College of Engineering and Technology

Project Work – Student Hand Book

Project Batch ID

CTB07806

Degree/ program	B.Tech	**Specialisation**	Computer Science & Engineering

Name of student	**Register Number**	**Department**	**Mobile Number**	**Email ID**
Siddharth Vats	RA2111003010606	CTECH	8928729440	ss3108@srmist.edu.in
Khayati Sharma	RA2111003010710	CTECH	8851202831	kp3185@srmist.edu.in

Working Title of the Project:	**AI–Powered Early Diagnosis And Personalised Health Recommendations For Coronary Artery Disease (CAD) Using Predictive Analytics**
Project Site / Location	Chennai
Name and address of the company / organisation (Applicable for projects with industry or industry support)	SRM IST, Kattankulathur, Chengalpattu District-603203

Supervision Team

	Supervisor	**Co-Supervisor**	**External Supervisor (If applicable)**
Name	Dr. M. Revathi		
Designation	Assistant Professor		
Department	Computing Technologies		
Campus	Kattankulathur		
Telephone	9894226698		
E-mail	revathim@srmist.edu.in		

College of Engineering and Technology

Project Work – Student Hand Book

Academic Year	2024-2025 (Even)	**Semester**	8
Course Code	**18CSP109L**	**Course Title**	Project

Mission Statement

Our mission is to leverage machine learning to enable early detection of heart disease and provide AI-assisted medication guidance, aiming to reduce mortality through accessible, preventive healthcare solutions.

Problem (or) Product Description:
Cardiovascular diseases (CVDs) remain one of the leading causes of death globally. In many cases, patients are diagnosed only at advanced stages due to the lack of timely screening and personalized medical guidance. Traditional diagnostic methods can be time-consuming, expensive, or inaccessible to populations in remote or underserved areas. Moreover, even after diagnosis, patients often struggle with understanding their condition or adhering to appropriate medication regimens due to lack of access to expert medical counsel.
This project addresses these critical gaps by building an AI-powered system that not only predicts the likelihood of heart disease using clinical data but also assists users with medication information and guidance through an AI-based chemist assistant module. The goal is to enable early diagnosis, improve medication adherence, and ultimately reduce preventable mortality due to heart-related illnesses.
Assumptions and Constraints
It is assumed that users will provide accurate health data and that the dataset used is a reliable representation of real-world cases. The system assumes users understand it offers supportive—not definitive—diagnoses. Constraints include limited access to diverse medical datasets, the need for high interpretability, and compliance with healthcare regulations. The AI chemist assistant is informational and not a substitute for professional advice.
Stakeholders
Key stakeholders include patients seeking early diagnosis, doctors using the tool for support, and healthcare institutions looking to enhance screening processes. Secondary stakeholders are researchers, software developers, academic evaluators, and regulatory bodies ensuring ethical deployment.

Division of work and contributors

Time period		Activities or components of the project	Name/Register Number of the Individual Contributor	Names/Register Number of the Joint Contributors
From Date	To Date			
15/11/24	15/11/24	Discussion on the title		KHAYATI SHARMA: RA2111003010710 SIDDHARTH VATS: RA2111003010606
17/11/24	17 /11/24	Confirmation of the title		KHAYATI SHARMA: RA2111003010710 SIDDHARTH VATS: RA2111003010606
12/12/24	14/12/24	Selection of Algorithm	SIDDHARTH VATS: RA2111003010606	
03/01/25	13/01/25	Literature Survey	KHAYATI SHARMA: RA2111003010710	
14/01/25	21/01/25	Discussion on Novelty		KHAYATI SHARMA: RA2111003010710 SIDDHARTH VATS: RA2111003010606
22/01/25	31/01/25	Setup of Code	SIDDHARTH VATS: RA2111003010606	
01/02/25	10/03/25	Development of Application Interface	SIDDHARTH VATS: RA2111003010606	
11/03/25	31/03/25	Backend Code Setup	SIDDHARTH VATS: RA2111003010606	
15/04/25	18/03/25	Research Paper Draft	KHAYATI SHARMA: RA2111003010710	
17/04/25	18/04/25	Research Paper submitted for Review	KHAYATI SHARMA: RA2111003010710	
18/04/25	19/04/25	Project Report Draft	KHAYATI SHARMA: RA2111003010710	
19/04/25	30/04/25	Project Report Verification		KHAYATI SHARMA: RA2111003010710 SIDDHARTH VATS: RA2111003010606

30/04/25	03/05/25	Paper submission	KHAYATI SHARMA: RA2111003010710	

Signature of the Supervisor

Summary record of major progress meetings with supervisors

Summary record of major progress meetings with supervisors			**Working title of dissertation/research project:** **AI–Powered Early Diagnosis And Personalised Health Recommendations For Coronary Artery Disease (CAD) Using Predictive Analytics**	
Meeting date & supervisors present	**Progress since last meeting**	**Agreed programme of work and target dates**	**Other issues, e.g. facilities, supervision, training needs, etc.**	**Date of next meeting**
15/11/25 Dr. Revathi M	Project title confirmation and finalized criteria for algorithm selection.	Develop prototype by 01 February 2024.	Novelty issues	25/01/25
25/01/25 Dr. Revathi M	Addition of Novelty to our project	Add new models and AI chemist to the website by 10 February 2025	Add models	17/02/2025
17/02/25 Dr. Revathi M	Addition of new models and a conversational chatbot	Research on the models used.	More knowledge on AI Chemist	23/03/25

Summary record of major progress meetings with supervisors

Summary record of major progress meetings with supervisors			**Working title of dissertation/research project:**	
Meeting date & supervisors present	**Progress since last meeting**	**Agreed programme of work and target dates**	**Other issues, e.g. facilities, supervision, training needs, etc.**	**Date of next meeting**
2/05/2025 Dr. Revathi M	Addressing feedback and revisions. Preparing for conference submission.	Plan presentation for the conference.	Discussing any last-minute adjustments or additions to the conference submission.	12/05/2025

Signature of the Supervisor

Worksheet / Data collection / Observation etc

For this project, data was collected from a publicly available and widely recognized dataset: the Cleveland Heart Disease dataset. This dataset is part of the UCI Machine Learning Repository and has been extensively used for research purposes in predictive analysis of cardiovascular diseases. It contains anonymized medical records of 303 patients, with each entry comprising 14 attributes related to clinical and diagnostic features relevant to heart health.

These attributes include age, sex, chest pain type, resting blood pressure, cholesterol levels, fasting blood sugar, resting electrocardiographic results, maximum heart rate achieved, exercise-induced angina, oldpeak (ST depression), slope of peak exercise ST segment, number of major vessels colored by fluoroscopy, thalassemia, and the target variable indicating the presence of heart disease.
To ensure consistency and accuracy, the data underwent preprocessing which included:

- Handling missing values and outliers.
- Encoding categorical variables into numerical formats.
- Scaling and normalization to prepare the features for model training.
- Target class binarization, where the 'target' field was converted into binary format: 0 for no heart disease and 1 for presence of heart disease.

This cleaned and structured data was then split into training and testing subsets using an 80:20 stratified ratio to maintain balance between the two classes in both subsets.

From the analysis of the dataset and model results, several key observations emerged:

1. **Class Imbalance Exists**: Although the dataset is relatively balanced, minor skewness in the number of healthy versus unhealthy individuals slightly affected model performance, necessitating the use of stratified splitting.
2. **Feature Importance**: Features such as chest pain type, thalassemia, and exercise-induced angina showed significant influence in predicting the presence of heart disease. This was particularly evident in models like Random Forest, which provided feature importance rankings.
3. **Model Performance**: The ensemble model using **Soft Voting Classifier**—combining Support Vector Machine, Logistic Regression, and Random Forest—demonstrated improved accuracy and

robustness compared to individual models. It leveraged the strengths of each algorithm to handle linear, non-linear, and categorical relationships in the data.

4. **Precision-Recall Trade-off**: The model exhibited a high **Precision-Recall AUC**, indicating good performance in detecting positive cases (patients with heart disease), which is crucial in healthcare where false negatives can be fatal.
5. **Real-World Applicability**: The model showed consistent prediction accuracy on unseen data, suggesting potential for real-world deployment as a diagnostic aid. However, medical validation and ethical considerations remain essential before clinical use.
6. **AI Chemist Assistant**: Based on the model's prediction, the system also mapped disease conditions to a curated list of commonly prescribed heart medications. Though currently static, this module can be expanded with integration to pharmaceutical databases for personalized recommendations.

Research Article with Journal Publication Details / Patent disclosure form with patent status

AI-Powered Early Diagnosis and Personalized Health Recommendations for Coronary Artery Disease (CAD) using Predictive Analytics

Siddharth Vats
Department of Computing Technologies, School of Computing SRM Institute of Science and Technology
Tamil Nadu, India
siddharthvats44@gmail.com

Khayati Sharma
Department of Computing Technologies, School of Computing SRM Institute of Science and Technology
Tamil Nadu, India
sharmakhayati0123@gmail.com

M. Revathi
Department of Computing Technologies, School of Computing SRM Institute of Science and Technology
Tamil Nadu, India
revathim@srmist.edu.in

***Abstract*—Cardiovascular diseases, especially coronary artery disease, have been the leading cause of deaths worldwide. Hence, the early detection and personalized treatment strategies for coronary artery disease (CAD) are now needed more than ever. This paper suggests an AI-driven framework based on machine learning algorithms for the timely risk detection of CAD based on 13 clinical parameters. The framework utilizes an ensemble learning method that includes logistic regression, random forest and support vector machine classifiers which improve the prediction accuracy. The proposed system also comes equipped with an AI chemist assistant which helps users specifically with medication-related queries, such as drug interactions, administration instructions and prescription explanations. This feature managed textual and visual inputs using Google's generative AI (Gemini), thus allowing users to upload photos or ask questions and receive instant assistance. Adding to the suite of features, the system also comes with a health recommendation engine, which provides users with certain lifestyle changes to be adopted based on the risk assessment, and a chatbot which answers more general questions that the user might have related to his diet, exercise, sleep or any symptoms being experienced. This approach provides an integrated preventative care solution by combining CAD risk prediction, AI-based medication assistance, a coronary health chatbot and lifestyle recommendations. Apart from improving accessibility, the suggested solution also improves patient education and enables more informed decision-making in cardiovascular disease care. The experimental results show the effectiveness of the ensemble model in risk prediction of CAD and AI chemist's capacity to help users with medication-related concerns. This integrated system aims to combine AI-based diagnostic techniques with medicative and general heart-health assistance, thus improving the efficiency and accessibility of healthcare services.**

***Keywords*—coronary artery disease (CAD), ensemble learning, generative AI, coronary health chatbot, AI medication assistant**

I. Introduction

Cardiovascular diseases (CVDs), and coronary artery disease (CAD) in particular, continue to be a major cause of morbidity and mortality worldwide [1]. Early diagnosis and risk stratification are essential to avoid serious complications. However, standard diagnostic procedures often involve long clinical assessment, advanced equipment, and specialist interpretation. Advances in artificial intelligence (AI) and machine learning (ML) have revolutionized predictive analytics into a valuable tool in medical diagnosis, with enhanced accuracy, efficiency, and accessibility [2].

The system in this case is built based on research with artificial intelligence to facilitate predictive analytics for coronary artery disease risk assessment and individualized health recommendations. To improve the prediction accuracy, the system integrates a hybrid ensemble approach that combines three distinct machine learning models: logistic regression, random forest, and a support vector machine (SVM). This combination leverages the strengths of each algorithm to produce more robust and reliable results. The models leverage clinical result data and integrate basic cardiovascular risk predictors to yield patient-specific risk measurements. This research goes beyond typical predictive modeling by incorporating an AI-enhanced chemist instrument that employs LLMs to facilitate users' understanding of medicines, drug interactions, and life changes. Integrating a multimodal AI system that can analyze text-based input and visual cues enhances user engagement and usability.

The primary objective of this research is to develop an AI-powered system capable of early diagnosis and risk assessment for Coronary Artery Disease (CAD) using predictive analytics. Additionally, the system integrates an AI Chemist Assistant to provide medication-related guidance and answer users' queries about drugs, side effects, and interactions. The proposed framework leverages machine learning algorithms for CAD prediction and natural language processing (NLP) for AI-driven patient interaction, enhancing accessibility and engagement in healthcare.

II. Related Work

Cardiovascular disease prediction has been a major focus of research with several studies using machine learning to improve diagnostic performance and detect diseases early. Many have attempted various models, datasets, and methods to enhance predictive systems. Seckeler and Hoke [3] presented a detailed overview of the epidemiology of rheumatic heart disease and its burden and long-term public health impact. Early detection remains a recurring theme in heart disease research, particularly within vulnerable and underserved populations. Several studies underscore the urgency of diagnosing cardiovascular conditions before they progress to severe stages. For instance, Gaziano et al. [4] raised concern over the growing burden of coronary artery disease (CAD) in low- and middle-income countries, drawing attention to how limited access to timely care often exacerbates patient outcomes. Their findings emphasize the need for scalable and proactive diagnostic solutions in such regions.

Acceptance Notification - IEEE ICCTDC 2025 External Inbox ×

Microsoft CMT <noreply@msr-cmt.org>
to me

Fri, Apr 25, 1:05 PM (4 days ago)

Dear Khayati Sharma

Paper ID / Submission ID : 1534

Title : AI-Powered Early Diagnosis and Personalized Health Recommendations for Coronary Artery Disease (CAD) using Predictive Analytics

Greeting from IEEE ICCTDC 2025

We are pleased to inform you that your paper has been accepted for the Presentation as a full paper for the– "2025 International Conference on Computing Technologies &Data Communication (ICCTDC), Hassan , Karnataka, India.

All accepted and presented papers will be submitted to IEEE Xplore for the further publication.

You should finish the registration before deadline, or you will be deemed to withdraw your paper:

Complete the Registration Process (The last date of payment Registration is 30 APRIL 2025)

Payment Links :

For Indian Authors: https://rzp.io/rzp/icctdc

For Foreign Authors: https://rzp.io/rzp/ICCTDCForeign

Further steps like IEEE PDF xpress and E copyright will be given later once registration is over after the deadline.

AGASTIYAR RESEARCH

1st ICCTDC - 2025

About the event

ICCTDC-25 ,Hasaan, India. Technically Co-Sponsored by IEEE Bangalore Section

Venue

Malnad College of Engineering, Hassan.

Starts at

4/07/2025

Ends at

5/07/2025

Show More

₹ 9000.00

Payment Successful!

Share Your Experience

Payment ID: pay_QOj7S1TcjBXQ3M

A confirmation email has been sent to you