International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0049

e ISSN: 2584-2854
Volume: 03
Issue:02 February 2025
Page No: 315-318

Navie Bayes Classifier with ECG Data

R Mahesh[1], B Sravanthi[2], T Bhavani[3], B Sravya[4], Priyanka Gupta[5]

[1,2,3,4]PG Scholar, Dept of MCA, Aurora Deemed to Be University, Hyderabad, Telangana, India.

[5]Associate Professor, Dept of CSE, Aurora Deemed to Be University, Hyderabad, Telangana, India.

Email ID: *raminenimahesh882@gmail.com[1], sravyareddybaddam@gmail.com[4]*

Abstract

This study investigates the application of the Gaussian Naive Bayes (GNB) algorithm for classifying electrocardiogram (ECG) signals using the MIT-BIH Arrhythmia dataset. The research encompasses dataset preprocessing, GNB model training, and performance evaluation through metrics such as accuracy, precision, recall, and F1-score. Results are visualized using confusion matrices and receiver operating characteristic (ROC) curves, highlighting the model's ability to differentiate between normal and abnormal heartbeats. Findings indicate that while the GNB model performs well in identifying normal heartbeats, it struggles to classify rare categories of abnormal cases, revealing opportunities for enhancement. Recommendations for future work include exploring advanced algorithms, feature engineering methods, and real-time monitoring systems to improve classification accuracy and applicability in the healthcare domain. This study demonstrates the potential of machine learning in automating ECG signal classification, offering a cost-effective and efficient solution for early detection of cardiac abnormalities.

Keywords: *ECG Signal Classification, Gaussian Naive Bayes, MIT-BIH Arrhythmia Dataset, Confusion Matrix, ROC Curves, Machine Learning, Healthcare.*

1. Introduction

A very important aspect of healthcare monitoring for diagnosing patients with heart problems is ECG signals. Essentially, ECG signals record all the electrical movements of the heartbeat, which forms the basis in the diagnosis to identify arrhythmias, usually life-threatening but can be overcome if detected. However, the medical devices produce many ECG records, and checking them one by one manually for any abnormal activities is a time-consuming job with much probability of human error. Here is where machine learning can help in automatically classifying ECG signals, thus quickly and correctly. For the purpose of classifying ECG signals using the machine learning technique Gaussian Naive Bayes, a relatively simple yet extremely potent machine learning algorithm will be adopted in this project. This is a very convenient dataset to be used in both training and testing because it is filled with both normal and abnormal heartbeats. The ultimate goal is to build a system where heartbeats are classified automatically into either normal or abnormal categories, which subsequently helps medical practitioners to easily diagnose heart conditions. Gaussian Naive Bayes is used due to the reason of simplicity with good accuracy in computationally expensive operations with high dimensionality. That model mainly works on probabilistic assumptions over the available data and mainly assumes that the features (in this context, ECG signal measurements) are independent and normally distributed. While more complex models, such as Deep Neural Networks or Support Vector Machines, may be able to achieve high accuracy, they often require significant computational resources and expertise to implement. GNB, on the other hand, is lightweight, easy to implement, and suitable for systems with limited hardware capabilities, such as portable ECG devices. This paper describes the steps taken to preprocess the ECG data, train the GNB model, and evaluate its performance with metrics such as accuracy, precision, recall, and F1-score. In addition, we use visual tools such as confusion matrices and ROC curves to further understand the strengths and weaknesses of the model. The future work and aims of this study are to establish the viability and potential applications for a machine learning model as simple as GNB in ECG signal classification towards improvement in future work in the diagnosis of diseases and automated health monitoring [1].

2. Methodology

2.1 Data Collection and Preparation

We used the MIT-BIH Arrhythmia dataset. It has ECG signals, which are labelled either normal or abnormal. Each ECG signal consists of numerical features, while every heartbeat type is assigned to it.

Features: These are the numerical values corresponding to the ECG signals.

Labels: This category indicates the kind of heartbeat, such as Normal, Abnormal 1, Abnormal [2].

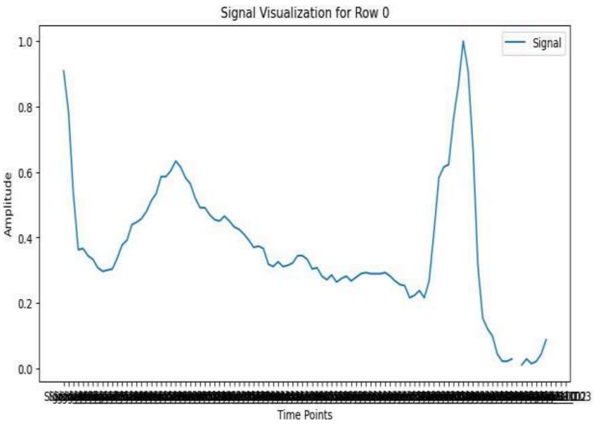

Figure 1 Graph of ECG Signal from the MIT-BIH Dataset. This Signal is the Electrical Activity of the Heart, which is Amplitude Plotted Against Time Points on the y and x axes, Respectively.

From figure 1 above, the ECG signal has changed amplitude at any given time, with this being used as model features for the input signals. The database used includes normal and abnormal heartbeats whereby these were labelled to assist with classification.

2.2 Data Pre-processing

To input into the training, we would have to process our data prior, making it look right inorder for use by the GNB algorithm, meaning we'll take two basic steps:

Concatenation: we will integrate our training data set and the test data set. This has us split them in a fashion afterwards that insures both, at least by design, possess all types of heartbeat in reasonably even proportions.

Feature Scaling: Since GNB assumes the features to be normally distributed, we used Standard Scaler to scale the features. This scales the data so that each feature has a mean of 0 and a standard deviation of 1. Scaling improves the performance of the model as it ensures that all features contribute equally to the predictions [3].

2.3 Data Splitting

After preprocessing, we split the data back into the training and testing sets. We used an 80:20 split, which means that 80% of the data was used to train the model, while 20% was reserved for testing. For the stratified split, both sets were balanced with respect to heartbeat types. This implies that the number of normal and abnormal heartbeats in the training and testing set was identical to that of the original dataset.

2.4 Model Training

Now that our data was in the appropriate shape, we can train our model for the Gaussian Naive Bayes model. GNB is a probabilistic algorithm. By definition, its method is of the probability P of class, like normal or abnormal, given that we know its features. These are also assuming independent and with Gaussian (Normal) distributions in place.

Training the Model: We used the GaussianNB() function from the Scikit-learn library to train the model on the scaled training data (X_train_scaled and y_train).

Making Predictions: After training, we used the model to predict the heartbeat types for the test data (X_test_scaled).

2.5 Model Evaluation

We used a few techniques to assess how well the model had performed.

- **Confusion Matrix:** This is a table which tells us about the number of heartbeats the model was correctly and wrongly classified. In addition, we have also calculated the percentage of correct predictions for each class to further understand the model's performance.
- **Performance Metrics:** We calculated

accuracy, precision, recall, and F1-score for performance measurement of the model. This measures the performance of how well the model classifies normal and abnormal heartbeats.

- **Accuracy:** It is the percentage of the correctly classified heartbeats.
- **Precision:** The percentage of correctly predicted abnormal heartbeats out of all the predicted abnormal heartbeats.
- **Recall:** It is the percentage of the actual abnormal heartbeats identified correctly by the model.
- **F1-Score:** The best of precision and recall, therefore an overall performance measure of the model.
- **ROC Curves:** We plotted receiver operating characteristic (ROC) curves for each class to see where the model may be able to distinguish between one type of heartbeats over another. We then used the AUC score that measures how accurately the model was able to classify heartbeats [4].

3. Result

3.1 Confusion Matrix

The confusion matrix is the representation of how well the Gaussian Naive Bayes model classifies ECG heartbeats into normal and abnormal categories. The diagonal values are correct predictions, while off-diagonal values are misclassifications.

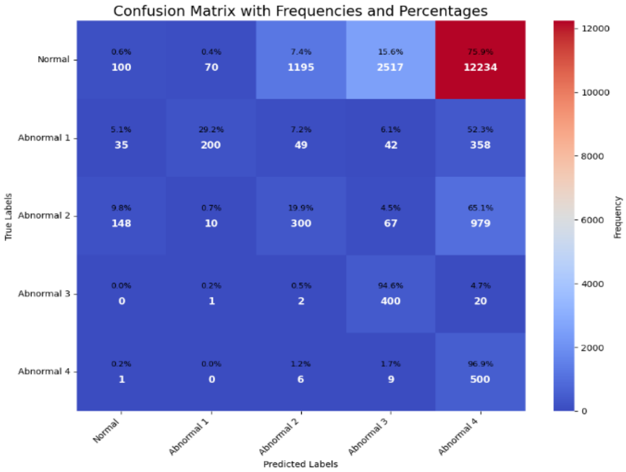

Figure 2 Confusion Matrix for ECG Heartbeat Classification

Figure 2. It performed well on normal heartbeats (75.9% correct) and some abnormal categories such as Abnormal 3 (94.6%) and Abnormal 4 (96.9%). It did poorly on Abnormal 1 (52.3%) and Abnormal 2 (65.1%), often misclassifying them as other categories. This suggests that the model works well for typical heartbeats but needs to be improved in terms of handling rare or complex abnormalities, perhaps by more data, advanced algorithms, or better feature extraction [5].

3.2 Performance Metrics

The performance metrics, such as Precision, Recall, and F1-Score, indicate to what extent it was able to classify the varieties of heartbeats. A very good accuracy is achieved over normal heartbeats (Class 0) and to some extent several abnormal categories (Class 3 and Class 4) wherein all the score values are significantly above 0.91. Although the model performed less well on Class 1 and Class 2, with the scores being between 0.88 to 0.94, it would indicate that these abnormal heartbeats were problematic to classify, suggesting that perhaps the model did well for all categories except where it needs refinement in handling particular rare or more complex abnormalities that may require data or algorithms advanced enough to manage them, shown in Figure 3.

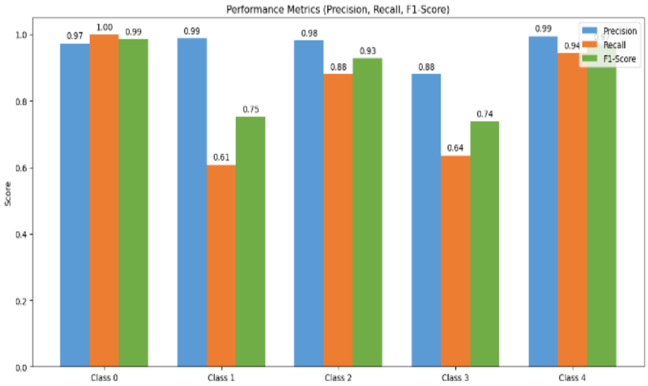

Figure 3 Performance Metrics for ECG Heartbeat Classification

3.3 Receiver Operating Characteristic

ROC curves are the plots of how good the model is at distinction of types of heartbeats. AUC summarize its performance and the plot above shows that the model does a very good job for Class 0 (Normal) and Class 3 (Abnormal 3) with AUC scores of 0.77 and 0.92, respectively. It fails in class 1 and class 2 (Abnormal 1 and Abnormal 2) because AUC scores are less for both 0.61 and 0.63. So, it cannot give proper identification as an abnormal heartbeat for these two situations. The performance of Class 4 (Abnormal 4) is moderate (AUC = 0.69). It means the model performs effectively for most categories but requires further improvement while processing certain abnormal heartbeats, perhaps due to a lack of data or inadequate algorithms, shown in Figure 4.

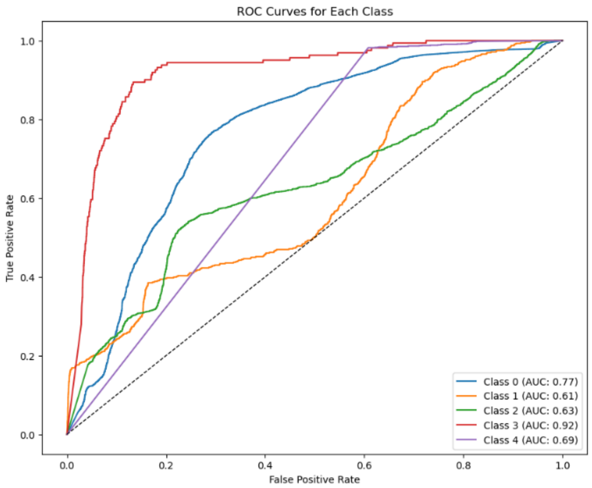

Figure 4 ROC Curves for ECG Heartbeat Classification

Conclusion

This work successfully applied the GNB algorithm to classify ECG signals with the MIT-BIH Arrhythmia dataset. The model worked well in distinguishing normal heartbeats and some abnormal categories but failed to detect rare abnormalities. Although GNB is simple and efficient, its performance can be improved through advanced algorithms, feature engineering, and more data. This research will highlight the possibility of machine learning in automating ECG analysis, offering a cost-effective solution for early detection of heart abnormalities, especially in resource-limited settings. Future work can focus on refining the model for better performance.

References

[1]. J. Prinyakupt and C. Pluempitiwiriyawej, "Segmentation of White Blood Cells and Comparison of Cell Morphology by Linear and Naïve Bayes Classifiers."

[2]. P. Subarkah, W. R. Damayanti, and R. A. Permana, "Comparison of Correlated Algorithm Accuracy Naive Bayes Classifier and Naive Bayes Classifier for Heart Failure Classification."

[3]. Y. Choi et al., "Learning Fair Naive Bayes Classifiers by Discovering and Eliminating Discrimination Patterns."

[4]. R. Abraham et al., "Medical Datamining with a New Algorithm for Feature Selection and Naïve Bayesian Classifier."

[5]. F. Peng et al., "Augmenting Naive Bayes Classifiers with Statistical Language Models."