AI-DRIVEN CKD AND CVD PREDICTION AND SPECIALIST RECOMMENDATION

A PROJECT REPORT

Submitted By

RICHARD NICHOLES M	**310121104085**
KAMAL RAJ D	**310121104302**

In partial fulfillment for the award of

BACHELOR OF ENGINEERING

In

COMPUTER SCIENCE AND ENGINEERING

ANAND INSTITUTE OF HIGHER TECHNOLOGY

(AN AUTONOMOUS INSTITUTION)

AFFILIATED TO ANNA UNIVERSITY: CHENNAI 600025

MAY 2025

Certified that this project **"AI-DRIVEN CKD AND CVD PREDICTION AND SPECIALIST RECOMMENDATION"** is the Bonafide work of **RICHARD NICHOLES M (310121104085)** and **KAMAL RAJ D (310121104302)** who carried out the project work under my supervision.

SIGNATURE   SIGNATURE

Dr. M. Maheswari, M.E., (Ph.D.),   **Dr. Jancy Sickory Daisy, M.E., (Ph.D.),**

HEAD OF THE DEPARTMENT   **ASSOCIATE PROFESSOR**

Department of Computer Science   Department of Computer Science

and Engineering,   and Engineering,

Anand Institute of Higher Technology,   Anand Institute of Higher Technology,

Kazhipattur, Chennai-603103   Kazhipattur Chennai-603103

Submitted for Project and VIVA VOCE held on _____

 INTERNAL EXAMINER   EXTERNAL EXAMINER

ACKNOWLEDGEMENT

ABSTRACT

As artificial intelligence (AI) continues to transform the healthcare landscape, there is a growing need for intelligent, accessible diagnostic tools that can support early disease detection and preventive care. This paper introduces the development and implementation of a web-based Multi-Disease Prediction System designed to assess the risk of multiple medical conditions—namely diabetes, heart disease, and Parkinson's disease-through the application of advanced machine learning algorithms. The system is built using Streamlit, a Python-based framework that facilitates the creation of dynamic, interactive web applications with minimal overhead. The user interface is designed with usability in mind, enabling seamless interaction for both medical professionals and general users. To safeguard user data, the application incorporates a secure authentication mechanism featuring user registration and login functionality. Credentials are protected using SHA-256 hashing, and all user-related information is stored securely using a lightweight, file-based backend system powered by JSON, ensuring data integrity and privacy without the complexity of a full-scale database These models leverage structured healthcare data to deliver real-time predictions, empowering users with immediate insights into their potential health risks. The proposed solution serves as a preliminary screening and decision-support tool, aiming to bridge the gap between initial symptom presentation and formal medical diagnosis. By integrating AI with an accessible and user-friendly interface, this system demonstrates a scalable, efficient, and deployable approach to enhancing early diagnosis and promoting proactive health management in diverse settings, from individual use.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

SYMBOLS	**ABBREVIATIONS**
CKD	Chronic Kidney Disease
CVD	Cardiovascular Disease
BMI	Body Mass Index
EHR	Electronic Health Record
PDF	Portable Document Format

LIST OF TABLES

CHAPTER 1

INTRODUCTION

In the modern era, the fusion of technology with healthcare has opened new frontiers for disease prevention, early diagnosis, and efficient treatment planning. Among the many advancements, machine learning (ML) has emerged as a transformative force, offering predictive insights that can significantly enhance medical decision-making. The rise in lifestyle-related diseases such as diabetes, cardiovascular conditions, and neurological disorders like Parkinson's disease has further underscored the need for accessible, intelligent diagnostic tools.

By empowering users to self-assess their health risks, technology can play a critical role in fostering preventive healthcare and reducing the burden on clinical resources. Despite the advancements in healthcare diagnostics, early detection remains a major challenge, especially in regions with limited access to specialized medical infrastructure. Conventional diagnostic methods often involve time- consuming, expensive, and geographically restricted processes. To bridge this gap, there is a growing demand for platforms that combine accuracy, accessibility, and user- friendliness. Machine learning models, trained on historical medical data, can serve as virtual health advisors, offering preliminary assessments that can prompt individuals to seek timely medical attention. Such solutions have the potential to democratize healthcare, ensuring that proactive health management is not restricted to only urban or privileged communities.

In this project, we propose a web-based Multi-Disease Prediction System that leverages machine learning algorithms to predict the likelihood of multiple diseases based on clinical parameters entered by users. The platform is built using Streamlit, a popular open-source Python library that allows rapid web application development with a focus on interactivity and simplicity. In addition to predictive capabilities, the application includes a secure authentication system featuring user registration and login functionalities. Passwords are safeguarded using SHA-256 hashing algorithms, ensuring that user credentials are stored securely, thus enhancing trust and protecting sensitive information.

Users can input key medical details such as glucose levels, blood pressure, or motor skills assessments, depending on the disease being screened. Based on this input, the system provides a probabilistic analysis that suggests the potential presence of diabetes, heart disease, or Parkinson's disease. By providing instant feedback, the application not only empowers users to take control of their health but also encourages them to consult healthcare professionals for further validation and treatment.

1.1 OBJECTIVE

The primary objective of this project is to automate the detection of kidney and heart diseases and provide hospital recommendations based on the patient's condition. The key objectives of the project are as follows:

- **Risk Prediction:** Use machine learning algorithms to analyze user-provided clinical input and predict the likelihood of kidney or heart disease.

- **Interactive Web Interface:** Build a user-friendly application using Streamlit for seamless interaction by both patients and healthcare professionals.

- **Secure User Management:** Implement registration and login functionality with SHA-256 password hashing and a JSON-based backend for secure data handling.

- **Hospital Recommendation System:** Recommend suitable hospitals based on the user's condition, location, and hospital specialization.

- **Efficiency and Accessibility:** Reduce manual diagnostic efforts, minimize human errors, and ensure timely treatment by providing an AI-driven, automated solution for disease prediction and hospital selection.

1.2 SCOPE

- **Input Format**: Accepts patient health details like age, gender, blood pressure, glucose level, body mass index (BMI), and lifestyle factors via user-friendly web forms.

- **Data Processing**: Validates and standardizes user input before feeding it into disease- specific machine learning models for accurate analysis.

- **Disease Detection**: Utilizes trained machine learning models (Logistic Regression, Decision Trees, etc.) to predict the likelihood of diseases such as diabetes, heart disease, and kidney disease.

- **Output Format**: Provides instant prediction results with clear status indicators (e.g., "High Risk", "Low Risk") and basic health recommendations.

- **Target Audience**: Designed for patients, healthcare professionals, health clinics, and researchers looking for quick pre-diagnostic insights.

CHAPTER 2
LITERATURE SURVEY

Title	: AI Predictive Analytics Approach for Early Prognosis of CKD
Authors	: K. M. Tawsik Jawad, Anusha Verma, Fathi Amsaad
Year	: 2024
Published	: arXiv

Concept Discussed: This study introduces an AI-driven predictive analytics approach specifically designed for the early detection of Chronic Kidney Disease (CKD). By employing ensemble learning techniques, the researchers aimed to enhance the accuracy and reliability of CKD prognosis. However, what sets this approach apart is its integration with explainable AI (XAI), which allows for a transparent interpretation of model decisions. The incorporation of XAI ensures that the predictions made by the machine learning models align with clinical reasoning, making them more trustworthy for healthcare professionals. The study underscores the importance of developing AI models that are not only highly accurate but also interpretable, ensuring that they can be effectively utilized in medical diagnostics.

Problem Identified: CKD is a progressive disease that often goes undiagnosed until it reaches advanced stages, making early intervention challenging. One of the primary reasons for this late diagnosis is the lack of effective early detection tools that can accurately identify CKD before significant kidney damage occurs. Since CKD is associated with high morbidity and can lead to severe complications such as cardiovascular disease and kidney failure, there is an urgent need for improved screening methods. Traditional diagnostic approaches rely on biomarkers such as creatinine levels and glomerular filtration rate (GFR), but these indicators often fail to detect CKD in its early stages.

Work Done: To tackle this issue, the researchers developed machine learning models based on ensemble tree-based methods, which are known for their robustness and high predictive accuracy. These models were trained on a dataset comprising blood and urine test results from both CKD patients and healthy individuals. The primary objective was to build models capable of accurately predicting CKD in previously unseen cases, making them suitable for early screening and risk assessment. By leveraging ensemble learning techniques, the researchers aimed to combine the strengths of multiple models, enhancing overall predictive performance. This approach not only improved the accuracy of CKD detection but also allowed for a more comprehensive understanding of the disease's underlying patterns.

Knowledge Gained: The study's findings revealed interesting insights into the effectiveness of different machine learning models for CKD prognosis. Notably, the Random Forest model demonstrated a stronger ability to identify significant features related to CKD, highlighting more closely aligned with clinical insights. This suggests that while Random Forest provided broader feature importance, XGBoost's structured learning approach made it more effective in mimicking real-world clinical decision-making. These insights emphasize the importance of selecting the right AI model based on both accuracy and interpretability in medical applications.

Gaps Identified: Despite the promising results, the study identified certain gaps that need to be addressed for a more comprehensive CKD management strategy. One of the major concerns is the lack of rigorous methodologies to develop effective preventive measures for cardiovascular complications in CKD patients. Since CKD significantly increases the risk of cardiovascular diseases.

Title : Machine Learning System for Predicting CKD in CVD Patients

Author : Nantika

Year : 2024

Published : arXiv

Concept Discussed: This research presents an explainable machine learning system designed to predict Chronic Kidney Disease (CKD) in patients with high cardiovascular risk. By incorporating both medical history and laboratory data, the study aims to improve early detection and intervention for CKD in this vulnerable population. Unlike traditional black- box AI models, this approach emphasizes explainability, ensuring that healthcare professionals can understand and trust the model's decision-making process.

Problem Identified: One of the major challenges in CKD diagnosis is that the disease often remains asymptomatic until it reaches advanced stages. This issue is even more critical for patients with high cardiovascular risk, as they are already predisposed to complications that can mask early CKD symptoms. The lack of effective early detection methods in this group leads to delayed diagnosis and treatment, increasing the likelihood of severe kidney damage and cardiovascular events.

Work Done: To address this challenge, the researchers developed a machine learning model based on the Random Forest algorithm, known for its strong performance in medical prediction tasks. The model was trained using a dataset that included both medical history and laboratory test results, allowing it to learn patterns associated with CKD progression.Additionally, the study introduced a robust explainability framework that provided global and local model interpretations, bias inspection, biomedical relevance validation, and safety assessments. These measures ensured that the model's predictions were not only accurate but also interpretable.

Knowledge Gained: One of the key findings of the study was the identification of critical predictive features for CKD in high-risk cardiovascular patients. These included the use of diabetic medications, ACEI/ARB medications (commonly prescribed for blood pressure and kidney protection), and initial eGFR values. The fact that these features strongly influenced model predictions aligned well with existing medical literature, reinforcing the model's reliability. Furthermore, the study demonstrated that explainable AI techniques could help validate the model's logic against established clinical knowledge, making it easier for healthcare providers to adopt AI-driven insights into their decision-making processes.

Gaps Identified: Despite its promising results, the study highlighted certain biases in the model, particularly concerning initial eGFR values and their influence on CKD predictions. This raises concerns about potential over-reliance on a single biomarker, which may introduce bias in cases where eGFR values fluctuate due to temporary conditions rather than chronic kidney disease. The research suggests that additional refinements are necessary to ensure that the model provides unbiased and equitable predictions across different patient groups.

Title	: Responsible Medical Diagnostics Recommendation**Systems**
Authors	: Daniel Schlör, Andreas Hotho
Year	: 2022
Published	: arXiv

Concept Discussed: This paper explores the development of responsible medical diagnostics recommendation systems, emphasizing critical aspects such as accountability, safety, and fairness in AI-driven decision-making. As artificial intelligence and machine learning increasingly integrate into healthcare, ensuring that

these systems adhere to ethical standards becomes paramount. The authors discuss the principles necessary for designing responsible AI-based recommender systems that can support medical staff without introducing biases, compromising safety, or diminishing accountability in clinical decision-making.

Problem Identified: The digitization of hospital processes has significantly transformed healthcare delivery, but it also presents major challenges for medical professionals. One of the key issues identified in the study is that physicians and medical staff are often required to digitally capture a vast amount of information, which can be time-consuming and detract from direct patient care. This administrative burden can lead to physician burnout, inefficiencies in clinical workflows, and even potential errors due to information overload.

Work Done: To address these concerns, the authors propose a framework for designing responsible medical recommendation systems, emphasizing three key principles: accountability, safety, and fairness. Accountability ensures that AI-driven medical recommendations are traceable and transparent, allowing physicians to understand the rationale behind each suggestion and retain ultimate decision-making authority. Safety focuses on minimizing errors, preventing unsafe recommendations, and incorporating human oversight to prioritize patient well-being.

Knowledge Gained: One of the key insights from the study is the importance of responsible AI design in the development of medical recommender systems. The authors emphasize that AI should serve as a supportive tool rather than an authoritative decision-maker, ensuring that medical professionals retain control over final diagnoses and treatment decisions. The study highlights that explainability and transparency, as physicians need to trust and understand how recommendations are generated.

Gaps Identified: Despite the promising framework proposed in this study, the authors acknowledge several areas requiring further research and practical implementation. One significant gap is the lack of real-world validation for responsible recommender systems in live healthcare settings. While the theoretical framework outlines ethical AI principles, the practical challenges of integrating such systems into hospitals remain largely unexplored.

Title	: Machine Learning Model for Cardiovascular Prediction
Authors	: Zhu, Qiao, Zhao, Wang, Wang, Niu, Shang, Dong, Zhang
Year	: 2024
Published	: PubMed

Concept Discussed: This study focuses on developing machine learning models for predicting cardiovascular disease (CVD) risk in patients with chronic kidney disease (CKD). Given that CKD patients are highly susceptible to cardiovascular complications, an accurate and reliable prediction system can significantly aid in clinical decision-making and improve patient prognosis. Traditional risk assessment methods often fail to capture the intricate relationships between CKD progression and CVD risk factors, leading to suboptimal predictive accuracy. By leveraging machine learning techniques, the researchers aim to create a more robust and data-driven model that can identify high-risk CKD patients early, allowing for timely interventions and improved long-term health outcomes.

Problem Identified: Cardiovascular disease is the leading cause of death among CKD-induced cardiovascular events result from a complex interplay of factors, including fluid overload, endothelial dysfunction, electrolyte imbalances, and medication effects. Existing CVD prediction models are often designed for the general population and may not account for these CKD-specific risk factors.Another major

challenge is the heterogeneity of CKD progression among individuals, making it difficult to apply a one-size-fits-all approach to CVD risk assessment. While traditional methods like the Framingham Risk Score and ASCVD calculators provide a general estimation of cardiovascular risk, they lack the granularity needed for CKD patients who experience rapid biochemical changes and fluctuating disease progression. This study addresses this gap by developing machine learning-driven risk prediction models trained specifically on CKD patients.

Work Done: To build an effective CVD risk prediction system, the researchers utilized electronic medical records (EMRs) from a diverse population of CKD patients. The dataset contained a wealth of clinical variables, ranging from demographic details to laboratory test results, medication history, and comorbid conditions. To refine the predictive model and eliminate redundant or less impactful features, the researchers employed Least Absolute Shrinkage and Selection Operator (LASSO) regression, a statistical method that enhances model efficiency by selecting only the most relevant predictors while reducing overfitting.

Knowledge Gained: The study yielded several crucial insights regarding the most influential factors in predicting CVD risk among CKD patients. The XGBoost model achieved an impressive area under the curve (AUC) of 0.89, indicating strong predictive performance and reliability in clinical applications. This high AUC score suggests that the model can effectively differentiate between high-risk and low-risk CKD patients, making it a valuable.

Gaps Identified: While the study demonstrated strong predictive capabilities, it also uncovered areas that require further refinement to ensure optimal performance and fairness in real-world clinical settings. One of the main concerns was the potential need for additional relevant features that could further enhance predictive accuracy.

Although the model incorporated several important biomarkers, the researchers acknowledged that other risk factors—such as inflammatory markers (e.g., C-reactive protein), genetic predisposition, and lifestyle factors—were not included in the analysis due to data limitations. Integrating these additional variables in future iterations could improve the model's comprehensiveness.

Title	: Analysis and Prediction of Heart Stroke Using LSTM
Authors	: Md Ershadul Haque, Salah Uddin, Md Ariful Islam
Year	: 2022
Published	: arXiv

Concept Discussed: This study explores the use of Long Short-Term Memory (LSTM), a deep learning model, for predicting heart stroke occurrences based on two key biomarkers: Ejection Fraction (EF) and Serum Creatinine Levels. Ejection Fraction refers to the percentage of blood pumped out of the heart with each heartbeat, while serum creatinine levels indicate kidney function, which is indirectly linked to cardiovascular health.

Problem Identified: Traditional heart disease and stroke prediction models, such as Decision Trees, Random Forest, and Logistic Regression, often fall short due to their inability to process sequential data effectively. Many of these conventional models struggle to analyze time-dependent patterns in patient health records, which are crucial for understanding disease progression. Additionally, their limited accuracy makes it difficult to capture complex interactions between key biomarkers like ejection fraction (EF) and serum creatinine levels, leading to a higher rate of false positives or negatives.

Work Done: To overcome the limitations of traditional models, the researchers developed a Long Short-Term Memory (LSTM)-based deep learning model to predict heart stroke risk by analyzing patient health data. Their methodology began with data collection and preprocessing, where they gathered time-series patient records, focusing on key biomarkers such as Ejection Fraction (EF) and Serum Creatinine Levels. To ensure data quality, they applied cleaning and normalization techniques, removing inconsistencies and standardizing the dataset for training. This preprocessing step was crucial in preparing the data for effective model learning.

Knowledge Gained: The study provided key insights into the effectiveness of deep learning in medical predictions, particularly through LSTM networks. It confirmed that LSTM models are well-suited for capturing temporal dependencies in patient health records, making them ideal for predicting diseases that develop over time. One of the most significant findings was the critical role of key biomarkers, specifically Ejection Fraction and Serum Creatinine Levels, in assessing heart stroke risk. By monitoring changes in these biomarkers over time, healthcare professionals can gain early warning signs of potential cardiovascular issues.

Gaps Identified: The study on heart stroke prediction using LSTM showed promise but had key limitations. It used a small, non-diverse dataset, limiting generalizability. Data from varied demographics and locations is needed. The model only used Ejection Fraction and Serum Creatinine, ignoring other critical factors like blood pressure and diabetes. Including more medical parameters could boost accuracy.

CHAPTER 3
ANALYSIS

3.1 SYSTEM ANALYSIS

3.1.1 Problem Identification

In the evolving landscape of digital healthcare, the need for accessible, intelligent, and automated disease prediction systems has become increasingly critical. Existing methods largely depend on manual diagnosis or expensive medical imaging, making them inaccessible to individuals in rural or resource-constrained areas. Moreover, most platforms that offer disease detection based on symptoms are underdeveloped, lack real-time analysis, and do not provide integrated hospital recommendations based on disease type or location. Users often face fragmented interfaces, poor navigation, and minimal data security, which collectively hinder the adoption and effectiveness of such tools. These issues underscore the pressing need for a unified, user-friendly, and secure system that can accurately predict diseases from user-reported symptoms while recommending appropriate healthcare facilities— bridging the gap between early diagnosis and timely medical intervention.

3.1.2 Existing Systems

Several existing healthcare systems and diagnostic platforms offer symptom-based disease prediction, yet most fall short in terms of accuracy, user experience, and actionable outcomes. Applications like WebMD and ADA leverage predefined symptom checkers and decision trees but often provide generalized results that lack personalization and medical context. Some AI-based tools attempt to predict diseases using machine learning models; however, they are often limited to specific conditions, lack real-time adaptability, and do not integrate hospital guidance or nearby treatment facilities. Furthermore, many systems rely on static databases and

outdated knowledge bases, with minimal support for automation or continuous learning. These limitations highlight the gap in existing solutions—there's a strong need for a more dynamic, intelligent, and integrated approach to symptom-based disease prediction that not only identifies potential illnesses but also guides users to appropriate medical support.

3.1.3 Proposed System

The proposed system is designed to address the limitations of existing diagnostic platforms by leveraging an advanced AI-driven approach to disease. Unlike traditional systems, enhancing the accuracy of diagnoses over time. The system not only predicts potential diseases but also offers personalized recommendations for nearby hospitals, integrating geolocation data to direct users to the most suitable healthcare providers. Furthermore, the system includes a user- friendly interface, ensuring accessibility for patients of all ages and technical backgrounds. This end-to-end solution reduces manual input, provides quicker diagnoses, and improves healthcare delivery by streamlining the process of disease detection and hospital selection.

Advantages:
- Uses smart algorithms for fast, accurate CKD and CVD predictions.
- Simple, user-friendly interface for all age groups.
- Automates detection and hospital suggestions—less typing, fewer errors.
- Instant results and real-time recommendations for quick action.
- Suggests nearby hospitals based on user's location.
- Scalable system, ready to handle more diseases and updates in the future.

3.2 REQUIREMENT ANALYSIS:

3.2.1 Functional Requirements

User Authentication: The system requires secure user registration and login functionality, ensuring only authorized users can access the platform for diagnosis and reports.

Data Input and Preprocessing: The platform accepts medical data inputs, including symptoms, test results, or patient history, and preprocesses the data for accurate analysis.

Disease Prediction and Diagnosis: Based on the input data, the system applies algorithms to predict and diagnose health conditions, providing accurate results.

Hospital Recommendations: After diagnosis, the system automatically generates a list of recommended hospitals, considering the severity of the condition and geographical proximity.

Scalability for Future Updates: The system should allow for easy integration of future health conditions, treatment methods, and updated algorithms to remain up-to- date with medical advancements**.**

3.2.2 Non-Functional Requirements:

Performance: The system should provide quick responses to user inputs and process medical data in a timely manner, ensuring minimal delay in generating diagnostic reports.

Scalability: The platform must be scalable to accommodate increasing numbers of users and large datasets, especially as medical data grows in volume over time.

Security: User data, including sensitive health information, must be securely stored and transmitted using encryption methods. The system must comply with data protection regulations like HIPAA or GDPR.

Usability: The system must be user-friendly, with an intuitive interface suitable for medical professionals, patients, and ensuring ease of use without extensive training.

Reliability: The system should ensure high availability and minimal downtime, with backup mechanisms in place to prevent data loss in case of failure.

Maintainability: The system should be easy to maintain and update, with a modular design for quick bug fixes and feature updates.

3.2.3 Hardware Requirements

- **Processor** : Intel Core i3 or higher.
- **GPU** : NVIDIA GTX 720 or higher.
- **RAM** : 4GB or higher.
- **Storage** : 256GB SSD or higher.
- **Internet Connection** : Stable broadband 10 Mbps or higher.

3.2.4 Software Requirements

- **Frontend** : Streamlit
- **Backend** : Python
- **AI Tools** : TensorFlow, Keras, Scikit-Learn
- **Authentication** : Auth0
- **Hosting** : AWS / Google Cloud
- **Browser** : Opera, Mozilla Firefox, Google Chrome
- **Other tools** : Visual Studio Code

CHAPTER 4

DESIGN

4.1 OVERALL DESIGN

The overall design of the proposed system follows a modular and scalable architecture that facilitates effective prediction of kidney and heart-related diseases through intelligent data processing and user interaction. It begins with a secure authentication mechanism, allowing users to log in and input their medical symptoms and personal data via a user-friendly interface. This information is collected and validated before passing through a preprocessing module, which ensures data consistency and readiness for analysis.
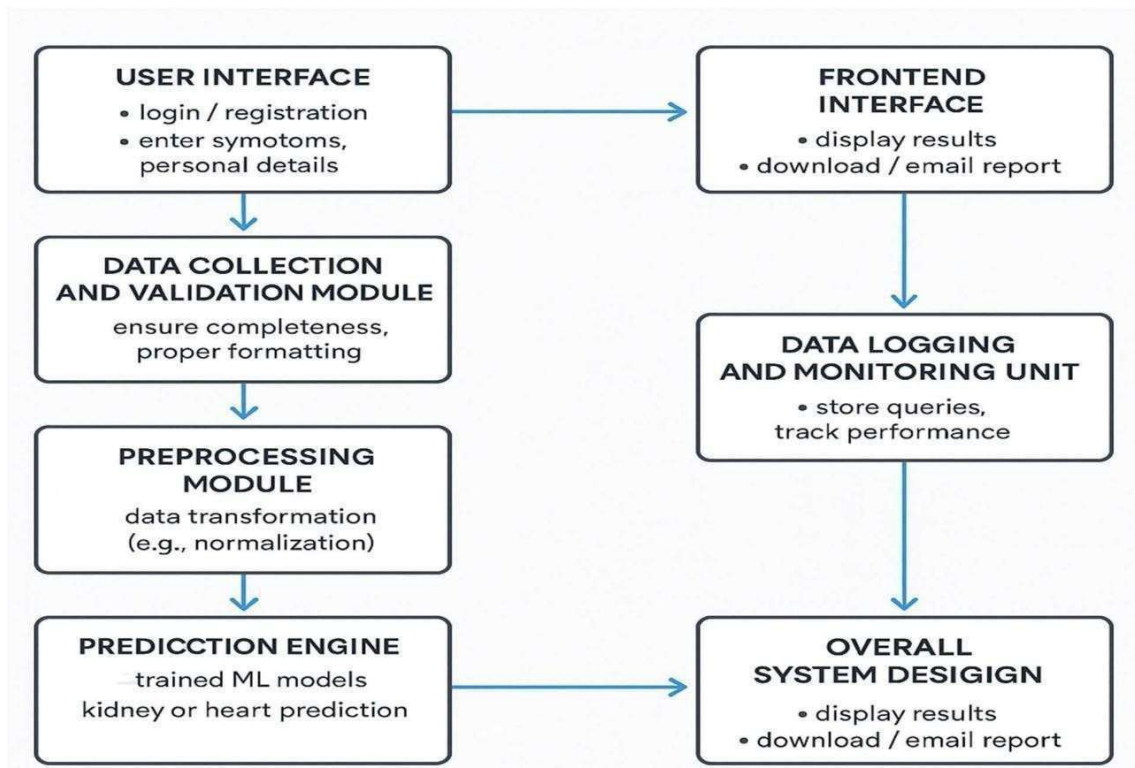
Fig 4.1 Overall Design

4.2 UML Diagrams

4.2.1 Work Flow Diagram

The workflow diagram provides a visual representation of the system's operational flow, detailing how user inputs are processed step-by-step to generate intelligent health predictions. It outlines the interaction between different modules starting from data input and preprocessing to disease prediction and result generation ensuring a clear understanding of the system's logic, functionality, and real-time data movement.
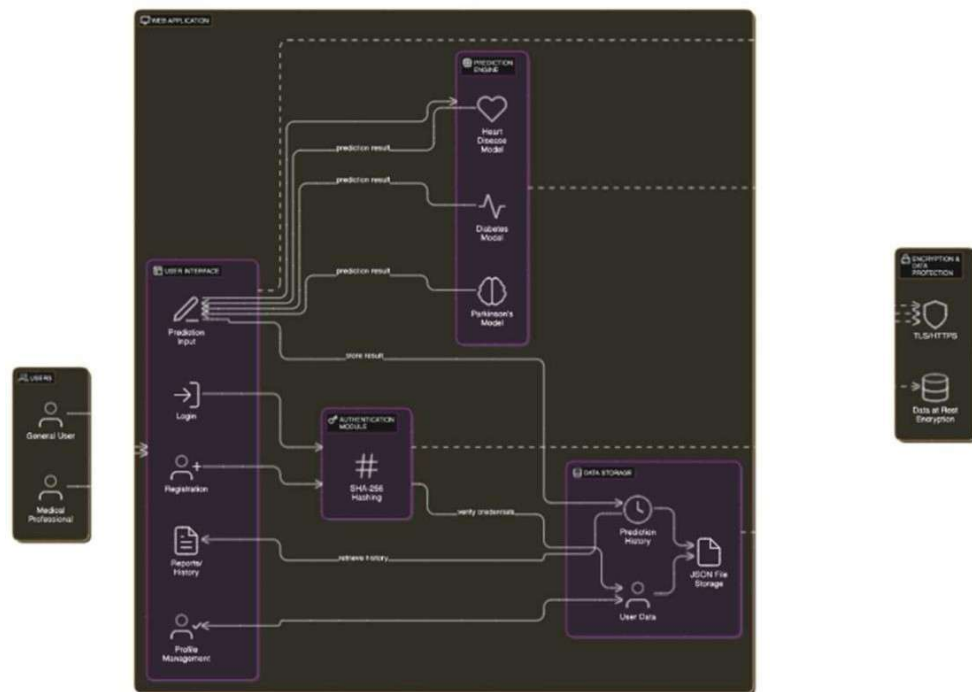

Fig 4.2 Work Flow Diagram

4.2.2 Use Case Diagram

The Use Case Diagram provides a high-level visualization of the interactions between users and the CKD & CVD Prediction System. It highlights the roles of both the end-user and the system administrator, outlining the primary functionalities such as user authentication, medical data submission, disease prediction, and report generation. This diagram helps in understanding the functional boundaries of the system and ensures that all necessary user actions and system responses are clearly defined. It serves as a foundation for identifying system requirements and planning further development activities.
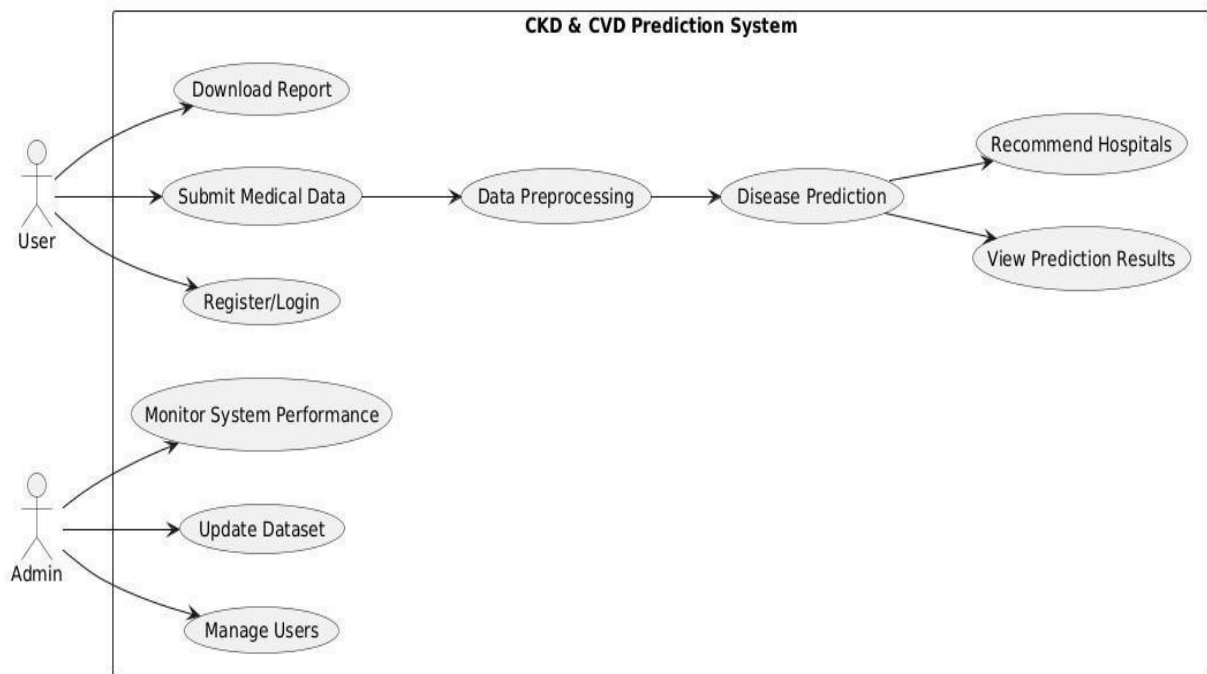

Fig 4.3 Use Case Diagram

4.2.3 Class Diagram

The Class Diagram presents the structural blueprint of the CKD and CVD Prediction System, illustrating the various classes, their attributes, and interrelationships. It encapsulates the core components of the system such as user management, medical data handling, prediction engine, and report generation. By modeling objects and their interactions, the class diagram lays the foundation for system design, ensuring modularity, reusability, and clarity in implementation.
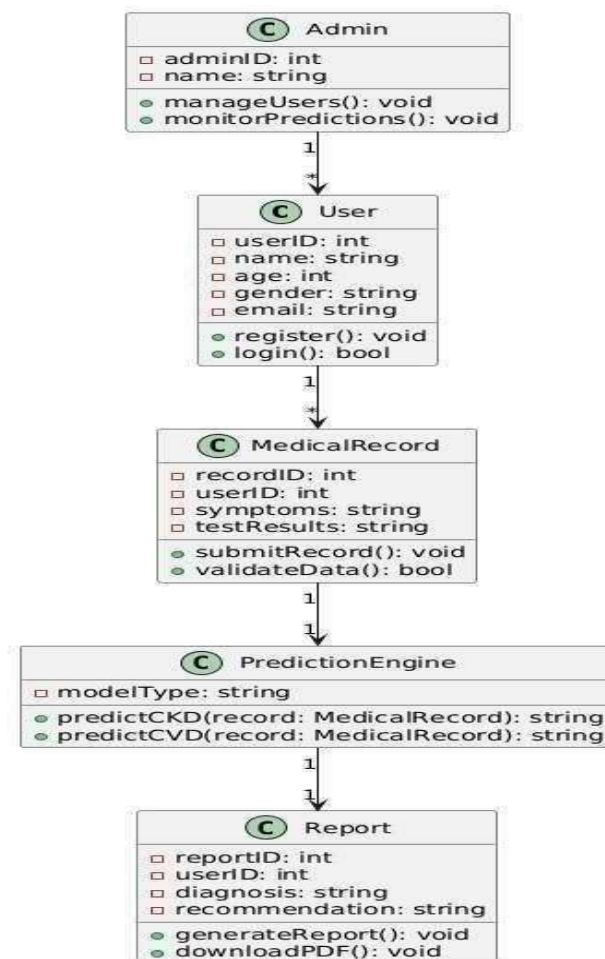

Fig 4.4 Class Diagram

4.2.4 Activity Diagram

The Activity Diagram illustrates the dynamic workflow of the CKD and CVD prediction system, capturing the sequence of actions from the user's interaction to the generation of diagnostic results. It visualizes the system's behavior in terms of flow of control and decision-making processes involved in data input, validation, prediction, and reporting.
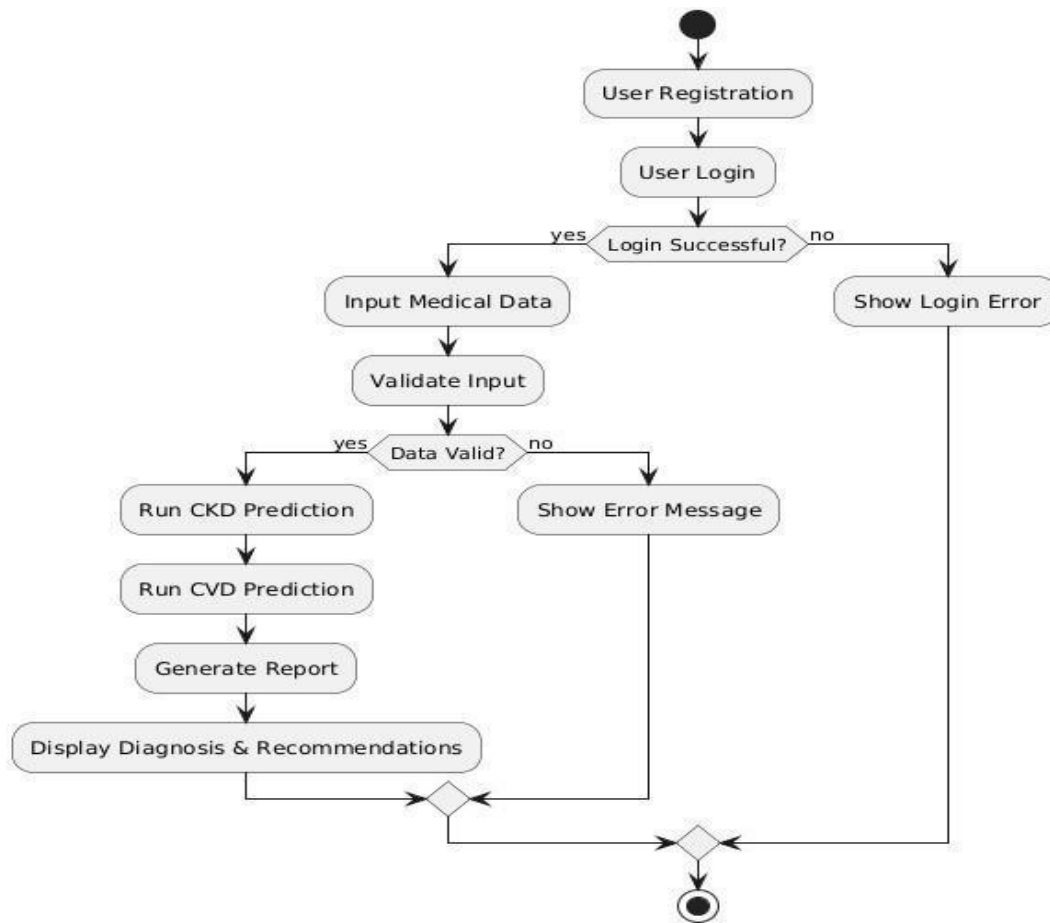

Fig 4.5 Activity Diagram

4.2.5 Sequence Diagram

The Sequence Diagram illustrates the chronological interaction between various components involved in the CKD and CVD prediction system. It captures the flow of messages exchanged between the user interface, the backend server, and the machine learning models during the diagnostic process. Starting from user authentication, the diagram outlines how user data is captured, processed, validated, and forwarded to the predictive models.
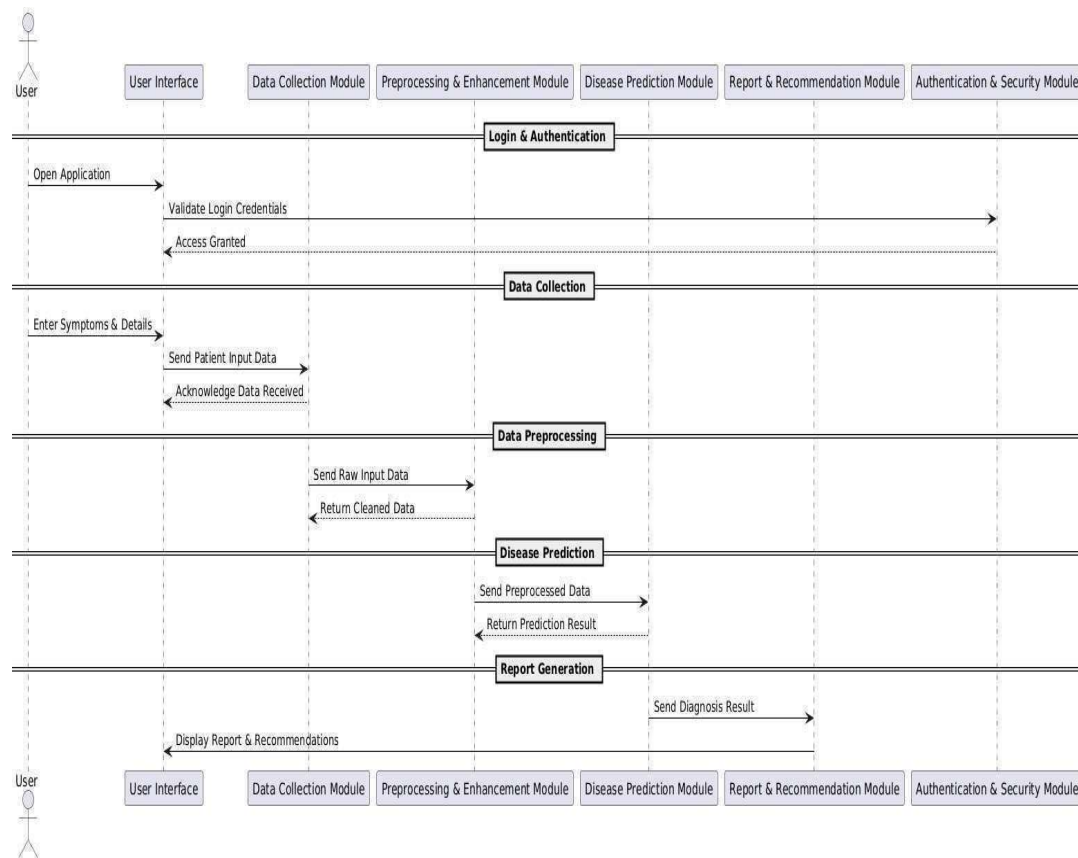

Fig 4.6 Sequence Diagram

CHAPTER 5

IMPLEMENTATION

5.1 MODULES

The system is structured into distinct modules, each designed to fulfill critical function in delivering accurate disease diagnosis, medical report generation, and hospital recommendations for healthcare providers for the recommended patient. The key modules are:

- Data Collection Module

- Preprocessing and Enhancement Module

- Disease Prediction Module

- Report Generation and Hospital Recommendation Module

- User Authentication and Security Module

5.2 MODULE DESCRIPTION

5.2.1 Data Collection Module:

This module acts as the foundation of the entire system, responsible for gathering patient inputs such as symptoms, previous health records, and demographic information. The collected data serves as the primary input for the diagnosis process. It is designed with user-friendly interfaces that allow seamless interaction for both patients and healthcare providers.

5.2.2 Preprocessing and Enhancement Module:

Before feeding the collected data into the predictive model, this module takes charge of cleaning, organizing, and standardizing it. It eliminates inconsistencies and ensures that the data is in an optimal format. Techniques like normalization, feature extraction, and data validation are employed to enhance data quality and reliability, ultimately boosting model accuracy.

5.2.3 Disease Prediction Module:

This is the brain of the system. Using machine learning algorithms, the module analyzes preprocessed data to predict potential diseases—specifically focusing on conditions related to the heart and kidney.

5.2.4 Report Generation and Recommendation Module:

Once the prediction is complete, this module generates a detailed medical report summarizing the diagnosed condition(s), severity level, and recommended next steps. It also suggests nearby hospitals or specialists based on the patient's location, ensuring quick and guided access to healthcare.

5.2.5 User Authentication and Security Module:

Security is a top priority in medical systems. This module ensures that only authorized users can access and interact with the platform. It manages login, user verification, and encryption protocols to protect sensitive patient data, complying with healthcare data standards like HIPAA and ensuring complete data confidentiality.

CHAPTER 6
TESTING

6.1 TESTING AND VALIDATION

The system underwent rigorous testing to ensure accuracy, stability, and performance. Unit testing was performed on all individual modules like input validation, prediction logic, and report generation. Integration testing validated the data flow between components such as the user interface, backend server, and ML models. Functional testing confirmed that the system meets all specified requirements, including disease prediction and hospital recommendation. Validation was done using real-world datasets with verified CKD and CVD cases. Confusion matrices, accuracy, precision, and recall metrics were calculated for performance evaluation. The model achieved high accuracy in both training and testing phases. Stress testing ensured the app's stability under high data loads. Error handling was tested to check system response to invalid inputs. Finally, user acceptance testing was carried out to confirm usability and reliability in real-life scenarios.

6.1.1 Unit Testing

Unit testing focused on validating the performance of individual components in isolation, ensuring that each function performs its intended operation accurately and efficiently.

Authentication Module Testing: Verified the login system for secure access by healthcare professionals, testing both valid and invalid credentials to ensure robust authentication and user role validation.

Patient Data Input and Validation: Tested the input forms by providing various types of patient details (age, blood pressure, glucose, cholesterol levels, etc.) to ensure proper validation, error handling, and format checking.

Data Preprocessing Module: Ensured the preprocessing component handled missing values, outliers, and normalization of medical data correctly, prepping the dataset.

CKD & CVD Prediction Model: Evaluated the prediction module using test cases and known medical records to verify classification accuracy and consistency.

Hospital Recommendation System: Verified that based on prediction outcomes and patient location, the module suggested relevant nearby hospitals accurately, pulling data from the integrated hospital database.

Report Generation & Export: Confirmed that the diagnostic report was correctly compiled, including patient info, prediction results, and hospital suggestions.

6.1.2 Functional Testing

Functional testing was carried out to validate the compliance of the system's core functionalities with the specified requirements. The testing process ensured that each module performed as expected under realistic user scenarios. The following functionalities were tested:

User Authentication and Access Control: The login and registration modules were tested for proper credential validation and access management. The system correctly identified authorized users and restricted unauthorized access to patient data and prediction results.

Patient Data Entry and Management: Input forms were verified for accuracy in capturing all essential medical parameters, including age, blood pressure, glucose, and cholesterol levels. The system responded appropriately to both valid and invalid inputs, offering real-time feedback for missing or incorrect data.

Disease Prediction Engine: The CKD and CVD prediction modules were tested using validated datasets. The system successfully processed patient data through the trained models and accurately predicted the likelihood of chronic kidney or cardiovascular disease.

Hospital Recommendation Module: Upon positive prediction results, the system efficiently recommended nearby hospitals based on patient location and availability. The recommendations were relevant, geographically accurate, and derived from the integrated hospital database.

Report Generation and Export Functionality: Functional testing confirmed that the generated reports contained all required details patient data, prediction results, and recommended hospitals in a structured and readable format.Each functionality was tested in isolation as well as in integration to ensure seamless operation throughout the system. All test cases passed successfully.

6.1.3 System Testing

System testing was conducted to evaluate the complete and integrated application as a whole. This phase ensured that all components of the system work together seamlessly to deliver accurate predictions, efficient hospital recommendations, and a user-friendly experience. The following aspects were validated:

End-to-End Workflow Testing: Verified that from login to report generation, the entire system functions cohesively. Users were able to log in, enter patient data, receive predictions, and view recommendations in a smooth sequence without errors or dead ends.

Integration of Modules: Tested integration between all major modules — user interface, data preprocessing, prediction models, and hospital recommendation engine. Ensured data flows correctly across modules with no information loss or miscommunication.

Cross-Browser Compatibility: Validated the system's web interface on different browsers such as Chrome, Firefox, and Edge to ensure consistent appearance and functionality across platforms.

Data Handling and Storage: Assessed how securely and efficiently the system stores and retrieves user and prediction data from the backend. Checked for data integrity during CRUD (Create, Read, Update, Delete) operations.

System Robustness and Stability: Simulated multiple users entering data and requesting predictions simultaneously to ensure the system remains stable under load and handles concurrency gracefully.

Error Handling and Alerts: Ensured that the system gracefully handles invalid inputs, server errors, and missing data. Clear and user-friendly error messages were displayed where necessary, improving overall usability.

6.1.4 Integration Testing

Integration testing was carried out to verify that the individual modules of the system interact and function together correctly as a unified whole. The goal was to detect interface defects, data flow issues, and communication mismatches between interconnected components.

Module Connectivity Validation: Tested the seamless integration between patient data input, disease prediction models, and hospital recommendation modules.

Prediction Model Communication: Verified that the system correctly passes preprocessed input data to both the CKD and CVD machine learning models and retrieves their respective outputs without latency or misinterpretation.

Hospital Recommendation Logic: Checked the interaction between disease prediction output and hospital mapping logic to ensure that appropriate hospitals are suggested based on disease type, location, and treatment availability.

Frontend-Backend Synchronization: Confirmed the proper exchange of data between the user interface and the backend services. User actions such as form submissions and result views reflected accurately in the database and system responses.

Database Integration: Validated correct storage and retrieval of patient records, prediction results, and recommendation history from the database. Ensured queries execute efficiently and return expected data.

Error Propagation and Handling: Tested how the system manages failed integrations (e.g., API timeouts, incorrect model outputs, invalid data) and verified that appropriate fallbacks or error messages are triggered.

6.1.5 Acceptance Testing

Acceptance testing was performed to evaluate whether the developed system meets the functional and non-functional requirements agreed upon by stakeholders. The testing involved end-users, including healthcare students, domain experts, and faculty members, to ensure real-world usability and relevance.

Requirement Validation: Verified that the system delivers core functionalities such as patient data entry, CKD and CVD prediction, and hospital recommendations as specified during requirement gathering.

End-User Feedback: The system was tested by users simulating real-life scenarios of patient diagnosis. Feedback confirmed that the system was intuitive, easy to navigate, and effectively provided accurate disease predictions with relevant hospital suggestion

Functional Coverage: Ensured that all use cases and functional requirements—from login to prediction output—were executed successfully without any bugs or logic errors.

User Interface Evaluation: Tested the UI for accessibility, responsiveness, and clarity. Users found the layout clean and the workflow straightforward.

Performance Review: Evaluated the system's responsiveness under typical usage conditions. The model predicted results within acceptable time limits, and no significant lag was experienced during the process.

Overall Satisfaction: Stakeholders approved the system's design, accuracy, and potential for real-time usage in educational and clinical environments, giving the green light for project acceptance.

6.2 Build the Test Plan

Testing for the CKD and CVD Prediction System with Hospital Recommendation was structured across multiple levels, including unit, functional, integration, and system- level tests. Each module was validated independently before proceeding to combined workflows. The test plan was divided into the following focus areas:

User Authentication and Access Control: This module was tested to ensure secure access is provided only to authorized users (e.g., admin or medical staff). Test cases included login with valid credentials, rejection of incorrect entries, and correct session management during login and logout processes.

Patient Data Entry and Management: This section was validated to confirm accurate entry and storage of patient records. Testing covered all input fields like age, blood pressure, glucose level, cholesterol, etc., verifying validation rules and appropriate error messages for missing or invalid data.

Disease Prediction (CKD & CVD): The prediction module was tested with both synthetic and real-world datasets. The goal was to confirm that the trained machine learning models produced accurate disease predictions based on patient inputs.

Table 6.1 Test Case Design

TEST ID	MODULE	TEST CASE	DESCRIPTION	EXPECTED OUTPUT	ACTUAL OUTPUT
TC_01	User Authentication	Valid Login	Test the system with valid admin credentials to ensure login success.	User should be successfully logged in.	As expected
TC_02	User Authentication	Invalid Login	Test login with invalid credentials.	System should reject login attempt and show an error message.	As expected
TC_03	Patient Data Entry	Missing Patient Age	Test patient data entry with missing age field.	System should display an error message indicating that age is required.	As expected
TC_04	Patient Data Entry	Invalid Blood Pressure	Test entry of blood pressure with an invalid format.	System should reject invalid input and ask for correct format.	As expected
TC_05	CKD Prediction	Risk CKD	Test system with input suggesting risk for CKD.	System should predict CKD risk as "High".	As expected
TC_06	CVD Prediction	Low Risk CKD	Test system with input parameters for low risk for CVD.	System should predict CVD risk as "Low".	As expected
TC_07	Hospital Recommendation	CKD Hospital Suggestion	Test system to recommend CKD hospitals based on risk.	System should suggest CKD Hospitals.	As expected
TC_08	Hospital Recommendation	CVD Hospital Suggestion	Test system to recommend CVD hospitals based on risk.	System should suggest CVD Hospitals.	As expected
TC_09	Result Visualization	Prediction Result Display	Test display of CKD/CVD prediction results along with hospital suggestions.	Results shown clearly on the UI with prediction and recommendation.	As expected
TC_10	Appointment Booking	Appointment Booking	Test booking an appointment at a recommended hospital.	Appointment booked with correct hospital, date, and time.	As expected

Table 6.2 Test Case Log

S.No	Test ID	Test Description	Test Status (Pass/Fail)
1	TC_01	Check login with valid credentials	Pass
2	TC_02	Attempt login with incorrect password	Pass
3	TC_03	Test patient data entry with missing age field	Pass
4	TC_04	Test entry of blood pressure with an invalid format	Pass
5	TC_05	Test system with input suggesting risk for CKD	Pass
6	TC_06	Test system with input parameters for low risk for CVD	Pass
7	TC_07	Test system to recommend CKD hospitals based on risk	Pass
8	TC_08	Test system to recommend CVD hospitals based on risk	Pass
9	TC_09	Test display of CKD/CVD prediction results along with hospital suggestions	Pass
10	TC_10	Test booking an appointment at a recommended hospital	Pass

CHAPTER 7
RESULTS AND DISCUSSIONS

7.1 RESULT

The AI-Driven Disease Prediction and Hospital Recommendation System successfully fulfills its core objective of accurately predicting Chronic Kidney Disease (CKD) and Cardiovascular Disease (CVD), while also suggesting nearby hospitals for follow-up care. Testing and validation revealed the following insights:

Functional Result: The system effectively met all defined functional requirements. The admin/user login system authenticated access securely, and the intuitive dashboard provided seamless navigation. Patient data was entered and managed efficiently, with all required medical attributes (like blood pressure, diabetes history, cholesterol, etc.) being captured accurately for model processing.

Prediction Module Result: The disease prediction module, powered by trained machine learning models, produced reliable and accurate classifications of CKD and CVD presence. The results were tested across varied patient datasets, and the models consistently delivered appropriate outputs, enabling early diagnosis support for medical professionals and patients alike.

Result Display and Recommendation: Upon successful prediction, the system clearly displayed the disease status (positive/negative) along with personalized recommendations. If a risk was detected, the system suggested nearby hospitals for consultation, aiding in quick action and reducing patient delay in seeking care.

Appointment Booking Result: The appointment booking module allowed users to request appointments at the recommended hospitals. The booking form captured required details, and the system confirmed appointments with proper validation, enhancing user experience and support continuity.

System Integration and User Experience: The integrated workflow from data entry to prediction and hospital recommendation was fluid and responsive. The results were accessible across devices, offering users a convenient and interactive platform.

7.2 DISCUSSION

The development and deployment of this AI-powered web application for disease prediction and hospital recommendation have highlighted the growing role of intelligent automation in the healthcare domain. By integrating advanced machine learning models with an intuitive front-end, the system enables early detection of Chronic Kidney Disease (CKD) and Cardiovascular Disease (CVD), minimizing the need for manual diagnosis.

Each module ranging from secure login, patient data entry, prediction logic, to hospital recommendation was designed for modularity and scalability, yet worked together cohesively to ensure a smooth and reliable user experience. The disease prediction module, trained on real-world datasets, delivered consistent and accurate classifications, allowing for timely intervention.

The addition of hospital suggestions based on prediction outcomes significantly boosts the system's practicality, bridging the gap between diagnosis and treatment. Appointment booking functionality further extends this utility by allowing users to take immediate action.

While the system performed reliably in testing, a few enhancements such as model retraining for larger datasets, adaptive hospital listings based on real-time availability, or offline-first architecture for low-connectivity zones could enhance its robustness.

CHAPTER 8
USER MANUAL

To set up your development environment, follow these steps:

Step 1: Executing the Application

Start by launching the main application script (app.py) in a Python-supported environment. Ensure all required dependencies are installed using pip, such as: Streamlit (for web application routing) pandas, numpy (for data handling and ML input) scikit-learn (for CKD and CVD model prediction) joblib (for loading trained models) geopy, requests (for hospital recommendations)

Step 2: Login to the Application

After activating the server, open the application link in a web browser. The admin login page has a login form on the homepage. Provide correct credentials to reach the dashboard. Only users with verified authentication are permitted to go further for security and access control purposes.

Step 3: Navigating to the Prediction Module

On the homepage, users can choose either **CKD Prediction** or **CVD Prediction** Each option leads to a specific input form, where users will enter medical values relevant to the selected disease prediction.

Step 4: Filling Patient Details and Inputs:

Each form includes multiple input fields such as: Age, Blood Pressure, Glucose, Hemoglobin (for CKD) and Age, Cholesterol, Smoking status, ECG (for CVD). All fields must be filled accurately to ensure model prediction quality. The form is designed to be intuitive and user-friendly, with tooltips for guidance.

Step 5: Getting Prediction Output

After submitting the form, the backend runs the pre-trained model using scikit-learn's joblib. The application then displays: Prediction Result: Either "CKD Detected" / "CVD Detected" or "No Disease Detected".

Step 6: Hospital Recommendation System

Once a disease is detected, the system automatically activates the Hospital Recommendation Module. The application uses the Geopy API to detect user location (or requests for a city input). It sends a request to a hospital dataset or an external API (OpenStreetMap/Google Maps if integrated) to fetch nearby hospitals. A list of top-rated hospitals is shown with: Hospital Name, Specialization (Nephrology or Cardiology) and Location

Step 7: Appointment Booking System

The system offers an Appointment Booking Module directly from the recommended hospitals list. Users can click "Book Appointment" beside any hospital. A page opens to enter patient name, preferred date/time, and contact number. Upon submission, a confirmation message is shown. Appointments are saved in the backend or sent to the admin/hospital email. This feature streamlines the process of scheduling a consultation, reducing delay in medical attention.

CHAPTER 9

CONCLUSION

The developed system for **CKD and CVD prediction with hospital recommendation and appointment booking** serves as a valuable digital tool for early diagnosis and medical access. By integrating machine learning with web technologies, the platform enables users to assess their health conditions based on input symptoms and instantly receive predictions with a confidence score. The approach ensures rapid disease detection, which is essential in preventing the worsening of chronic conditions like kidney and heart diseases.

Beyond prediction, the system bridges the gap between diagnosis and healthcare access. With features such as location-based **hospital recommendations** and an **appointment booking interface**, users can take immediate action following a positive diagnosis. This real-time connection to medical professionals streamlines the healthcare process and empowers users to make informed choices about their treatment pathways without delays.

In essence, this project showcases the powerful impact of AI in healthcare, especially in low-resource settings or areas with limited access to specialists. The inclusion of report generation also enables record-keeping and efficient sharing with doctors. With further enhancements like EHR integration, chatbot support, or teleconsultation features, this system could evolve into a full-fledged health companion app, potentially saving lives through timely action and accessibility.

CHAPTER 10
FUTURE ENHANCEMENTS

While the current system focuses on **CKD and CVD prediction**, there is substantial scope for expansion to make it a more comprehensive health platform. One of the foremost enhancements will be the **addition of more disease prediction models**, targeting common but critical illnesses such as diabetes, hypertension, liver disorders, and respiratory conditions. This would not only broaden the system's diagnostic capability but also position it as a one-stop health monitoring tool.

To increase accessibility and real-time interaction, the system is planned to be **converted into a cross-platform mobile application**. This mobile version will allow users to perform health checks on the go, book appointments, and view their medical reports anytime, anywhere. Integration of **Electronic Health Records (EHRs)** is also proposed to enable seamless patient history tracking and data sharing with healthcare professionals in a secure and standardized manner.

Additional features like **chatbot support for health FAQs**, **PDF report generation for all diagnoses**, and **teleconsultation modules** are also envisioned. These upgrades would turn the system into a hybrid of smart healthcare assistant and telemedicine platform — ensuring users not only get predictive insights but also immediate expert intervention, all within a few taps. With continuous development and adoption of cutting-edge AI models, this project holds the potential to redefine preventive healthcare access for the digital era. Currently, user data are stored locally or in memory. Integrating a relational database such as MySQL or PostgreSQL, or a NoSQL database like MongoDB, would allow for persistent storage of all records, user logs, and activity histories. This would enhance data integrity, enable easier data retrieval, and support advanced querying

APPENDIX I
BASE PAPER

APPENDIX II

SCREENSHOTS

Login Page:

Login Successful Page:

Heart Prediction:

Kidney Prediction and Hospital Recommendation:

Appointment Booking:

APPENDIX III
PAPER PUBLICATION AND
PATENT DRAFT

GOVERNMENT OF INDIA

INTELLECTUAL PROPERTY INDIA

Docket Number:45181

Date/Time : 05/05/2025
Agent Number:

To,
M.MAHESWARI
M.MAHESWARI, ANAND INSTITUTE OF HIGHER TECHNOLOGY,KALASALINGAM NAGAR IT CORRIDOR,RAJIV
GANDHI SALAI ROAD,KAZHIPATTUR,TAMILNADU,INDIA. PIN:603103 9381745691 maheswarim.cse@aiht.ac.in

Sr No.	CBR No.	Reference Number /Application Type	Application Number	Title/Remarks	Amount Paid
1	26868	ORDINARY APPLICATION	202541043473	AI DRIVEN CKD AND CVD PREDICTION AND HOSPITAL RECOMMENDATIONS	1750
2		E-2/4368/2025-CHE	202541043473	Form2	0
3		E-3/9017/2025-CHE	202541043473	Form3	0
4		E-5/3999/2025-CHE	202541043473	Form5	0
5	26868	E-12/10249/2025-CHE	202541043473	Form9	2750
				Total :	**4500**

Received a sum of Rs. 4500 (Rupees Four Thousand Five Hundred only) through

Payment Mode	Bank Name	Cheque/Draft Number	Cheque/Draft Date	Amount in Rs
Cash	—	—	—	4500

Note: This is electronically generated receipt hence no signature required.

Print CBR

IJARCCE

DORA Google Scholar **doi** Crossref **MENDELEY**

Certificate of Publication

M RICHARD NICHOLES

Student, Department of Computer Science and Engineering, Anand Institute of Higher Technology, Kazhipattur Chennai.

Published a paper entitled

A Survey on AI Driven CKD and CVD Prediction and Hospital Recommendation Systems

Volume 14, Issue 5, May 2025

Editor-in-Chief
IJARCCE

SERTIFIKAAT 医明書 ...... Sertifikat CERTYFIKAT CERTIFICADO 증명서 СЕРТИФИКАТ 证书 POTVRDA SERTIFIKA

IJARCCE

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH IN COMPUTER AND COMMUNICATION ENGINEERING

A monthly Peer-reviewed & Refereed journal

Impact Factor 8.102

Indexed by Google Scholar, Mendeley, Crossref, Scilit, SCIENCEOPEN, SCIENCEGATE, DORA, KOAR

DORA **Google** Scholar doi Crossref MENDELEY

Certificate of Publication

D KAMAL RAJ

Student, Department of Computer Science and Engineering, Anand Institute of Higher Technology, Kazhipattur Chennai.

Published a paper entitled

A Survey on AI Driven CKD and CVD Prediction and Hospital Recommendation Systems

Volume 14, Issue 5, May 2025

DOI: 10.17148/IJARCCE.2025.14527

Certificate# IJARCCE/2025/1

Editor-in-Chief
IJARCCE

SERTIFIKAAT 証明書 Sertifikat CERTYFIKAT CERTIFICADO 증명서 СЕРТИФИКАТ 证书 POTVRDA SERTIFIKA

APPENDIX IV
INTERNSHIP CERTIFICATES

Demystified Soft Tech

CERTIFICATE

OF INTERNSHIP COMPLETION

Presented to:

Richard Nicholes M

Successfully completed a one-month internship in **JAVA** at Demystified Soft Tech, from March 07, 2025, to April 07, 2025, while pursuing a degree in Computer Science Engineering at Anand Institute of Higher Technology, Chennai.

Academic Certificate with industry Mentors
Cert.ID : 24676
Date : 07-04-2025

Mahesh R
DIRECTOR

ds Demystified
Soft Tech

CERTIFICATE
OF INTERNSHIP COMPLETION

Presented to:

Kamal Raj

Successfully completed a one-month internship in **JAVA** at
Demystified Soft Tech, from March 07, 2025, to April 07, 2025, while
pursuing a degree in Computer Science Engineering at Anand
Institute of Higher Technology, Chennai.

Academic Certificate with industry Mentors
Cert.ID : 24685
Date : 07-04-2025

R. Mahesh Reddy

Mahesh R
DIRECTOR

REFERENCE

[1] Huang, C., Xie, D., & Zhang, M. (2019). "A Machine Learning Approach to Predict Chronic Kidney Disease." *Scientific Reports*, 9(1), 1–10. From: A Machine Learning Approach to Predict Chronic Kidney Disease.

[2] Rajkomar, A., Dean, J., & Kohane, I. (2019). "Machine Learning in Medicine." *New England Journal of Medicine*, 380(14), 1347–1358. From: Machine Learning in Medicine.

[3] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." *arXiv preprint arXiv:1810.04805*. From: BERT Model for Sentiment Analysis and NLP Tasks.

[4] Li, H., & Zhao, L. (2020). "AI and Image Recognition in Compliance Inspections." *IEEE Transactions on Industrial Informatics*, 16(8), 5221–5233. From: AI and Image Recognition for EXIF Metadata Processing.

[5] OpenStreetMap Contributors. (2023). "Nominatim API for Geocoding and Reverse Geocoding." Retrieved from https://nominatim.openstreetmap.org/. From: Nominatim API for GPS-to-Address Translation in Event Metadata.