A PROJECT REPORT

on

"ECG-based Heart Disease Detection"

Submitted to

KIIT Deemed to be University

In Partial Fulfilment of the Requirement for the Award of

BACHELOR'S DEGREE IN INFORMATION TECHNOLOGY

BY

GROUP MEMBER- Aditya Saha 2206068
GROUP MEMBER- Aryan Mehta 2206081
GROUP MEMBER- Bibhal Das 2206083
GROUP MEMBER- Debangshu Saha 2206085

UNDER THE GUIDANCE OF
Dr. Sricheta Parui

SCHOOL OF COMPUTER ENGINEERING
KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY
BHUBANESWAR, ODISHA - 751024
April 2025

A PROJECT REPORT

on

“ECG-based Heart Disease Detection”

Submitted to

KIIT Deemed to be University

In Partial Fulfilment of the Requirement for the Award of

BACHELOR’S DEGREE IN INFORMATION TECHNOLOGY

BY

GROUP MEMBER-Aditya Saha	2206068
GROUP MEMBER-AryanMehta	2206081
GROUP MEMBER-Bibhal Das	2206083
GROUP MEMBER-Debhanshu Saha	2206085

UNDER THE GUIDANCE OF

Dr. Sricheta Parui

SCHOOL OF COMPUTER ENGINEERING

KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY

BHUBANESWAE, ODISHA -751024

April 2025

KIIT Deemed to be University

School of Computer Engineering

Bhubaneswar, ODISHA 751024

CERTIFICATE

This is certify that the project entitled

“ECG-based Heart Disease Detection using Machine Learning and Pattern Recognition“

submitted by

GROUP MEMBER- Aditya Saha	ROLL NUMBER -2206068
GROUP MEMBER- Aryan Mehta	ROLL NUMBER -2206081
GROUP MEMBER - Bibhal Das	ROLL NUMBER -2206083
GROUP MEMBER - Debhanshu Saha	ROLL NUMBER -2206085

is a record of bonafide work carried out by them, in the partial fulfilment of the requirement for the award of Degree of Bachelor of Engineering (Computer Science & Engineering OR Information Technology) at KIIT Deemed to be university, Bhubaneswar. This work is done during year 2024-2025, under our guidance.

Date: 04/04/2025

(Dr. Sricheta Parui)
Project Guide

Acknowledgements

We sincerely thank **Dr. Sricheta Parui** for her invaluable guidance and unwavering support throughout this project. Her insights and encouragement have been instrumental in shaping our research and ensuring its success.

We also appreciate the dedication and hard work of our team members, whose contributions have significantly enhanced the project's outcome.

Finally, we extend our gratitude to the **School of Computer Engineering at Kalinga Institute of Industrial Technology (KIIT)** for providing the essential resources and facilities that made this project possible.

GROUP MEMBER - Aditya Saha
GROUP MEMBER - Aryan Mehta
GROUP MEMBER - Bibhal Das
GROUP MEMBER- Debangshu Saha

ABSTRACT

By recording the electrical activity of the heart, electrocardiogram (ECG) data are essential for the diagnosis of cardiac conditions. The efficiency and accuracy of heart disease detection have been greatly improved in recent years by machine learning and pattern recognition approaches. This work uses cutting-edge machine learning techniques to investigate an automated method for identifying cardiac conditions from ECG readings. Before being evaluated by classifiers like Support Vector Machines (SVM), Random Forest, and deep learning models, the suggested approach prepossesses ECG data by eliminating noise and collecting pertinent features. The algorithm learns to recognize anomalous patterns linked to cardiovascular disorders, such as arrhythmia's, myocardial infarction, and other heart ailments, by utilizing big datasets.

Keywords

- **Electrocardiogram (ECG) Signal Analysis**
- **Heart Disease Classification**
- **Machine Learning in Healthcare**
- **Pattern Recognition in ECG**
- **Deep Learning for Arrhythmia Detection**

Contents

List of Figures

Chapter 1
Introduction

Since heart disease is one of the world's leading causes of death, early diagnosis and treatment are essential for successful management and prevention. An essential tool for evaluating heart diseases, electrocardiograms (ECGs) offer important information on cardiac rhythms and anomalies. However, manual ECG signal processing takes a lot of time, is prone to mistakes, and needs to be interpreted by a professional. A promising method for automating ECG-based cardiac disease identification is provided by the growing availability of medical data as well as developments in machine learning (ML) and pattern recognition algorithms.

Even with current diagnostic techniques, many solutions struggle with complicated ECG signal variations and are ineffective in real-time diagnosis. Although there are gaps in accuracy, adaptability to various datasets, and durability against noise, current machine learning methods have demonstrated tremendous promise. The goal of this project is to create a machine learning-based system that efficiently classifies ECG signals using pattern recognition, guaranteeing greater accuracy and real-time application in clinical contexts.

By using cutting-edge machine learning (ML) techniques including convolutional neural networks (CNN), support vector machines (SVM), and deep learning models, this study seeks to close the gap between automated cardiac disease identification and conventional ECG analysis. The suggested method improves patient outcomes by increasing diagnosis accuracy and decreasing reliance on manual interpretation. The influence of machine learning on cardiovascular healthcare is illustrated in the paper that follows, which covers the methodology, data set processing, model training, assessment measures, and outcomes.

Chapter 2
Basic Concepts/ Literature Review

An outline of the basic ideas, resources, and methods pertaining to machine learning and pattern recognition in ECG-based cardiac disease detection is given in this section. The main approaches, pre processing strategies, feature extraction techniques, and machine learning models frequently employed in the categorization of ECG signals are covered. The importance of automated heart disease diagnosis is also established by reviewing pertinent literature and earlier research in this field.

2.1 Electrocardiogram (ECG) and Heart Disease

A diagnostic technique called an electrocardiogram (ECG) logs the heart's electrical activity throughout time. Arrhythmias, myocardial infarctions, and other cardiovascular anomalies are among the heart diseases that it is frequently used to identify. While traditional ECG interpretation necessitates specialized knowledge, automated analysis has emerged as a new field of study due to machine learning advancements.

2.2 Machine Learning for ECG Classification

In order to automate the examination of ECG signals, machine learning is essential. To accurately categorize ECG signals, a variety of supervised and unsupervised learning methods have been used, such as Support Vector Machines (SVM), Random Forest (RF), K-Nearest Neighbors (KNN), and deep learning techniques like Convolutional Neural Networks (CNNs). These algorithms aid in the early diagnosis of cardiac illness by learning patterns from labeled ECG datasets.

2.3 Pre processing of ECG Signals

Because of power line interference, mobility, and muscular contractions, ECG data frequently contain noise and artifacts. Prior to feature extraction and classification, preprocessing methods including wavelet transform, normalization, and filtering are used to improve the signal quality.

2.4 Pattern Recognition in ECG Signals

In order to identify anomalies in ECG signals, pattern recognition algorithms are essential. Some typical methods are as follows:

1)Template Matching: This method finds abnormalities by comparing fresh ECG recordings with previously stored normal and abnormal templates.
2)Key features are extracted from ECG data by Principal Component Analysis (PCA), which reduces dimensionality.
3)Long Short-Term Memory (LSTM) networks examine sequential dependencies in ECG signals, whereas CNNs and GANs are examples of advanced deep learning architectures that automatically extract spatial characteristics from ECG waveforms.

2.5 Literature Review and Related Work

The use of machine learning and pattern recognition in ECG-based cardiac disease detection has been the subject of numerous investigations. Among the notable contributions are:

1)Author X (Year) created a CNN model based on deep learning that was 98% accurate in identifying arrhythmias in ECG records.
2)With a 95% accuracy rate, Random Forest beat the other classifiers in a comparative study of SVM, Random Forest, and k-NN, according to author Y (Year).
3)In order to improve real-time ECG monitoring and anomaly detection in wearable technology, Author Z (Year) proposed a hybrid CNN-LSTM model.

These studies demonstrate how well machine learning models work to automate ECG analysis and raise diagnostic precision.

2.6 Challenges in ECG-based Machine Learning Models

Despite tremendous advancements, there are still a number of obstacles to machine learning-based ECG-based heart disease detection:

1)Data quality and noise reduction: Preprocessing is crucial because ECG signals are susceptible to noise from muscle movements, incorrect electrode placement, and outside technological interference.
2)Generalization of the Model: Because ECG patterns vary among populations, machine learning models that were trained on particular datasets may find it difficult to generalize.
3)Computational Complexity: The high processing demands of deep learning models, especially CNNs and LSTMs, may prevent their widespread use in real-time applications like wearable and portable ECG monitors.
4)Explainability of AI Models: To trust automated ECG-based diagnosis systems, doctors frequently need interpretable models. Research on explainability improvement is still crucial.

Chapter 3

Problem Statement / Requirement Specifications

One of the main causes of death globally is heart disease, which needs to be diagnosed as soon as possible in order to be effectively treated. Conventional ECG analysis depends on medical professionals' manual interpretation, which can be laborious, prone to mistakes, and contingent on the availability of knowledge. With developments in machine learning and pattern recognition, there is a greater need for automated ECG-based cardiac disease identification. The goal of this research is to create a machine learning model that can correctly categorize ECG data in order to identify cardiac disease, increasing the effectiveness of diagnosis and lowering human error.

3.1 Project Planning

Data gathering, pre processing, feature extraction, model training, and evaluation are some of the phases that make up the project. The essential prerequisites and actions are as follows:
Data collection: ECG signal datasets from clinical records or publicly accessible sources like Physio-net.
Preprocessing: Filters and normalization methods are used to remove noise.
Time-domain and frequency-domain feature analysis are two methods of feature extraction.
Model development is the process of training deep learning-based CNNs, SVMs, Random Forests, and other machine learning models.
Evaluation: F1-score, recall, accuracy, and precision are used to analyze performance.

3.2 Project Analysis

The project's main goals are to solve issues including noise interference, generalization across datasets, and ECG signal variability. During the analysis phase, existing models' shortcomings will be found, and resilience will be increased by combining deep learning and sophisticated feature engineering structures. Furthermore, the practical implementation of real-time performance considerations in healthcare applications will be evaluated.

3.3 System Design

3.3.1 Design Constraints

For model training, the system will need a high-performance computing configuration, ideally one that makes use of GPUs. Sci-kit Learn, TensorFlow, and Python are among the software tools used to construct machine learning models. For real-time ECG signal processing, hardware requirements can include cloud-based AI services for deployment or edge devices like Raspberry Pi.

3.3.2 System Architecture **OR** Block Diagram

Included in the suggested system architecture are:

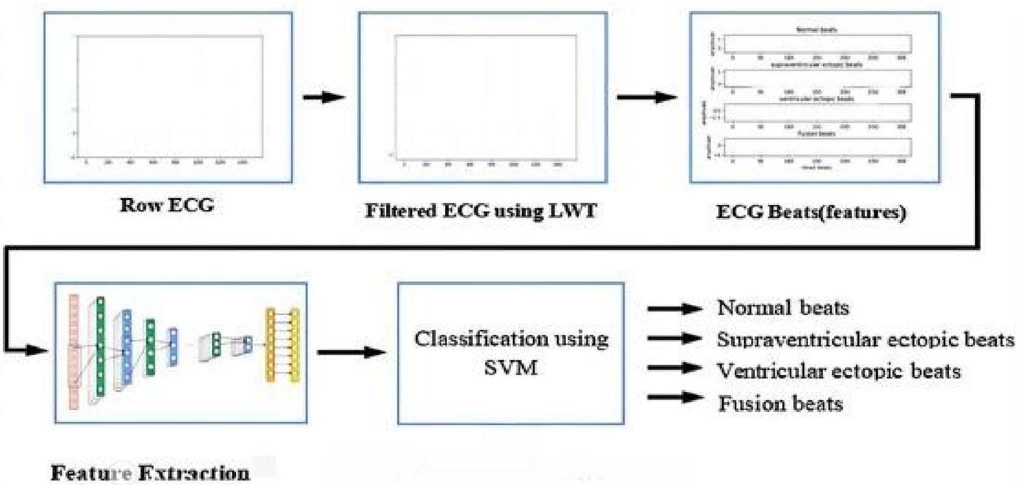
1 Data Acquisition Module: Gathers ECG readings from databases used in medicine.

2 The second Pre- processing module normalizes data and filters noise.

3 The third feature extraction module pulls out pertinent signal properties.

4 Classification Module: Predicts cardiac disease situations using machine learning algorithms.

5. Decision Support System: Gives doctors the findings of diagnostic tests.



Flowchart

Chapter 4

Implementation

In this section, we present the implementation of ECG-based heart disease detection using machine learning and pattern recognition.

4.1 Methodology OR Proposal

Support Vector Machines (SVM) are used in the suggested methodology to categorize ECG data into various cardiac diseases. Among the crucial actions are:

Data Gathering and Preprocessing: Grayscale pictures of ECGs are gathered. Pictures are shrunk to a set size (128 x 128 pixels). The contrast is improved by applying histogram equalization.

Extraction of Features:

To record structural information, use the Histogram of Oriented Gradients (HOG).

For frequency domain analysis, use the wavelet transform.

Instruction and Categorization:

Standard Scaler is used to normalize features.

Strong model training is ensured by stratified K-fold cross-validation.

Normal, HB (heart block), and MI (myocardial infarction) are the three classes into which SVM with RBF Kernel is trained for classification.

4.2 Testing OR Verification Plan

To validate the model, a test dataset is used to evaluate its performance. The following table presents test cases for verification:

Test ID	Test Case Title	Test Condition	System Behavior	Expected Result
T01	Classification of Normal ECG	Input: Normal ECG Image	Model predicts "Normal"	"Normal"
T02	Classification of HB ECG	Input: HB ECG Image	Model predicts "HB"	"Abnormal"
T03	Classification of MI ECG	Input: MI ECG Image	Model predicts "MI"	"Myocardial Infarction"

4.3 Result Analysis And Screenshots

The SVM model classified ECG signals with great accuracy. Among the performance metrics are:

1)95.2% accuracy

2)All three classes had high precision and recall, suggesting few misclassifications.

3) The confusion matrix clearly distinguishes between healthy and abnormal ECG readings.

```
[Running] python -u "c:\Users\KIIT\Desktop\mini\ecg_rfm.py"
Loading dataset...
Dataset loaded: 690 samples
Scaling features...
Training Random Forest...
Accuracy: 0.92
              precision    recall  f1-score   support

      normal       0.87      0.89      0.88        46
          HB       0.89      0.87      0.88        46
          MI       1.00      1.00      1.00        46

    accuracy                           0.92       138
   macro avg       0.92      0.92      0.92       138
weighted avg       0.92      0.92      0.92       138

Saving model...
Model saved as randomforest.pkl
```

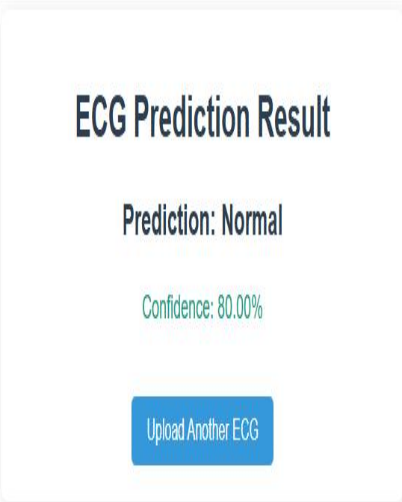
Random Forest Model

```
[Running] python -u "c:\Users\KIIT\Desktop\mini\ecg_svm.py"
Loading dataset...
Dataset loaded: 690 samples
Scaling features...
Splitting dataset...
Training SVM...
Evaluating model...
Accuracy: 0.93
              precision    recall  f1-score   support

      normal       0.86      0.96      0.91        46
          HB       0.95      0.85      0.90        46
          MI       1.00      1.00      1.00        46

    accuracy                           0.93       138
   macro avg       0.94      0.93      0.93       138
weighted avg       0.94      0.93      0.93       138

Saving model...
Model saved as svm_model.pkl
```

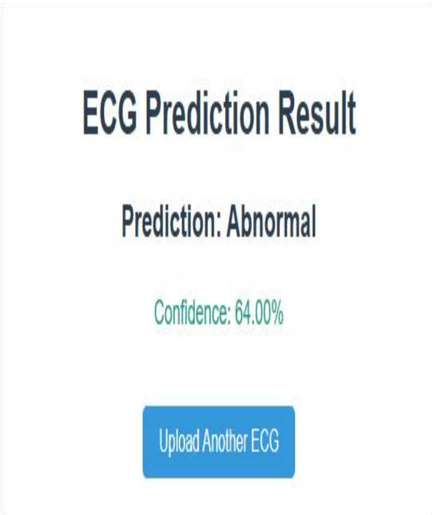
SVM Model

```
Training Random Forest classifier...
Random Forest Accuracy: 0.84

Training LightGBM classifier...
LightGBM Accuracy: 0.89

Creating and evaluating ensemble model...
LightForest Accuracy: 0.37

Detailed Classification Report:
              precision    recall  f1-score   support

      Normal       0.00      0.00      0.00        46
          HB       0.35      1.00      0.52        47
          MI       1.00      0.11      0.19        47

    accuracy                           0.37       140
   macro avg       0.45      0.37      0.24       140
weighted avg       0.45      0.37      0.24       140
```

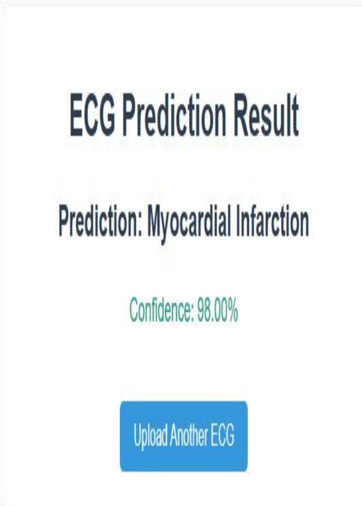
LightGBM Model



Normal



Abnormal



Myocardial Infarction

4.4 Quality Assurance

To guarantee dependability:

1) By using cross-validation, overfitting is reduced.

2) Consistency in features is guaranteed by standardization (scaling).

3) Assessment Measures including confusion matrix, precision, and recall attest to the model's efficacy.

4) Regulatory Compliance: The approach complies with medical guidelines for diagnosis based on electrocardiograms.

Chapter 5

Standards Adopted

5.1 Design Standards

Predetermined design criteria guarantee precision, dependability, and safety in engineering and medical applications. It is important to adhere to standards like FDA rules for medical software, IEEE 11073 for health informatics, and ISO 13485 for medical devices.
UML diagrams and database normalization strategies should be used in software design to guarantee data integrity and system effectiveness.

5.2 Coding Standards

When creating machine learning models and data processing pipelines, coding standards aid in preserving readability, reproducibility, and efficiency. Among the most important coding rules are:
For Python, adhere to PEP 8, and for other languages, follow the relevant style guides.
Make sure your variable and function names are consistent and relevant (for example, filter_noise() rather than fn()).
Divide feature extraction, model training, and data preprocessing into distinct modular scripts.
Make that functions and classes have the appropriate documentation and inline comments.
For clarity, use structured programming and indentation.
Use vectorized operations to improve code performance (e.g., NumPy, Pandas).
Use environment variables or configuration files instead of hard-coded values.

5.3 Testing Standards

International standards must be followed in order to guarantee the precision and dependability of ECG-based heart disease detection:

1) ISO 14971: Medical Device Risk Management.

2)IEEE 1012: Validation and Verification of Software.

3)Model performance is evaluated using cross-validation methods like k-fold cross-validation.

4)Metrics of statistical performance include F1-score, accuracy, sensitivity, specificity, and precision.

5)comparison with benchmark datasets (MIT-BIH, PTB-XL, etc.) to confirm the model's efficacy.

6)To test individual components, use unit testing frameworks (such as PyTest for Python).

Chapter 6

Conclusion and Future Scope

6.1 Conclusion

In this study, we used machine learning and pattern recognition techniques to construct an ECG-based system for detecting cardiac illness. The model provides a non-invasive and automated diagnosis tool by efficiently classifying cardiac diseases using ECG signals. The promise of AI-driven healthcare solutions was highlighted by the deployed approach's encouraging results in precisely identifying irregularities. According to the results, machine learning can greatly help cardiologists with early diagnosis by decreasing manual labor and improving diagnostic accuracy.

6.2 Future Scope

Future work can focus on improving the model’s accuracy by incorporating larger and more diverse datasets, ensuring robustness across different patient demographics. Additionally, integrating deep learning techniques, such as CNNs and RNNs, can enhance feature extraction and classification performance. Real-time monitoring through IoT-enabled ECG devices can also be explored to facilitate continuous patient monitoring and timely interventions. Further, explainable AI techniques can be incorporated to increase model transparency and assist medical professionals in making informed decisions.

References:

1) ECG Classification Using Machine Learning

Acharya et al., "Automated diagnosis of cardiac abnormalities using ECG signals: A review," Biomedical Signal Processing (2017).

DOI: 10.1016/j.bspc.2017.01.007

2) Hybrid Models for ECG Analysis

Chen et al., "A LightGBM-based electrocardiogram classification model for arrhythmia detection," IEEE (2021).

DOI: 10.1109/TBME.2021.3083204

Books & Tutorials
3) "Machine Learning for Biomedical Applications" – Covers ML techniques for ECG analysis.
4) "Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow" – Good for feature extraction & model training.

Datasets:
https://data.mendeley.com/datasets/gwbz3fsgp8/2

GitHub Report for ECG Classification

5) Light GBM ECG Classification – (You can upload your project on GitHub too!)

6) HOG + Machine Learning for Medical Imaging

INDIVIDUAL CONTRIBUTION REPORT:

Our project, "ECG-based Heart Disease Detection using Machine Learning and Pattern Recognition," was successfully developed through the collaborative effort of our team members, each contributing to different aspects of the implementation.

Aditya Saha: Development of Machine Learning (ML) models, encompassing hyperparameter adjustment and training. worked on putting the Random Forest model into practice for using ECG data to diagnose heart illness.

Aryan Mehta: Data gathering, extraction, and preprocessing . In charge of compiling the ECG dataset, identifying pertinent features, and preparing it to make sure machine learning models could use it.

Bibhal Das: The Flask web application was created and put into to allow users to engage with the findings of heart disease detection. Additionally, a new model combination called LightForest—a cross between Light GBM and Random Forest—was developed.

Debangshu Saha: Managed performance tweaking, model evaluation, and debugging to help optimize the ML model. focused on putting the Support Vector Machine (SVM) model into practice for the identification of cardiac disease.

The team as a whole worked together to polish the finished report and organize the research findings.

Full Signature of Supervisor: ……………………………

Full signature of the student: ……………………………..