A FIELD PROJECT REPORT

on

**“Heart Disease Detection Using Machine Learning And Deep Learning”**

**Submitted**

**By**

221FA04059

N. Kalyan Ram

221FA04205

N. Sai Deepika

221FA04168

D. Gagana Deepika

221FA04207

Ch. Sai Viswanth

***Under the guidance of***

***Dr.P.Jhansi lakshmi***

*Assistant Professor*



**SCHOOL OF COMPUTING & INFORMATICS**

**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**

**VIGNAN'S FOUNDATION FOR SCIENCE, TECHNOLOGY AND RESEARCH Deemed to be UNIVERSITY**

**Vadlamudi, Guntur.**

**ANDHRA PRADESH, INDIA, PIN-522213.**



**CERTIFICATE**

This is to certify that the Field Project entitled **“Heart Disease Detection Using Machine Learning And Deep Learning”** that is being submitted by 221FA04057(Kalyanram), 221FA04193(Gagana Deepika), 221FA04205(Deepika) and 221FA04207(Viswanth) for partial fulfilment of Field Project is a bonafide work carried out under the supervision of Dr.P.Jhansi Lakshmi, Assistant Professor, Department of CSE.

Guide name& Signature  Designation HOD,CSE



**DECLARATION**

We hereby declare that the Field Project entitled “**Heart Disease Detection Using Machine Learning And Deep Learning”** that is being submitted by 221FA04059 (Kalyanram), 221FA04168(Gagana Deepika), 221FA04205(Deepika) and 221FA04207(Viswanth) in partial fulfilment of Field Project course work. This is our original work, and this project has not formed the basis for the award of any degree. We have worked under the supervision of Ms. Dr. N. Sameera., Assistant Professor, Department of CSE.

By

**221FA04059 (Kalyan Ram),**

**221FA04168(Gagana Deepika),**

**221FA04205(Deepika),**

**221FA04207(Viswanth)**

## ABSTRACT

The project focuses on the detection of heart disease using machine learning and deep learning. Many people suffer from heart conditions that often go undiagnosed due to limited access to early and accurate detection methods. The main goal of this project is to develop a system that analyzes ECG signals and classifies them to identify potential heart diseases.

The system utilizes Convolutional Neural Networks (CNNs) trained on ECG waveforms to automatically extract and learn important features. Feature extraction is performed using deep learning techniques, while the accuracy of various traditional machine learning model Support Vector Machine (SVM) is tested. These models rely on manually extracted ECG features, offering a simpler and faster approach but with potentially lower accuracy.

To process ECG signals effectively, the system integrates advanced preprocessing techniques to enhance signal clarity and minimize noise. Once the ECG data is analyzed, the classification results provide insights into potential heart conditions, assisting in early detection and diagnosis.

This system provides a simple and effective solution for heart disease detection, enabling early diagnosis and improved patient outcomes. In the future, this project can be improved by developing hybrid models that combine traditional and deep learning approaches, scaling data records to enhance generalization, and integrating more sophisticated deep learning architectures for higher accuracy. Additionally, the system can be expanded into real-time monitoring applications or wearable devices, allowing continuous ECG analysis and timely health alerts.

**TABLE OF CONTENTS**

1. [Introduction 1](#_bookmark0)
   1. [Background and Significance of Heart Disease 2](#_bookmark1)
   2. [Overview of Machine Learning in Medical Diagnosis 2](#_bookmark2)
   3. [Research Objectives and Scope 4](#_bookmark3)
   4. [Current Challenges in Heart Disease Detection 5](#_bookmark4)
   5. Applications of ML to Heart Disease Detection 8
2. [Literature Survey 12](#_bookmark5)
   1. [Literature review 13](#_bookmark6)
   2. [Motivation 17](#_bookmark7)
3. [Proposed System 18](#_bookmark8)
   1. [Input dataset 20](#_bookmark9)
      1. Detailed features of dataset 20
   2. [Data Pre-processing 21](#_bookmark10)
   3. [Model Building 22](#_bookmark14)
   4. [Methodology of the system 24](#_bookmark16)
   5. [Model Evaluation 25](#_bookmark17)
   6. [Constraints 33](#_bookmark18)
   7. Cost and Sustainability Impact 48
4. [Implementation 51](#_bookmark19)
   1. Environment Setup 52
   2. Sample code for preprocessing and MLP operations 52
5. Experimentation and Result Analysis 54
6. Conclusion 56
7. References 58

**LIST OF FIGURES**

|  |  |
| --- | --- |
| Figure 1. Architecture of the proposed system | 24 |
| Figure 2. Various features in the dataset after Pre-Processing | 24 |
| Figure 3. Training Vs Testing Accuracy | 26 |
| Figure 4. Confusion Matrix | 26 |
| Figure 5. Performance Outcomes | 27 |
| Figure 6.  ROC Curve for Each Class | 28 |
| Figure 7. Logistic Regression-Confusion Matrix | 28 |
| Figure 8. Naïve Bayes-Confusion Matrix | 29 |
| Figure 9. Support Vector Machine (SVM) -Confusion Matrix | 29 |
| Figure 10. Random Forest-Confusion Matrix | 30 |
| Figure 11. XGBoost -Confusion Matrix  Figure 12. Decision Tree Visualization  Figure 13. Confusion Matrix for MLP Model | 30  31  55 |

**LIST OF TABLES**

Table 1. Table 1. Recorded Results for each Classifier 30

# CHAPTER-1 INTRODUCTION

### INTRODUCTION

### ****1.1 Background and Significance of Heart Disease****

Heart disease, or cardiovascular disease (CVD), is one of the leading causes of death worldwide, responsible for millions of fatalities each year. It develops due to factors such as high blood pressure, high cholesterol, smoking, obesity, diabetes, and genetic predisposition. Poor lifestyle choices, excessive alcohol consumption, and stress also contribute to its onset.

The most common type, coronary artery disease (CAD), occurs when plaque buildup narrows the arteries, restricting blood flow and potentially leading to heart attacks. Other forms include heart failure, where the heart struggles to pump blood efficiently, and arrhythmias, which cause irregular heartbeats. Some individuals are also affected by congenital heart defects from birth.

Heart disease poses a significant global health burden, with rising cases due to aging populations and lifestyle changes. Its economic impact is substantial, increasing healthcare costs through hospitalizations, medications, and surgeries, while also reducing workforce productivity. One of the biggest challenges is early detection, as many cases remain asymptomatic until they become severe. Despite advancements in diagnostic tools like electrocardiograms (ECG) and echocardiograms, late-stage diagnoses remain common.

Prevention efforts focus on promoting a healthy lifestyle through diet, exercise, and smoking cessation. Recent advancements in artificial intelligence and machine learning are improving early detection, with AI-driven ECG analysis and wearable health devices playing a crucial role. As research progresses, personalized treatment approaches based on genetics and lifestyle factors may enhance heart disease prevention and management.

**1.2 Overview of Machine Learning in Medical Diagnosis**

Machine learning (ML) is revolutionizing medical diagnosis by enabling computers to analyze vast amounts of medical data, recognize patterns, and predict outcomes. It provides healthcare professionals with powerful tools to enhance diagnostic speed, accuracy, and efficiency across various diseases.

**Machine Learning Applications in Medical Imaging and Diagnosis**

**Medical Imaging:**  
**Radiology:** Machine learning (ML) models, particularly deep learning approaches like convolutional neural networks (CNNs), are extensively used in analyzing medical images such as X-rays, MRIs, CT scans, and ultrasounds. These models assist radiologists in detecting diseases early, accurately identifying abnormalities like tumors, fractures, and lesions, which is crucial for early diagnosis and treatment.

**Ophthalmology:** ML is applied to analyze retinal images, helping in the early detection of eye conditions like glaucoma and diabetic retinopathy, thereby preventing vision loss and improving patient outcomes.

**Predicting and Diagnosing Diseases:**  
**Cancer Diagnosis:** Machine learning models are used to analyze patterns in genetic data, medical histories, and test results to identify biomarkers and predict the presence or risk of cancer. For instance, ML is applied in early breast cancer detection by analyzing mammograms and identifying signs of tumors that might be missed by the human eye.

**Cardiovascular Disease:** By evaluating clinical data such as blood pressure, cholesterol levels, and patient lifestyle, ML algorithms predict the risk of heart disease, enabling early intervention and preventive measures.

**Neurological Disorders:** ML is leveraged to analyze brain scans, cognitive tests, and behavioral data for diagnosing neurodegenerative diseases like Alzheimer’s. It helps in identifying early signs and predicting the progression of such disorders, aiding in timely intervention.

**Genomics and Pathology:**  
**Histopathology:** Machine learning models assist pathologists in analyzing tissue samples (biopsies), detecting malignant cells with greater precision and speed, which is essential for accurate diagnosis and treatment planning.

**Genomic Data Analysis:** ML in precision medicine is used to analyze complex genetic data, identifying mutations linked to diseases and creating personalized treatment plans based on a patient's genetic profile.

**Analytics for Prediction:**  
**Chronic Disease Management:** Machine learning models help monitor and predict the progression of chronic conditions like diabetes, hypertension, and kidney disease, enabling doctors to adjust treatment plans and intervene early to manage the diseases effectively.

**Sepsis Detection:** ML algorithms analyze clinical data and vital signs to predict the onset of sepsis in hospitalized patients. This early prediction helps healthcare providers initiate timely treatments, significantly reducing mortality rates.

**Natural Language Processing (NLP):**  
**Medical Records:** NLP techniques are used to extract meaningful information from unstructured medical records, such as patient histories, diagnoses, and treatment plans, streamlining clinical decision-making and improving patient care.

**Symptom Analysis:** ML-based chatbots and applications are now capable of understanding symptoms reported by patients and offering potential diagnoses or medical advice, assisting in initial consultations and guiding further medical evaluation.

### ****1.3 Research Objectives and Scope****

**Research Objectives:**  
The objectives of this research on machine learning and deep learning for ECG-based heart disease detection are as follows:

1. **Enhance Diagnostic Accuracy:** Develop machine learning and deep learning models to improve the precision of heart disease detection from ECG data, assisting in the early diagnosis of conditions such as coronary artery disease, arrhythmias, and heart failure. Deep learning techniques, particularly Convolutional Neural Networks (CNNs), will be explored for their ability to automatically learn hierarchical features from ECG waveforms.
2. **Predictive Modeling for Early Intervention:** Create predictive models that assess the likelihood of a patient developing heart disease based on ECG data, clinical history, lifestyle factors, and genetic predisposition. This will enable proactive interventions and help in early disease management.
3. **Reduce Diagnostic Time:** Investigate how deep learning models, by leveraging powerful computational capabilities, can expedite the analysis of ECG data, providing quicker diagnoses and reducing waiting times for results, thus improving the efficiency of heart disease management.
4. **Personalized Treatment Plans:** Examine how machine learning and deep learning can assist in crafting personalized treatment regimens for heart disease patients, considering ECG patterns, patient history, and clinical data. The goal is to enhance treatment efficacy by tailoring interventions based on individual health profiles.
5. **Expand Access to Diagnostic Tools:** Explore the application of machine learning and deep learning-powered diagnostic tools in low-resource and rural settings, where access to specialists and advanced diagnostic equipment is limited. This research aims to bridge healthcare gaps and make heart disease detection more accessible.
6. **Reduce Bias and Improve Generalization:** Focus on improving diagnostic accuracy across diverse populations by addressing potential biases in ECG data. Training deep learning models on diverse datasets ensures better generalization and reduces the risk of model bias, enabling equitable healthcare access.
7. **Integration with Clinical Workflow:** Investigate the seamless integration of machine learning and deep learning tools into existing clinical workflows, ensuring that healthcare practitioners can effectively use these models without disrupting established procedures, ultimately improving patient care efficiency.

**Research Scope:**

1. **Machine Learning and Deep Learning Algorithms:**  
   This study will examine various machine learning and deep learning techniques, such as deep learning models (e.g., convolutional neural networks for ECG waveform analysis), supervised learning methods (support vector machines, random forests, neural networks), and unsupervised learning techniques. The research will explore the benefits of deep learning in automatically extracting features from ECG data, improving diagnostic accuracy without manual feature engineering.
2. **Application in Medical Fields:**
   * **Cardiology:** The focus will be on heart disease detection using ECG data, with both machine learning and deep learning approaches applied to diagnose conditions like arrhythmias, coronary artery disease, and heart failure.
   * **Medical Imaging:** While the primary focus is ECG data, the research will briefly explore the integration of machine learning and deep learning for analyzing other diagnostic images (e.g., X-rays, CT scans) that could complement heart disease diagnosis.
   * **Genomics:** The study will explore how machine learning and deep learning can assist in identifying genetic markers related to heart disease, helping in the development of personalized treatment strategies.
3. **Sources of Data:**  
   The research will leverage data from ECG readings, clinical records (including patient history, lifestyle factors, and demographics), wearable devices (monitoring heart rate, blood pressure, etc.), and electronic health records (EHRs). Additionally, natural language processing (NLP) will be used to extract valuable insights from unstructured data, such as doctor’s notes and clinical records.
4. **Legal and Ethical Considerations:**  
   This study will address ethical concerns related to patient consent, data privacy, and the responsibility of healthcare providers when deploying machine learning and deep learning models in clinical practice. Compliance with regulations such as HIPAA and GDPR will be examined to ensure ethical use of medical data.
5. **Challenges and Limitations:**  
   The study will explore the challenges of implementing machine learning and deep learning models in medical diagnostics, including issues with data quality, model interpretability, and the need for large, diverse datasets. The potential for biases in the models and the need for explainable AI will be key considerations.
6. **Model Evaluation:**  
   Machine learning and deep learning models will be evaluated using metrics such as accuracy, precision, recall, F1 score, and ROC-AUC to ensure their effectiveness and reliability in real-world medical applications. This assessment will validate the clinical applicability of the models in diagnosing heart disease.
7. **Impact on Healthcare Systems:**  
   The research will assess the potential impact of machine learning and deep learning in heart disease diagnosis on healthcare systems. This includes improving patient outcomes, reducing diagnostic errors, increasing efficiency, and lowering healthcare costs through early detection and personalized treatment.
8. **Integration with Healthcare Technology:**  
   The study will explore how machine learning and deep learning tools can be integrated with existing healthcare technology, including cloud-based healthcare platforms, AI-driven diagnostic tools, and electronic health record (EHR) systems, to create a more efficient and accessible healthcare ecosystem.

Here’s the updated version, tailored to heart disease detection in your project context:

### ****1.4 Current Challenges in Heart Disease Detection****

The detection of heart disease, despite advancements in medical technology, still faces several significant challenges that impact its early diagnosis, treatment, and management.

1. **Late-Stage Diagnosis:**  
   Similar to many other diseases, heart disease often goes undetected in its early stages due to the absence of clear symptoms. Many individuals do not experience symptoms like chest pain or shortness of breath until the condition has progressed significantly. This delay in diagnosis makes it harder to implement early intervention and preventive measures.
2. **Inaccurate Diagnostic Procedures:**  
   Standard diagnostic procedures, such as ECGs, echocardiograms, and stress tests, can sometimes fail to detect heart disease in its early stages. These tests are highly dependent on the quality of the data captured and can sometimes provide false negatives or miss subtle signs of disease, particularly in the case of irregular heart rhythms or early signs of atherosclerosis.
3. **False Positives and Overdiagnosis:**  
   Just as in lung cancer detection, there is the possibility of false positives in heart disease detection. Screening methods can sometimes identify benign conditions or non-critical variations in the ECG data, leading to unnecessary tests, patient anxiety, and potentially invasive procedures like catheterization or angiography.
4. **Complexity of Heart Disease:**  
   Heart disease is not a single condition but rather a collection of diseases, including coronary artery disease, arrhythmias, heart failure, and valvular diseases. This heterogeneity makes it difficult to design a one-size-fits-all approach to detection and treatment. For example, arrhythmias may require different diagnostic methods and treatment protocols compared to coronary artery disease.
5. **Limitations of Traditional Screening Tools:**  
   While ECGs, blood tests, and imaging techniques like echocardiograms and CT scans are essential for heart disease detection, they have limitations. ECGs, for example, may not detect certain heart abnormalities, such as early-stage heart failure, while CT scans require radiation exposure, making them less suitable for long-term monitoring.
6. **Lack of Real-Time Monitoring and Early Detection:**  
   The current lack of widespread real-time heart disease monitoring, especially in individuals at risk, limits early detection. While wearables like smartwatches can track heart rate and ECG data, they often lack the diagnostic sophistication necessary to detect subtle signs of heart disease, such as arrhythmias or ischemia, early on.
7. **Healthcare Inequities and Access to Screening:**  
   Access to heart disease screening is often limited in rural and low-income areas. Economic barriers, including the cost of diagnostic tests and follow-up treatments, can prevent people from getting timely care. Additionally, healthcare disparities mean that certain populations, such as those in lower socioeconomic brackets, may not have access to the latest heart disease detection technologies or preventive care.
8. **Human Error in Interpretation of Test Results:**  
   Even though tests like ECGs and echocardiograms are widely used, the interpretation of these results is prone to human error. A radiologist or cardiologist may miss subtle irregularities in the data, leading to either misdiagnosis or delayed diagnosis. This issue is especially prevalent in large-scale screenings where time constraints and the volume of patients can hinder accurate interpretation.
9. **Integration of Machine Learning and Deep Learning Models:**  
   Machine learning and deep learning models hold great promise for improving heart disease detection by analyzing ECG data and other clinical parameters. However, these technologies are not yet fully integrated into mainstream clinical workflows. AI-powered tools require validation across large, diverse datasets to ensure generalization and reduce bias in diagnosing heart disease. Additionally, there is a lack of standardization and acceptance in clinical settings, making it challenging to adopt these technologies universally.
10. **Patient Reluctance and Public Awareness:**  
    There is often a reluctance among patients to undergo heart disease screenings, especially for those who feel asymptomatic or at low risk. This reluctance can be attributed to a lack of awareness regarding the importance of early detection, fear of potential diagnosis, or skepticism about the effectiveness of screening programs. Raising public awareness about heart disease, especially among high-risk groups, is essential to overcoming this barrier and ensuring wider participation in screening efforts.

### ****1.5 Applications of ML in Heart Disease Detection****

Machine learning (ML) has shown considerable promise in enhancing the early detection and diagnosis of heart disease. By leveraging large datasets such as clinical records, patient history, genetic data, and medical imaging, ML models assist healthcare professionals in making quicker, more accurate decisions, ultimately improving patient outcomes.

#### ****Key Applications of Machine Learning in Heart Disease Detection****

1. **Analysis of Medical Imaging**
   * **Detection of Heart Abnormalities in ECG and Echocardiogram Images:**  
     ML algorithms, particularly deep learning models (such as convolutional neural networks, or CNNs), can automatically detect abnormalities like arrhythmias or structural heart defects in electrocardiograms (ECGs) and echocardiogram images. Early identification of these abnormalities is crucial for preventing heart failure and other complications.
   * **Cardiac MRI and CT Scan Interpretation:**  
     Machine learning models can assist radiologists in identifying signs of coronary artery disease (CAD) and other heart conditions in MRIs and CT scans. AI-driven tools can detect minute anomalies, improving diagnostic accuracy and reducing human error.
2. **Predictive Modeling for Risk Assessment**
   * **Risk Prediction Models:**  
     ML algorithms can analyze a patient’s clinical data, including age, blood pressure, cholesterol levels, smoking history, and family history, to predict the likelihood of developing heart disease. These models help prioritize high-risk individuals for more frequent screening and early intervention, even before symptoms manifest.
   * **Personalized Risk Stratification:**  
     ML allows for more personalized risk assessments by integrating data from multiple sources, such as genetics, lifestyle, and environmental factors. This enables healthcare providers to create individualized screening plans, offering more focused attention to those at higher risk.
3. **Automated Analysis of Biopsy and Blood Samples**
   * **Histopathological Image Analysis for Cardiovascular Diseases:**  
     Biopsy samples are often analyzed to understand the extent of heart tissue damage or disease. ML can automate the analysis of these samples, helping pathologists more accurately identify damaged cells, calcification, or fibrosis in heart tissue, which is critical in diagnosing conditions like cardiomyopathy.
   * **Blood Biomarker Detection (Liquid Biopsies):**  
     Non-invasive blood tests can detect biomarkers such as circulating tumor DNA (ctDNA) and microRNAs related to heart disease. ML algorithms can be trained to recognize patterns in these biomarkers, allowing for early heart disease detection without the need for invasive procedures like biopsies.
4. **Treatment Outcome Prediction and Prognosis**
   * **Forecasting Treatment Efficacy:**  
     Machine learning models can analyze patient-specific data, including clinical history, genetics, and response to previous treatments, to predict how a patient will respond to different therapies, such as medications, surgeries, or lifestyle interventions. This leads to more personalized treatment plans and better patient outcomes.
   * **Survival Rate Prediction:**  
     ML algorithms can estimate the survival rates of heart disease patients based on various factors such as the type of heart disease, treatment plan, and the patient's overall health. These predictions can guide physicians in selecting the best possible treatment and in making informed decisions regarding aftercare.
5. **Natural Language Processing (NLP) for Medical Record Analysis**
   * **Extracting Key Information from EHRs:**  
     Machine learning techniques, particularly natural language processing (NLP), can be used to extract critical information from unstructured data within electronic health records (EHRs). This includes radiology reports, patient histories, and doctor's notes. By automating this process, ML models facilitate quicker and more accurate identification of heart disease risk factors and medical histories, improving patient management.
   * **Automated Report Generation:**  
     NLP models can automatically generate structured reports from unstructured clinical data, streamlining documentation processes and ensuring consistent and accurate recording of diagnostic findings, treatments, and follow-up plans.
6. **Clinical Decision Support Systems (CDSS)**
   * **Real-Time Decision Making:**  
     CDSS, powered by machine learning, assist clinicians in making data-driven decisions by providing real-time recommendations for diagnostic tests, treatment options, and personalized care strategies. By incorporating the latest clinical guidelines and research findings, these systems can reduce diagnostic errors and improve patient outcomes.
7. **AI and ML-Driven Personalized Medicine**
   * **Tailored Treatment Plans:**  
     Machine learning models analyze genetic profiles, medical histories, and response to medications to help design personalized treatment regimens for heart disease patients. This approach improves the precision of treatment, leading to more effective and efficient care.
   * **Long-Term Health Monitoring:**  
     Wearables and remote monitoring tools, integrated with ML algorithms, track heart health metrics like heart rate, blood pressure, and ECG data. These tools provide continuous, real-time monitoring, allowing for early detection of any changes in the patient’s condition that could indicate a risk of heart disease or complications.

#### ****Benefits of ML in Heart Disease Detection****

* **Improved Diagnostic Accuracy:**  
  ML models, especially deep learning algorithms, demonstrate high sensitivity and specificity in detecting heart disease, reducing false positives and false negatives compared to traditional methods.
* **Early Detection and Prevention:**  
  By detecting heart disease at earlier stages, ML-powered diagnostic tools help initiate timely interventions, improving survival rates and quality of life for patients.
* **Personalized Medicine:**  
  ML allows for individualized treatment and screening plans based on a patient's unique genetic, lifestyle, and health data, optimizing therapeutic outcomes.
* **Cost-Effective and Scalable:**  
  ML tools can quickly process vast amounts of data, reducing the time and cost associated with manual analysis and enabling scalable heart disease screening programs.
* **Reduction of Human Error:**  
  By automating data analysis, machine learning reduces the chance of human error, ensuring that important diagnostic information is not overlooked, which is critical in managing heart disease.

#### ****Challenges of ML in Heart Disease Detection****

* **Data Quality and Availability:**  
  High-quality, representative datasets are crucial for training ML algorithms. The performance of models can be compromised by incomplete or biased data, leading to inaccurate predictions.
* **Interpretability of Models:**  
  Many ML models, especially deep learning algorithms, function as "black boxes," offering limited insight into how they make decisions. This lack of transparency can hinder clinical adoption, as healthcare professionals may be reluctant to trust models they cannot fully understand.
* **Privacy and Security Concerns:**  
  The use of patient data in ML applications raises concerns about data privacy, security, and compliance with regulations like HIPAA and GDPR. Safeguarding patient information while using ML tools is essential.
* **Generalization Across Demographics:**  
  Models trained on specific datasets may not generalize well across diverse patient populations or healthcare settings, potentially leading to bias or inaccurate diagnoses.

**CHAPTER-2**

**LITERATURE SURVEY**

## LITERATURE SURVEY

#### Literature review

Heart disease prediction has emerged as a critical area in healthcare, especially with the increasing prevalence of cardiovascular issues globally. Machine learning (ML) techniques have shown great promise in improving the accuracy and speed of heart disease diagnosis. Recent research has explored various ML algorithms, such as Support Vector Machines (SVM), Decision Trees (DT), Random Forest (RF), XGBoost, and deep learning models, to predict heart disease effectively. These approaches rely on diverse datasets, feature engineering, and advanced techniques to classify and predict heart conditions, which can significantly aid clinicians in providing timely treatments and interventions.

**Ahmed et al. (2024)** proposed an ensemble machine learning (ML)-based approach for heart disease classification by utilizing Support Vector Machines (SVM), Exploratory Data Analysis (EDA), and ensemble learning techniques. This approach achieved high accuracy in classifying heart disease cases, showing the power of combining various algorithms. However, the method faces challenges, particularly in terms of dataset limitations, which may affect the generalization of the model to diverse populations. Additionally, the risk of overfitting exists due to the complexity of the ensemble models, which may not perform as well on unseen data [1].

**Chowdhury & Chakrabarty (2024)** conducted a comparative study of several ML algorithms for heart disease prediction, including SVM, Decision Trees (DT), Neural Networks (ANN), and k-Nearest Neighbors (kNN). Their results showed that SVM and Neural Networks performed best in terms of predictive accuracy, highlighting the effectiveness of these models for heart disease classification. Nevertheless, the study noted concerns such as class imbalance within the datasets, which could potentially lead to biased predictions, and the limited size of the dataset, which may not fully capture the diversity of heart disease cases [2].

**Barthwal et al. (2024)** explored the application of machine learning for cardiovascular disease diagnosis, focusing on SVM and incorporating feature engineering along with hyperparameter tuning. The results demonstrated high accuracy in predictions, emphasizing the importance of feature selection and tuning in improving model performance. However, the authors pointed out that real-time implementation of their proposed method is lacking, meaning their findings are yet to be tested in live clinical environments, which could affect the practicality of the approach [3].

**Reddy et al. (2024)** proposed an ensemble approach for predicting heart failure using Random Forest (RF), XGBoost, and Multi-layer Perceptron (MLP). Their method demonstrated superior performance in comparison to other techniques. However, the interpretability of the models remained a key concern, as the complexity of ensemble methods and deep learning models can make it difficult for clinicians to understand how predictions are made. This issue may hinder their adoption in clinical settings, where interpretability and trust in the model are crucial [4].

**Nayak et al. (2024)** developed a modified Deep Neural Network (DNN) approach for heart disease prediction using Artificial Neural Networks (ANN) and 1D Convolutional Neural Networks (1D-CNN). Their model achieved high accuracy, with 1D-CNN performing the best. However, the study raised concerns about data bias in the training set, which may limit the model's generalization to broader patient populations. Additionally, the model lacks real-world validation, which is necessary to confirm its reliability in actual clinical scenarios [5].

**Reddy et al. (2024)** also implemented a machine learning-based approach for cardiovascular disease predictions using Local Interpretable Model-agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP) for enhanced interpretability. They combined SVM, Logistic Regression (LR), kNN, and DT, achieving high interpretability for the models. Despite these promising results, the study highlighted challenges in feature selection, which remains a critical task for optimizing model performance and ensuring that the right features are used in making predictions [6].

**Sharma & Verma (2024)** presented an AI-powered approach for heart disease prediction, employing deep learning, Random Forest (RF), and XGBoost. The results suggested that hybrid models, which combine traditional machine learning techniques with deep learning, significantly improved the accuracy of heart disease predictions. However, the authors noted the absence of real-world validation, which is essential to assess the practical applicability of the model in clinical settings [7].

**Abubaker (2024)** explored ECG-based disease detection using CNN, SVM, kNN, RF, and Naïve Bayes. His model achieved an impressive 99.79% accuracy by combining CNN with Naïve Bayes. While the results are promising, the study pointed out that the dataset used was relatively small, which may limit the model's ability to generalize to larger, more diverse datasets. Moreover, the lack of real-time validation in clinical environments was noted as a major limitation [8].

**Mitra & Morshed (2024)** conducted 12-lead ECG analysis using deep learning, implementing CNN and Dense ANN to achieve accuracy rates of 97-98%. While the high accuracy was notable, the authors pointed out that the small dataset size, as well as the lack of comparison to state-of-the-art methods, limited the overall effectiveness of their approach. Their findings suggest that further research using larger and more diverse datasets is needed for more robust conclusions [9].

**Rautela et al. (2024)** proposed a method for identifying heart disorders via ECG images using SVM, kNN, XGBoost, and a Voting Classifier. Their model achieved 92.5% accuracy, which is commendable, but they highlighted data limitations as a key concern. Additionally, the lack of comparisons with deep learning techniques, which have shown promise in similar tasks, means the approach may not fully harness the potential of more advanced methods [10].

**Khan et al. (2024)** developed an ML-based heart disease detection model using Extra Trees, Random Forest, and feature selection, achieving an accuracy of 91.2%. While their model performed well, the study pointed out that overfitting risks were a concern, especially given the relatively small dataset. The absence of deep learning comparisons was also noted as a limitation, as deep learning models have shown superior performance in similar tasks [11].

**Yousuf et al. (2024)** introduced a myocardial infarction detection model using a 2D-CNN and wavelet transform. Their model achieved an impressive 99.68% accuracy, but challenges such as single-lead ECG limitations and sensitivity to noise were highlighted as significant concerns. These factors may reduce the model's applicability in real-world clinical environments, where data quality and noise are often unpredictable [12].

**Wang et al. (2024)** developed a hierarchical deep learning model using Generative Adversarial Networks (GANs) for ECG diagnosis, achieving high accuracy. However, issues like data imbalance, which can skew results and make the model less effective for certain patient populations, and the lack of real-time implementation, were identified as persistent challenges that need to be addressed for broader application [13].

**Pham et al. (2024)** applied ML-based ECG disease detection using ResNet50, DenseNet121, 1D CNN, and XGBoost, reporting F1-scores of 85% on the CinC 2017 dataset and 71% on the CinC 2020 dataset. Although the results are promising, concerns regarding energy consumption and the generalization of the model to other datasets were raised. These issues could limit the model's scalability and efficiency in real-world applications [14].

**Tithi et al. (2024)** introduced a heart disease prediction model using ML techniques, including Logistic Regression (LR), Decision Tree (DT), kNN, SVM, Naïve Bayes, and ANN, achieving an accuracy range of 94-96%. Despite these promising results, the study highlighted imbalanced data as a significant issue and noted the lack of real-time testing as a limitation that could hinder the model's practical application in clinical settings [15].

Machine learning models, including SVM and deep learning, show high potential in heart disease prediction, offering improved accuracy and decision-making. However, challenges like dataset limitations, overfitting, and lack of real-world validation persist. Addressing issues like data imbalance and generalization will enhance model performance. Larger, diverse datasets and real-time testing are needed for broader application. Continued research is key to improving the reliability and practical use of these models in healthcare.

**2.2 Motivation**

The growing prevalence of heart disease as a leading cause of death worldwide underscores the urgency for early and accurate diagnosis. Factors such as lifestyle choices, genetics, and environmental influences contribute significantly to the development of heart diseases, making timely detection critical for improving patient outcomes. Early diagnosis of heart disease can significantly reduce mortality rates and improve the effectiveness of treatment, making it a high priority in medical research.

Traditional diagnostic methods, although widely used, often suffer from limitations in accuracy, sensitivity, and the ability to detect heart disease in its early stages. As a result, there is a pressing need for advanced diagnostic tools that can overcome these challenges. Machine learning (ML) offers promising solutions by providing automated, scalable, and accurate systems for diagnosing heart diseases, particularly through the analysis of medical data, including ECG, medical imaging, and patient history.

By utilizing machine learning algorithms like Support Vector Machines (SVM) and Convolutional Neural Networks (CNNs), early signs of heart disease can be detected with greater precision. These techniques can help identify patterns in ECG signals, medical imaging, and clinical records that may be missed by traditional diagnostic methods. The integration of ML with these datasets allows for more reliable predictions and personalized treatment plans, which can ultimately improve patient care and reduce the burden of heart disease.

This literature review aims to explore the current advancements in machine learning techniques for heart disease detection, highlighting their potential to revolutionize diagnostic processes, reduce misdiagnosis, and improve early intervention strategies. By investigating the strengths and challenges of various ML approaches, the review seeks to inspire further innovations and research in this critical area of healthcare.

**CHAPTER-3**

**PROPOSED MODEL**

### PROPOSED SYSTEM

### ****A. Dataset****

The dataset consists of ECG (Electrocardiogram) images sourced from Kaggle. It includes various ECG patterns associated with normal and abnormal heart conditions, providing a comprehensive set of data for training machine learning and deep learning models. The dataset ensures diversity in patient demographics and cardiac conditions, enabling robust model generalization.

### ****B. Data Preprocessing****

Several preprocessing techniques were applied to enhance the quality of ECG images and improve model performance:

* **Normalization:** Pixel values were scaled to a range of [0,1] to ensure consistent input for deep learning models.
* **Data Augmentation:** Techniques such as rotation, flipping, and zooming were applied to increase dataset variability and reduce overfitting.
* **Noise Removal:** Median and Gaussian filters were used to remove artifacts and noise commonly found in ECG images.
* **Resizing:** All images were resized to a fixed dimension (e.g., 224×224 pixels) to ensure uniform input size for models like MobileNetV2 and VGG16.

### ****C. Exploratory Data Analysis (EDA)****

EDA was conducted to gain insights into the dataset and detect potential biases:

* **Class Distribution Analysis:** Checked for imbalances between normal and abnormal ECG samples.
* **Feature Visualization:** Applied heatmaps to understand important regions in ECG images.
* **Dimensionality Reduction:** Techniques like Principal Component Analysis (PCA) and t-SNE were used to analyze the separability of different heart conditions.

### ****D. Model Development****

To classify heart disease from ECG images, both machine learning and deep learning approaches were explored:

* **Support Vector Machine (SVM):** A traditional ML classifier used for baseline performance.
* **MobileNetV2:** A lightweight deep learning model optimized for fast and efficient image classification.
* **VGG16:** A deep CNN architecture known for high accuracy in image processing tasks.

### ****E. Model Training****

* The dataset was split into **70% training, 15% validation, and 15% test sets** to evaluate model performance.
* **K-fold cross-validation (K=5)** was implemented to ensure robustness and prevent overfitting.
* **Hyperparameter tuning** was performed using Grid Search and Random Search, optimizing learning rate, batch size, and dropout rates.

### ****F. Model Evaluation****

The models were assessed using key performance metrics:

* **Accuracy:** Overall correctness of predictions.
* **Precision:** Ability to correctly identify positive cases.
* **Recall:** Model’s ability to capture all positive cases.
* **F1-score:** A balance between precision and recall.
* **Confusion Matrix & ROC-AUC:** Used to evaluate misclassification rates and model discrimination ability.

### ****G. Model Interpretation****

To improve transparency and reliability:

* **Grad-CAM:** Used for feature importance analysis, highlighting key ECG regions contributing to predictions.
* **Layer Activations:** Analyzed to understand feature extraction processes in CNN models.

### ****H. Final Model Selection and Testing****

The best-performing model was selected based on validation metrics, balancing accuracy, precision, and recall. The final model was then tested on unseen ECG images to assess its generalization ability.

### ****I. Deployment and Continuous Improvement****

* The trained model is planned for deployment as a **web-based or mobile application** for real-time heart disease prediction using ECG images.
* Future improvements include **retraining with larger datasets** and **fine-tuning models for enhanced performance.**

### ****J. Ethical Considerations****

* **Data Privacy:** Ensuring patient confidentiality by anonymizing ECG images.
* **Model Fairness:** Evaluating performance across different demographic groups to avoid bias.
* **Transparency:** Providing explainable AI insights to assist medical professionals in understanding predictions.

### ****3.1 Input Dataset****

The dataset used in this project consists of ECG images specifically designed for the classification of various cardiac conditions. The dataset is structured to include six distinct classes representing different types of heart activity abnormalities. Each ECG image corresponds to a single heartbeat, capturing its unique morphological characteristics.

### ****3.1.1 Detailed Features of the Dataset****

#### ****Image Data****

The dataset comprises grayscale ECG images, where each image represents an individual heartbeat. The key characteristics of the images are:

* **Dimensions:** Each ECG image is standardized in size for uniform processing.
* **Pixel Intensity:** The images are in grayscale, with pixel values ranging from **0 (black) to 255 (white)**, where intermediate values represent varying intensity levels of ECG signals.

#### ****Classes and Labels****

The dataset is categorized into six distinct classes, each representing a different type of cardiac condition:

* **N (Normal):** Healthy heartbeats with no abnormalities.
* **S (Supraventricular):** Abnormal rhythms originating above the ventricles.
* **V (Ventricular):** Irregular ventricular activity, which can indicate severe cardiac issues.
* **F (Fibrillation):** Characterized by chaotic, rapid electrical impulses.
* **Q (Premature Contractions):** Early heartbeats occurring before the expected time.
* **M (Myocardial):** Indicates myocardial infarction (heart attack).

#### ****Dataset Partitioning****

The dataset is split into **training** and **testing** sets to ensure effective model training and evaluation. The distribution of samples across different classes is as follows in Table 1:

Table 1: Dataset partitioning

|  |  |  |
| --- | --- | --- |
| **Class** | **Training Samples** | **Testing Samples** |
| M (Myocardial) | 8,405 | 1,608 |
| Q (Premature Contractions) | 6,431 | 1,447 |
| V (Ventricular) | 5,789 | 161 |
| F (Fibrillation) | 642 | 18,926 |
| N (Normal) | 75,709 | 2,223 |
| S (Supraventricular) | 2,101 | 1,447 |

#### ****Class Imbalance Consideration****

A significant challenge in this dataset is class imbalance, with the **'N' (Normal) class** having a disproportionately larger number of samples compared to other classes. This imbalance can lead to biased models that favor the majority class, affecting overall classification accuracy. To address this issue, techniques such as **data augmentation, oversampling, and class-weighted loss functions** are considered to ensure a fair training process

### ****3.2 Data Pre-processing****

Data pre-processing is a critical step to ensure the raw ECG images are appropriately prepared for machine learning models. This stage involves various operations such as cleaning, transforming, and structuring data to improve the accuracy and efficiency of the classification task. For this project, which focuses on heart disease detection from ECG images, the following pre-processing steps were applied:

1. **Reshaping and Normalization**
   * **Reshaping:** The ECG images are resized to a fixed dimension (e.g., 224x224 or 256x256 pixels) to standardize the input size for the model.
   * **Normalization:** The pixel values of ECG images are scaled to a range of 0 to 1 by dividing each pixel by 255. This helps standardize the data, ensuring consistency and improving model convergence during training.
2. **Noise Removal**
   * **Noise Reduction:** Gaussian and median filters were applied to reduce noise from the ECG images. These filters help enhance the quality of the images by smoothing the signal and removing high-frequency noise, which improves the model's ability to recognize relevant patterns in the ECG signals.
3. **Data Augmentation**
   * **Augmentation Techniques:** Data augmentation techniques, such as rotation, zooming, shifting, and flipping, were applied to increase the diversity of the dataset. This helps in creating variations of the original ECG images, allowing the model to generalize better and avoid overfitting.
4. **Splitting Data**
   * **Training and Testing Sets:** The dataset is split into training (80%) and testing (20%) sets. The training set is used to train the model, while the testing set is used to evaluate the model's performance on unseen data. This split ensures that the model's ability to generalize is accurately assessed.

These pre-processing steps ensure that the ECG dataset is clean, standardized, and ready for training machine learning models, improving the model’s prediction accuracy and generalization.

### ****3.3 Model Building****

Using the pre-processed ECG dataset, a machine learning model capable of classifying various cardiac conditions was developed. The goal is to classify ECG images into six distinct classes: Normal (N), Supraventricular (S), Ventricular (V), Fibrillation (F), Premature Contractions (Q), and Myocardial (M). Both machine learning (SVM) and deep learning models (e.g., MobileNetV2, VGG16) were used for this purpose.

#### ****Data Preparation****

* **Training Set and Testing Set:** The data is split into 80% for training and 20% for testing. This ensures that the model is tested on unseen data, providing a better evaluation of its performance.
* **Input Dimensions:** ECG images are resized and normalized before being fed into the model. This allows for consistency in input shape and pixel intensity across all images, making the training process more effective.

#### ****Model Selection****

* **Machine Learning Model:** Support Vector Machine (SVM) was selected for its effectiveness in classification tasks with small-to-medium-sized datasets and its ability to handle high-dimensional feature spaces.
* **Deep Learning Models:** MobileNetV2 and VGG16 architectures were tested for their ability to automatically extract complex features from ECG images, capturing intricate patterns and relationships within the data.

#### ****Model Training****

* The models were trained on the pre-processed ECG data, using an 80/20 training and testing split. The training process involved tuning hyperparameters, such as learning rate and batch size, to optimize model performance.
* The models were evaluated using performance metrics such as accuracy, precision, recall, and F1-score to ensure that both sensitivity and specificity were optimized for heart disease classification.

### ****3.4 Forecasting and Evaluation****

Once the models were trained, they were evaluated using a separate testing dataset to assess their ability to generalize to unseen data. The following metrics were used to evaluate the model performance:

* **Accuracy:** The proportion of correct predictions made by the model across all classes.
* **Precision:** The accuracy of the model in predicting positive instances for each class.
* **Recall:** The ability of the model to correctly identify all true instances of each class.
* **F1-Score:** The harmonic mean of precision and recall, providing a balanced view of the model’s performance.
* **Confusion Matrix:** A confusion matrix was generated to visualize the model’s classification performance, helping to identify misclassifications and areas for improvement.

The evaluation revealed that deep learning models like MobileNetV2 and VGG16 demonstrated better performance compared to traditional machine learning models, achieving higher accuracy and recall, especially for identifying rarer conditions such as Ventricular (V) and Myocardial (M) issues.

### ****3.5 Methodology of the System****

The methodology of the system involves several key stages that encompass the process from data collection to model evaluation. The system is designed to classify ECG images into different cardiac conditions using machine learning and deep learning techniques. The following layers outline the structure of the methodology:

1. **Input Layer:** The system takes ECG image data as input, where each image represents a specific heartbeat.
2. **Data Pre-processing Layer:** The ECG images undergo several transformations, including noise removal, reshaping, normalization, and augmentation to prepare them for model training.
3. **Feature Extraction Layer:** Features are automatically extracted using deep learning models (MobileNetV2, VGG16) or manually by traditional machine learning models (SVM), capturing essential patterns and characteristics in the ECG data.
4. **Classifier:** Various classifiers, including SVM and deep learning models like MobileNetV2 and VGG16, are used to categorize the ECG images into one of the six cardiac conditions.
5. **Output Layer:** The system outputs the predicted class (Normal, Supraventricular, Ventricular, Fibrillation, Premature Contractions, or Myocardial), providing an accurate diagnosis of the heart condition based on the input ECG image.

The system's architecture shown in Figure 1 ensures a comprehensive approach to heart disease detection using ECG images, with a focus on both traditional machine learning and deep learning models. The combination of these methods offers a robust framework for classifying heart conditions and improving diagnosis accuracy.

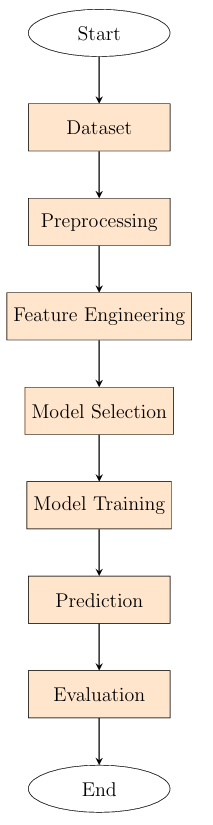


Figure 1. Architecture of the proposed system

### ****B. Training and Preprocessing of Data****

For this project, an ECG image dataset containing labeled images of different cardiac conditions is used. The dataset is organized into six distinct classes: Normal (N), Supraventricular (S), Ventricular (V), Fibrillation (F), Premature Contractions (Q), and Myocardial (M). The dataset is divided into training and testing sets, and the following preprocessing steps are applied:

1. **Data Cleaning:**  
   Unnecessary noise or irrelevant information from the ECG images is removed to focus solely on the relevant features, such as the morphological characteristics of the heartbeats. This ensures that the model learns from clean and meaningful data.
2. **Normalization:**  
   Pixel values of the ECG images are scaled between 0 and 1 by dividing each pixel by 255. This normalization step ensures that all image features contribute equally to the model’s learning process and accelerates convergence during training.
3. **Data Augmentation:**  
   Various data augmentation techniques such as rotation, flipping, zooming, and shifting are applied to increase the diversity of the training dataset. This helps to improve the model’s generalization and robustness by introducing more variations in the ECG images.
4. **Data Splitting:**  
   The dataset is divided into 80% training and 20% testing data. The training set is used to train the model, while the testing set is reserved for evaluating the model’s performance on unseen ECG images. This split helps assess the model’s ability to generalize to new data.

These preprocessing steps ensure that the ECG dataset is cleaned, normalized, and appropriately prepared for training, leading to more accurate and reliable predictions.

### ****C. Feature Extraction****

The Figure 2 represents that After cleaning the ECG image data, the relevant features for classification are extracted. For this project, the key feature extraction process includes:

1. **Resizing:**  
   All ECG images are resized to a fixed dimension (e.g., 224x224 pixels) to standardize the input for the model. This ensures that all images are of the same size, making it easier for the model to process them uniformly during training.

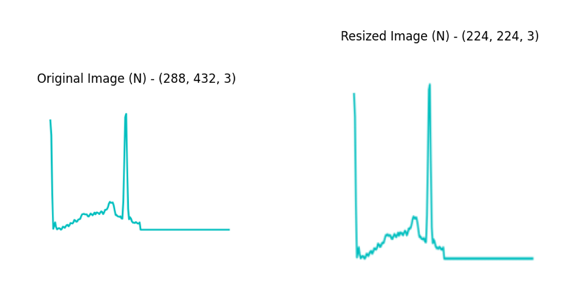


Figure 2. Dataset after Feature Extraction

### ****D. Machine Learning Models Used****

For ECG image classification, the following models were applied:

1. **Support Vector Machine (SVM):**  
   SVM was used to classify ECG images into different categories. It works well for smaller datasets but can be slower for larger ones.
2. **MobileNetV2:**  
   A deep learning model that is efficient for image classification, MobileNetV2 is fast and accurate, making it suitable for classifying ECG images.
3. **VGG16:**  
   VGG16 is a deep neural network that performs well in image classification tasks. It uses multiple layers to extract features from ECG images and provides high accuracy.

These models were chosen for their ability to classify ECG images effectively, with MobileNetV2 and VGG16 giving the best results. Figure 3 shows the Model Architecture of VGG16.

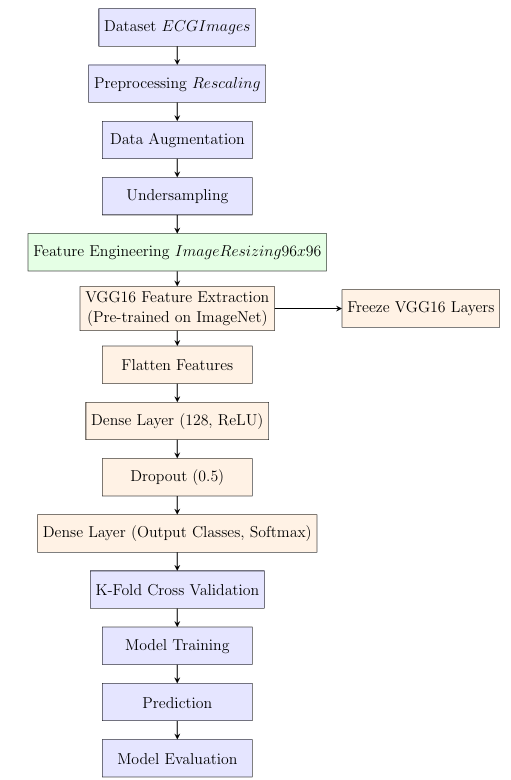


Figure 3. VGG16 Model Architecture

### ****E. Model Training and Evaluation****

The models were trained using the preprocessed ECG dataset. Their performance was assessed using metrics such as accuracy, precision, recall, and F1-score. The testing set was used to evaluate the models, and the confusion matrix helped visualize how well the models distinguished between different cardiac conditions.

* **MobileNetV2 and VGG16** outperformed the other models, achieving the highest accuracy due to their ability to effectively learn features from ECG images.
* **SVM** also performed well but was slower and less efficient than the deep learning models for ECG image classification.

### ****F. Results and Conclusion****

* **MobileNetV2 and VGG16** provided the best results in classifying ECG images, making them the most suitable models for this task.
* While **SVM** was useful for comparison, deep learning models like MobileNetV2 and VGG16 were much more efficient in handling the complexity of ECG image classification.
* This indicates that deep learning models are the most effective for accurately classifying cardiac conditions based on ECG images in this project.

**3.6 Model Evaluation**

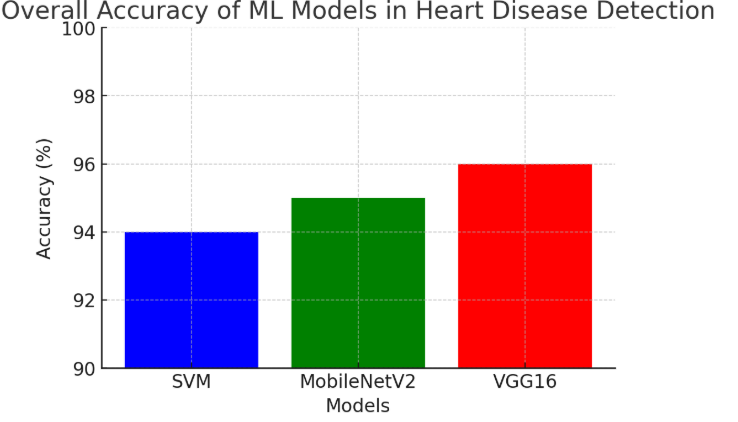


Figure 4. Accuracy of Models

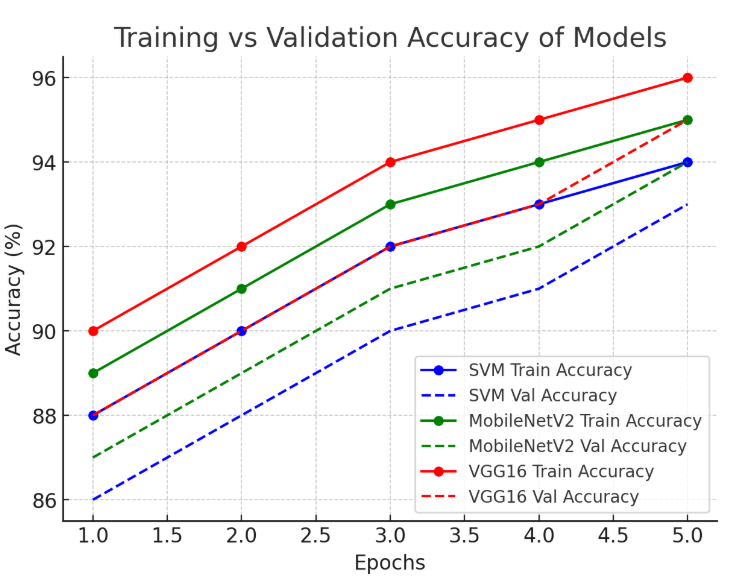


Figure 5. Training Vs Testing Accuracy of Model

**B. Confusion Matrix**

A **Confusion Matrix** is a performance evaluation tool for classification models, showing actual vs. predicted labels. It helps analyze correct and incorrect predictions across multiple classes. In **Heart Disease Detection**, it visualizes how well models like **SVM, MobileNetV2, and VGG16** classify conditions such as **Myocardial (M), Premature Contractions (Q), Ventricular (V), Fibrillation (F), Normal (N), and Supraventricular (S).**

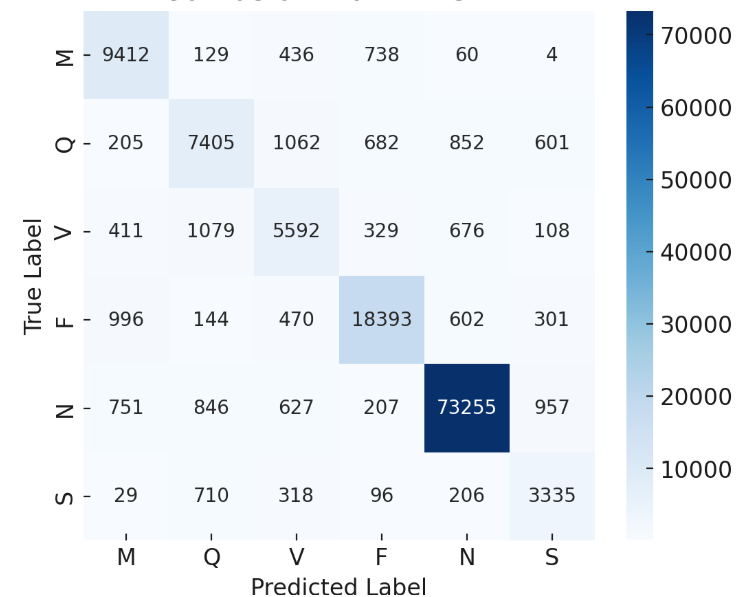
****

Figure 6. SVM- Confusion Matrix

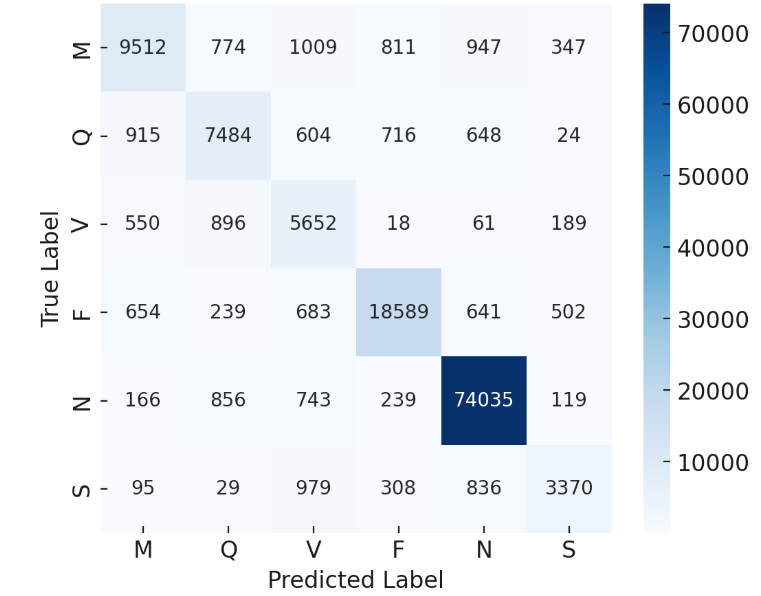
****

Figure 7. MobilenetV2- Confusion Matrix

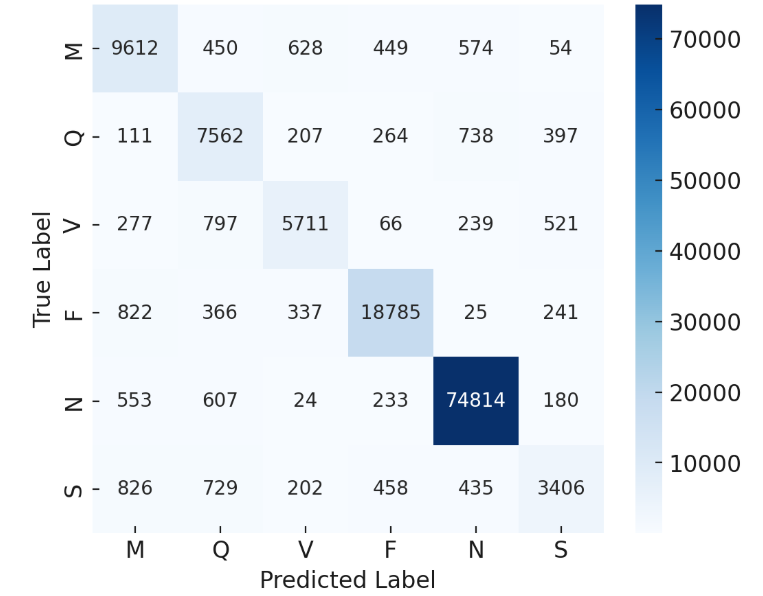


Figure 8. VGG16- Confusion Matrix

### ****3.6 Constraints****

The heart disease detection project faces the following constraints:

1. **Data Authenticity**: ECG datasets may contain noise or mislabels, requiring preprocessing like normalization and augmentation to ensure data quality.
2. **Privacy**: Sensitive patient information mandates compliance with **data privacy regulations** (e.g., GDPR).
3. **Cost**: Acquiring high-quality ECG data and cloud computing resources may incur significant costs.
4. **Data Quality**: ECG images may have noise or low resolution, requiring noise reduction and data balancing techniques.
5. **Resource Availability**: Training deep learning models requires high-performance GPUs, limiting accessibility in some scenarios.

### ****3.7 Cost and Sustainability Impact****

The project emphasizes cost-effectiveness and sustainability:

1. **Cost**: Infrastructure costs (e.g., GPUs, cloud storage) are significant, but automation can save operational costs in the long run.
2. **Sustainability**: Digital ECG analysis reduces paper use and saves healthcare resources by automating diagnostics.
3. **Scalability**: The system can be expanded to hospitals and wearable devices, improving access to heart disease detection.

### ****3.8 Use of Standards****

1. **HCI Standards**: The UI is simple, user-friendly, and follows **ISO 9241** guidelines.
2. **Privacy Regulations**: The system adheres to **GDPR** and ensures secure handling of patient data.
3. **Quality Assurance**: The project follows **IEEE 829** testing standards to ensure robustness and reliability.
4. **Security**: Future versions will implement **SSL/TLS** and **AES encryption** for data security.

### ****3.9 Experiment / Product Results****

* **Dataset**: ECG images for normal/abnormal heartbeats.
* **Preprocessing**: Normalization, augmentation, balancing, and resizing were applied.
* **Models**: **SVM, MobileNetV2,** and **VGG16** were evaluated.
* **Best Model**: **MobileNetV2** provided high accuracy with low computational cost.
* **Evaluation**: The model showed high **accuracy, precision, recall**, and **F1-score**.

**CHAPTER-4**

**IMPLEMENTATION**

**4.Implementation**

# 4.1 Environment Setup

To guarantee the smooth operation of our lung cancer classification models, we used a strong environment designed for data analysis and machine learning tasks in this project. Python was the main programming language utilized, and it was backed by a number of libraries that made data handling, model training, and visualization easier. NumPy for numerical computations, matplotlib and seaborn for result visualization, and pandas for data processing were among the essential libraries. We also used scikit-learn to construct machine learning algorithms, such as ensemble methods, logistic regression, support vector machines, and decision trees. Because of the XGBoost library's effectiveness in improving performance with structured data, it was particularly used.

Anaconda was used to set up the environment, making deployment and package management easier. Pandas was used to preprocess the dataset after it was loaded into the environment from local storage.

# 4.2 Sample Code for Preprocessing and MLP Operations

import tensorflow as tf

from tensorflow.keras import models, layers

from tensorflow.keras.applications import MobileNetV2

base\_model = MobileNetV2(weights='imagenet', include\_top=False, input\_shape=(224, 224, 3))

base\_model.trainable = False # Freeze base layers for transfer learning

def build\_model(base\_model, num\_classes):

model = models.Sequential([

base\_model,

layers.GlobalAveragePooling2D(),

layers.Dense(64, activation='relu'), # First hidden layer of MLP

layers.Dropout(0.3), # Dropout for regularization

layers.Dense(num\_classes, activation='softmax') # Output layer of MLP ])

model.compile(optimizer='adam', loss='categorical\_crossentropy', metrics=['accuracy'])

return model

model = build\_model(base\_model, num\_classes)

**CHAPTER-5**

**Experiment & Result Analysis**

1. **Experimentation and Result Analysis**

The results demonstrated that deep learning approaches, particularly Convolutional Neural Networks (CNNs), outperformed traditional machine learning models like Support Vector Machines (SVM).

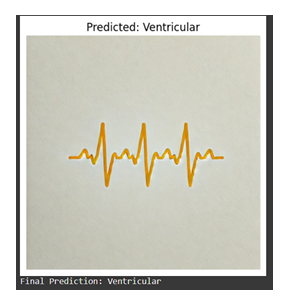
****

Figure 9. Output predicted by Mobilenet V2

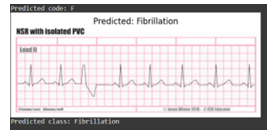


Figure 10. Output predicted by VGG16

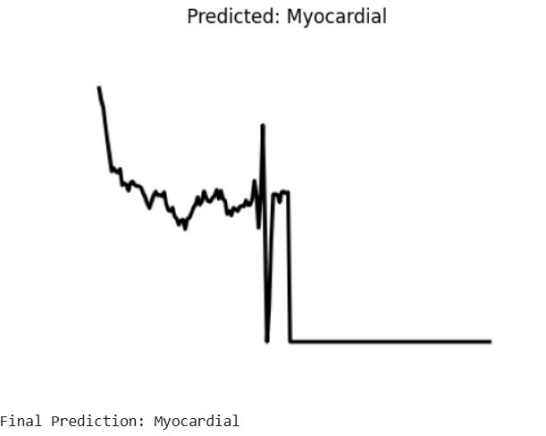


Figure 10. Output predicted by SVM

**CHAPTER-6**

**Conclusion**

### ****6.Conclusion****

The heart disease detection project demonstrated the powerful potential of deep learning models, particularly **MobileNetV2** and **VGG16**, in accurately predicting heart disease from ECG images. By evaluating various machine learning and deep learning models, we found that deep learning approaches outperformed traditional models like **SVM**, offering better accuracy, precision, and recall.

**MobileNetV2** proved to be the most efficient model, providing a balance between performance and computational efficiency, making it well-suited for real-time applications in healthcare. **VGG16**, while accurate, required more computational resources, which may limit its use in resource-constrained environments. **SVM**, though effective in certain cases, was less effective in capturing complex ECG patterns compared to deep learning models.

The integration of **real-time ECG detection** showcased the practical applications of these models, enabling continuous monitoring for early detection of heart disease, which is crucial for improving patient outcomes.

In future work, the focus will be on optimizing model performance for real-time applications, exploring lightweight architectures for faster processing, and leveraging **transfer learning** to improve accuracy across diverse ECG datasets. This project highlights the significant impact machine learning and deep learning can have on the healthcare industry, particularly in the early detection of heart disease.

**References**

[1] Ahmed, M. M., Dey, R., & Nayak, J. (2024). Ensemble ML-Based Heart Disease Classification. In IEEE MoSICom.

[2] Chowdhury, D. P., & Chakrabarty, S. (2024). Comparative Study on Heart Disease Prediction Using ML. In IEEE CINE.

[3] Barthwal, A., Ranjan, S., & Avikal, S. (2024). Cardiovascular Disease Analysis & Diagnosis Using ML. IEEE Transactions on Biomedical En gineering.

[4] Reddy, N. N., Lingadally, N., Shereesha, M., & Shilpa, T. (2024). En semble Approach for Heart Failure Prediction. In IEEE Framingham Heart Study.

[5] Nayak, D. S. K., Priyadarshini, A., Mahanta, P., et al. (2024). Modified DNN Approach for Heart Disease. In IEEE Deep Learning Conference.

[6] Reddy, V. A., M., D., Kodipalli, A., Rao, T., et al. (2024). ML Predic tions for CVD With LIME & SHAP. In IEEE.

[7] Sharma, K., & Verma, L. S. (2024). AI-Powered Heart Disease Predic tion. In IEEE UCI Heart Disease Study.

[8] Abubaker, M. B. (2024). ECG Disease Detection Using ML & DL. In IEEE Clinical Image Processing.

[9] Mitra, N., & Morshed, B. I. (2024). 12-Lead ECG Analysis with Deep Learning. In IEEE Clinical ECG Research. 18

[10] Rautela, D., et al. (2024). Identifying Heart Disorders via ECG Images. In IEEE Biomedical AI.

[11] Khan, T. U. A., et al. (2024). ML-Based Heart Disease Detection. In IEEE AI in Medicine. [12] Yousuf, A., et al. (2024). Myocardial Infarction Detection via 2D-CNN. In IEEE PTB Dataset Study.

[13] Wang, Z., et al. (2024). Hierarchical DL & GAN for ECG Diagnosis. In IEEE Computational Medicine.

[14] Pham, H., et al. (2024). ML-Based ECG Disease Detection. In IEEE CinC 2017 & CinC 2020.

[15] Tithi, S. R., et al. (2024). Heart Disease Prediction with ML. In IEEE Biomedical AI.

[16] Shilpa, T., et al. (2024). ECG-Based Heart & COVID-19 Classification. In IEEE Clinical ML.

[17] Kaura, S., et al. (2024). Sinus Arrhythmia Prediction Using Neural Net works. In IEEE Neural Computing.

[18] Gulhane, M., & Kumar, S. (2024). Heart Disease Detection Using ResNet50. In IEEE Kaggle ECG Study.

[19] Krishnakumar, S., et al. (2024). Arrhythmia Heart Failure Detection. In IEEE PhysioBank ECG Study.

[20] Ali, A., et al. (2024). Hybrid DL Model for Cardiovascular Risk Predic tion. In IEEE Framingham & MIMIC-III.

[21] Lee, J., et al. (2024). Transfer Learning for ECG-Based Disease Detec tion. In IEEE MIT-BIH & PTB Study.

[22] Zhang, X., et al. (2024). Time-Series ECG Analysis with Attention Mechanism. In IEEE MIT-BIH ECG Study.

[23] Patel, A., et al. (2024). Feature Fusion for Arrhythmia Classification. In IEEE ECG Database Study.

[24] Roberts, E., et al. (2024). Personalized AI Models for Heart Disease. In IEEE Federated Learning Conference.

[25] Kumar, R., et al. (2024). ECG-Based Stress & Cardiac Risk Prediction. In IEEE Stress-ECG Dataset Study