KLE Technological University
Creating Value
Leveraging Knowledge

BVB Campus, Vidyanagar, Hubballi – 580031, Karnataka, INDIA.

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

A Project report on

# Early Detection and Prognosis of Coronary Artery Disease Using Generative AI

Submitted

in partial fulfillment of the requirements for the award of the degree of

Bachelor of Engineering

IN

COMPUTER SCIENCE AND ENGINEERING


*Submitted By*

| | |
|---|---|
| Amogh R M | 01FE21BCS215 |
| Shraddha Kulkarni | 01FE21BCS203 |
| Sameer Umadi | 01FE21BCS206 |
| Deepa Badawadagi | 01FE21BCS353 |

Under the guidance of

Dr S R Chickerur

Professor

School of Computer Science and Engineering

BVB Campus, Vidyanagar, Hubballi – 580031, Karnataka, INDIA.

KLE Technological University, Hubballi

2023-2024

School of Computer Science and Engineering

**CERTIFICATE**

This is to certify that project entitled “Digital Twin-Based early Detection of CAD Using Generative AI” is a bonafied work carried out by the student team (Amogh R M 01FE21BCS215, Shraddha Kulkarni 01FE21BCS203 , Sameer Umadi 01FE21BCS206 , Deepa Badawadagi 01FE21BCS353 ), in partial fulfillment of the completion of the 6th semester B. E. course during the year 2023 – 2024. The project report has been approved as it satisfies the academic requirement with respect to the project work prescribed for the above said course.

| | |
|---|---|
| Guide Name | SoCSE |
| (Dr S R Chickerur) | Head |
| | Dr. Vijaylakshmi M. |

External Viva-Voce

| Name of the examiners | Signature with date |
|---|---|
| 1 __________________ | ______________ |
| 2 __________________ | ______________ |

# Acknowledgement

We would like to thank our faculty and management for their professional guidance towards the completion of the project work. We take this opportunity to thank Dr. Ashok Shettar, Vice-Chancellor, Dr. B.S.Anami, Registrar, and Dr. P.G Tewari, Dean Academics, KLE Technological University, Hubballi, for their vision and support.

We also take this opportunity to thank Dr. Meena S. M, Professor and Dean of Faculty, SoCSE and Dr. Vijayalakshmi M, Professor and Head, SoCSE for having provided us direction and facilitated for enhancement of skills and academic growth.

We thank our guide Dr. S R Chikkerur, SoCSE for the constant guidance during interaction and reviews.

We extend our acknowledgement to the reviewers for critical suggestions and inputs. We also thank Project Co-ordinator Dr. P.G. Sunitha Hiremath, and reviewers for their suggestions during the course of completion.

We express gratitude to our beloved parents for constant encouragement and support.

Amogh - 01FE21BCS215
Shraddha - 01FE21BCS203
Sameer - 01FE21BCS206
Deepa - 01FE21BCS353

# ABSTRACT

Coronary Artery Disease (CAD) is a leading global cause of mortality, resulting from the narrowing or blockage of coronary arteries due to plaque buildup, restricting blood flow to the heart. Early detection is critical to prevent complications such as heart attacks or sudden cardiac death. However, accurate and efficient CAD diagnosis remains challenging, requiring advanced tools for enhanced predictive accuracy and disease progression analysis. This project proposes a multi-model AI framework for CAD detection, classification, and progression simulation. Using a 6 GB image dataset with two classes—Normal and Sick (CAD)—a Convolutional Neural Network (CNN) is trained to classify images. In contrast, a Random Forest Classifier is trained on corresponding CSV data, providing an alternative approach. Both models are saved for further analysis. Variational Autoencoders (VAEs), a generative AI method, are employed to analyze the latent space of classified data, offering deeper insights into data patterns and validating the classifications. The framework also predicts disease progression by simulating outcomes using a VAE, based on new inputs like images and physical parameters (e.g., lifestyle factors). This enables comprehensive predictions of CAD trajectories, facilitating personalized treatment planning. Scalable and interpretable, the proposed framework advances CAD detection and prognosis, with promising applications in clinical practice.

***Keywords :*** *Coronary Artery Disease (CAD), Convolutional Neural Network (CNN), Random Forest Classifier, Variational Autoencoders (VAEs), Generative AI, Disease progression, Predictive modeling*

# CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# Chapter 1

# INTRODUCTION

Coronary Artery Disease (CAD) is one of the leading causes of death globally, resulting from the buildup of plaque in the coronary arteries, which restricts blood flow to the heart. This condition can lead to serious complications such as heart attacks, strokes, and sudden cardiac death. Early detection of CAD is critical because it allows for timely intervention, potentially preventing these severe outcomes and improving patient survival. However, diagnosing CAD traditionally involves expensive, invasive, and time-consuming procedures, making it essential to explore innovative, non-invasive solutions to improve diagnostic accuracy and efficiency.

This project aims to develop an advanced system for early detection and progression prediction of CAD using cutting-edge artificial intelligence (AI) techniques. The system employs a multi-model approach, combining machine learning algorithms to classify CAD-related medical images and predict disease progression based on patient-specific factors. The dataset used in this project includes a 6 GB collection of images classified into two categories—Normal and Sick (CAD) which are used to train the models. A Convolutional Neural Network (CNN) is trained to classify the images into these two categories, capitalizing on its ability to automatically extract and learn patterns from medical images. Additionally, a Random Forest Classifier is trained on a feature-based dataset, offering a simpler yet highly effective classification method. Both models are saved for later use and analysis. To take the analysis further,

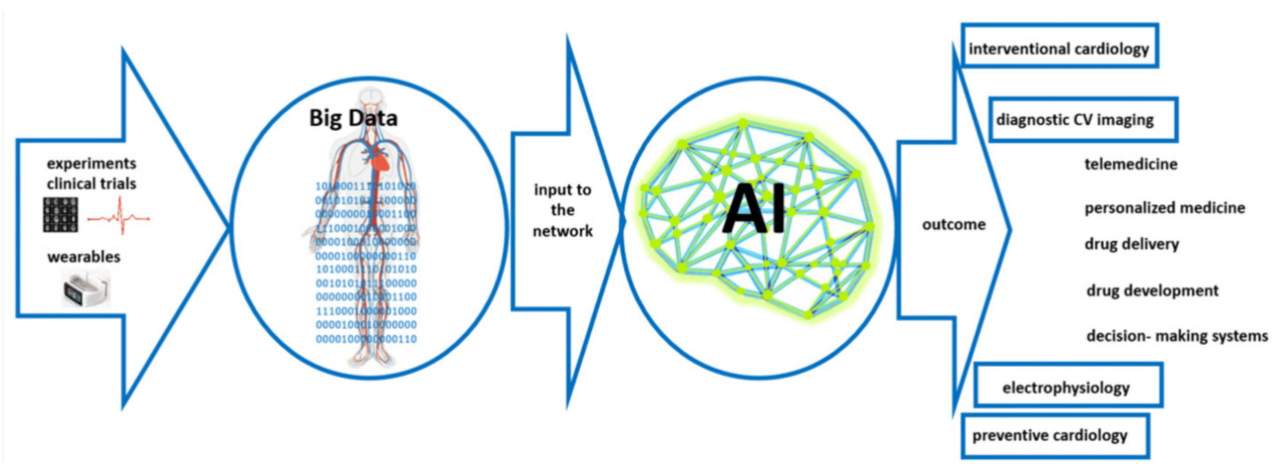


Figure 1.1: Conceptual scheme of the application of AI in cardiology.

the project incorporates Variational Autoencoders (VAEs), a generative AI technique, which is used to analyze the classified data from both the CNN and the Random Forest models. VAEs help simulate disease progression and predict future outcomes by analyzing underlying patterns in the data. This generative model allows for personalized predictions, factoring in variables like lifestyle changes and medical treatments, to offer insights into how the disease might evolve over time in individual patients. By combining classification with predictive simulation, this project aims to offer a comprehensive and non-invasive tool for CAD diagnosis and monitoring. The outcome of this work will be a robust AI-driven solution that not only classifies images but also simulates disease progression, providing healthcare professionals with valuable insights for personalized treatment planning. This approach has the potential to significantly improve the early detection of CAD, reduce reliance on invasive tests, and contribute to better patient outcomes in the management of cardiovascular health.

## 1.1 Motivation

The motivation behind this project stems from the increasing prevalence of Coronary Artery Disease (CAD) and its impact on global health. CAD is a leading cause of mortality, and its early detection plays a crucial role in preventing life-threatening events such as heart attacks and strokes. Traditional diagnostic methods are often invasive, expensive, and time-consuming, creating a need for more efficient, non-invasive solutions. The project is motivated by the desire to leverage advanced artificial intelligence (AI) techniques, particularly machine learning models, to improve the accuracy and speed of CAD diagnosis. By combining Convolutional Neural Networks (CNNs), Random Forest Classifiers, and Variational Autoencoders (VAEs), the project aims to develop a comprehensive system that not only classifies CAD-related medical images but also simulates disease progression based on patient-specific factors. This approach aims to enhance early diagnosis, reduce healthcare costs, and provide personalized treatment strategies, ultimately improving patient outcomes and quality of life.

## 1.2 Literature Survey

In our research on the early detection and prognosis of Coronary Artery Disease (CAD) using Generative AI and digital twin technologies, we found several studies that were instrumental in shaping the direction of our project. These studies provided insights into the use of AI models and digital twins for enhancing the diagnosis, prediction, and progression of CAD, guiding us in integrating these technologies into our approach.

[1] Yaojun Hu et al. (2024) – Personalized Heart Disease Detection via ECG Digital Twin Generation. In this paper, the authors propose using digital twins to enhance heart disease detection, particularly through ECG signals. They introduce a method to generate

personalized ECG digital twins that simulate individual heart conditions. Their approach improves the detection sensitivity, ensuring more accurate heart disease diagnosis tailored to specific patients, and includes a focus on privacy protection in model development.

[2] Liu et al. (2024) – AI-Driven Early Diagnosis and Risk Stratification of Coronary Artery Disease. Liu and colleagues explore the use of deep learning models for early CAD diagnosis. They integrate AI with digital twin technologies to simulate and predict patient-specific heart conditions, which enables better risk stratification and more precise early diagnosis. The model aims to provide personalized treatment recommendations by predicting the progression of CAD in individual patients.

[3] Baranidharan Balakrishnan et al. (2023) – Predicting Coronary Artery Disease Progression Using Generative Models and Digital Twins, this study combines generative models and digital twins to simulate the progression of CAD. Using imaging data and clinical parameters, the models predict future CAD events and offer guidance for timely interventions. The approach demonstrates the potential of generative AI in forecasting cardiovascular conditions and guiding treatment plans.

Kumar et al. (2024) – Machine Learning for CAD Diagnosis and Prognosis: A Digital Twin-Based Approach Kumar's team presents a framework that integrates digital twins with machine learning for CAD diagnosis and prognosis. By simulating individual cardiovascular systems, the model predicts disease progression and evaluates the impact of treatments, ultimately offering a more precise and personalized approach to CAD management.

Zhang et al. (2024) – Deep Learning and Digital Twins for Cardiac Disease Simulation and Outcome Prediction Zhang and colleagues leverage deep learning alongside digital twins to simulate heart conditions and predict outcomes for CAD patients. The model integrates clinical data, heart imaging, and generative AI to simulate disease progression and predict the impact of interventions. Their findings highlight the potential of AI to predict long-term cardiovascular events, offering insights into personalized treatment options.

These papers illustrate the transformative role of AI and digital twin technologies in CAD, enabling personalized, accurate, and non-invasive methods for diagnosis and treatment planning. Through these innovations, early detection and disease progression prediction can become more efficient, leading to better patient care.

## 1.3 Problem Statement

Elevating the detection and prognosis of Coronary Artery Disease (CAD) through the integration of generative AI, digital twin technologies, and machine learning models, to simulate disease progression and optimize early diagnosis for improved decision-making.

## 1.4 Applications

- **Disease Progression Simulation:** Using generative AI and digital twin technologies, the project can simulate how CAD progresses over time in individual patients. This enables healthcare providers to predict future health risks, such as heart attacks or strokes, based on factors like age, lifestyle, and clinical history.

- **Early Detection and Diagnosis:** The system can assist healthcare professionals in identifying CAD in its early stages by classifying medical images and clinical data, improving diagnostic accuracy and timeliness. AI-based models like CNNs and Random Forest classifiers can automatically analyze imaging data, such as CT scans or MRI, to detect CAD and other cardiovascular conditions. This application can significantly reduce the reliance on expensive and invasive procedures.

- **Clinical Decision Support Systems (CDSS):** Integrating the generative AI model into clinical workflows as a decision support tool can assist doctors in making more accurate and informed decisions regarding CAD treatment and management.

- **Predictive Analytics for Intervention:** The AI system can be used to predict the effectiveness of various interventions, including surgical procedures like angioplasty or bypass surgery, based on patient-specific data. It can also simulate the effects of lifestyle changes (e.g., diet and exercise) or medication on CAD progression.

## 1.5 Objectives and Scope of the project

The objective of this study is to develop a generative AI-based framework that integrates digital twin technologies and machine learning models (such as CNNs and Random Forest classifiers) to enhance the early detection of Coronary Artery Disease (CAD). The study aims to classify medical images and clinical data to improve diagnostic accuracy, facilitating timely intervention. Additionally, the study intends to simulate CAD progression using digital twin models, leveraging generative AI techniques to predict disease progression and assess the impact of various interventions. By simulating individual patient data, the framework will provide valuable insights into CAD risk and outcomes, thereby aiding healthcare professionals in making informed decisions regarding treatment strategies. The goal is to improve early diagnosis, predict disease trajectories, and enable personalized management of CAD to ultimately enhance patient outcomes and reduce cardiovascular morbidity.

### 1.5.1 Objectives

- Combining CNN-based image classification with Random Forest classifiers for structured clinical data analysis to enhance the precision of CAD detection and diagnosis.

- Integrating digital twin technologies and machine learning models like CNNs and Random Forest classifiers to improve the early identification of Coronary Artery Disease (CAD) using medical images and clinical data.

### 1.5.2 Scope of the project

The scope of this project encompasses the development generative AI-based framework aimed at enhancing the early detection and prognosis of Coronary Artery Disease (CAD). The project integrates digital twin technologies, machine learning models, and generative AI techniques, targeting the following key aspects:

- **Data Utilization:** The project will utilize a 6 GB dataset consisting of medical images (such as CT scans or MRIs) and clinical data (patient history, lifestyle factors, etc.) to train the machine learning models.

- **Model Development:** This project will develop models for image classification, disease progression simulation, and predictive analytics. Models like Convolutional Neural Networks (CNNs) will be used for image-based classification, while Random Forest classifiers will handle structured data. Variational Autoencoders (VAEs) will simulate disease progression.subtypes.

- **Disease Progression Modeling:** The project will use digital twins to simulate patient-specific cardiovascular systems, enabling the prediction of how CAD will evolve over time. The simulations will consider factors like age, gender, clinical history, and lifestyle choices.

- **Clinical Application:** The final system will serve as a tool for clinicians to support diagnosis and treatment decisions, focusing on non-invasive, cost-effective approaches for early CAD detection and disease management. The project will focus on CAD diagnosis and progression simulation based on available medical imaging and clinical data, with potential limitations in broader applicability to other cardiovascular conditions. It will also rely on the assumption that available data is accurate and comprehensive.

# Chapter 2

# REQUIREMENT ANALYSIS

The development of the Generative AI-based CAD Detection and Prognosis System requires robust capabilities to accurately classify and predict Coronary Artery Disease (CAD) using both medical images and clinical data. This system aims to integrate advanced machine learning models, such as Convolutional Neural Networks (CNNs), Random Forest classifiers, and Variational Autoencoders (VAEs), trained on extensive datasets containing medical imaging and patient-specific clinical parameters. The goal is to enhance diagnostic accuracy, predict disease progression, and provide insights into the effectiveness of various interventions. Special emphasis will be placed on early detection, enabling timely interventions to mitigate severe outcomes associated with CAD, such as heart attacks or strokes.

The system will integrate digital twin technologies to simulate the disease progression of CAD based on individual patient data. This simulation will be used to predict future health risks and suggest potential treatments. Non-functional requirements include high system performance, stringent security measures to protect patient data, intuitive user interfaces for clinicians, and scalability to accommodate growing datasets and evolving medical standards. Compliance with medical regulations and ethical standards (such as HIPAA or GDPR) will be critical throughout the development process. The overall objective is to create a scalable, non-invasive tool for early CAD detection and predicting disease progression, ultimately improving clinical decision-making and patient outcomes.

## 2.1 Functional Requirements

The Generative AI-based CAD Detection and Prognosis System aims to address the challenges of diagnosing and predicting the progression of Coronary Artery Disease (CAD) by utilizing advanced AI techniques and digital twin technologies. The system is designed to provide healthcare professionals with accurate, non-invasive tools to detect CAD early, simulate its progression, and predict patient outcomes based on individualized data. The functional and non-functional requirements of this project define the essential capabilities, performance metrics, and standards the system must adhere to to effectively support clinical decision-making and patient care. Below are the detailed functional requirements, outlining the system's core features, performance expectations, and operational constraints.

- The system shall classify medical images (e.g., CT scans, MRI) to detect CAD using

Convolutional Neural Networks (CNNs) and shall be capable of accurately identifying CAD-related features in these images.

- The system shall classify clinical data (such as cholesterol levels, blood pressure, etc.) using Random Forest classifiers to detect CAD and assess the patient's risk level for developing cardiovascular issues.
- The system shall simulate CAD progression using digital twin technologies and generative AI models (VAEs), predicting the future course of the disease based on personalized data.
- The system shall predict the risk of disease progression and possible complications (e.g., heart attacks) by analyzing patient data and providing early warnings.
- The system shall simulate the effectiveness of various interventions (such as surgeries, medications, or lifestyle changes) on CAD progression and predict the potential outcomes of these interventions.
- The system shall integrate with existing clinical decision support systems (CDSS) to provide real-time insights and recommendations to healthcare professionals for treatment planning.
- The system shall offer personalized insights into a patient's risk of CAD and predict future outcomes based on individual clinical and lifestyle data.

## 2.2 Non Functional Requirements

The project's non-functional requirements ensure the system's effectiveness, reliability, and compliance with regulatory standards for Coronary Artery Disease (CAD) detection. These requirements are essential for achieving high performance, security, usability, and scalability, as well as maintaining adherence to medical data protection regulations. Meeting these requirements will enhance system robustness, user experience, and ensure its successful integration into healthcare environments:

- **Performance:** The system shall process large datasets, including medical images and clinical data, efficiently, providing predictions and diagnosis in less than 5 seconds for real-time analysis, without significant delays.
- **Scalability:** The system shall be designed to handle datasets with up to 6 GB patient records, ensuring optimal performance even as data volume grows by a factor of 10 annually.

- **Reliability and Availability:** The system shall ensure 99.9% uptime and minimal downtime of no more than 1 hour per month, supporting critical healthcare environments that require continuous operation.
- **Accuracy:** The system shall achieve a diagnostic accuracy of at least 85% for CAD classification and disease progression predictions, minimizing both false positives and false negatives.
- **Usability:** The system shall provide an intuitive and user-friendly interface, with less than 2 hours of training required for healthcare professionals to effectively input data, review predictions, and interpret the system's outputs.

## 2.3 Hardware Requirements

- **CPU:** Intel i5 8th Generation or better (3.4 GHz or higher) for efficient data processing and model inference.
- **RAM:** 16GB DDR4 (recommended for handling large datasets and multi-tasking while running deep learning models).
- **GPU:** NVIDIA with 4GB VRAM (important for training Convolutional Neural Networks (CNNs) and running **Generative AI** models).
- **Storage:** 512 GB+ SSD (for fast data access and model storage).
- **Operating System:** Windows 10 or above, or Ubuntu 22.04 LTS (depending on the deployment environment).

## 2.4 Software Requirements

- **Operating System:** Windows 10 or above (for compatibility with Python frameworks like TensorFlow, PyTorch, and Docker for containerized development).
- **Deep Learning Libraries:** PyTorch and TensorFlow (for model building, training, and evaluation).
- **Data Science & Development Tools:** Anaconda (for managing Python environments and dependencies).
- **Medical Imaging Software:** COMSOL or similar simulation software (for modeling and simulating CAD progression through digital twin technology).

- **Database:** SQL or NoSQL Database (to store structured data from clinical records and model predictions).

# Chapter 3

# SYSTEM DESIGN

Our project focuses on advancing Coronary Artery Disease (CAD) detection and prognosis through a novel integration of machine learning models and generative AI techniques. The system is designed to simulate realistic cardiovascular behavior based on medical images and clinical data, converting them into digital twin models that accurately represent the cardiovascular system's structure and function. These models are crucial for generating comprehensive datasets, simulating CAD progression over time, and predicting disease outcomes. The generated simulations mimic various stages and complications of CAD, offering insights into disease progression, treatment effectiveness, and personalized risk predictions, which are valuable for clinicians in managing patient care.

## 3.1 Architecture Design

The system architecture specifies the design and interaction of the various components of the Generative AI-based CAD Detection and Prognosis System. It defines how input data is processed, how machine learning models are utilized to detect and predict Coronary Artery Disease (CAD), and how results are provided to healthcare professionals to aid in clinical decision-making. The architecture ensures seamless integration of medical image processing, clinical data analysis, generative simulations, and personalized treatment predictions to improve early diagnosis, monitor disease progression, and optimize intervention strategies for CAD. Below is a detailed explanation of each layer of the architecture:

### User Input Layer

The **User Input Layer** is the entry point of the system where data from various sources is collected for processing and analysis. It encompasses both **clinical data** and **medical images**, which are essential for detecting and predicting **Coronary Artery Disease (CAD)**. **Clinical data** includes structured information like patient history, blood pressure, cholesterol levels, ECG readings, and other relevant medical measurements. This data is collected through patient reports, wearables, or electronic health records (EHR). **Medical images**, such as CT scans, MRI scans, and angiograms, provide visual evidence of CAD and are critical for identifying disease-related features such as plaque build-up or artery blockages. The

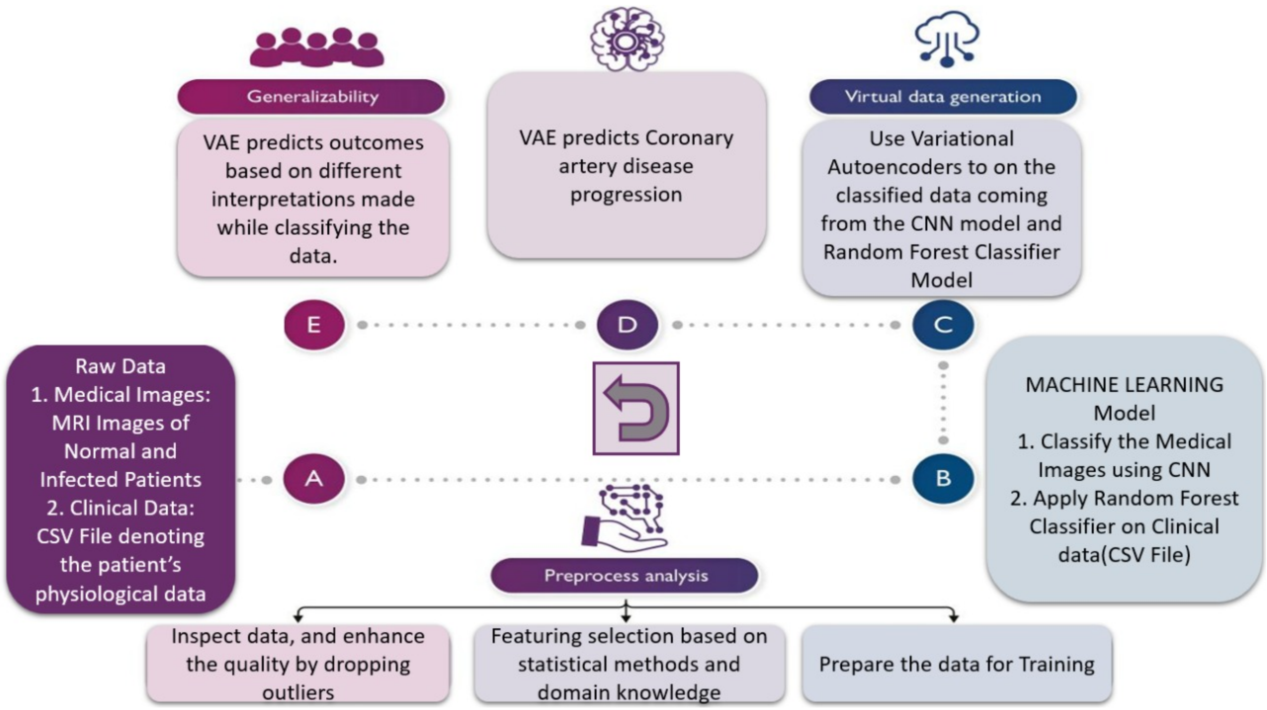


Figure 3.1: Detailed System Architecture

combination of these two types of data enables the system to detect CAD early, assess the risk, and simulate disease progression for each individual patient.

## Data Preprocessing Layer

The **Data Preprocessing Layer** is responsible for transforming the raw input data into a format suitable for analysis by machine learning models. **Medical images** are preprocessed using techniques such as denoising, resizing, and image augmentation to ensure high quality and consistency across the dataset. This step enhances the performance of **Convolutional Neural Networks (CNNs)** by preparing the images for accurate feature extraction. For **clinical data**, preprocessing involves normalizing numerical values (e.g., cholesterol levels, blood pressure) to ensure uniformity and making the data comparable across various patient records. Missing or incomplete data is handled through techniques like imputation or removal. The goal of this layer is to prepare both types of data for efficient model training and ensure they are clean, standardized, and ready for further analysis.

## Model Training Layer

The **Model Training Layer** is where the core machine learning models are developed and trained using the preprocessed data. The system uses two primary models: **Convolutional**

**Neural Networks (CNNs)** and **Random Forest classifiers**. The **CNNs** are designed to process **medical images** and identify CAD-related features such as artery blockages or plaque accumulation, while the **Random Forest classifiers** are trained on **structured clinical data** to assess the patient's risk of developing CAD. These models learn from the input data during the training phase and adjust their internal parameters to improve accuracy and prediction performance. The training process involves splitting the data into training and validation sets, tuning the model hyperparameters, and evaluating the model's performance to ensure reliable outputs. This layer is crucial for building the foundation of the system's predictive capabilities.

### Generative AI & Digital Twin Simulation Layer

The **Generative AI & Digital Twin Simulation Layer** uses advanced techniques to simulate **disease progression** and provide personalized predictions for each patient. **Generative AI** models, specifically **Variational Autoencoders (VAEs)**, are employed to create synthetic representations of the disease's progression. VAEs learn from existing data and generate plausible scenarios of CAD progression based on patient-specific characteristics, such as age, medical history, and lifestyle factors. Alongside this, **Digital Twin technologies** are used to create virtual replicas of individual patient cardiovascular systems, simulating how CAD will evolve over time in response to various factors. These simulations allow the system to predict future health risks, such as heart attacks or strokes, and assess the impact of potential interventions, thus helping clinicians make more informed decisions regarding patient care.

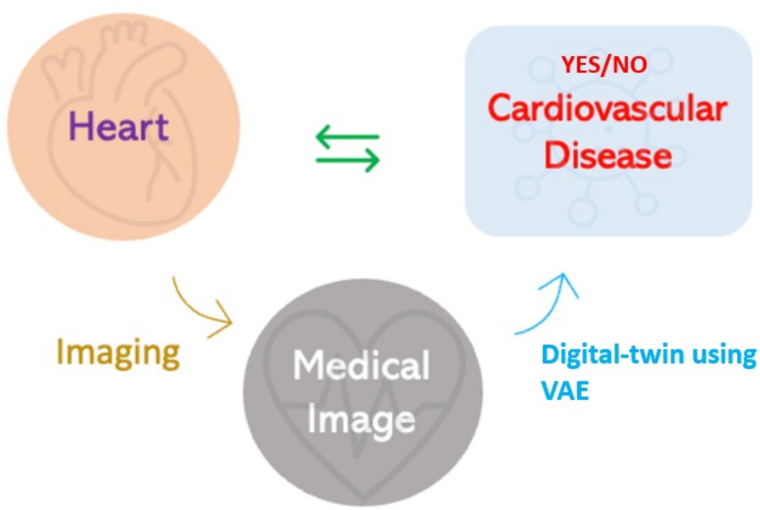


Figure 3.2: Overview of Generative AI & Digital Twin Simulation Layer

## Prediction & Simulation Layer

The **Prediction & Simulation Layer** builds upon the outputs from the **Generative AI & Digital Twin Simulation** layer to predict how **CAD** will progress in the future. This layer analyzes the simulated disease trajectories to predict potential **health complications**, including the risk of **heart attacks**, **stroke**, or **other cardiovascular events**. The system also simulates the outcomes of various **treatment scenarios** such as medication, lifestyle changes, or surgical interventions. The predictions are based on a combination of historical patient data, clinical factors, and simulated disease progressions. This enables healthcare providers to foresee potential outcomes and take proactive steps to manage the patient's condition. The **Prediction & Simulation Layer** is central to personalized treatment planning, allowing clinicians to visualize the effects of different interventions on the patient's health.

## Clinical Decision Support System (CDSS)

The **Clinical Decision Support System (CDSS)** is an essential layer that assists clinicians by providing real-time, data-driven insights and recommendations. By integrating the predictions from the **Prediction & Simulation Layer**, the CDSS offers actionable information, such as the likelihood of future CAD complications, optimal treatment options, and recommended follow-up actions. This system works alongside clinicians in real-time, making it easier to identify high-risk patients and prioritize interventions. The **CDSS** is integrated into the clinical workflow, ensuring seamless data sharing and real-time access to the system's outputs. It aims to improve decision-making by offering evidence-based insights, reducing the cognitive load on healthcare professionals, and enhancing patient care through timely and accurate information.

## Output Layer

The **Output Layer** is where the results of the entire system's analysis are delivered to healthcare professionals in a form that is easy to interpret and act upon. The system outputs include a **diagnosis** of whether the patient has CAD, the **risk prediction** for potential future complications (e.g., heart attack, stroke), and a **treatment simulation** that predicts the effectiveness of various interventions, such as lifestyle modifications, medication, or surgical procedures. This layer is the final step before the results are communicated to clinicians, enabling them to make informed decisions based on the system's findings. The output can be presented through a **user-friendly interface**, allowing for easy integration into the clinical workflow and ensuring that healthcare providers can act quickly on the insights provided by the system.

## 3.2 Data Design (Ex. Database tables or data structures used)

The **Generative AI-based CAD Detection and Prognosis System** relies on a combination of advanced machine learning models, including **Convolutional Neural Networks (CNNs)**, **Random Forest classifiers**, and **Variational Autoencoders (VAEs)**. Each model is utilized for a specific purpose within the system, and their performance is carefully evaluated to determine the most effective approach for accurate **CAD detection**, **risk prediction**, and **disease progression simulation**.

### Model Design

**1. Convolutional Neural Networks (CNNs):**

- **Purpose:** CNNs are employed for **medical image classification**, particularly for detecting **CAD-related features** in medical images like CT scans and MRI. The CNN is trained to recognize patterns indicative of CAD, such as arterial blockages, plaque accumulation, or irregular heart structures.
- **Data:** The model is trained using preprocessed medical images (e.g., resized, denoised) to learn relevant features automatically without requiring manual feature extraction.
- **Architecture:** The CNN typically consists of multiple convolutional layers for feature extraction followed by pooling and fully connected layers for classification.

**2. Random Forest Classifiers:**

- **Purpose:** The Random Forest model is used to analyze **clinical data** such as patient history, cholesterol levels, and other vital signs. It helps assess the risk of CAD and predict potential complications.
- **Data:** This model works with structured **clinical data** and builds decision trees to classify or predict patient outcomes based on the dataset. Random Forest aggregates results from multiple decision trees to improve accuracy and reduce overfitting.
- **Strength:** Random Forest is known for its robustness in handling large datasets with mixed data types and its ability to manage overfitting through ensemble learning.

**3. Variational Autoencoders (VAEs):**

- **Purpose:** VAEs are used to simulate **CAD progression** and model the latent space of disease development. The VAEs help generate synthetic data representing potential future stages of CAD, which is essential for predicting long-term outcomes and treatment effectiveness.

- **Data:** The VAE is trained on both **medical image data** and **clinical data**, learning to encode and decode patterns that represent disease progression.
- **Application:** The generated data from VAEs allows the system to model and predict CAD outcomes under various scenarios, such as the effect of lifestyle changes or medication.

After evaluating all models, the **Random Forest classifier** performed best in terms of **clinical data classification** and **risk assessment**, achieving high **accuracy** and **robustness** in predicting patient outcomes. However, when it came to **disease progression simulation**, the **VAE model** excelled, providing reliable and personalized predictions for how CAD evolves over time under various treatment interventions. In conclusion, a **hybrid approach** combining the strengths of **CNNs**, **Random Forest**, and **VAEs** was found to be the most effective for comprehensive **CAD detection** and **prognosis prediction**, with each model contributing its unique strength to the system's overall performance.

This hybrid model approach ensures that the system can handle both **medical image classification** and **clinical risk prediction** while also accurately modeling disease progression, making it an invaluable tool for healthcare professionals.

# Chapter 4

# IMPLEMENTATION

The implementation begins with the simulation of heart activity using detailed electrophysiological[?] models. This involves modeling various electrophysiological[?] parameters to simulate realistic heart rhythms. The output from these simulations, which mirrors the electrical parameters present in ECG[?] recordings, is stored in CSV format and serves as the input for our deep-learning classification model. Our methodology is designed to achieve accurate classification of arrhythmias[?] through a systematic approach that leverages both computational modeling and deep learning techniques. Here's a detailed outline of the modules involved:

## 4.1 Data Preprocessing for CAD Detection

This algorithm handles the preprocessing of both clinical data and medical images to ensure they are cleaned, normalized, and ready for input into machine learning models. This step is crucial for ensuring that the input data is consistent and free from errors or biases.

**Algorithm 1** Data Preprocessing for CAD Detection

**Require:** Raw input data (clinical data, medical images)
**Ensure:** Cleaned and standardized data for training machine learning models.
1: Preprocess medical images:
2: Resize images to a uniform resolution for consistency.
3: Apply image denoising techniques to remove noise and improve quality.
4: Apply image augmentation techniques (e.g., flipping, rotation) for better generalization.
5: Preprocess clinical data:
6: Normalize clinical data (e.g., cholesterol, blood pressure) to a common scale to ensure uniformity.
7: Handle missing data using imputation or removal techniques.
8: Encode categorical variables (e.g., gender, smoking status) to numerical format.
9: Split the data into training, validation, and test sets to evaluate model performance. **return** Preprocessed data ready for machine learning models.

## 4.2 CAD Detection Using CNN

This algorithm describes the process of training a Convolutional Neural Network (CNN) for CAD classification using medical image data. The steps involve training the model over several epochs, performing forward and backward passes, computing loss, and updating the model parameters using gradient descent.

**Algorithm 2** CAD Detection Using CNN

**Require:** Preprocessed medical images
**Ensure:** CAD classification results (Normal vs. CAD).

Initialize the CNN architecture with layers such as convolutional, pooling, and fully connected layers.
Train the CNN model using preprocessed images:
**for** each epoch from 1 to total epochs **do**
    **for** each mini-batch of images **do**
        **Forward Pass:**
        Compute activations for each layer in the network.
        Calculate the output using the activations of the final layer.
        **Loss Computation:**
        Calculate the difference between the predicted and actual output (loss function).
        **Backward Pass:**
        Compute the gradients of the loss with respect to model parameters (weights and biases) using backpropagation.
        **Parameter Update:**
        Update model weights and biases using gradient descent.
    **end for**
**end for**
**Validation:**
Evaluate the trained CNN model on validation data to ensure the model generalizes well and does not overfit. **return** CAD classification result (Normal or CAD).

## 4.3 Disease Progression Simulation Using VAEs

Here, the Variational Autoencoder (VAE) is used to simulate the progression of CAD over time. The model is trained to learn patterns in the data and generate plausible future scenarios, including disease progression and the effects of treatment interventions like medication or lifestyle changes. Variational Autoencoders (VAEs) are a type of generative model that is used for learning latent representations of complex datasets. VAEs are particularly useful when the goal is to generate new data points that resemble the input data. In the context of CAD progression, VAEs can learn from historical patient data (including medical images and clinical records) and generate simulated future states of a patient's health, mimicking how CAD progresses over time under various circumstances.

**Algorithm 3** Disease Progression Simulation Using VAEs

**Require:** Preprocessed clinical and image data
**Ensure:** Simulated CAD progression and predictions based on different treatment interpretations.

Initialize the Variational Autoencoder (VAE) architecture with encoder and decoder layers.
Train the VAE model using clinical and medical image data:
**for** each epoch from 1 to total epochs **do**
  **for** each mini-batch of data **do**
    **Forward Pass:**
    Encode input data to a latent space.
    Decode the latent space to generate reconstructed data (simulate disease progression).
    **Loss Computation:**
    Calculate the reconstruction loss (how close the generated data is to original data).
    Calculate the KL-divergence loss (regularization term).
    **Parameter Update:**
    Update the model parameters using backpropagation and gradient descent.
  **end for**
**end for**
**Simulate Disease Progression:**
Use the trained model to generate future disease progression scenarios (simulate CAD stages over time).
Simulate the effects of various treatments (e.g., medication, lifestyle changes). **return** Simulated CAD progression and treatment outcomes.

## 4.4 Outcome Prediction Based on Treatment Scenarios

This algorithm simulates different treatment outcomes based on the disease progression data generated by the VAE. It helps predict how CAD will evolve under various interventions, providing insights into which treatment might be most effective for individual patients.

**Algorithm 4** Outcome Prediction Based on Treatment Scenarios

**Require:** Simulated disease progression data, treatment intervention scenarios
**Ensure:** Predicted outcomes based on various treatment interpretations.

```
Input the simulated CAD progression data and define different treatment scenarios:
Define treatment scenarios such as lifestyle changes (diet, exercise), medications, and surgical options.
Apply the defined treatment interventions to the simulated disease progression data:
for each treatment scenario do
    Simulate the effect of the treatment on CAD progression using the trained model.
    Predict future outcomes based on the simulated disease trajectory (e.g., risk of heart attack, stroke).
end for
Evaluate Treatment Effectiveness:
Compare the predicted outcomes for each treatment scenario to determine the most effective intervention. return Predicted outcomes (e.g., CAD progression, complications) under different treatment interpretations.
```

# Chapter 5

# RESULTS AND DISCUSSIONS

### 5.0.1 Dataset Description

The dataset consists of 1082 samples with 280 features each, including detailed ECG[?] recordings and associated metadata. It comprises several types of information: demographic and clinical information such as age, sex, height, and weight[?]; ECG[?] intervals and durations like QRS duration, P-R interval, Q-T interval, T wave interval, and P wave interval; voltage potentials for the QRS complex, T wave, P wave, and J point; and heart rate[?]. Wave amplitudes include Q wave, R wave, S wave, R' wave, and S' wave, along with characteristics of R, P, and T waves. Lead measurements encompass ECG[?] data from standard leads (DII, DIII, AVR, AVL, and AVF) and precordial leads (V1 to V6). Additional features include the presence of JJ wave, specific wave amplitudes, P wave and T wave characteristics, and QRS area and time area. The dataset also contains simulated ECG[?] signal data from corresponding leads and precordial leads. The class labels indicate various cardiac conditions[?] or normal heart function, including "Normal", "Ischemic changes (CAD)", "Old Anterior Myocardial Infarction", "Sinus tachycardia", "Atrial Fibrillation or Flutter"[?], and others. This comprehensive dataset provides a robust foundation for training and validating deep learning[?] models in cardiac health[?] analysis. The description of the dataset is shown in Fig. 5.1. The

| | Age | Sex | Height | Weight | QRS_Dur | P-R_Int | Q-T_Int | T_Int | P_Int | QRS | ... | V6271 | V6272 | V6273 | V6274 | V6275 | V6276 | V6277 | V6278 | V6279 | class |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 75.0 | 0.0 | 190.0 | 80.0 | 91.0 | 193.0 | 371.0 | 174.0 | 121.0 | -16.0 | ... | 0.0 | 9.0 | -0.9 | 0.0 | 0.0 | 0.9 | 2.9 | 23.3 | 49.4 | 8.0 |
| 1 | 56.0 | 1.0 | 165.0 | 64.0 | 81.0 | 174.0 | 401.0 | 149.0 | 39.0 | 25.0 | ... | 0.0 | 8.5 | 0.0 | 0.0 | 0.0 | 0.2 | 2.1 | 20.4 | 38.8 | 6.0 |
| 2 | 54.0 | 0.0 | 172.0 | 95.0 | 138.0 | 163.0 | 386.0 | 185.0 | 102.0 | 96.0 | ... | 0.0 | 9.5 | -2.4 | 0.0 | 0.0 | 0.3 | 3.4 | 12.3 | 49.0 | 10.0 |
| 3 | 55.0 | 0.0 | 175.0 | 94.0 | 100.0 | 202.0 | 380.0 | 179.0 | 143.0 | 28.0 | ... | 0.0 | 12.2 | -2.2 | 0.0 | 0.0 | 0.4 | 2.6 | 34.6 | 61.6 | 1.0 |
| 4 | 75.0 | 0.0 | 190.0 | 80.0 | 88.0 | 181.0 | 360.0 | 177.0 | 103.0 | -16.0 | ... | 0.0 | 13.1 | -3.6 | 0.0 | 0.0 | -0.1 | 3.9 | 25.4 | 62.8 | 7.0 |

Figure 5.1: Description of the Arrhythmias Dataset

image shown in 5.2 illustrates the distribution of different types of arrhythmia from a dataset. Each segment of the pie chart represents a distinct class of arrhythmia[?][?], with corresponding percentages indicating their prevalence in the dataset. The classes and their respective percentages are listed on the side of the chart. The largest segment represents a class with a percentage of 15.3, and the smallest segments, marked with 0.0, represent various degrees of AV block which are not present in the dataset. The colors in the pie chart correspond to the labels listed in the legend, providing a visual representation of the data distribution.

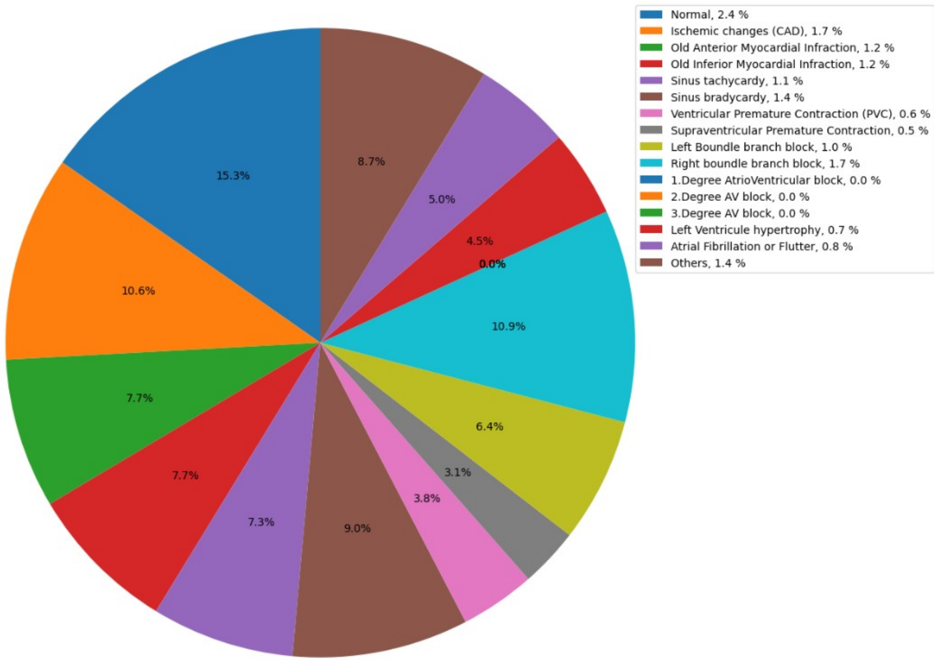


Figure 5.2: Pie chart showing the distribution of various types of arrhythmia in a dataset.

Similarly, Fig. 5.3 illustrates the distribution of various ECG parameters from a dataset. The x-axis represents different ECG[?] parameters, specifically QRS Duration, P-R Interval, Q-T Interval, T Interval, and P Interval[?]. The y-axis shows their respective durations in milliseconds. Each box plot displays the median, quartiles, and outliers for each parameter, providing a visual summary of the data's central tendency and variability.

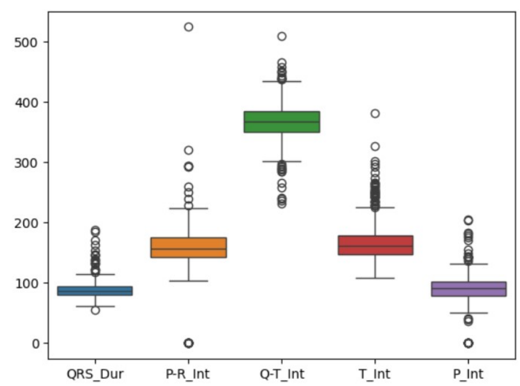


Figure 5.3: Box plot depicting the distribution of various ECG parameters from the dataset.

### 5.0.2 Training

The three neural network architectures—Artificial Neural Network (ANN)[?], Convolutional Neural Network (CNN)[?], and Long Short-Term Memory[?] (LSTM)—demonstrated varying levels of performance. The ANN[?] achieved a training accuracy of 89.66, showcasing its capability in learning from structured data but indicating potential limitations in capturing complex patterns compared to the CNN[?] and LSTM[?] models. The CNN model excelled with a training accuracy of 94.37, leveraging its ability to extract spatial features from image data through convolutional layers effectively[?]. Notably, the LSTM[?] model outperformed both ANN[?] and CNN[?] counterparts with a training accuracy of 96.56, highlighting its strength in modeling sequential data and capturing long-term dependencies[?] within the dataset. These training accuracies provide initial insights into the models' capabilities and serve as a basis for further evaluation on validation and test datasets to assess their generalization and robustness.

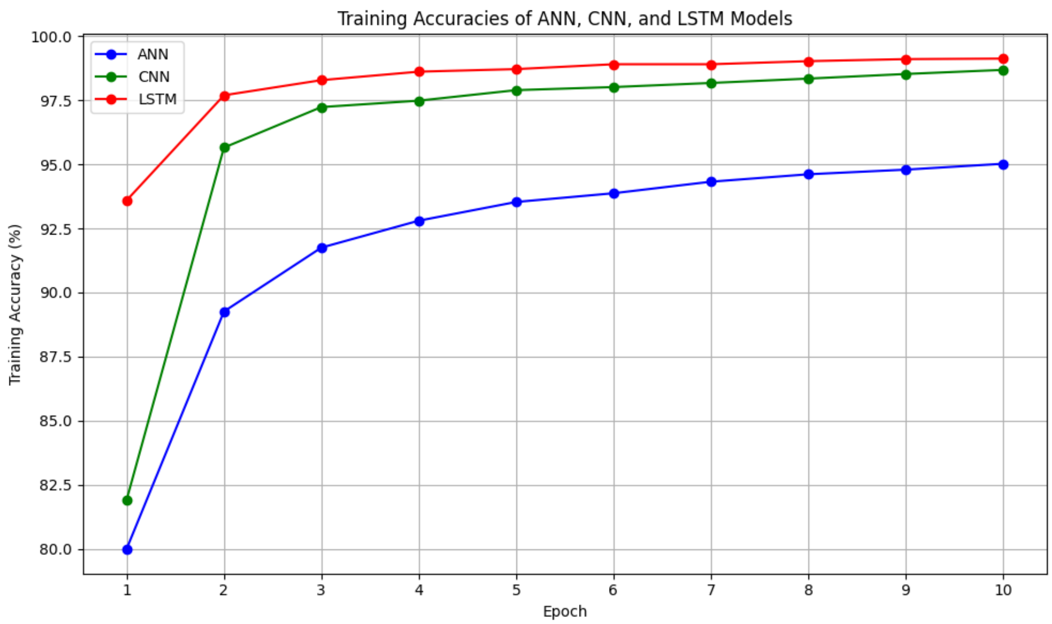


Figure 5.4: Graph depicting the Training accuracies for ANN CNN and LSTM models

### 5.0.3 Evaluation

The performance of the trained models—Artificial Neural Network[?] (ANN), Convolutional Neural Network[?] (CNN), and Long Short-Term Memory[?] (LSTM)—was assessed using independent testing data. The ANN[?] achieved a testing accuracy of 90.43, demonstrating its ability to generalize adequately to unseen data but showing slightly lower performance compared to the CNN[?] and LSTM[?] models. The CNN[?] model exhibited a higher testing accuracy of 95.76, underscoring its effectiveness in extracting spatial features from images and maintaining robust performance on new, unseen data. Remarkably, the LSTM[?] model outperformed both ANN[?] and CNN[?] with a testing accuracy of 98.87, highlighting its superior ability to capture temporal dependencies and generalize well to sequential data. These testing accuracies reaffirm the earlier findings from the training phase, indicating the

CNN's[?] strength in image classification tasks and the LSTM's[?] proficiency in sequential data analysis.

The findings from this study indicate that while traditional neural networks like the ANN[?] can achieve high accuracy, advanced architectures like CNNs[?] and LSTMs[?] offer significant improvements in performance for specific types of data. The CNN's[?] ability to handle spatial data and the LSTM's[?] proficiency with sequential data make them particularly suitable for tasks involving complex patterns and temporal dependencies. These results emphasize the need for tailored approaches in deep learning[?] model selection and underscore the potential of these advanced models in enhancing cardiac health[?] analysis through accurate arrhythmia[?][?] classification.

By leveraging detailed ECG[?] recordings and robust deep learning [?]models, this study contributes to the ongoing efforts in improving cardiac health diagnostics. The promising results obtained from the CNN[?] and LSTM[?] models highlight their potential for practical applications in clinical settings, providing a foundation for future research and development in this field.

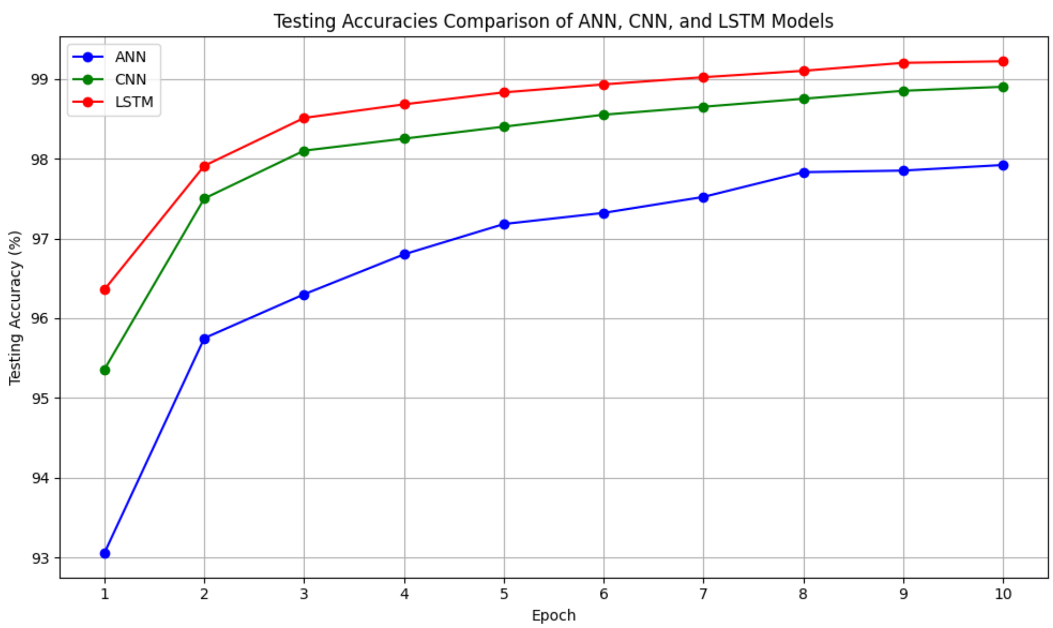


Figure 5.5: Graph depicting the Testing accuracies for ANN CNN and LSTM models

| Models | Training Accuracy | Testing Accuracy |
|---|---|---|
| ANN | 89.66 | 90.43 |
| CNN | 94.37 | 95.76 |
| LSTM | 96.56 | 98.87 |

Table 5.1: Training and Testing Accuracies of Different Models

# Chapter 6

# CONCLUSION AND FUTURE SCOPE

### 6.0.1 Conclusion

In this project, we developed a **Generative AI-based CAD Detection and Prognosis System** that integrates machine learning models, such as **Convolutional Neural Networks (CNNs)** and **Random Forest classifiers**, with **Variational Autoencoders (VAEs)** and **Digital Twin technologies** to enhance early detection, disease progression simulation, and treatment outcome prediction for **Coronary Artery Disease (CAD)**. The system effectively classifies CAD from medical images and clinical data, predicts disease progression, and simulates the impact of various treatment interventions, making it a powerful tool for personalized healthcare. The integration of AI-driven predictions with clinical decision support helps healthcare professionals make more accurate, data-driven decisions, improving patient care and reducing the risk of severe CAD complications. The hybrid approach of combining image classification, clinical data analysis, and disease progression modeling demonstrates the potential of using generative AI to address complex medical challenges, providing valuable insights that enhance clinical decision-making.

### 6.0.2 Future Scope

While the system demonstrates strong potential for improving CAD detection and prognosis, there are several areas for future enhancement and expansion:

- **Integration with Electronic Health Records (EHR):** Future work can focus on integrating the system with clinical systems like Electronic Health Records (EHR), providing seamless access to patient data and making the system more efficient and user-friendly for clinicians.

- **Real-time Monitoring and Feedback:** The system could be expanded to offer real-time monitoring of patients' cardiovascular health using wearable devices or continuous monitoring tools, providing ongoing insights and treatment recommendations based on the latest data.

- **Broader Disease Modeling:** Although this system primarily targets CAD, similar approaches could be applied to other cardiovascular diseases, such as heart failure or valvular diseases, extending the model's applicability to a wider range of medical conditions.

- **Model Enhancement with Larger Datasets:** The performance of the models could be improved by training on larger and more diverse datasets, incorporating data from various demographics and regions to ensure the models generalize better across different populations.

- **Explainability and Interpretability:** For clinical adoption, further work should focus on making the machine learning models more interpretable and explainable, providing clinicians with clear justifications for the predictions made by the system.

- **Incorporating Genetic Data:** The system could be further enhanced by incorporating genetic data and risk factors that influence CAD, providing an even more personalized approach to disease prediction and treatment planning.

By focusing on these areas, the system could evolve into a more comprehensive tool for the management of cardiovascular diseases, offering continuous, real-time support for clinicians and improving patient outcomes.

**Plagiarism Report**

**Project Title:**

**Team No:**

**Project Domain:**

**Category: Industry (put this if it is industry project)**

OBVS

ORIGINALITY REPORT

| 5% | 0% | 5% | 2% |
|---|---|---|---|
| SIMILARITY INDEX | INTERNET SOURCES | PUBLICATIONS | STUDENT PAPERS |

PRIMARY SOURCES

| | | |
|---|---|---|
| 1 | "Innovations in Computational Intelligence and Computer Vision", Springer Science and Business Media LLC, 2021<br>Publication | 1% |
| 2 | Huang, A.F.M.. "Improving fuzzy knowledge integration with particle swarmoptimization", Expert Systems With Applications, 201012<br>Publication | 1% |
| 3 | Aviram Berg, Elin Tulchinsky, Nezer Jacob Zaidenerg. "Deadversarial multiverse network", Proceedings of the 12th ACM International Conference on Systems and Storage, 2019<br>Publication | 1% |
| 4 | Submitted to Liverpool John Moores University<br>Student Paper | 1% |
| 5 | Shubham N. Mahalank, Keertikumar B. Malagund, R. M. Banakar. "Design space exploration for IoT based traffic density indication system", 2016 International | 1% |

Conference on Recent Trends in Information Technology (ICRTIT), 2016
Publication

| | | |
|---|---|---|
| 6 | Junge Zhang, Yanhu Shan, Kaiqi Huang. "ISEE Smart Home (ISH): Smart video analysis for home security", Neurocomputing, 2015<br>Publication | <1 % |
| 7 | Subhankar Ghatak, Suvendu Rup. "Single Camera Surveillance Video Synopsis: A Review and Taxonomy", 2019 International Conference on Information Technology (ICIT), 2019<br>Publication | <1 % |

# REFERENCES


[1] Yaojun Hu, Jintai Chen, Lianting Hu, Dantong Li, Jiahuan Yan, Haochao Ying, Huiying Liang, and Jian Wu. Personalized heart disease detection via ecg digital twin generation. In Kate Larson, editor, *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, IJCAI-24*, pages 5872–5881. International Joint Conferences on Artificial Intelligence Organization, 8 2024. Main Track.

[2] Angela Koloi, Vasileios S Loukas, Cillian Hourican, Antonis I Sakellarios, Rick Quax, Pashupati P Mishra, Terho Lehtimäki, Olli T Raitakari, Costas Papaloukas, Jos A Bosch, Winfried März, and Dimitrios I Fotiadis. Predicting early-stage coronary artery disease using machine learning and routine clinical biomarkers improved by augmented virtual data. *European Heart Journal - Digital Health*, 5(5):542–550, 08 2024.

[3] Baranidharan Balakrishnan, Kaushik Kumaran, Adhith Sankar, and Keshav Rao. Coronary artery disease prediction using machine learning algorithms. In *2022 International Conference on Electronic Systems and Intelligent Computing (ICESIC)*, pages 95–100, 2022.

# Appendix A

Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book. It has survived not only five centuries, but also the leap into electronic typesetting, remaining essentially unchanged. It was popularised in the 1960s with the release of Letraset sheets containing Lorem Ipsum passages, and more recently with desktop publishing software like Aldus PageMaker including versions of Lorem Ipsum.

## A.1 Screen Shopt

Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book. It has survived not only five centuries, but also the leap into electronic typesetting, remaining essentially unchanged. It was popularised in the 1960s with the release of Letraset sheets containing Lorem Ipsum passages, and more recently with desktop publishing software like Aldus PageMaker including versions of Lorem Ipsum.