots for R_amp, QRS_width_ms, RMS by label

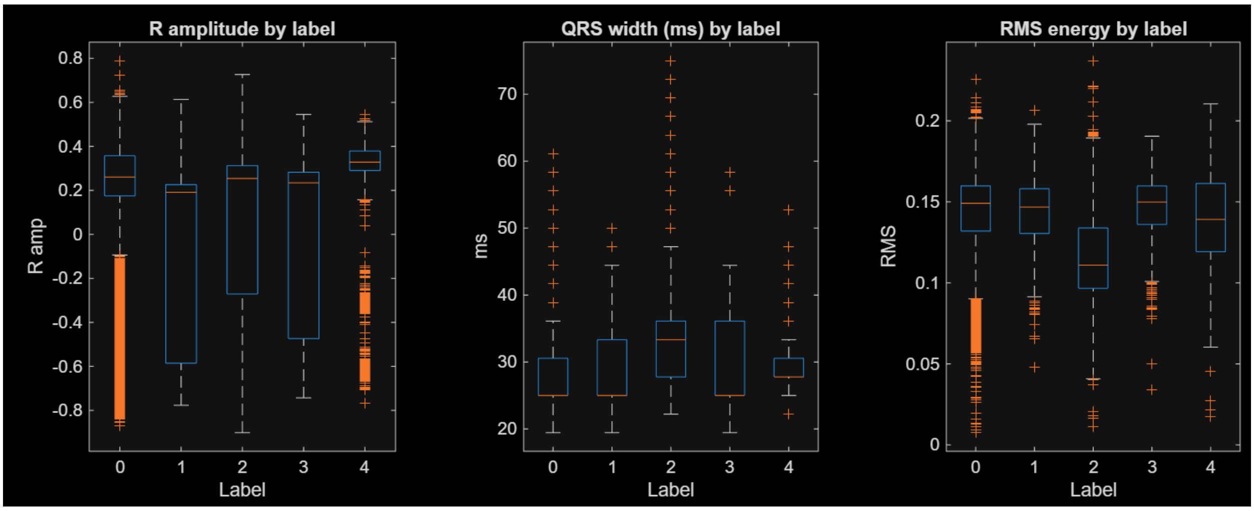
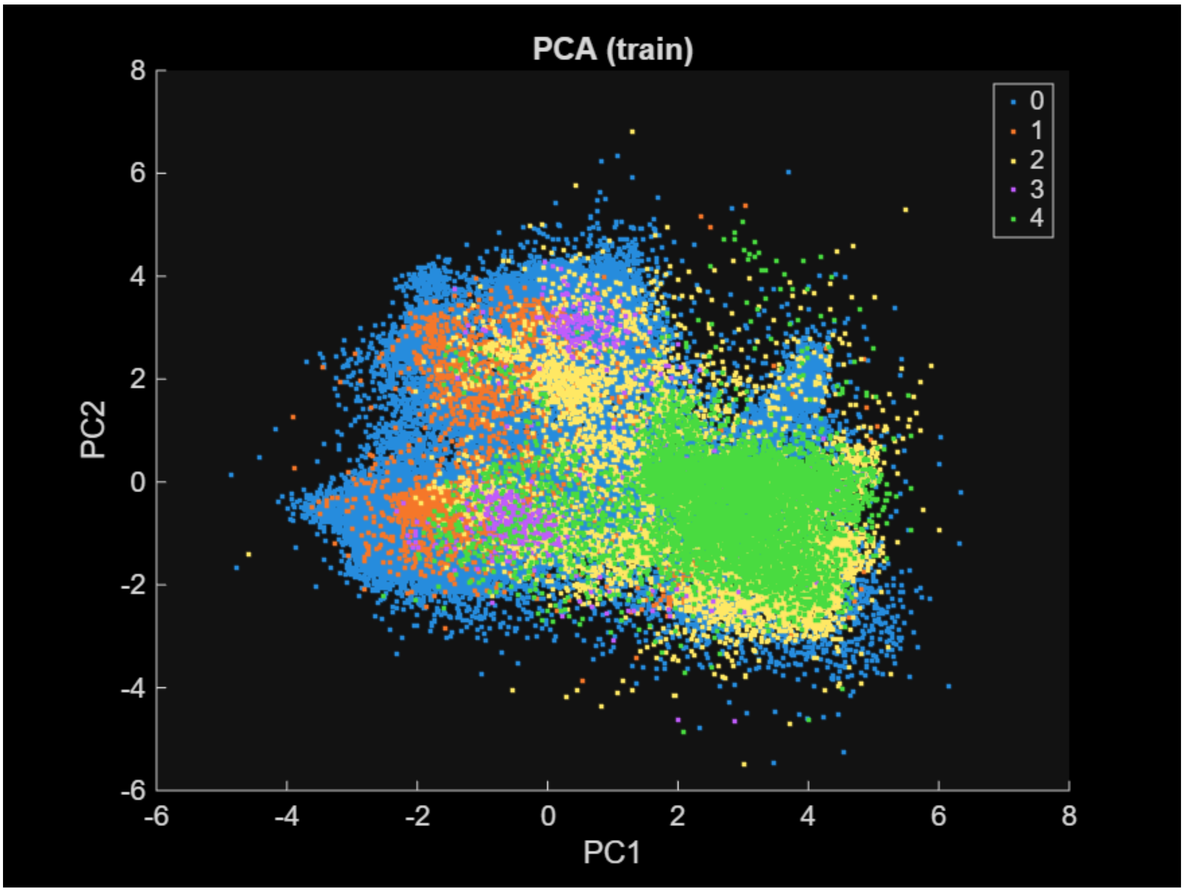


Figure 3:PCA 2-D embedding coloured by label



5. Observation

- The detector reliably returns an index for every snippet, aided by fallback to global maximum, giving 100% detection. However, localization precision (centering) is moderate (~60% within +-20 samples).
- **Class 4** exhibits consistently sharp, high-amplitude, and wide QRS complexes leading to excellent localization and classification accuracies (~96% central accuracy; F1=0.76).
- **Class 0** (normal beats) is also well handled (F1=0.95).
- **Classes 1 and 3** show very poor localization/classification. Causes: class imbalance (few examples), overlapping morphology with other classes, low R peak prominence in many examples, and possibly inconsistent snippet centering.
- The classification demonstrated that simple DSP features carry substantial discriminative power, a linear SVM reaches ~90% overall accuracy, mostly driven by separation of major classes (0 and 4).
- The nearly identical train/test localization accuracies (~60%) indicate consistent behaviour across splits and no obvious overfitting of the detector's fixed rule.

6. Limitations and Future Work

Limitations

- Dataset is segmented: no continuous timestamps; HR/HRV cannot be computed here.
- Simple prominence-based detection can mislabel high-contrast artifacts as R peaks.
- Class imbalance limits classifier performance on minority classes.

Future Work and improvement

- Test on continuous MIT-BIH records with annotation files (PhysioNet) to compute true sensitivity and PPV with ±150 ms tolerance.
- Implement a full Pan–Tompkin’s pipeline or wavelet-based detector to improve QRS discrimination.
- Use template matching or a small classifier to pick the correct candidate among several peaks.
- Per-class parameter tuning (prominence, center window) or a label-aware detector may significantly reduce mis localizations for problematic classes.
- Consider Deep learning models (1-D CNN) trained end-to-end if labelled continuous records are available.

7. Conclusion

We implemented a concise DSP pipeline in MATLAB that localizes R peaks inside isolated ECG snippets using band-pass filtering and prominence-based peak selection. The detector returns an index for every snippet and yields ~60% within-snippet localization accuracy (±20 samples) on both train and test splits, with substantial variation across beat classes. Simple DSP features combined with linear SVM achieve ~90% accuracy for beat classification, illustrating the value of traditional signal processing features. The project provides a clear, reproducible baseline and identifies practical directions for improving localization and minority-class classification.

8. References

1. J. Pan and W. J. Tompkins, “A Real-Time QRS Detection Algorithm,” IEEE Transactions on Biomedical Engineering, 1985.
2. PhysioNet / MIT-BIH Arrhythmia Database. https://physionet.org/content/mitdb/1.0.0/
3. Kaggle ECG Heartbeat Categorization dataset (derived from MIT-BIH).