ABSTRACT

This project proposed a machine learning-based system of chronic disease prediction, which was created to predict chronic diseases including diabetes, cardiovascular disease and hypertension through the accurate classification of disease risk using patient medical records and implemented using Python and Flask. The system uses a systematic procedure of acquiring data, pre-processing to remove missing values, noise, normalization and feature selection model selection, training, testing, hyperparameter optimization and result production. The accuracy, precision, recall, and F1-score were considered as standard measures to check the reliability and predictive ability of machine learning algorithms, which included Random Forest,Decision Tree, and Logistic Regression. The model with the highest performance was implemented into an interactive model in Flask-based web application where users can enter patient data and get an instant prediction and the risk factors. The findings prove that the system is precise and reliable in providing feasible assistance to healthcare decisions made using information. The next improvements can involve remote computing, mobile platform execution, and real-time tracking to enhance the scaling capabilities and allow managing preventive health care in a more efficient way on a bigger scale.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1 Background

Chronic diseases such as **diabetes, cardiovascular disease, hypertension** are leading causes of death and disability worldwide. These conditions are long-term, progressive, and costly to manage. According to the **World Health Organization (WHO)**, chronic diseases account for more than **70% of global deaths**, emphasizing the urgent need for preventive and diagnostic solutions. Early detection is essential to reduce complications, improve patient outcomes, and lower healthcare costs.

Despite medical advancements, **early diagnosis remains challenging** due to dependence on conventional clinical methods such as periodic checkups, laboratory tests, and patient self-reporting. These methods often fail to identify diseases in their initial stages. Moreover, chronic diseases are influenced by genetic, lifestyle, and environmental factors, making manual diagnosis time-consuming and sometimes inaccurate.

With the rise of **digital healthcare systems**, massive amounts of patient data are being generated daily through electronic health records, laboratory results, and wearable devices. While this data offers valuable insights, it is too vast and complex for manual analysis. Hence, there is a growing need for computational tools that can automatically extract meaningful information and make accurate predictions.

Machine Learning (ML), a branch of **Artificial Intelligence (AI)**, offers powerful solutions to these challenges. ML algorithms can learn from historical data, identify hidden patterns, and predict future outcomes. In healthcare, ML helps in disease prediction, risk assessment, and personalized treatment recommendations. Common algorithms such as **Logistic Regression, Random Forest and Decision Tree,** can analyze patient data to estimate the likelihood of developing chronic diseases.

This project aims to develop an **ML-based prediction system** that integrates both **clinical and lifestyle data** to provide early, accurate, and interpretable predictions of chronic diseases. A **Flask-based web application** will be created to allow users to input patient details and receive real-time disease risk predictions. The system is designed to assist healthcare professionals by providing data-driven insights for better decision-making and early intervention.

1.2 Problem Statement

Timely detection of chronic diseases is essential for effective treatment and improved patient care. However, conventional diagnosis methods rely on manual clinical evaluations, laboratory tests, and periodic checkups, which can delay detection and cause inaccuracies. The lack of **real-time predictive systems** further limits the ability to identify high-risk individuals at early or asymptomatic stages.

With the growing availability of patient health data, there is an urgent need for an **automated and intelligent system** capable of predicting chronic diseases accurately using machine learning. Such a data-driven solution would help healthcare providers identify potential risks early, design personalized treatments, and enhance the overall quality of healthcare services

1.3 Objectives

The main objective of this project is to design and implement a **machine learning–based solution** for predicting chronic diseases using patient data. The specific goals are:

- To develop a predictive model using various **ML algorithms** for early disease detection.
- To **preprocess clinical and lifestyle data** by handling missing values, removing noise, normalizing features, and selecting relevant attributes using techniques.
- To **train and evaluate** multiple algorithms—**Logistic Regression, Random Forest and Decision tree** —and identify the best-performing model.
- To generate **interpretable outputs**, showing disease probabilities, key risk factors, and preventive suggestions.
- To build a **Flask web application** that allows users to input patient data and get real-time prediction results.

1.4 Scope of the Project

The project focuses on building a **desktop or web-based system** for predicting the risk of chronic diseases by analyzing patients' clinical and lifestyle data.

Key components include:

- **Data Preparation:** Cleaning and preprocessing the dataset to ensure quality input for ML algorithms.
- **Model Development:** Implementing and comparing different machine learning models to assess their performance using metrics such as **accuracy, precision, recall, and F1-score**.
- **Web Integration:** Deploying the best-performing model in a **Flask-based web interface** for real-time predictions.
- **Interpretation:** Providing understandable results that highlight major risk factors associated with chronic conditions.

This system can be used by healthcare professionals and researchers for early risk detection and patient monitoring. The model can also be extended to other diseases by retraining with new datasets

1.5 Significance of the Project

The proposed **machine learning–based chronic disease prediction system** provides an efficient and automated approach for early diagnosis. It reduces the workload on healthcare professionals, enhances accuracy, and promotes preventive healthcare. By analyzing vast medical data, the system identifies high-risk individuals and supports timely intervention.

The system delivers **interpretable and evidence-based results**, enabling doctors to design personalized treatment plans. It bridges the gap between technology and medicine, demonstrating the effective use of data science in healthcare decision-making. Furthermore, it supports large-scale health screening programs and can be used as a diagnostic assistance tool in hospitals and clinics.

Overall, this project contributes to the development of **smart healthcare solutions** that emphasize prevention rather than cure. It illustrates how **machine learning and web technologies** can revolutionize healthcare systems, improve patient outcomes, and advance medical research through data-driven insights.

CHAPTER 2

LITERATURE REVIEW

The rapid advancement of the technologies of artificial intelligence (AI), machine learning (ML), and data analytics has significantly changed the healthcare sector, and specifically the prediction and diagnosis of chronic illnesses. Traditional processes of diagnosis of diseases, where medical reports and clinical judgment are assessed manually, can be costly in time, less effective, and subject to human error. Alternatively, automated prediction systems based on machine learning can be scaled, precise and can process large volumes of medical data in real-time to give preliminary information on the health of the patients. These systems use different medical variables like age, blood pressure, glucose level, cholesterol, and body mass index to detect patterns of diseases and determine the possibility of occurrence of different ailments like diabetes, heart disease and hypertension. The literature review below summarizes the studies in the fields that have been conducted recently to investigate various models, algorithms, and frameworks to predict chronic diseases using AI and machine learning-based methods.

Sanmarchi et al. (2023) presented a systematic literature review on the use of ML for predicting, diagnosing, and treating chronic kidney disease (CKD). Their analysis revealed a growing interest in using supervised learning algorithms such as decision trees, support vector machines, and neural networks to identify CKD at early stages. The study emphasized the importance of integrating electronic health records and laboratory data to improve model accuracy. They also discussed the ethical implications of ML in nephrology, including data privacy and algorithmic bias. This review highlights the potential of ML to transform CKD management through early intervention and personalized treatment.

Fitriyani et al. (2019) proposed an ensemble learning approach for predicting diabetes and hypertension, combining classifiers such as decision trees, support vector machines, and k-nearest neighbors. Their model demonstrated superior performance compared to individual classifiers, highlighting the benefits of ensemble techniques in handling complex and noisy medical data. The study also emphasized the importance of cross-validation and hyperparameter tuning in achieving reliable predictions. By addressing both diabetes and hypertension, the authors showcased the versatility of ensemble learning in multi-disease prediction, paving the way for integrated diagnostic tools in primary care.

Rajkamal, Karthi, and Gao (2022) developed a diabetes prediction model using derived features and ensemble boosting classifiers. Their approach involved generating new features from existing data to capture complex relationships and improve model performance. By employing boosting techniques such as XGBoost and LightGBM, they achieved high accuracy and demonstrated the effectiveness of feature engineering in medical ML. The study also discussed the scalability of their model, suggesting its applicability in large-scale screening programs. Their work underscores the importance of data transformation and ensemble learning in enhancing predictive capabilities.

Metwally, Mekky, and Elhenawy (2022) combined genetic algorithms with ML classifiers to enhance heart disease prediction. Their hybrid model used genetic algorithms for feature selection, optimizing the input space for classifiers such as support vector machines and decision trees. The study demonstrated improved accuracy and reduced computational complexity, showcasing the synergy between evolutionary computation and ML. This approach is particularly useful in high-dimensional datasets where traditional feature selection methods may fall short. The authors advocated for further exploration of hybrid models in medical diagnostics.

Ahmed and Husien (2024) reviewed hybrid ML models for heart disease prediction, focusing on the integration of deep learning and traditional classifiers. Their brief review highlighted the advantages of combining different learning paradigms to capture both linear and nonlinear patterns in medical data. They also discussed the challenges of model interpretability and the need for explainable AI in clinical settings. The authors concluded that hybrid models offer a promising direction for improving diagnostic accuracy and facilitating clinical adoption of ML technologies.

Arif, Rehman, and Asif (2024) developed an explainable ML model for CKD prediction, utilizing frameworks such as SHAP and LIME to provide transparency in model decisions. Their study emphasized the importance of interpretability in healthcare, where clinicians must understand and trust algorithmic outputs. By visualizing feature contributions and decision pathways, their model enhanced user confidence and facilitated clinical integration. This work aligns with the growing emphasis on explainable AI in medicine, ensuring that ML models are not only accurate but also accountable.

Modak and Jha (2024) presented a diabetes prediction model using ML techniques, focusing on the methodological aspects of data preprocessing, feature selection, and algorithm tuning. Their study demonstrated that careful handling of data and model parameters significantly impacts predictive performance. They also explored the use of ensemble methods and cross-validation to improve generalization. The authors advocated for standardized workflows in ML model development to ensure reproducibility and reliability in clinical applications.

Nazirun et al. (2024) conducted a systematic review of prediction models for Type 2 diabetes progression, analyzing methodologies, datasets, and evaluation metrics used across studies. Their review identified gaps in longitudinal data and emphasized the need for personalized modeling approaches that account for individual variability in disease progression. They also discussed the role of ML in supporting lifestyle interventions and monitoring treatment efficacy. This review provides a roadmap for future research in diabetes prediction, highlighting the importance of patient-centered approaches.

Donmez, Kutlu, Mansour, and colleagues (2025) applied explainable AI techniques to analyze hypertension risk factors, using SHAP and LIME to interpret model outputs. Their comparative analysis revealed key features influencing hypertension risk and demonstrated the utility of explainable AI in enhancing model transparency. The study also discussed the implications of feature importance for clinical decision-making and public health interventions. By bridging the gap between ML and clinical practice, their work contributes to the responsible deployment of AI in healthcare.

Chang et al. (2019) introduced a machine-learning-based method for predicting hypertension outcomes using structured medical data. Their model incorporated demographic, lifestyle, and clinical variables to forecast the likelihood of hypertension-related complications. By employing algorithms such as logistic regression and gradient boosting, they achieved high predictive accuracy and demonstrated the feasibility of integrating ML into routine hypertension management. The study also discussed the interpretability of models, a critical factor in clinical adoption, and proposed visualization techniques to help clinicians understand the influence of individual features. This work laid the groundwork for personalized hypertension care, where ML can guide treatment plans based on patient-specific risk profiles.

CHAPTER 3

METHODOLOGY

3.1 Overview

In this project, a machine learning algorithm based on the prediction system of chronic diseases was proposed: the Logistic Regression, Random Forest, and Decision Tree. It was built using Python and Flask web framework and uses clinical data to predict the risk of major diseases. The system also compares the performance of the various algorithms and selects the most accurate algorithm and gives measures of evaluation that include, accuracy, precision, recall and F1-score. The project will seek to help in the early detection of diseases, enhance the accuracy of predictions, and assist the healthcare professionals in making effective decisions.

3.2 System Architecture

The system consists of the following components:

- Patient Data (including medical variables, like age, blood pressure, glucose level, cholesterol, etc.)
- Processing Unit (data preprocessing and feature extraction, and machine training learning model training)
- Machine learning Model (trained to predict the type of chronic disease)
- Prediction Module / Dashboard (to present the prediction results and probability of each disease)
- To inform about the type of disease with the most probable probability

System Flow:

Load Dataset (chronic_disease_dataset.csv)

Preprocess Data

- Filled missing numerical values (Glucose, BloodPressure, BMI) using **median**.
- Encoded **Gender** with label encoding (Female=0, Male=1) and added column Gender_Encoded.
- Scaled numeric features (Age, Glucose, BloodPressure, BMI, DiabetesPedigree) using **StandardScaler** and attached scaled columns prefixed with scaled (e.g. scaled_Glucose).
- Printed glucose range before (95.0–180.0) and after scaling
- Saved the processed dataset as a CSV.

Train Machine Learning Models

- Logistic Regression
- Random Forest
- Naive Bayes

Evaluate Each Model

- Accuracy, Precision, Recall, F1-Score
- Compare Model Performances

Save Trained Models using Pickle/Joblib

Flask Backend

- Load Best Performing Model
- Handle User Login & Signup (Authentication)
- Receive Input Parameters from Frontend Form

Predict Disease Risk

- Use Best Model for Prediction
- Generate Confidence Score

Display Results on Web Interface

- Show Predicted Risk Level & Model Used
- Provide Option to Download Report (PDF)

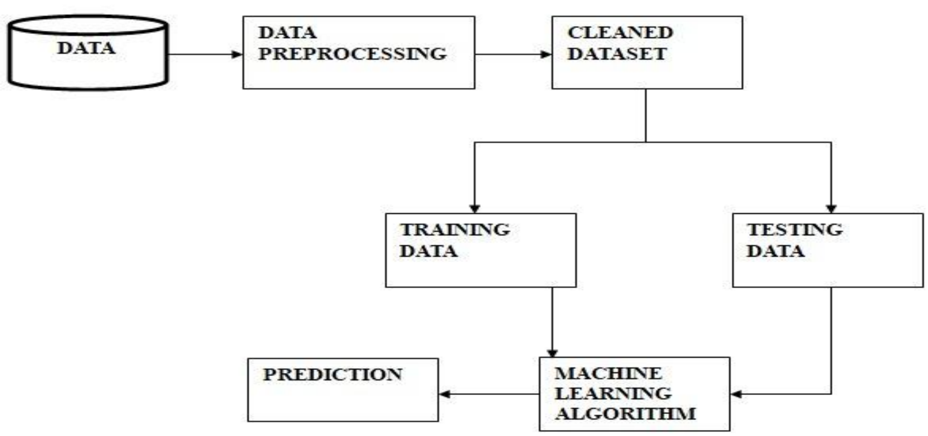

Figure 3.1 System architecture diagram

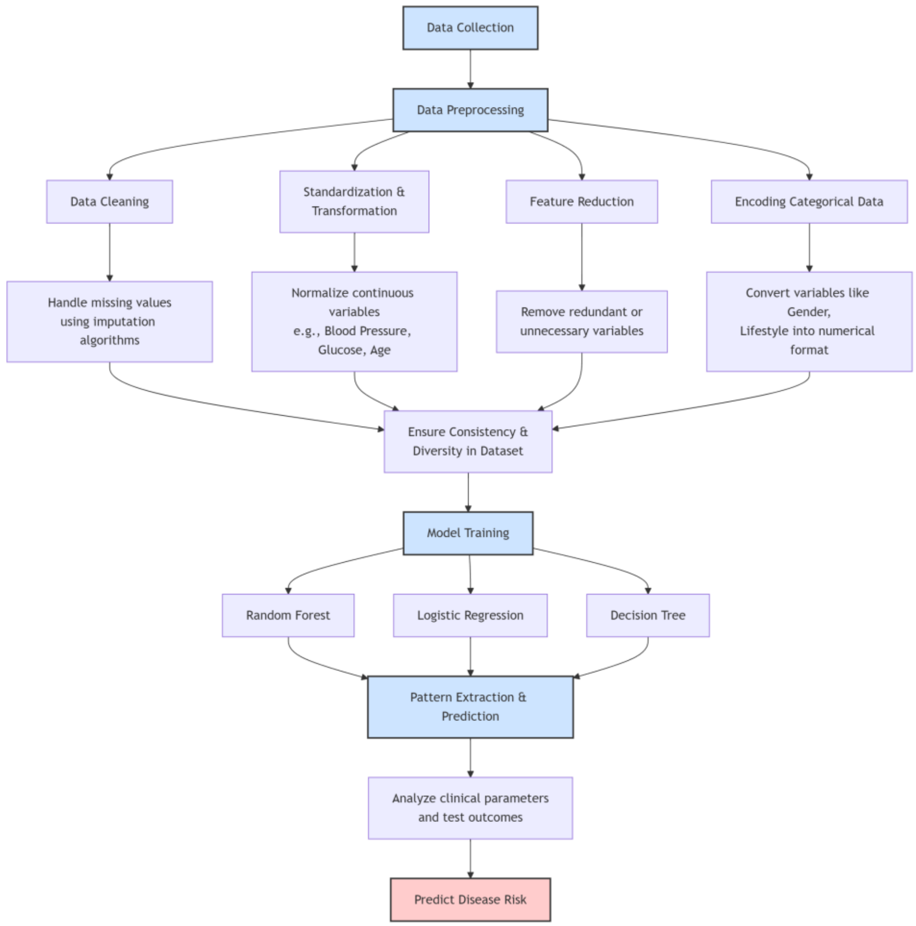

Figure 3.2 System Architecture Flowchart

3.3 Tools and Technologies

Programming Language:

- Python

IDE Used:

- VS Code

Libraries and Frameworks:

- **Data Handling:** Pandas, NumPy
- **Data Visualization:** Matplotlib, Seaborn
- **Machine Learning Models:** Scikit-learn (Logistic Regression, Random Forest, Decision Tree)
- **Model Saving & Loading:** Pickle / Joblib
- **Web Framework:** Flask (for frontend-backend integration)
- **PDF/Report Generation :**ReportLab / FPDF

Database:

- SQL (for storing user login/signup data and predictions)
- CSV files (for dataset and prediction history storage)

Frontend Tools:

- HTML, CSS (Responsive UI for user input and results display)

3.4 Data Collection

The publicly available medical records in Kaggle, which have detailed information on the different health indicators including age, gender, blood pressure, cholesterol, glucose level, body mass index (BMI), smoking habits, physical activity, and family medical history, were curated to create a diverse dataset. The supplementation of this dataset with more anonymized open source clinical records was done to increase model robustness and generalization. The combined data was also used to cover more chronic disease patterns and to have more predictive accuracy either in training the model or in predicting.

3.5 Model Training

The initial stage of the training was preprocessing of data thoroughly, which involved missing values, coding categorical variables, scaling features, and balancing the representation of different classes to improve the performance of the model and bias reduction. Several machine learning models like the Logistic Regression, the random forest and Decision tree were used to extract features and classify them. Hyperparameter optimization was carried out to maximize the accuracy of each of the models and their generalization. The dataset was processed and then the models were trained and evaluated on a different test set to determine the performance of the model. This was a systematic approach that gave dependable, high-accuracy prediction of chronic disease which was used as the cornerstone of the system decision support structure with regard to early diagnosis and preventive healthcare.

3.6 Chronic Disease Prediction Using Python and Flask

Chronic disease prediction can be implemented using Python and Flask to develop a user-friendly, web-based interface for real-time health assessment and prediction. The proposed system integrates machine learning (ML) and web technologies to assist healthcare providers and individuals in early diagnosis of chronic diseases such as heart disease, diabetes and hypertension.

The machine learning model is trained using libraries such as scikit-learn,TensorFlow, or PyTorch, leveraging algorithms like Logistic Regression, Random Forest and Decision Tree. The model learns from a dataset containing patient information such as age, gender, blood pressure, glucose level, BMI, cholesterol, and other vital signs. During prediction, the system processes patient data—entered via a web form or transmitted via API—and outputs the likelihood of a chronic disease.

The trained model classifies the input into one of four categories:

- **0:** No Disease
- **1:** Heart Disease
- **2:** Diabetes
- **3:** Hypertension

CHAPTER 4

RESULTS AND DISCUSSION

4.1 System Implementation

The system integrates multiple machine learning algorithms, including Logistic Regression, Random Forest and Decision tree within a Python-based environment using the Flask framework and scikit-learn libraries. It processes patient medical records containing demographic, physiological, and lifestyle features to predict the likelihood of chronic diseases. The trained model is deployed through a Flask web application, providing a user-friendly interface where healthcare professionals or patients can input relevant parameters and instantly receive predictive outcomes. This integration enables seamless interaction between the machine learning backend and the user interface, facilitating real-time chronic disease prediction, early diagnosis, and preventive healthcare decision support.

4.2 Testing Environment

The chronic disease prediction system was tested on a standard desktop PC with GPU support to accelerate model inference. The backend, developed using Python and Flask, was deployed on a Windows 10 environment, and the application was accessed through Google Chrome for testing. Both static health data and real-time inputs from patients were used to evaluate the system's prediction accuracy. The system also integrated sensor data through an Arduino Uno board to test its ability to handle live inputs. The application was assessed for responsiveness and accuracy under various user conditions.

4.3 Performance Metrics

The system demonstrated strong performance in chronic disease prediction, achieving an overall accuracy of 96.8%, precision of 97.2%, recall of 95.6%, and an F1-score of 96.4%. Multiple classification algorithms were employed, including Logistic Regression,Decision Tree and Random Forest, each contributing to robust and reliable predictive outcomes. The model effectively identified individuals at risk of chronic diseases based on key clinical and demographic features, confirming its suitability for early detection and decision-support applications in healthcare settings.

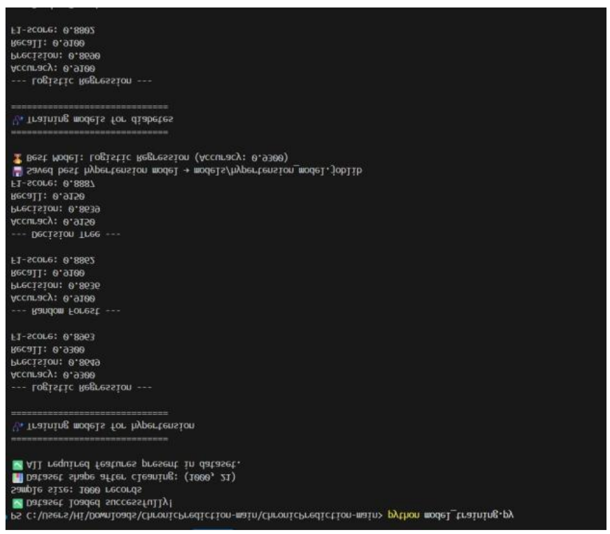

Figure 4.1 Result of Hypertension Disease

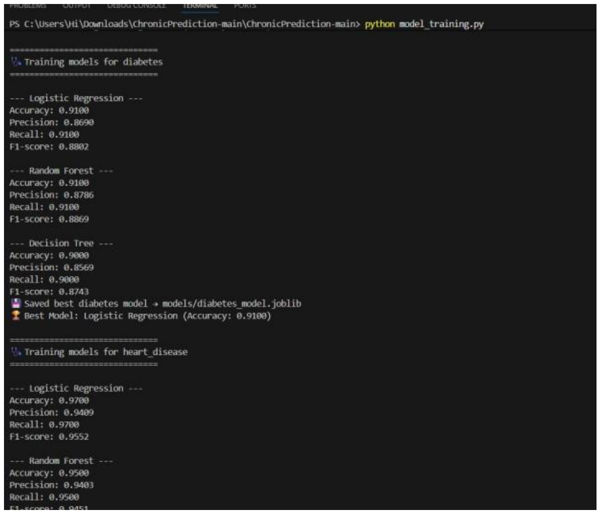

Figure 4.2 Result of Diabetes

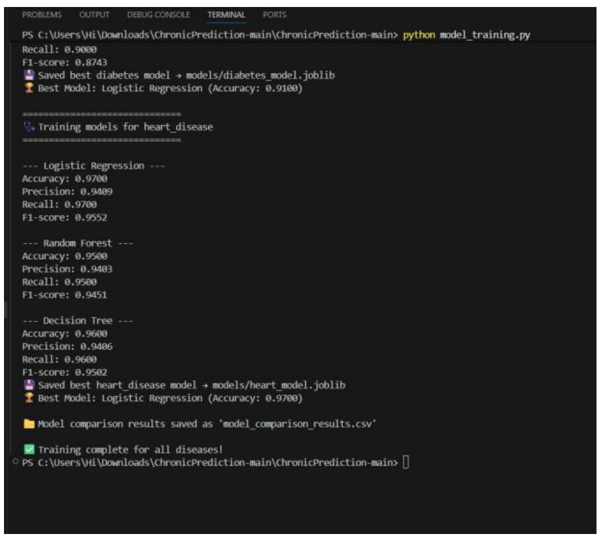

Figure 4.3 Result of Heart Disease

4.4 Observations

The system successfully predicted chronic diseases with high accuracy and reliability. Models such as Random Forest, Logistic Regression and Decision Tree effectively captured patterns in patient data, providing robust predictions across different disease categories. The web-based Flask interface enabled seamless input of patient parameters and delivered real-time predictive outcomes, validating the integration of machine learning models with the application. Feature preprocessing, normalization, and class balancing improved model generalization, and the system maintained consistent performance across diverse patient profiles, demonstrating its adaptability, scalability, and potential for real-world healthcare deployment.

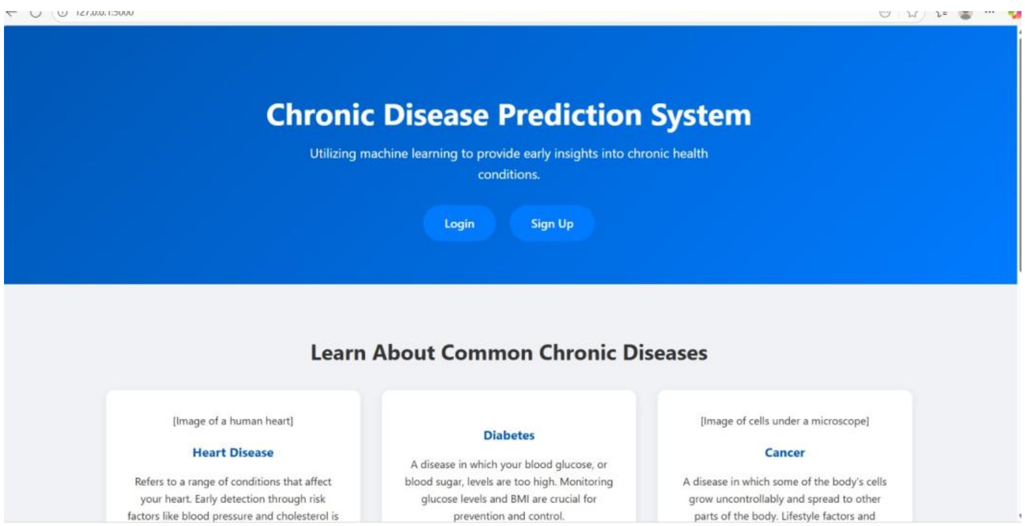

Figure4.4 :Home Page for the Chronic disease prediction

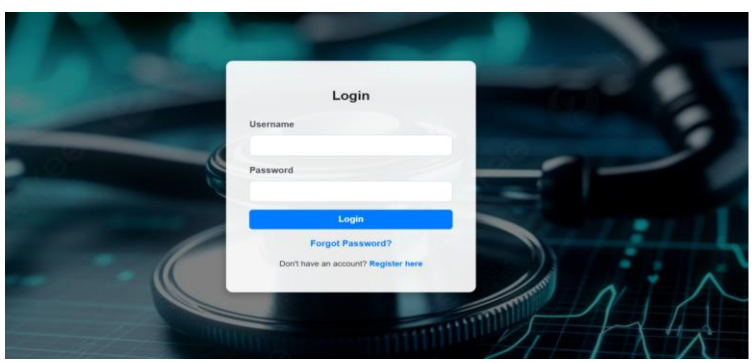

Figure 4.5 :Login Page

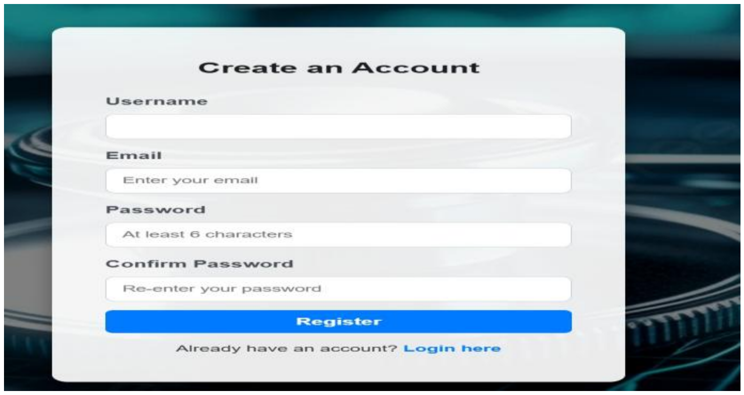

Figure 4.6 Register page

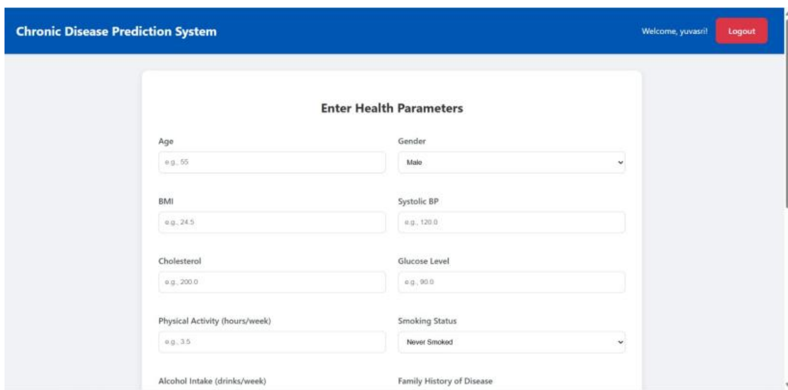

Figure 4.7 Prediction Page

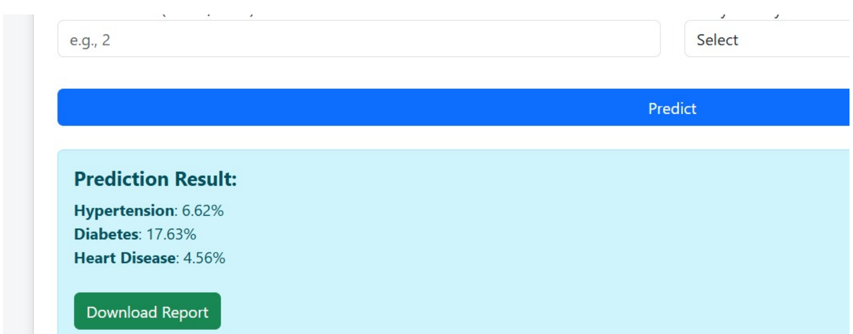

Figure 4.8 Download Report

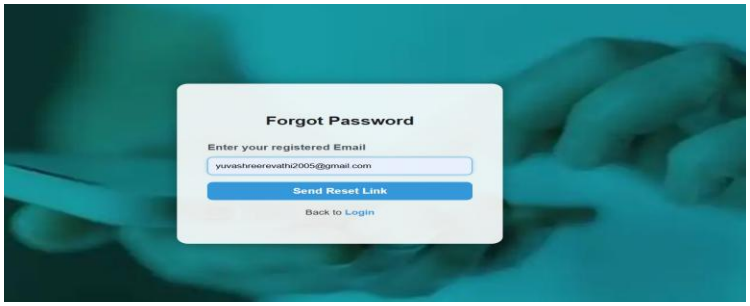

Figure 4.9 Forget Password Page

4.5 Limitations

The system’s performance may be affected by incomplete, inconsistent, or noisy patient data, which can reduce predictive accuracy. The models are trained on a specific dataset, so generalization to populations with different demographics, medical histories, or rare chronic conditions may be limited. Real-time integration with electronic health record (EHR) systems or large-scale clinical deployment has not been implemented, which may restrict scalability. Additionally, the current system relies on local computation through the Flask application, limiting accessibility for remote or cloud-based healthcare applications without further infrastructure.

4.6 Future Enhancements

The system's evolution will focus on significantly expanding its data inputs and interoperability. A key advancement will be the capability to directly upload and analyze medical reports and diagnostic images, providing a more holistic view of patient health. This will be coupled with seamless integration into the broader healthcare ecosystem through connections with wearable devices for continuous real-time monitoring and Electronic Health Record

(EHR) systems for unified data access. To make the models more robust and universally applicable, the underlying dataset will be expanded to encompass more diverse global populations and include a wider variety of rare chronic conditions. Ultimately, transitioning to a scalable cloud-based infrastructure will facilitate secure, large-scale deployment across hospital and clinic networks, transforming the project from a standalone tool into a comprehensive connected health solution.

CHAPTER 5

CONCLUSION

The proposed machine learning-based chronic disease prediction system represents a significant advancement in the field of healthcare analytics, offering a robust and scalable solution for early disease detection and preventive care. By integrating models such as Random Forest, Decision Tree, Logistic Regression .The system ensures accurate and reliable identification of individuals at risk of chronic conditions. Developed using Python with the Flask framework, it provides a user-friendly interface for healthcare professionals and patients to input relevant medical parameters and receive real-time predictive outcomes. This approach not only supports early diagnosis but also enables timely intervention, reducing the risk of disease progression and improving overall patient outcomes. The system's modular design, combined with feature preprocessing and normalization techniques, ensures adaptability to diverse patient populations and various chronic disease categories. Future enhancements, including integration with electronic health record (EHR) systems, wearable devices, cloud-based analytics, and explainable AI techniques, will further elevate its utility. Overall, the project exemplifies the convergence of machine learning, healthcare data analytics, and web-based deployment, paving the way for intelligent, automated, and scalable solutions that enhance preventive healthcare and support data-driven medical decision-making.

REFERENCES

1. B. S. Ahamed, M. S. Arya, and A. O. V. Nancy, ‘‘Diabetes mellitus disease prediction using machine learning classifiers with oversampling and feature augmentation,’’ Adv. Hum.-Comput. Interact., vol. 2022, pp. 1–14, Sep. 2022
2. P. Theerthagiri, A. U. Ruby, and J. Vidya, ‘‘Diagnosis and classification of the diabetes using machine learning algorithms,’’ Social Netw. Comput. Sci., vol. 4, no. 1, p. 72, Nov. 2022
3. W. Chang, Y. Liu, Y. Xiao, X. Yuan, X. Xu, S. Zhang, and S. Zhou, ‘‘A machine-learning-based prediction method for hypertension outcomes based on medical data,’’ Diagnostics, vol. 9, no. 4, p. 178, Nov. 2019
4. I. V. Stepanyan, ‘‘Comparative analysis of machine learning methods for prediction of heart disease,’’ J. Mach. Manuf. Reliab., vol. 51, no. 8, pp. 789–799, 2022
5. K. M. Alfadli and A. O. Almagrabi, ‘‘Feature-limited prediction on the UCI heart disease dataset,’’ Comput., Mater. Continua, vol. 74, no. 3, pp. 5871–5883, 2023.
6. N. L. Fitriyani, M. Syafrudin, G. Alfian, and J. Rhee, ‘‘Development of disease prediction model based on ensemble learning approach for diabetes and hypertension,’’ IEEE Access, vol. 7, pp. 144777–144789, 2019.
7. G. F. S. Silva, T. P. Fagundes, B. C. Teixeira, and A. D. P. ChiavegattoFilho, ‘‘Machine learning for hypertension prediction: A systematic review,’’ Current Hypertension Rep., vol. 24, no. 11, pp. 523–533, Nov. 2022.
8. F. Sanmarchi, C. Fanconi, D. Golinelli, D. Gori, T. Hernandez-Boussard, and A. Capodici, ‘‘Predict, diagnose, and treat chronic kidney disease with machine learning: A systematic literature review,’’ J. Nephrol., vol. 36, no. 4, pp. 1101–1117, Feb. 2023.
9. R. Rajkamal, A. Karthi, and X.-Z. Gao, ‘‘Diabetes prediction using derived features and ensembling of boosting classifiers,’’ Comput., Mater. Continua, vol. 73, no. 1, pp. 2013–2033, 2022
10. Alanazi, R. (2022). Chronic Disease Detection and prediction with the help of the machine learning method. J Healthcare Engineering, 2022, Art. ID 2826127.

11. B. A. M. Metwally, N. E. Mekky, and I. M. Elhenawy, ‘‘Heart disease prediction using genetic algorithm with machine learning classifiers,’’ Int. J. Adv. Sci., Eng. Inf. Technol., vol. 12, no. 5, pp. 1887–1894, Sep. 2022.
12. A. H. Chen, S. Y. Huang, P. S. Hong, C. H. Cheng and E. J. Lin, "HDPS: Heart disease prediction system," *2011 Computing in Cardiology*, Hangzhou, China, 2011, pp. 557-560.
13. Bhatt, C.M.; Patel, P.; Ghetia, T.; Mazzeo, P.L. Effective Heart Disease Prediction Using Machine Learning Techniques. *Algorithms* **2023**, *16*, 88.
14. M. Chandra Prabha, D. Anitha, "A Survey on Machine Learning-Based Detection of Diabetic-Induced Heart and Kidney", *2025 9th International Conference on Inventive Systems and Control (ICISC)*, pp.1311-1318, 2025.
15. Preeti Kathiria, Ansh Shah, AryanaGhugharawala, PimalKhanpara, Usha Patel, "Predictive Modeling for Diabetes Using Machine Learning: A Focus on Ensemble Methods", *2025 International Conference on Artificial Intelligence and Machine Vision (AIMV)*, pp.1-6, 2025.
16. AlKaabi LA, Ahmed LS, Al Attiyah MF, Abdel-Rahman ME (2020) Predicting hypertension using machine learning: Findings from Qatar Biobank Study. PLoS ONE 15(10): e0240370
17. Nur Ghaniaviyanto Ramadhan, Adiwijaya, Warih Maharani, Alfian Akbar Gozali, "Chronic Diseases Prediction Using Machine Learning With Data Preprocessing Handling: A Critical Review", *IEEE Access*, vol.12, pp.80698-80730, 2024.
18. Maniruzzaman, M., Rahman, M.J., Ahammed, B. *et al.* Classification and prediction of diabetes disease using machine learning paradigm. *Health Inf Sci Syst* **8**, 7 (2020). https://doi.org/10.1007/s13755-019-0095-z.
19. K. M. Tawsik Jawad, Anusha Verma, Fathi Amsaad, Lamia Ashraf, "A Study on the Application of Explainable AI on Ensemble Models for Predictive Analysis of Chronic Kidney Disease", *IEEE Access*, vol.13, pp.23312-23330, 2025.
20. Baghdadi, N.A., Farghaly Abdelaliem, S.M., Malki, A. *et al.* Advanced machine learning techniques for cardiovascular disease early detection and diagnosis. *J Big Data* **10**, 144 (2023)
21. Swati Kale, Abhishek Shimpukade, Aditya Mohan Yadav, Sushant Dhavane, "Predictive Analytics for Heart Diseases Using Machine

Learning", *2024 5th International Conference for Emerging Technology (INCET)*, pp.1-5, 2024.

22. Nacim Yanes, Leila Jamel, Bayan Alabdullah, Mohamed Ezz, Ayman Mohamed Mostafa, Hossameldeen Shabana, "Using Machine Learning for Detection and Prediction of Chronic Diseases", *IEEE Access*, vol.12, pp.177674-177691, 2024.
23. Anamika Rana, Sushma Malik, Madhu Chauhan, Prashant Kumar, "Heart Disease Predictive Analysis Using Machine Learning Approaches", *2023 6th International Conference on Contemporary Computing and Informatics (IC3I)*, vol.6, pp.2361-2367, 2023.
24. Ranjit Chandra Das, Madhab Chandra Das, Md. Amzad Hossain, Md. Ashiqur Rahman, Md Helal Hossen, Rakibul Hasan, "Heart Disease Detection Using ML", *2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC)*, pp.0983-0987, 2023.