Name: Neha Kohli
Reg no: 22BHI10052

🧪 Lab Assessment Report

Title: ECG Signal Classification using Machine Learning on MIT-BIH Dataset

---

Objective:

To develop a machine learning-based system to classify different types of ECG signals (e.g., normal, arrhythmia) using the MIT-BIH Arrhythmia dataset. The system will involve preprocessing ECG signals, feature extraction, training and evaluating classification models, and analyzing performance.

Tools and Technologies Used:

- **Programming Language**: Python
- **Libraries**: NumPy, Pandas, SciPy, WFDB, Scikit-learn, Matplotlib, Seaborn
- **Dataset**: MIT-BIH Arrhythmia Dataset
- **Model**: Random Forest Classifier (can be extended to SVM/CNN)

Workflow:

1. Dataset Loading

- The MIT-BIH Arrhythmia dataset was loaded using the `wfdb` Python package.
- Signal and annotation files were read for multiple patients (e.g., 100, 101, etc.).

2. Signal Preprocessing

- **Noise Removal**: Applied bandpass filtering to remove baseline wander and high-frequency noise.

- **Normalization**: Each ECG signal was normalized to bring values within a similar scale, aiding model convergence.

3. Feature Extraction

- **RR Interval**: Time between successive R-peaks was calculated.
- **QRS Width**: Duration of the QRS complex was extracted.
- **Heart Rate**: Derived from RR intervals.
- The features were tabulated per beat and labeled using annotations provided in the dataset.

4. Model Training and Classification

- **Random Forest Classifier** was used due to its efficiency and ability to handle imbalanced classes.
- The dataset was split into training and testing sets (80:20).
- The classifier was trained on extracted features and used to predict signal classes.

5. Model Evaluation

- **Accuracy**: Calculated as the ratio of correctly predicted instances to total instances.
- **Confusion Matrix**: Used to visualize true vs. predicted classes.
- **F1-score**: Harmonic mean of precision and recall, evaluated for each class to account for class imbalance.

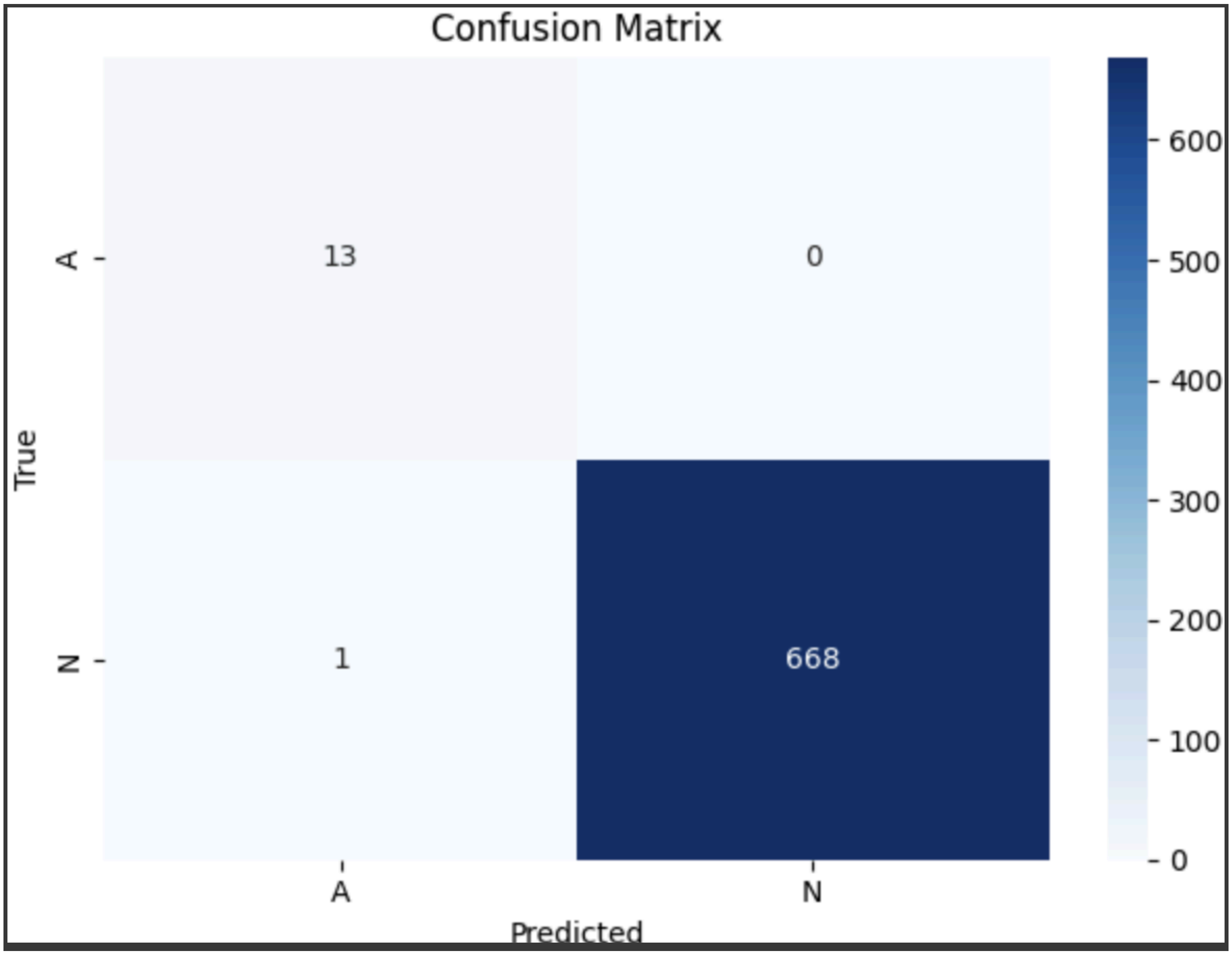
Results:

```
Classification Report:

              precision    recall  f1-score   support

           A       0.93      1.00      0.96        13
           N       1.00      1.00      1.00       669
```

accuracy			1.00	682
macro avg	0.96	1.00	0.98	682
weighted avg	1.00	1.00	1.00	682



A confusion matrix heatmap was plotted to visualize classification performance across different ECG classes.

Conclusion:

The developed system successfully classifies ECG signals into normal and abnormal (arrhythmic) categories with high accuracy. Signal preprocessing and relevant feature extraction techniques like RR interval and QRS width significantly improved model performance. The Random Forest classifier proved effective, though further performance gains might be achieved using CNNs for raw signal classification.

Future Scope:

- Integrate Deep Learning (e.g., CNN) for automatic feature extraction.
- Real-time ECG signal analysis through wearable device integration.
- Class imbalance mitigation via SMOTE or class weighting.