TRIBHUVAN UNIVERSITY

INSTITUTE OF ENGINEERING

Kathmandu Engineering College

Department of Computer Engineering

MAJOR PROJECT FINAL REPORT

ON

DIABETES PREDICTION SYSTEM USING SVM

[Code no: CT 755]

Submitted By:

Deephang Thegim KAT077BCT021

Dipesh Awasthi KAT077BCT023

Esparsh Tamrakar KAT077BCT024

Pankaj Karki KAT077BCT040

Kathmandu, Nepal

FALGUN 2081

TRIBHUVAN UNIVERSITY

INSTITUTE OF ENGINEERING

Kathmandu Engineering College

Department of Computer Engineering

MAJOR PROJECT FINAL REPORT

ON

DIABETES PREDICTION SYSTEM USING SVM

[Code no: CT 755]

PROJECT REPORT SUBMITTED TO THE DEPARTMENT OF COMPUTER ENGINEERING IN PARTTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE BACHELOR OF ENGINEERING

Submitted By:

Deephang Thegim	KAT077BCT021
Dipesh Awasthi	KAT077BCT023
Esparsh Tamrakar	KAT077BCT024
Pankaj Karki	KAT077BCT040

Kathmandu, Nepal

FALGUN 2081

ACKNOWLEDGEMENT

We would like to express heartfelt gratitude to our teachers for providing us this golden opportunity to get ourselves involved in this Major Project. We are very thankful to the Department of Computer Engineering for helping us grow our programming skills and broadening our knowledge.

We would like to thank our project supervisor **Er. Nabin Neupane** for guiding us and assisting us in our project. We would also like to give special thanks to all faculty teachers, HOD of computer department **Er. Sudeep Shakya**, Deputy HOD **Er. Kunjan Amatya** and major project coordinator **Er. Sarad Chandra Joshi** whose encouragement and guidance have made us perform this project.

There are many people involved in this project to whom we are deeply indebted. Lastly, we would give our sincere thanks to all those from whom we have learnt so much either from informal conversations or meetings.

ABSTRACT

Diabetes is a prevalent metabolic disorder characterized by high blood sugar levels due to ineffective insulin secretion or use. Early detection and effective management are essential to minimize complications and improve quality of life. The System was developed to address these challenges by leveraging machine learning and deep learning techniques for accurate prediction and detection. The diabetes prediction model uses RBF Kernel Support Vector Machine (SVM), trained on a dataset of 100,000 medical records, with key parameters like Age, Gender, Hypertension, Heart Disease, Smoking, BMI, HbA1c Level, and Blood Glucose Level. Data preprocessing, normalization, and feature selection were applied, followed by hyperparameter tuning. For Foot Ulcer Detection, Custom Convolutional Neural Network (CNN) is employed, with preprocessing steps like resizing and augmentation on 1055 annotated diabetic foot ulcer images. The system also detects Diabetes Retinopathy using ResNet50 CNN, trained on 35,126 fundus retinal images for accurate detection. Additionally, the system integrates features such as Medicine Suggestions, Yoga Animations, Blogs, Fitness Tracker, Risk Calculator, Food Nutrition and Exercise Suggestions, Chatbot, and Chat-with-Doctor to provide comprehensive diabetes management and support. The system was developed to improve the quality of care and enhance the patient experience in managing diabetes and its complications.

Keywords: *Diabetes, Insulin, Machine Learning, Deep Learning, SVM, CNN*

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AI Artificial Intelligence

BMI Body Mass Index

CNN Convolutional Neural Network

CVD Cardio Vascular Disease

GLM General Logistic Model

HA1C Hemoglobin A1c

KNN K-Nearest Neighbors

LR Logistic Regression

ML Machine Learning

SVM Support Vector Machine

RBF Radial Basis Function

T2D Type 2 Diabetes

CHAPTER 1: INTRODUCTION

1.1 BACKGROUND THEORY

Diabetes is a chronic metabolic disorder characterized by prolonged high blood sugar levels due to inadequate insulin secretion or ineffective utilization. It is classified into Type 1 diabetes, where the pancreas produces little or no insulin, and Type 2 diabetes, where the body becomes insulin-resistant or does not produce enough insulin. Additionally, gestational diabetes occurs during pregnancy, posing health risks for both mother and child. The global rise in diabetes prevalence has led to severe complications, including cardiovascular diseases, kidney failure, neuropathy, retinopathy, and lower-limb amputations. While genetic predisposition plays a role, modifiable factors such as poor diet, obesity, hypertension, physical inactivity, and stress significantly contribute to disease onset and progression.

With advancements in Artificial Intelligence (AI) and Machine Learning (ML), healthcare is undergoing a transformation by leveraging data-driven, automated solutions for early diagnosis and disease management. Support Vector Machines (SVMs) have demonstrated high accuracy in predicting diabetes by analysing key health parameters such as Age, Gender, Hypertension, Heart Disease, Smoking History, BMI, HbA1c Level, and Blood Glucose Levels. Data preprocessing techniques, including normalization, feature selection, and hyperparameter tuning, improve model efficiency, enabling precise detection and timely medical intervention. Furthermore, AI models continuously learn from real-world patient data, refining predictions and assisting healthcare professionals in risk assessment and treatment planning.

Diabetic complications such as foot ulcers and retinopathy pose significant risks to patients, often leading to severe disabilities if not detected early. Convolutional Neural Networks (CNNs) have emerged as powerful tools in medical imaging, allowing for precise detection of these complications. Diabetic foot ulcers, a major cause of lower-limb amputations, can be identified at early stages using CNN models trained on annotated foot images. Preprocessing techniques such as image augmentation, resizing, and contrast enhancement further enhance model accuracy. Similarly, CNNs have proven highly effective in detecting diabetic retinopathy by analysing retinal fundus images to identify early signs of vascular damage. AI-driven screening improves accessibility and accuracy, particularly in remote healthcare settings, reducing the risk of blindness through timely treatment.

Beyond early detection and diagnosis, AI-powered technologies are transforming diabetes management through digital health solutions. Wearable devices such as Continuous Glucose Monitors (CGMs) are also solution HBA1c testing enable real-time tracking of blood sugar levels and provide instant alerts for abnormal fluctuations but don't diagnose Diabetes directly. AI-driven decision support systems analyse glucose patterns and offer personalized recommendations for diet, exercise, and medication adjustments. Additionally, modern diabetes management platforms integrate AI-enhanced features such as chatbots for instant guidance, fitness trackers for monitoring physical activity, yoga animations for holistic wellness, and virtual coaching for continuous healthcare support. These advancements empower individuals to take a proactive role in managing their condition while reducing the strain on healthcare providers.

The integration of AI, ML, and digital health technologies is revolutionizing diabetes care, making it more efficient, accessible, and personalized. Future innovations, such as AI-driven insulin delivery systems, predictive analytics for diabetes complications, and telemedicine-enabled remote diagnostics, are expected to further enhance patient outcomes. By leveraging these cutting-edge technologies, diabetes management is evolving into a more comprehensive and patient-centric approach, ultimately improving the quality of life for millions worldwide.

1.2 PROBLEM STATEMENT

Diabetes is rapidly emerging as a significant public health threat in Nepal as well as the world, with a growing number of individuals remaining undiagnosed until they develop severe complications. Delayed diagnosis and poor management of diabetes can lead to life-threatening conditions such as heart attack, stroke, kidney failure, diabetic foot ulcers, and diabetic retinopathy. Traditional diagnostic methods often require specialized medical expertise and can be time-consuming, making early detection and timely intervention challenging, especially in regions with limited healthcare access. Additionally, the detection of diabetic foot ulcers and diabetic retinopathy demands specialized knowledge and advanced imaging techniques, which may not be readily available in many healthcare facilities.

There is an urgent need for an automated system that can accurately predict diabetes and detect its complications at an early stage, reducing the risk of severe health outcomes. An AI-driven solution capable of diagnosing diabetes, identifying foot ulcers and retinopathy through image analysis, and offering personalized recommendations can significantly enhance patient care. Moreover, diabetes management involves complex lifestyle modifications, including diet, medication adherence, and physical activity, which can be overwhelming for patients without tailored guidance. Many individuals struggle to interpret and apply medical advice effectively, leading to suboptimal disease control. While lifestyle interventions such as yoga have been proven beneficial in regulating blood sugar levels, they are often overlooked or not integrated into conventional treatment plans.

A holistic approach that combines advanced medical diagnostics with AI-powered ulcer and retinopathy detection, personalized recommendations, and lifestyle management is essential to improving diabetes care. By incorporating early detection, complication prevention, and evidence-based lifestyle interventions, such a system can provide a comprehensive solution to help individuals manage their condition more effectively, ultimately enhancing their quality of life and reducing the burden on healthcare systems.

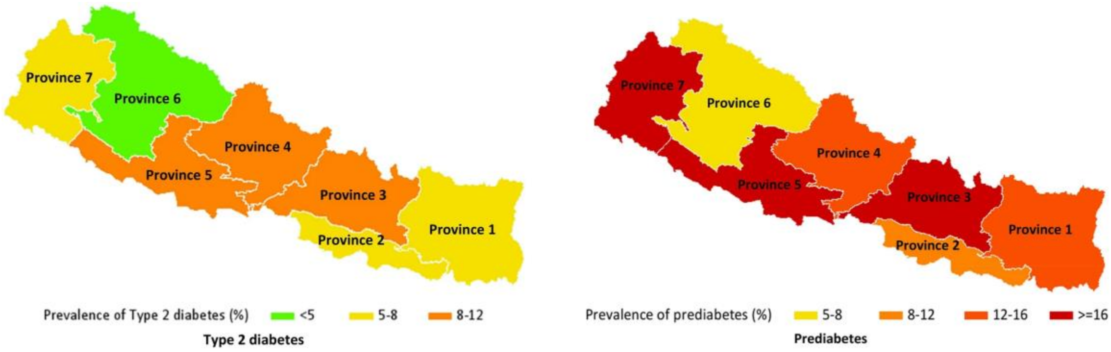


Figure 1: Diabetes in Nepal [1]

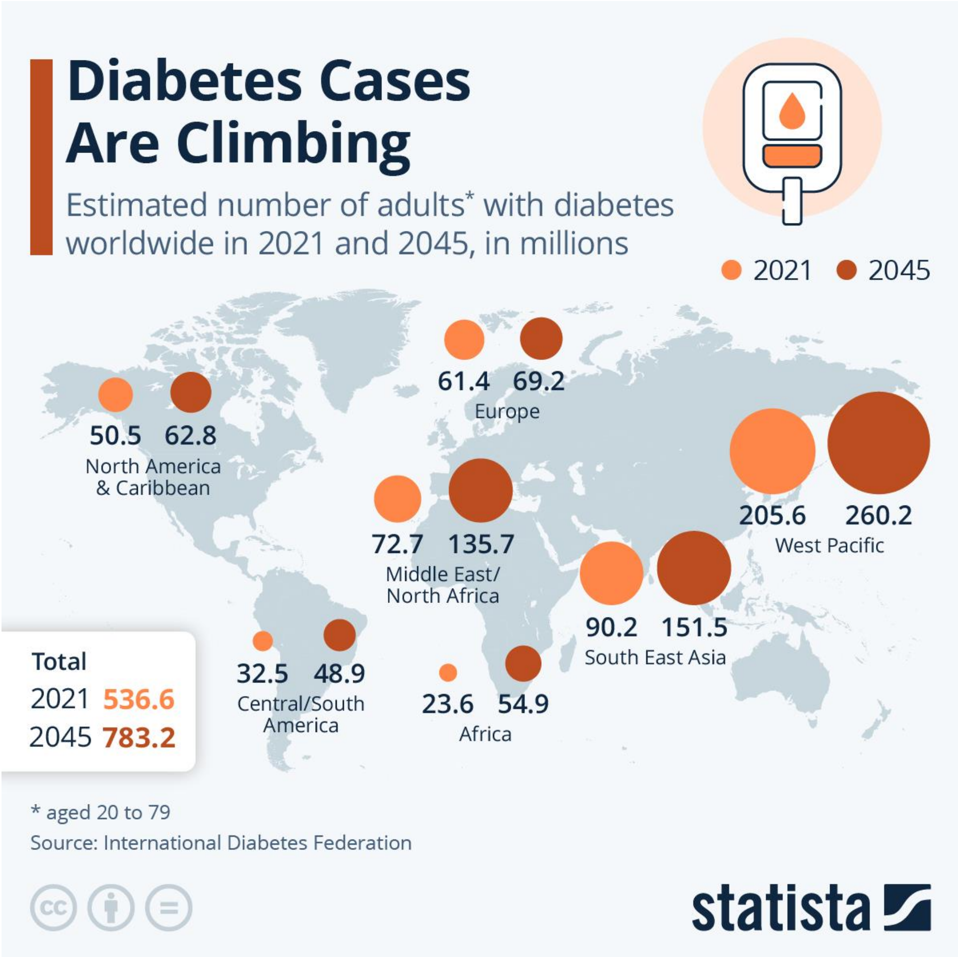


Figure 2: Diabetes in the World [2]

1.3 OBJECTIVES

The objectives of our project are:

- Predict diabetes of users/patients with SVM, detect Diabetic Foot Ulcer and Diabetic Retinopathy with CNN.
- Enhanced User Engagement through integrated features such as Medicine Suggestions, Yoga Animations, Blogs, Fitness Tracker, Risk Calculator, Food Nutrition and Exercise Suggestions, Chatbot, and Chat-with-Doctor.

1.4 SCOPE OF THE PROJECT

There are many individuals who are unaware of being diabetic. With our system, they can easily find out by simply entering some required parameters. In the future, it can also be used in pharmacies or hospitals to predict diabetes in patients. Many people are also unaware of the classification, diet, medications, and complications related to diabetes. This system can help users by providing information about all aspects of diabetes management. Our system is a comprehensive platform designed for early diabetes detection and management, utilizing non-invasive methods and machine learning to analyse clinical parameters for accessible and reliable predictions. In addition to diabetes prediction, it also detects diabetic foot ulcers using image analysis with CNN. It further extends its functionality by detecting Diabetic Retinopathy, utilizing Convolutional Neural Network (CNN) trained on retinal fundus images for early identification of retinal damage caused by diabetes. Beyond diagnosis, our system offers features like blogs, yoga animations, doctor consultations, lifestyle recommendations, and medication guidance, empowering users with knowledge and tools for a healthier lifestyle. The platform serves as a screening tool for healthcare professionals and is scalable for community health programs, especially in underserved areas. Furthermore, our system contributes to research by analysing anonymized data to improve predictive models. Its modular design ensures adaptability for future enhancements, such as advanced algorithms, multilingual support, and additional features, making it a globally applicable solution for better diabetes management and prevention.

1.5 APPLICATIONS

The applications of our project are:

Hospital and Clinical Use: The system can be deployed in hospitals and clinics as a quick diagnostic tool to identify diabetes risk in patients during routine checkups. It assists healthcare providers in making informed decisions and prioritizing care for at-risk individuals.

Health Camps and Community Screening: The model can be used in health camps and rural outreach programs to screen large populations for diabetes. This ensures early detection and intervention, even in underserved areas with limited access to medical facilities.

Personal Health Monitoring: Individuals can use the system to monitor their health by assessing diabetes risk based on their clinical parameters. This empowers users to take proactive steps, such as consulting a doctor or adopting healthier habits, before symptoms worsen.

Integration with Digital Health Platforms: The system can be integrated into existing digital health applications or wearable devices, offering users continuous monitoring and real-time insights about their diabetes risk.

Educational and Preventive Campaigns: Governments and NGOs can leverage the system for educational campaigns about diabetes awareness and prevention. By offering an easy-to-use predictive tool, they can encourage individuals to take control of their health early.

Support for Lifestyle Adjustments: With integrated features like yoga animations, blogs, and diet recommendations, the system supports users in making lifestyle changes that can help prevent or manage diabetes effectively.

CHAPTER 2: LITERATURE REVIEW

Diabetes mellitus is one of the metabolic disorders that results in blood glucose levels that are higher than they should be. Insulin regulates the flow of sugar from the blood into our cells, where it can be stored or utilized as fuel. Because of diabetes, the body is unable to create enough insulin or to utilize it effectively [3]. Type-2 diabetes is the most common type of diabetes and accounts for around 90% of all diabetes occurrences. Family history, age, obesity, the distribution of body fat, race, gestational diabetes, and lifestyle are some of the prevalent risk factors [4]. Increasing dietary fiber consumption lowers the risk of developing diabetes. Additionally, diabetes mellitus can appear as a secondary disorder brought on by a primary illness like pancreatic disease, a hereditary trait like myotonic dystrophy, or medications like glucocorticoids, while pregnancy-related gestational diabetes is a transient disorder.

Artificial intelligence (AI) models are valuable tools that can identify diabetes [5]. AI techniques can be used to deduce and make clinical decisions by using machine learning (ML) to optimize the knowledge base in order to reveal clinically relevant information [6]. Diabetes classification can be done using machine learning. In machine learning, classification uses a specific dataset to model, with the goal of obtaining certain anticipated class labels. The classification requires training a large number of datasets. The seriousness of diabetes has given rise to it becoming a topic for research, with researchers discovering different methods to face the situation. This research aims to develop a machine learning model using Support Vector Machine (SVM) for the efficient classification of diabetes mellitus.

Early prediction and diagnosis of those at high risk of developing diabetes will play a crucial role in lowering the burden of this disease. The creation of simple, non-invasive tools to classify diabetes mellitus based on readily recognizable socio-demographic, lifestyle, and physical characteristics can increase awareness of the presence of such a condition on time.

Several studies have been carried out using machine learning for diagnosis. A study by Kumari and Chitra (2013) on the application of SVM in the classification of diabetes mellitus underlines the importance of ML in clinical decision-support systems. An accuracy of 78% on the dataset used was obtained. The research did not, however, subject the dataset to the feature subset selection process, which impacted negatively on the accuracy, making the system less reliable as the number of false diagnoses was high.

A precise and quick methodology for diagnosing diabetes mellitus was developed [7]. The method used five standard classifications to categorize patients with diagnoses based on nine non-invasive and easily ascertainable clinical features, including race, age, gender, Body Mass Index (BMI), hypertension, histories of cardiovascular disease, family histories of diabetes, physical activity, job stress, and salt stress. The results showed that the J48 decision tree classification, which is a popular decision tree algorithm that creates decision trees by recursively partitioning data based on attribute values, had the highest value in comparison to other classifiers and may be used to screen patients.

In a study by Ekong (2023), an evaluation of the effectiveness of ML algorithms toward the early prediction of cardiovascular diseases (CVD) was carried out. SVM, Random Forest (RF), and K-Nearest Neighbor (KNN) were evaluated, and KNN emerged as the best model in both training and test performances and was recommended for the early diagnosis of CVD. This showed how effective ML algorithms could be used in classification.

Ekong and Udo used an ML-based model for the prediction of fasting blood sugar levels to help in controlling CVD [8]. The General Logistic Model (GLM) was adopted for the prediction of Fasting Blood Sugar levels. Performance analysis results showed effective prediction with an accuracy of 70%. Maniruzzaman developed a machine learning-based system for predicting diabetic conditions using Logistic Regression (LR) to identify the risk factors [9]. Some classifiers were adopted to predict diabetes in patients using three types of partition protocols. It was concluded that some factors constitute risk factors for diabetes and that with an accuracy of 94.25%, the combined LR-based and RF-based classifier was better than using the algorithms individually.

In another study by Ekong, a supervised ML model was used for the effective classification of patients with Covid-19 symptoms based on a Bayesian belief network [10]. The dataset was gathered from experts' review of symptoms that suspected cases exhibited, and it was found to be consistent with reviews of the symptoms that Covid-

19 patients exhibit. The model yielded 98% accuracy, showing a significant classification accuracy of patients with Covid-19.

Further study by Miao involved the development of a machine learning-based model to assess the risk of developing CVD due to type-2 diabetes (T2D) [11]. More so, a machine learning model to predict diabetes was built [10]. They used LightGBM and CatBoost to develop their model based on health habits, and a predictive accuracy of 84.96% was obtained, leading to the deduction that health habits could be used to a good extent to determine the likelihood of someone suffering from type-2 diabetes in the future.

Prasanna Venkatesan researched the traditional classification algorithms and neural network-based ML for the diabetes dataset using different performance metrics [12]. They used K-Nearest Neighbor, Multilayer Perceptron, and other ML algorithms to classify and predict patients who suffer or may likely suffer from diabetes later in life.

Nair and Ruqaiya evaluated different machine learning algorithms for the classification of diabetes mellitus to determine which one best fit that classification need [13]. They concluded that among Logistic Regression, Naive Bayes, SVM, and Decision Trees algorithms, the Naïve Bayes algorithm produced the best accuracy of 79.8%.

Diabetic foot ulcers (DFUs) and diabetic retinopathy (DR) are severe complications of diabetes that require early detection to prevent irreversible damage. CNN-based deep learning models have been extensively used in the medical field to analyse images and detect these complications.

Diabetic retinopathy (DR) is a major cause of vision loss in diabetic patients, caused by damage to the blood vessels in the retina. Convolutional Neural Networks (CNNs) have been widely used for DR detection by analysing retinal fundus images for abnormalities such as microaneurysms, haemorrhages, and exudates. A study by Miao (2022) developed a CNN-based DR detection system, achieving 94% accuracy using a dataset of retinal images. Similarly, deep learning models like ResNet, VGG, and InceptionNet have been used to enhance DR detection accuracy, helping in the automated grading of DR severity levels.

For diabetic foot ulcer (DFU) detection, CNN-based models analyse ulcer images to classify and localize affected areas. A study by Prasanna Venkatesan confirmed that CNN-based ulcer detection significantly outperforms traditional ML classifiers. Researchers have integrated deep learning with clinical parameters, such as glucose levels and patient history, to improve diagnostic performance for both DR and DFU.

All the literature reviewed showed the efficacy of machine learning in disease diagnoses and in the classification of diabetes mellitus, but neither their classification nor predictive accuracy was sufficient to make them sufficiently reliable, a factor that is the most important consideration in clinical decision support systems, or they were not directly addressing the subject of this research. The low classification accuracy in some was partly due to their skipping of important machine learning procedures, among other reasons. Additionally, there has been limited research integrating diabetes prediction, diabetic retinopathy detection, and ulcer detection into a unified system. It is this gap that this research intends to bridge by improving diabetes prediction using SVM, ulcer detection using CNN, and diabetic retinopathy detection using CNN, ensuring a more comprehensive healthcare solution for diabetic patients.

2.1 EXISTING SIMILAR WEBSITES

Some of the systems similar to our project are:

- **mySugr:** Founded in 2012, mySugr is a digital health company that aims to simplify life with diabetes. It gives people with diabetes the tools, know-how, and confidence to ease the complexity of their daily diabetes routine.

- **Mayo Clinic:** Mayo Clinic is a website that Offers a symptom checker and detailed information on diabetes symptoms, diagnosis, and treatment. It also provides resources for managing diabetes.

- **Glucose Buddy:** Glucose Buddy is a popular diabetes management app designed to help individuals monitor and manage their blood glucose levels, medications, meals, and physical activities. It aims to provide users with an all-in-one solution for tracking their diabetes-related health data, offering insights and tools to manage their condition effectively.

2.2 LIMITATIONS OF EXISTING SYSTEMS

Existing systems mainly focus on monitoring diabetes but are often designed for foreign countries, making their diet and medication recommendations less suitable for users in Nepal. These systems also lack diabetic foot ulcer detection and diabetic retinopathy detection, crucial for preventing complications like infections and blindness. Diabetic retinopathy, a major cause of blindness in diabetic patients, can be detected through Convolutional Neural Network (CNN), yet many systems don't include this feature. Furthermore, these systems do not provide holistic care such as medicine suggestions, personalized diet plans, yoga animations, fitness tracking, blogs, chatbots, or the ability to chat with doctors, which are essential for comprehensive diabetes management. Without these features, existing systems fail to address all aspects of diabetes care, including prediction, complication detection, and lifestyle support.

2.3 SOLUTION PROVIDED BY OUR SYSTEM

Our system is a comprehensive, all-in-one solution dedicated to diabetes prediction, management, and care, providing a holistic approach to early detection, complication prevention, and management. Our system provides Diabetes Prediction, Diabetic Foot Ulcer Detection, and Diabetic Retinopathy detection. It offers features such as Medicine Suggestions, Yoga Animations, Blogs, Fitness Tracker, Risk Calculator, Food Nutrition and Exercise Suggestions, Chatbot, and Chat-with-Doctor.

CHAPTER 3: METHODOLOGY

3.1 PROCESS MODEL

INCREMENTAL PROCESS MODEL

Incremental Model is a process of software development where requirements divided into multiple standalone modules of the software development cycle. In this model, each module goes through the requirements, design, implementation and testing phases. Every subsequent release of the module adds function to the previous release. The process continues until the complete system achieved.

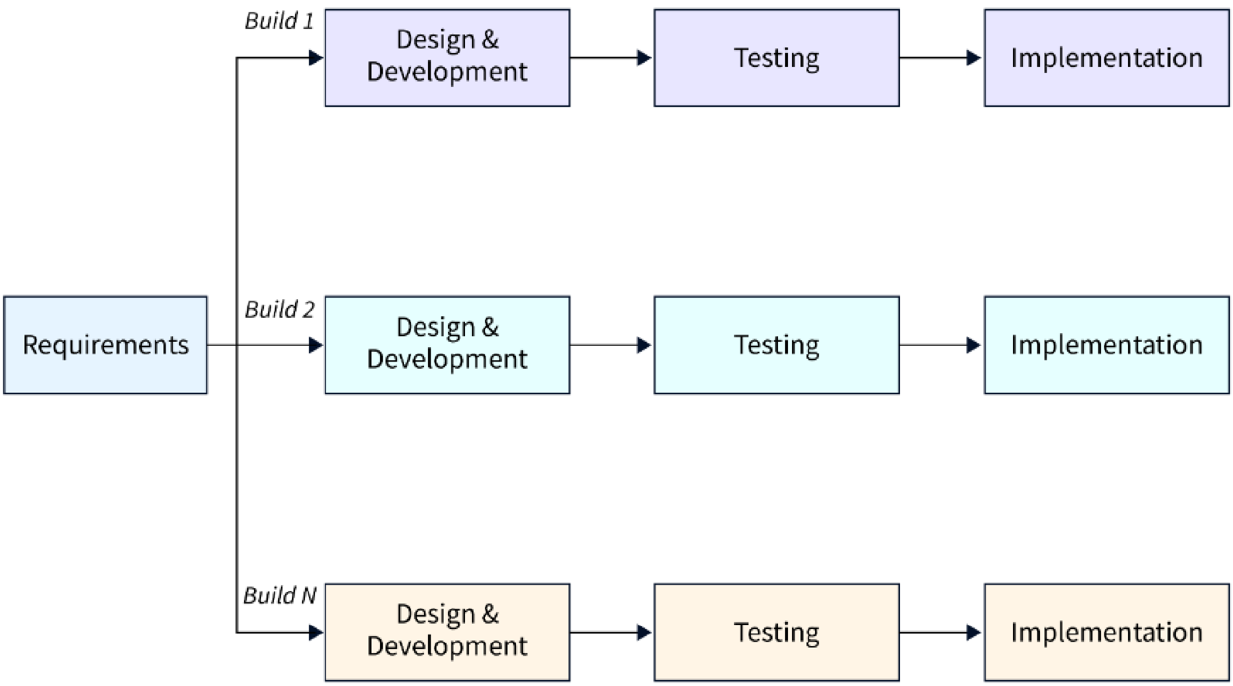


Figure 3: Incremental Process Model

The various phases of Incremental Process Model are as follows:

1. Requirement analysis: Our team began by conducting a comprehensive requirement analysis to define the system's goals and features. The primary objective was to develop an accurate, user-friendly platform capable of predicting diabetes, detecting diabetic foot ulcers, and identifying diabetic retinopathy. We aimed to provide a holistic solution that also includes features such as Blogs, Yoga Animations, Fitness Tracker, Risk Calculator and Medicine Suggestion, Chatbot and Chat-with-Doctor. We gathered a medical dataset of 100,000 data containing clinical parameters such as age, gender, BMI, blood glucose, and HbA1c levels for Diabetes Prediction, 1055 data for Foot Ulcer Detection, 35,126 data for Diabetes Retinopathy from Kaggle. Additionally, we integrated a CNN-based module for diabetic foot ulcer detection and diabetic retinopathy detection. We chose the Incremental Model for systematic progress and iterative enhancements to both the predictive models and auxiliary features.

2. Design & Development: Our team designed a modular system architecture for flexibility and scalability, consisting of three main components: a preprocessing module, a machine learning module for diabetes prediction, and auxiliary modules for features like foot ulcer and retinopathy detection. The preprocessing module handled data normalization, missing values, and feature selection. For diabetes prediction, we selected the Support Vector Machine (SVM) model with an RBF kernel, which was well-suited for handling non-linear data relationships. We also developed a Custom CNN-based Diabetic Foot Ulcer detection module and a ResNet50 CNN-based Diabetic Retinopathy Detection module, which allowed us to analyze retinal images for microaneurysms, hemorrhages, and exudates. Along with these, we integrated features like blogs, yoga animations, fitness tracking, and medication guidance to support users in managing their condition holistically.

3. Testing: Our team conducted rigorous testing to ensure the reliability and efficiency of the system. The SVM model achieved an desired accuracy along with Diabetic Foot Ulcer and Diabetic Retinopathy evaluated based on metrics like Accuracy, Precision, Recall, F1 Score and ROC. The CNN-based ulcer detection model was able to identify Ulcer and No Ulcer Foot effectively, while the retinopathy detection model showed acceptable accuracy in identifying retinal complications. Integration testing ensured that all system components, including the predictive models, ulcer and retinopathy detection, and auxiliary features, worked seamlessly together. Usability testing was carried out to ensure the platform was intuitive and user-friendly. During this phase, we resolved minor bugs, improving the system’s overall performance.

4. Implementation: After completing the testing phase, our team trained the SVM model on the pre-processed dataset to accurately predict diabetes. We also trained the CNN-based modules for diabetic foot ulcer and diabetic retinopathy detection using annotated images. Once the models were fine-tuned, we integrated them with the system's auxiliary features, creating a unified platform. The implementation phase focused on delivering a responsive system that could effectively predict diabetes, detect diabetic foot ulcers, identify diabetic retinopathy, and provide users with personalized care through yoga animations, fitness tracking, medication guidance, and more. Following successful deployment, the system provided a comprehensive solution for diabetes prediction, ulcer detection, retinopathy detection, and lifestyle management. Our team plans to enhance the system further by expanding the dataset, improving retinopathy detection accuracy, and exploring ensemble learning techniques to increase the overall performance of the platform.

3.2 SYSTEM BLOCK DIAGRAM

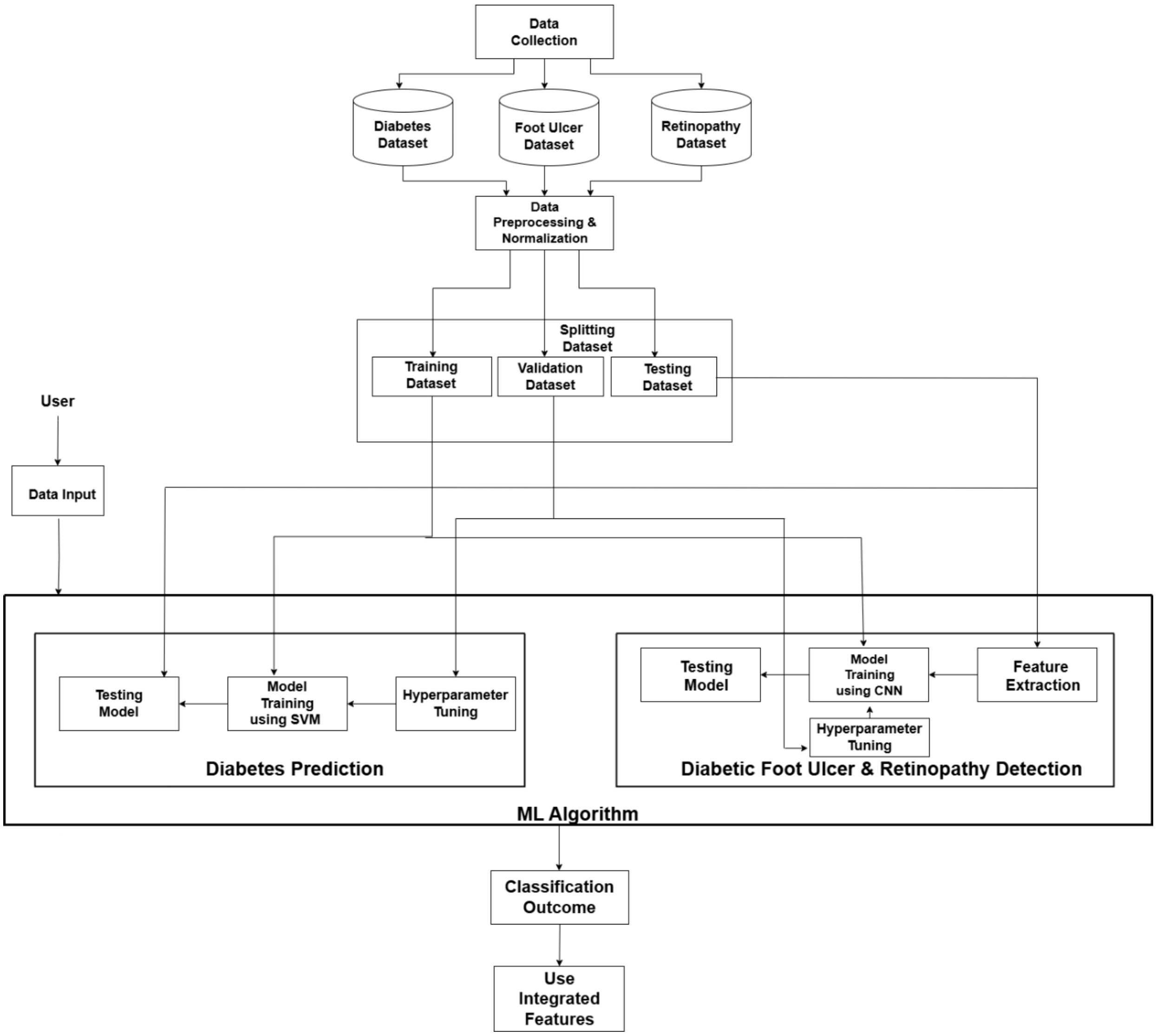


Figure 4: System Block Diagram

Explanation:

Data Collection

We initiated our project by gathering a robust and diverse dataset to train our models. For diabetes prediction, we sourced a dataset from Kaggle that contained key medical parameters such as Age, Gender, Hypertension, Heart Disease, Smoking, BMI, HbA1c level, and Blood Glucose Level. These parameters were carefully chosen for their relevance in predicting diabetes and formed the foundation of our machine learning analysis. Additionally, for foot ulcer detection, we acquired a dataset of foot ulcer images across various severities of ulcers from Kaggle. For diabetic retinopathy detection, we collected high-resolution retinal images, which were also labelled according to the severity of diabetic retinopathy (from No DR to Proliferative DR) from Kaggle. Finally, The Datasets were stored in Diabetes Dataset, Foot Ulcer Dataset and Retinopathy Dataset, respectively.

```
diabetes_prediction_dataset.csv ×
diabetes_prediction_dataset.csv > data
You, 2 weeks ago | 1 author (You)
gender,age,hypertension,heart_disease,smoking_history,bmi,HbA1c_level,blood_glucose_level,diabetes
Female,80,0,1,never,25.19,6.6,140,0
Female,54,0,0,unknown,27.32,6.6,80,0
Male,28,0,0,never,27.32,5.7,158,0
Female,36,0,0,current,23.45,5,155,0
Male,76,1,1,current,20.14,4.8,155,0
Female,20,0,0,never,27.32,6.6,85,0
Male,44,1,1,never,25.11,6.5,200,1
Female,79,0,0,unknown,23.86,5.7,85,0
Male,42,0,0,never,33.64,4.8,145,0
Female,32,0,0,never,27.32,5,100,0
Female,53,0,0,never,27.32,6.1,85,0
Female,54,0,0,former,54.7,6,100,0
Female,78,0,0,former,36.05,5,130,0
Female,67,0,0,never,25.69,5.8,200,0
Female,76,0,0,unknown,27.32,5,160,0
Male,78,0,0,unknown,27.32,6.6,126,0
Male,15,0,0,never,30.36,6.1,200,0
Female,42,0,0,never,24.48,5.7,158,0
Female,42,0,0,unknown,27.32,5.7,80,0
Male,37,0,0,former,25.72,3.5,159,0
Male,40,0,0,current,36.38,6,90,0
Male,5,0,0,unknown,18.8,6.2,85,0
Female,69,0,0,never,21.24,4.8,85,0
```

Table 1: Diabetes Dataset in .csv format

Figure 5: Foot Ulcer Dataset

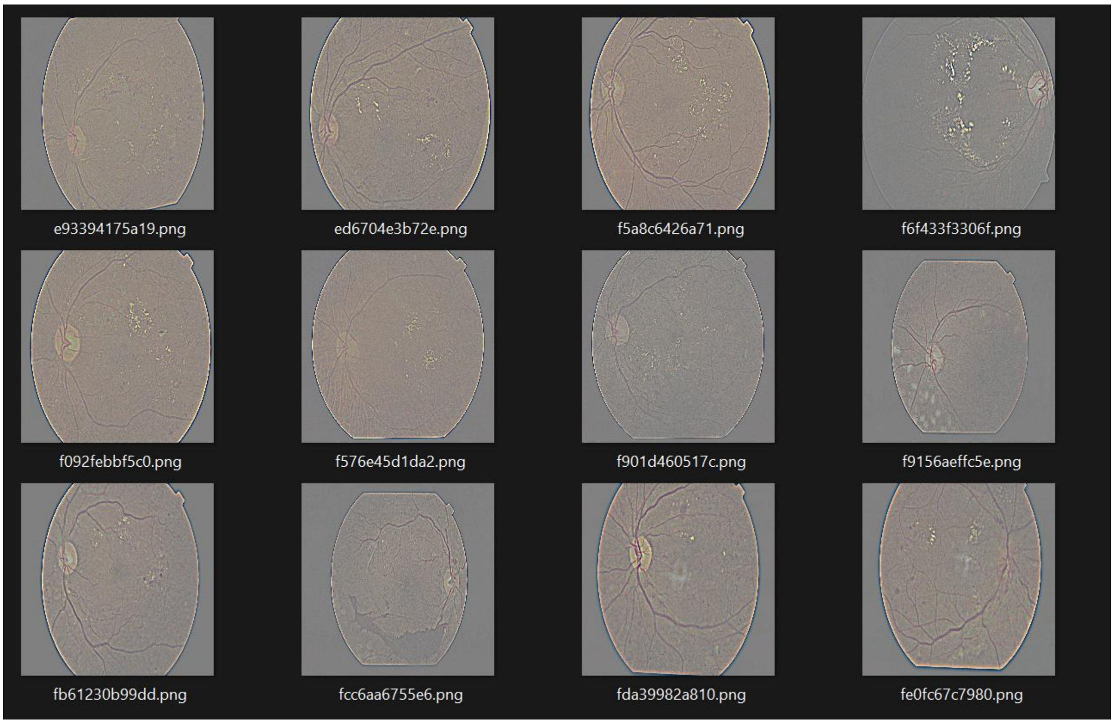


Figure 6: Diabetes Retinopathy Dataset

Data Preprocessing & Normalization

For data preprocessing, we applied normalization to ensure all features in the diabetes dataset were on a consistent scale, which improved the model's performance and stability. We handled missing values via imputation or removal to preserve data integrity. Feature selection was performed to retain only the most relevant medical parameters, eliminating irrelevant or redundant features. For the foot ulcer images, preprocessing included resizing, grayscale conversion, and contrast enhancement to improve the quality of the images for feature extraction. The diabetic retinopathy images underwent similar preprocessing steps, including resizing and normalization, to standardize the dataset before training.

Splitting Dataset

The dataset was split into Training, Validation, and Testing Datasets in the ratio of 70:15:15. The training dataset was used to train the machine learning models, while the testing dataset was used for evaluating the models' generalization. The validation dataset was used to fine-tune model parameters and avoid overfitting.

Model	Training	Validation	Testing
RBF Kernel SVM	70,000	15,000	15,000
Custom CNN	739	158	158
ResNet50	24,588	5,269	5,269

Table 2: Splitting of Datasets

Diabetes Prediction Using RBF Kernel SVM

Model Training Using SVM

We selected Support Vector Machine (SVM) with the Radial Basis Function (RBF) kernel for the classification task. The model was trained using the training dataset, learning the complex relationships between the features to classify patients as diabetic or non-diabetic based on medical parameters. The RBF kernel was particularly effective in capturing non-linear relationships in the dataset.

Hyperparameter Tuning

To optimize the performance of the SVM model, we conducted hyperparameter tuning. Key parameters such as the regularization parameter (C=1) and the kernel-specific parameter (gamma=scale) were fine-tuned. This optimization step was crucial for improving the model's accuracy and ensuring it struck a balance between overfitting and underfitting, which ultimately enhanced the predictive capability of the model.

Testing Model

Once the SVM model was trained and fine-tuned, it was tested on the testing dataset to evaluate its performance. We used evaluation metric like Accuracy, Precision, Recall F1 Score and ROC.

Ulcer Detection Using Custom CNN

Feature Extraction for Ulcer Detection

For foot ulcer detection, feature extraction was done through the Custom CNN model. The model automatically learned important features such as texture, shape, and patterns that are indicative of ulcer severity. Preprocessing steps such as resizing, grayscale conversion, and contrast enhancement were crucial to improving the quality of the images, which, in turn, enhanced the model's ability to extract meaningful features from the data.

Hyperparameter Tuning

To improve the CNN model's performance, we employed hyperparameter tuning. Important parameters such as the number of convolutional layers, kernel sizes, learning rate=0.0001, and batch size were adjusted to find the optimal configuration. Additionally, dropout layers were added to reduce the risk of overfitting and improve the model's generalizability across new, unseen data.

Model Training Using CNN

The CNN model was trained using labelled foot ulcer images, where it learned to detect different features of foot ulcers. The model utilized multiple convolutional layers, pooling layers, and fully connected layers to extract features from the images and classify them accurately. The CNN model proved highly effective in distinguishing between healthy skin.

Testing Model

After training, the CNN model was evaluated on the testing dataset. We assessed the model's performance using Accuracy, Precision, Recall, F1 Score and ROC. The model demonstrated high accuracy in detecting foot ulcers at different stages, making it a reliable tool for healthcare professionals.

Classification Outcome for Foot Ulcer Detection

The CNN-based ulcer detection model performed excellently in identifying diabetic foot ulcers. It successfully distinguished them from healthy skin. This model's early detection capabilities can significantly improve patient outcomes by enabling timely intervention.

Diabetic Retinopathy Detection Using ResNet50 CNN

Feature Extraction for Retinopathy Detection

For diabetic retinopathy detection, feature extraction was conducted on retinal images using the ResNet50 CNN model. The model automatically identified and learned important features such as microaneurysms, haemorrhages, and exudates, which are characteristic of different stages of diabetic retinopathy. Preprocessing steps like resizing, grayscale conversion, and contrast normalization were applied to enhance the quality of the retinal images for better feature extraction.

Hyperparameter Tuning

To optimize the CNN model's performance, hyperparameter tuning was performed. Key parameters such as the number of convolutional layers, kernel sizes, learning rate=0.0001, and batch size were fine-tuned. Dropout layers were added to prevent overfitting and ensure that the model could generalize well to unseen retinal images.

Model Training Using CNN

The CNN model was trained on labelled retinal images to classify Diabetic Retinopathy into five stages: No DR, Mild DR, Moderate DR, Severe DR, and Proliferative DR. Using multiple convolutional layers and pooling operations, the CNN learned to recognize key features from retinal images that corresponded to each stage of Diabetic Retinopathy.

Testing Model

After training, the CNN model was tested on the testing dataset. We evaluated its performance using Accuracy, Precision, Recall, F1 Score and ROC, which confirmed that the model could classify the severity of diabetic retinopathy with high accuracy.

Classification Outcome for Diabetic Retinopathy Detection

The CNN-based retinopathy detection model accurately classified retinal images into five categories: No DR, Mild DR, Moderate DR, Severe DR, and Proliferative DR. This high accuracy in classification allows for early intervention and better management of diabetic retinopathy, potentially preventing severe complications such as blindness.

Data Input

Before the Classification Outcome phase, the Data Input Block allows users to input their medical data and images for diabetes prediction and ulcer/retinopathy detection.

For diabetes prediction, users can input essential medical parameters, such as Age, Gender, Hypertension, Heart Disease, Smoking, BMI, HbA1c level, and Blood Glucose Level. These inputs are then processed through the trained SVM model, which provides a prediction regarding the likelihood of the user being diabetic or non-diabetic.

For foot ulcer detection, users are required to upload an image of their foot, which is processed by the trained CNN model. The model classifies the ulcer into different stages based on the image provided.

For diabetic retinopathy detection, users upload a retinal image, which is fed into the trained CNN model. The model evaluates the image and classifies the presence and severity of diabetic retinopathy into stages ranging from No DR to Proliferative DR.

Use Integrated Features

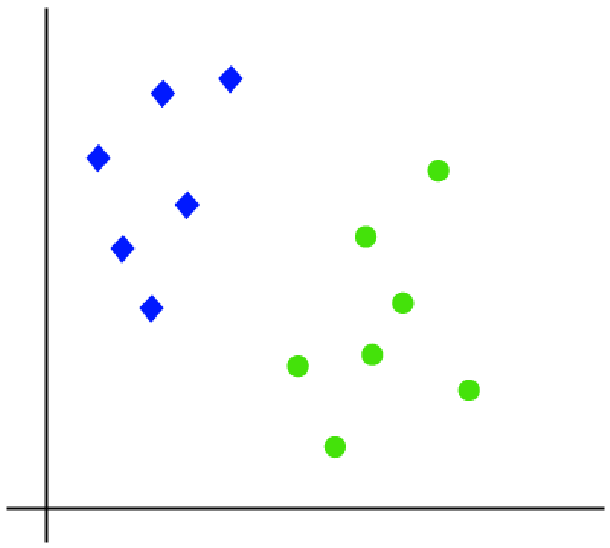
Users can use other integrated features of our system such as Medicine Suggestions, Yoga Animations, Blogs, Fitness Tracker, Risk Calculator, Food Nutrition and Exercise Suggestions, Chatbot, and Chat-with-Doctor that helps in holistic Diabetes care and management.

3.3: ALGORITHM

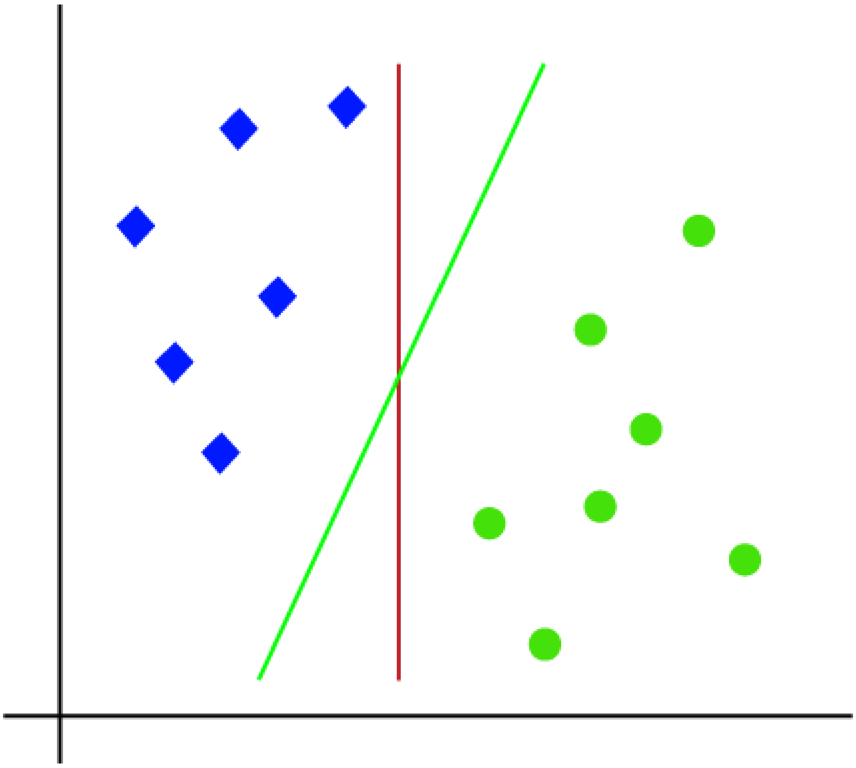
3.3.1: SUPPORT VECTOR MACHINE(SVM)

Support Vector Machine (SVM) is a supervised machine learning algorithm used for both classification and regression. Though we say regression problems as well it's best suited for classification. The main objective of the SVM algorithm is to find the optimal hyperplane in an N-dimensional space that can separate the data points in different classes in the feature space. The hyperplane tries that the margin between the closest points of different classes should be as maximum as possible. The dimension of the hyperplane depends upon the number of features. If the number of input features is two, then the hyperplane is just a line. If the number of input features is three, then the hyperplane becomes a 2-D plane. It becomes difficult to imagine when the number of features exceeds three.

The working of the SVM algorithm can be understood by using an example. Suppose we have a dataset that has two tags (green and blue), and the dataset has two features x1 and x2. We want a classifier that can classify the pair (x1, x2) of coordinates in either green or blue. Consider the below image:



So, as it is 2-d space so by just using a straight line, we can easily separate these two classes. But there can be multiple lines that can separate these classes. Consider the below image:



Hence, the SVM algorithm helps to find the best line or decision boundary; this best boundary or region is called as a hyperplane. SVM algorithm finds the closest point of the lines from both the classes. These points are called support vectors. The distance between the vectors and the hyperplane is called as margin. And the goal of SVM is to maximize this margin. The hyperplane with maximum margin is called the optimal hyperplane.

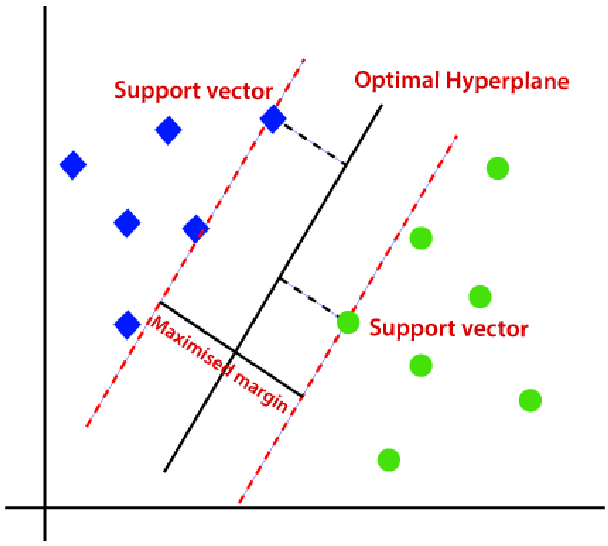


Figure 7: SVM Decision Boundary and Support Vectors [14]

SUPPORT VECTOR MACHINE (SVM) for Diabetes Prediction

Step 1: Problem Formulation

The goal of SVM is to find a hyperplane that separates the two classes (diabetic and non-diabetic) with the largest margin in the feature space. Given a dataset with n samples:

$$D = \{(\mathbf{x}_i, y_i)\}_{i=1}^{n}, \mathbf{x}_i \in \mathbb{R}^d, y_i \in \{-1,1\}$$

Where:

- $\mathbf{x}_i$ = feature vector (e.g., glucose, BMI, age, etc.)
- y_i = class label (+1 for diabetic, -1 for non-diabetic)

Step 2: Linear SVM

For linearly separable data, the hyperplane is defined as:

$$\mathbf{w} \cdot \mathbf{x} + b = 0$$

Where:

- $\mathbf{w}$ = weight vector (perpendicular to the hyperplane)
- b = bias (distance from origin)

The objective is to maximize the margin between the hyperplane and the nearest data points (support vectors). The margin is given by $\frac{2}{\|\mathbf{w}\|}$. Maximizing the margin is equivalent to minimizing $\|\mathbf{w}\|^2$. The optimization problem becomes:

$$\min_{\mathbf{w},b} \frac{1}{2} \|\mathbf{w}\|^2 \ \text{ subject to } \ y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1, \forall i$$

Step 3: Soft Margin SVM (For Non-Linearly Separable Data)

When data is not linearly separable, slack variables ξ_i are introduced:

$$\min_{\mathbf{w},b,\xi} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^{n} \xi_i \ \text{ subject to } \ y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 - \xi_i, \xi_i \geq 0, \forall i$$

Where C is the regularization parameter balancing margin size and misclassification tolerance.

Step 4: Kernel Trick for Nonlinear Relationships

Since the data in this project is non-linearly separable, the RBF kernel is applied to transform the feature space into a higher-dimensional space where the data becomes linearly separable. The kernel function computes the similarity between two points $\mathbf{x}_i$ and $\mathbf{x}_j$ as:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\gamma \left\|\mathbf{x}_i - \mathbf{x}_j\right\|^2\right)$$

Where:

- $\gamma =$ kernel parameter controlling the influence of individual points.

The optimization problem in the dual form becomes:

$$\max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j)$$

$$\text{subject to } \ 0 \leq \alpha_i \leq C, \sum_{i=1}^{n} \alpha_i y_i = 0$$

Where α_i are the Lagrange multipliers.

Step 5: Decision Function

Once the optimization is solved, the decision function for a new input $\mathbf{x}$ is:

$$f(\mathbf{x}) = \text{sign}\left(\sum_{i=1}^{n} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b\right)$$

The sign of $f(\mathbf{x})$ determines whether $\mathbf{x}$ is classified as diabetic $(+1)$ or non-diabetic (-1).

Step 6: Hyperparameter Optimization

Key hyperparameters tuned during the model training were:

1. **C (Regularization parameter): Controls** the trade-off between achieving a low error on the training data and maintaining a large margin.
2. $\boldsymbol{\gamma}$ **(Kernel parameter):** Defines how far the influence of a single training example reaches in the feature space.

Grid search and cross-validation were used to find the optimal values for C and γ.

The SVM algorithm with the RBF kernel successfully captured the non-linear relationships in the diabetes dataset. It transformed the input features into a higher-dimensional space where a separating hyperplane was identified. This approach, combined with proper hyperparameter tuning, achieved an accuracy of 96.14%, making it an effective tool for early diabetes prediction.

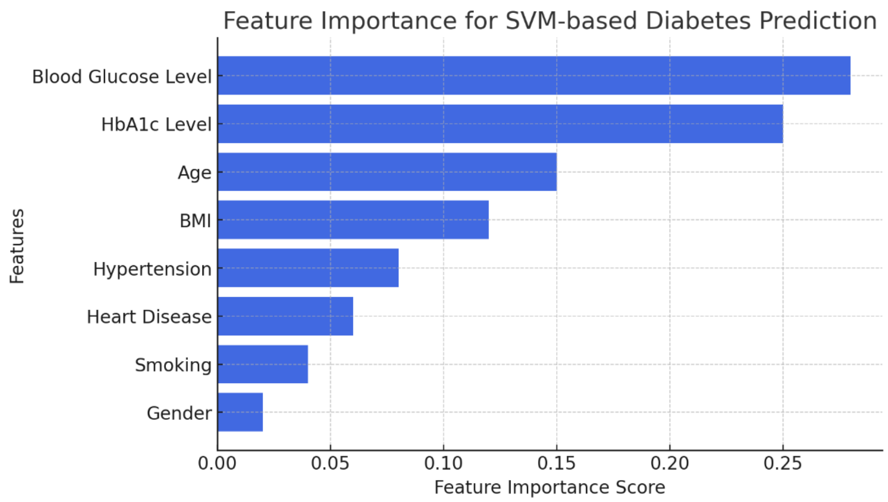


Figure 8: Feature Importance for SVM-Based Diabetes Prediction

The graph above illustrates the importance of various features in the SVM model for predicting diabetes. The x-axis represents the "Mean Decrease in Accuracy," which indicates how much the model's performance decreases when a specific feature is randomly permuted. The features are listed on the y-axis.

Observation from graph:

Blood Glucose Level

- This is the most significant predictor of diabetes in the model.
- Higher blood glucose levels strongly correlate with diabetes risk.

HbA1c Level

- The second most important feature.
- HbA1c measures average blood sugar over time, making it a reliable indicator of diabetes.

Age

- Older individuals have a higher risk of developing diabetes.
- The model assigns considerable importance to this factor.

BMI (Body Mass Index)

- Higher BMI is associated with increased diabetes risk.
- Obesity is a well-known risk factor for insulin resistance.

Hypertension (High Blood Pressure)

- Moderately important in the prediction.
- People with high blood pressure often have a higher risk of diabetes.

Heart Disease

- Also has moderate importance.
- Diabetes and heart disease are closely linked due to shared risk factors.

Smoking

- Has relatively low importance in the model.

- While smoking is linked to several health issues, its direct impact on diabetes may not be as strong.

Gender

- The least important factor in the model.
- While diabetes prevalence may differ by gender, it does not significantly affect prediction in this case.

3.3.2: CONVOLUTIONAL NEURAL NETWORK (CNN)

A Convolutional Neural Network (CNN) is a type of artificial neural network (ANN) that is specifically designed to handle image data. CNNs are inspired by the structure of the human visual cortex and have a hierarchical architecture that allows them to extract features from images at different scale.

The working of the CNN algorithm can be understood by the below figure:

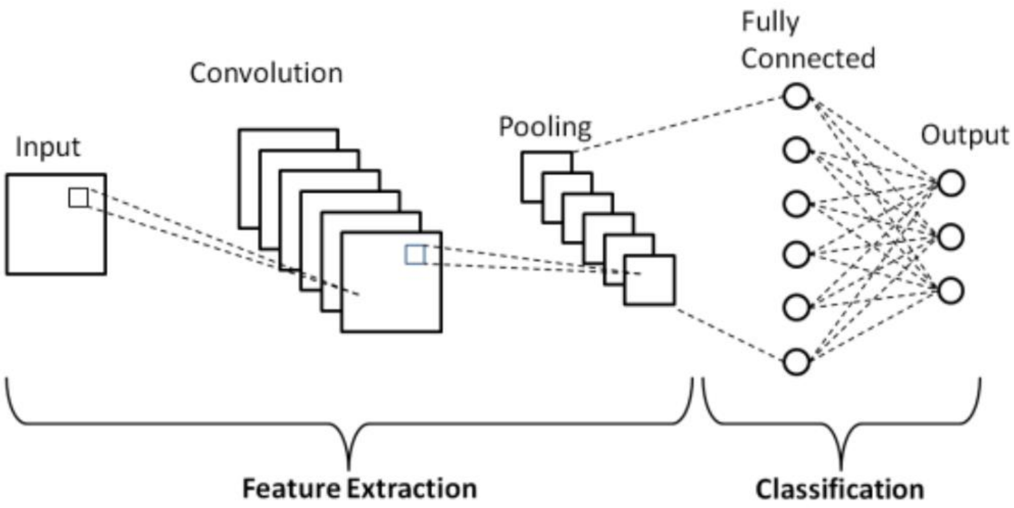


Figure 9: Convolutional Neural Network (CNN) [15]

Convolutional Neural Network (CNN) for Foot Ulcer and Diabetes Retinopathy Detection

Convolutional Neural Networks (CNNs) are a class of deep learning models particularly effective for image analysis tasks, such as detecting foot ulcers in medical images. The following steps outline how a CNN works for foot ulcer detection:

1. Input Layer

- The input to the CNN is a pre-processed foot image, typically resized to a fixed dimension (e.g., 224x224 pixels) and normalized to ensure consistency.
- Each image is represented as a 3D tensor (height × width × channels), where channels correspond to color components (e.g., RGB).

2. Convolutional Layers

- Convolutional layers apply a set of learnable filters (kernels) to the input image to extract spatial features such as edges, textures, and patterns.
- Each filter slides over the image, performing element-wise multiplication and summation to produce a feature map.
- Multiple filters are used to capture different features, and the depth of the output feature maps corresponds to the number of filters.

Mathematical Operation:

$$\text{Feature Map}(x, y) = \sum_{i=0}^{k-1} \sum_{j=0}^{k-1} \text{Input}(x+i, y+j) \cdot \text{Filter}(i, j) + \text{Bias}$$

- k : Filter size (e.g., 3×3 or 5×5).

3. Activation Function

- After convolution, a non-linear activation function (e.g., ReLU) is applied to introduce non-linearity and enable the network to learn complex patterns.

$$\text{ReLU}(x) = \max(0, x)$$

4. Pooling Layers

- Pooling layers (e.g., max pooling or average pooling) down sample the feature maps, reducing their spatial dimensions while retaining the most important information.
- This helps reduce computational complexity and prevents overfitting.
- For example, max pooling selects the maximum value from a local region of the feature map.

5. Fully Connected Layers

- After several convolutional and pooling layers, the high-level features are flattened into a 1D vector and passed through fully connected (dense) layers.
- These layers combine the extracted features to make a final prediction (e.g., ulcer present or not and No DR to Proliferative DR).

6. Output Layer

- The output layer uses a SoftMax activation function for multi-class classification or a sigmoid activation function for binary classification (e.g., ulcer vs. no ulcer and No DR to Proliferative DR).
- The output represents the probability of the input image belonging to each class.

7. Loss Function and Optimization

- A loss function (e.g., binary cross-entropy for binary classification) measures the difference between the predicted and actual labels.
- The model is optimized using gradient-based algorithms like Adam or SGD to minimize the loss.

8. Training Process

- The CNN is trained on a labelled dataset of foot and retinopathy images, where each image is annotated as either containing an ulcer or not and No DR to Proliferative DR.

- During training, the model learns to adjust its weights and biases to improve prediction accuracy.

9. Evaluation

- The trained CNN is evaluated on a separate test dataset using metrics such as Accuracy, Precision, Recall, F1 Score and ROC.

Custom CNN Model Architecture

Custom CNN was used for Diabetic Foot Ulcer Detection in our project that was customized and designed as per our need.

Key Features:

- **Depth:** 6 convolutional layers followed by dense layers.
- **Pooling:** MaxPooling layers reduce spatial dimensions, improving computational efficiency.
- **Dropout:** Prevents overfitting by randomly disabling neurons during training.
- **Activation Function:** ReLU for hidden layers, Sigmoid for binary classification.

Architecture:

1. **Convolutional Layers:**
    - **First Layer:** 3×3 Conv2D (32 filters) + ReLU
    - **Second Layer:** 2×2 MaxPooling
    - **Third Layer:** 3×3 Conv2D (64 filters) + ReLU
    - **Fourth Layer:** 2×2 MaxPooling
2. **Fully Connected Layers:**
    - **Flatten:** Converts feature maps into a 1D vector.
    - **Dense Layer (128 neurons, ReLU):** Extracts high-level features.
    - **Dropout (0.5):** Prevents overfitting.

- ◦ **Dense Output Layer (Sigmoid):** Outputs probability for ulcer classification.

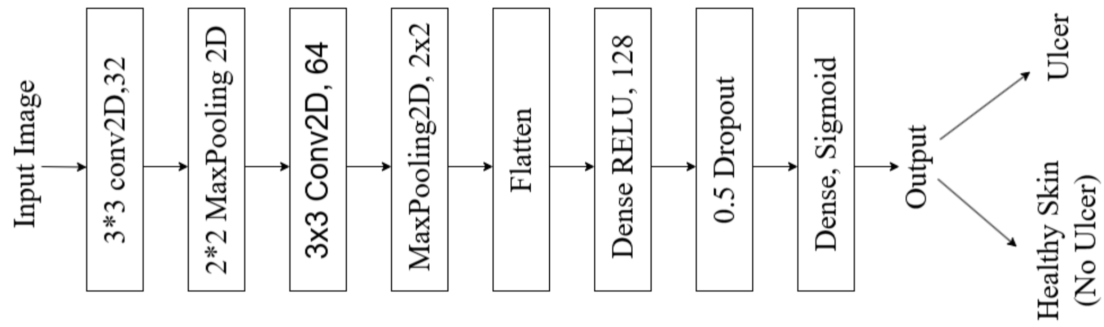


Figure 10: Custom CNN Model Architecture

ResNet50 Model Architecture

ResNet50 was used for Diabetic Retinopathy Detection in our project that follows Bottleneck Architecture. ResNet50 (Residual Network-50) is a deep convolutional neural network (CNN) with 50 layers, designed to tackle the vanishing gradient problem using residual learning. It is widely used for image classification, object detection, and feature extraction.

Key Features:

- **Depth:** 50 layers (48 convolutional layers, 1 max-pooling layer, 1 fully connected layer).
- **Residual Blocks:** Uses skip connections (identity shortcuts) to bypass layers, allowing smoother gradient flow.

Architecture:

- **Initial Layers:** 7×7 convolution + max pooling
- **Four Residual Stages:**

    Conv2_x: 3 bottleneck blocks

    Conv3_x: 4 bottleneck blocks

    Conv4_x: 6 bottleneck blocks

Conv5_x: 3 bottleneck blocks

Fully Connected Layer: 1000 classes (ImageNet)

Activation Function: ReLU

Pooling: Global Average Pooling (GAP) before the fully connected layer

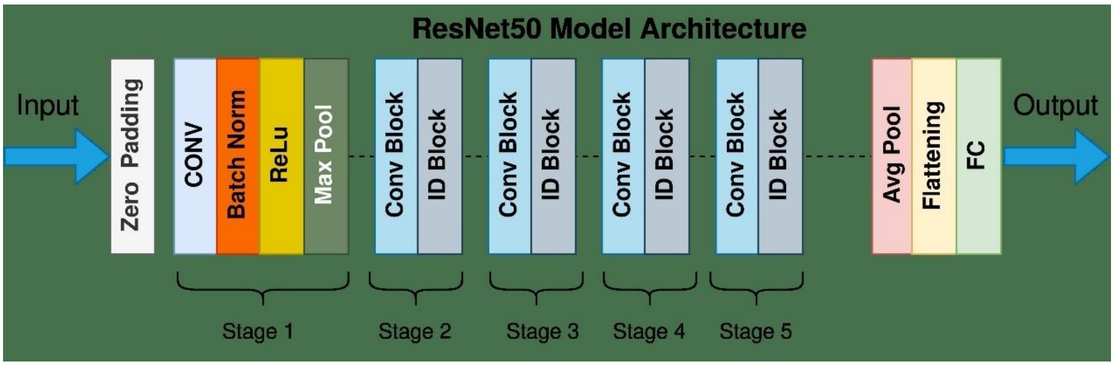


Figure 11: ResNet50 Model Architecture

3.4: FLOWCHART

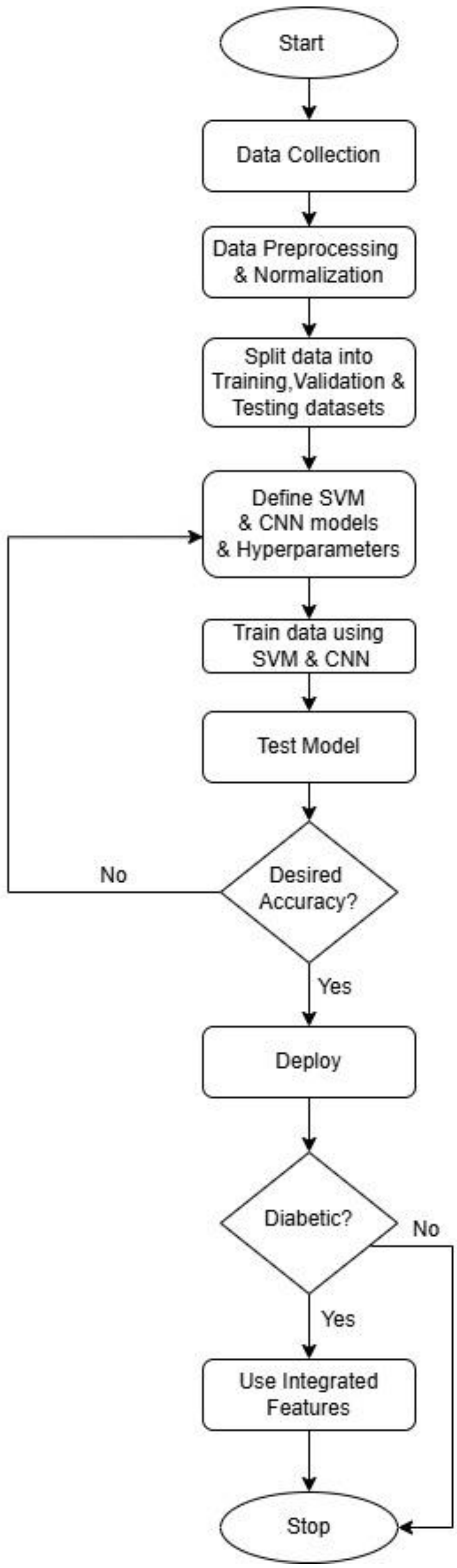


Figure 12: Flowchart

3.5 UML DIAGRAMS

3.5.1 USE CASE

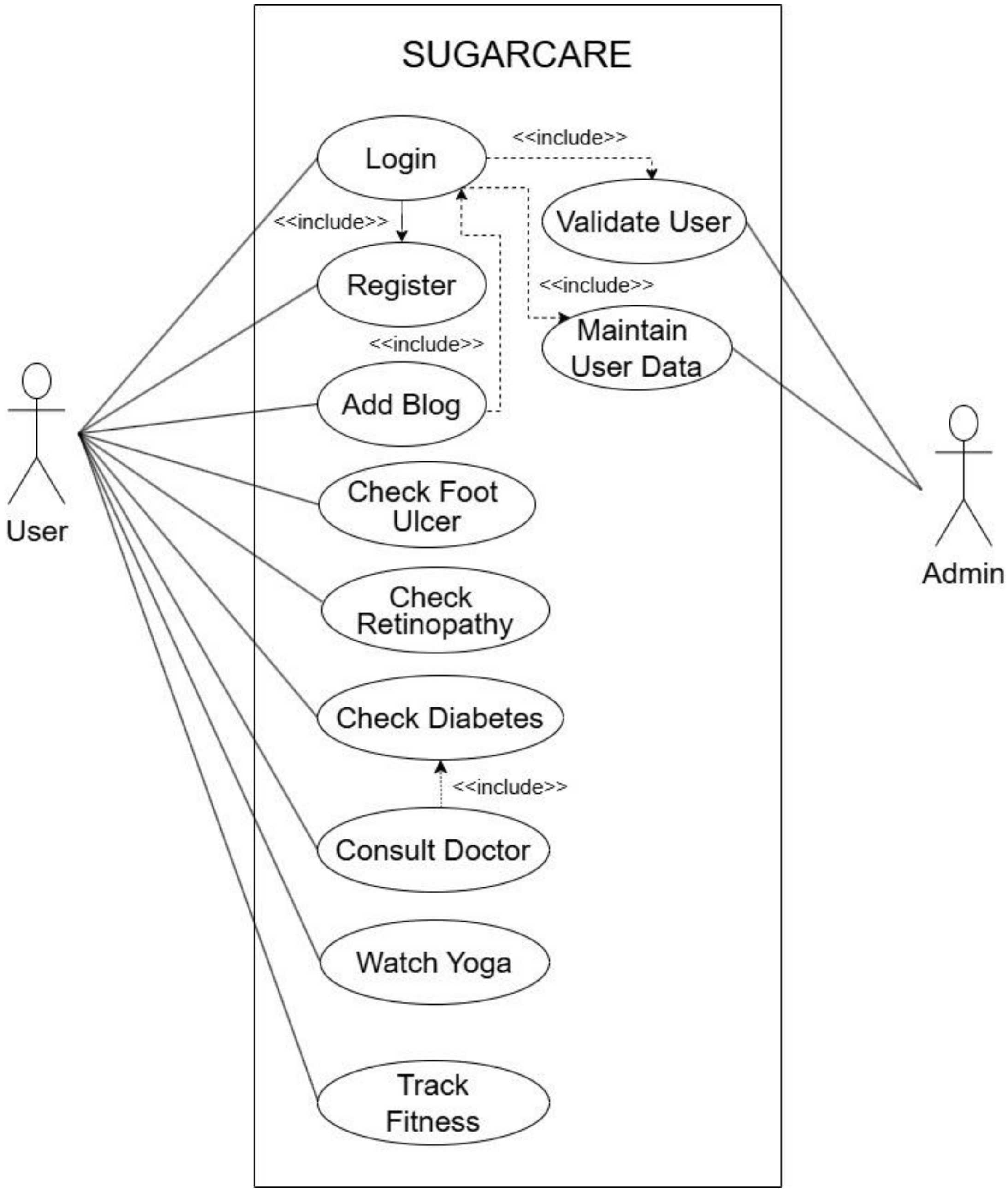


Figure 13: Use Case Diagram

3.5.2 ACTIVITY DIAGRAM

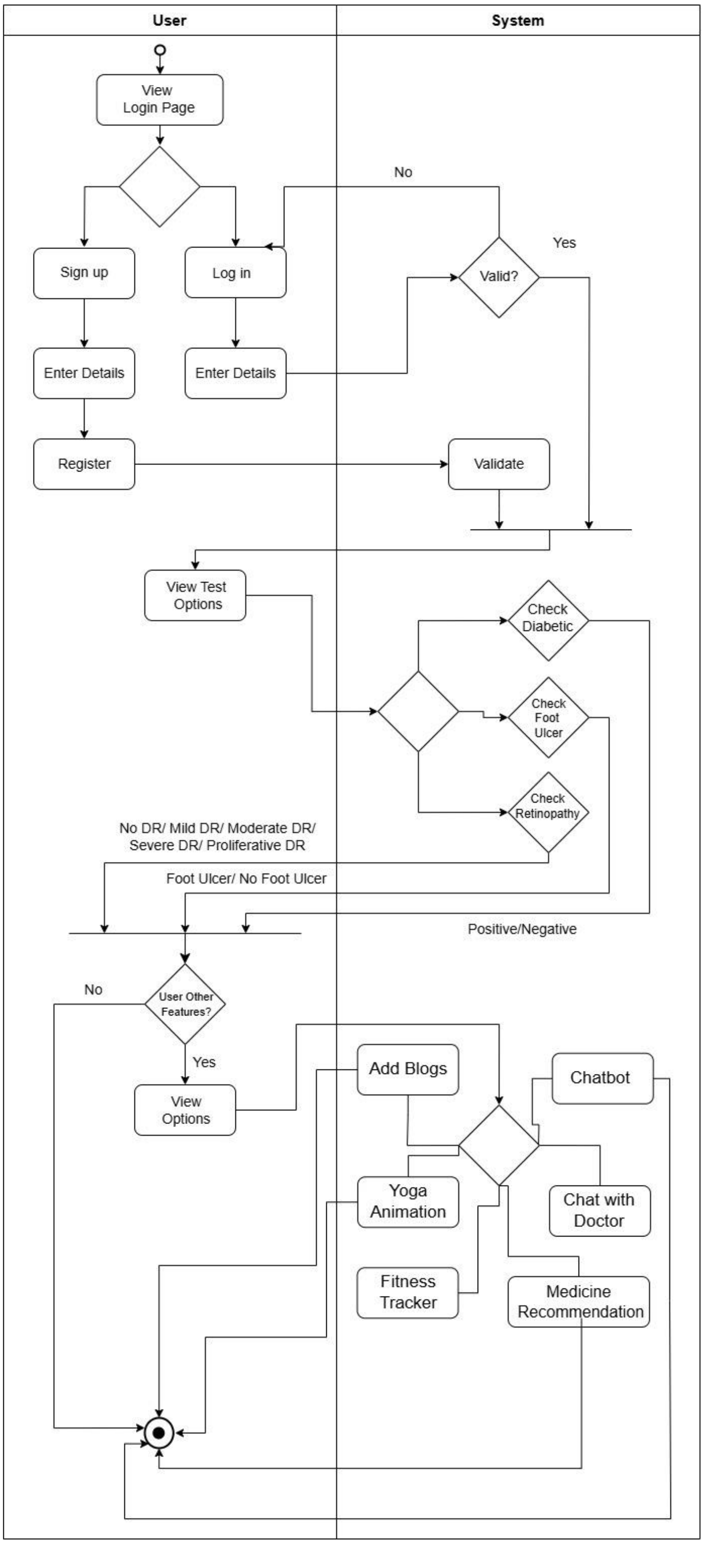


Figure 14: Activity Diagram

3.5.3 SEQUENCE DIAGRAM

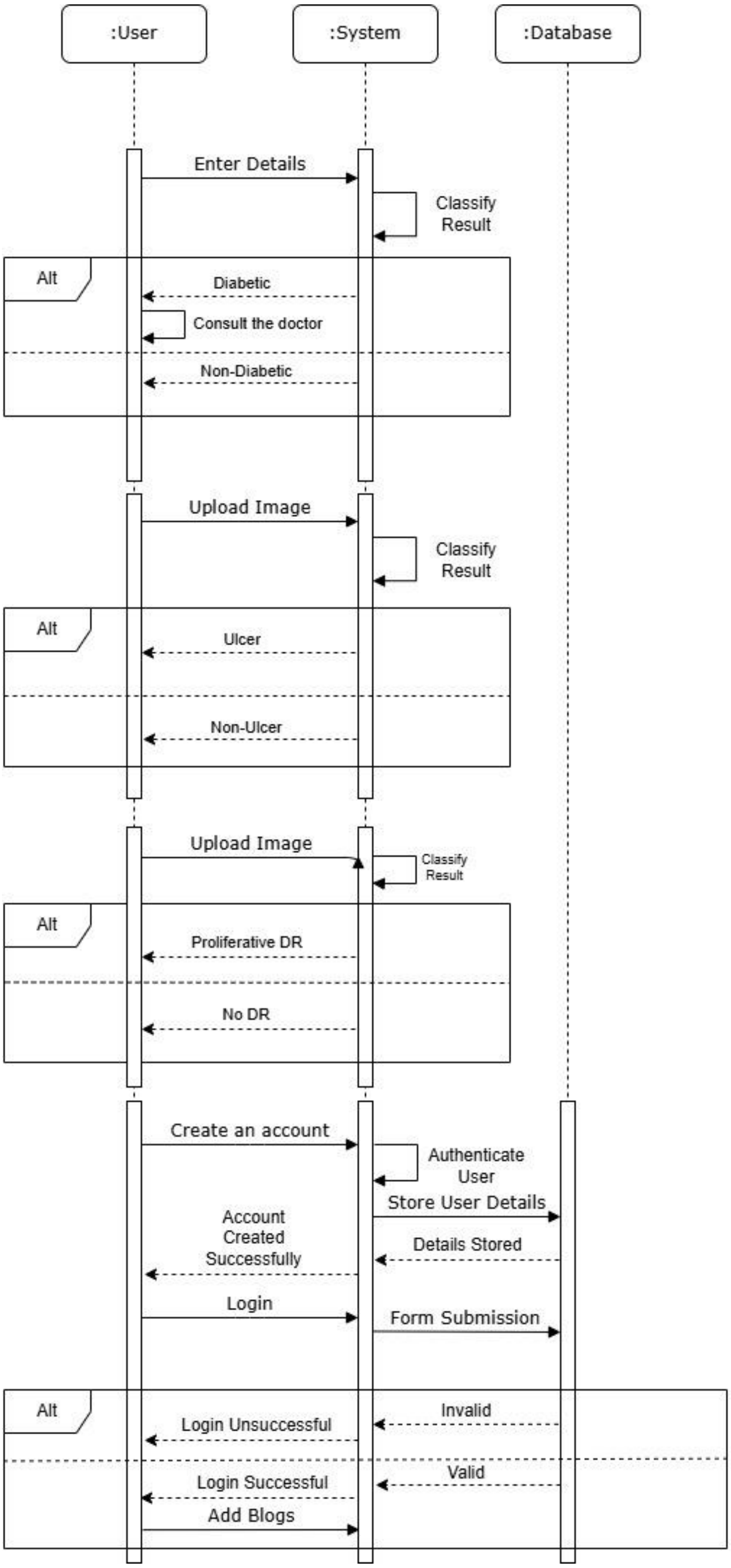


Figure 15: Sequence Diagram

3.6 TOOLS USED

1. PYTHON

Python is a widely used programming language known for its simplicity and versatility. In this project, Python was used for developing machine learning models, implementing algorithms, and handling data processing tasks.

2. NumPy

NumPy is a library for numerical computing and handling multi-dimensional arrays. It was used for performing mathematical operations and efficiently managing large medical datasets.

3. Pandas

Pandas is a data manipulation and analysis library. It was used for preprocessing the diabetes dataset, cleaning missing values, and structuring data for machine learning models.

4. Scikit-Learn:

Scikit-Learn is a machine learning library that provides tools for classification, regression, and clustering. It was used to implement the Support Vector Machine (SVM) for diabetes prediction.

5. Bootstrap:

Bootstrap is a front-end framework for building responsive web applications. It was used to design the user interface of the Our System, ensuring accessibility across devices.

6. TensorFlow

TensorFlow is an open-source deep learning framework developed by Google. It was used to train and deploy the Convolutional Neural Network (CNN) model for foot ulcer detection.

7. OpenCV

OpenCV is a computer vision library used for image processing. It was used to preprocess foot ulcer and retinopathy images, including resizing, augmentation, and noise reduction before feeding them into the CNN model.

8. Keras

Keras is a high-level deep learning API built on TensorFlow. It was used to design, train, and fine-tune the CNN architecture for detecting foot ulcers in medical images.

9. PIL (Pillow)

PIL (Pillow) is a Python Imaging Library for handling and processing images. It was used for loading, resizing, and standardizing medical images before training the deep learning models.

10. pytorch:

PyTorch is an open-source deep learning framework known for its dynamic computational graph and GPU acceleration. It was used for training Convolutional Neural Network (CNN) models for diabetic retinopathy detection.

11. Blender:

Blender is an open-source 3D creation software. It was used for creating 3D animations for yoga.

12. Visual Studio Code:

Visual Studio Code (VS Code) is a lightweight and powerful code editor. It was used for writing, debugging, and managing the development of the project efficiently

3.7 VERIFICATION AND VALIDATION

Stage	Task	Verification	Validation
Data Collection	Collect medical data	Verified data sources (Kaggle) for Diabetes Dataset (100,000), Foot Ulcer Dataset (1000), Retinopathy Dataset (35,126) and ensured the datasets are complete and accurate.	Validated that collected parameters (age, gender, hypertension, heart disease, smoking, BMI, HbA1c level, blood glucose level) are clinically relevant. Ensured high-quality ulcer and retinal images are included.
Data Preprocessing	Normalize and cleaned structured and image data	Verified normalization scales, check for missing values, ensure image resizing, contrast enhancement, and augmentation are applied correctly.	Validated that preprocessing improves model accuracy without distorting clinical information. Ensure images retain diagnostic clarity.
Dataset Splitting	Split data into Training (70,000, 739, 24588), Validation (15000, 158, 5269) and Testing (15000, 158, 5269) datasets for SVM, Custom CNN and ResNet50 CNN, respectively.	Verified that the splitting follows the standard ratio (70:15:15).	Validated that the split supports meaningful evaluation of real-world predictive performance.
Model Training (RBF Kernel SVM)	Train SVM model with RBF kernel	Verified that the model uses the correct hyperparameters (C and gamma) as per optimization.	Validated that the model fits the training data and correctly identifies complex relationships.
Model Training (Custom CNN)	Train Ulcer Foot Detection model using Custom CNN Model	Verified that the CNN architecture is implemented correctly with appropriate layers and activation functions.	Validated that the CNN achieves high sensitivity and specificity in detecting foot ulcers.
Model Training (ResNet50 CNN)	Train Diabetic Retinopathy	Verified that the CNN is structured with	Validated that the model accurately classifies

	Detection model using ResNet50 CNN	convolutional layers, transfer learning, and proper preprocessing.	retinopathy severity (mild, moderate, severe, proliferative).
Hyperparameter Tuning	Optimize *C* and *gamma* (SVM) and CNN parameters	Verified that value of C & Gamma for SVM and Learning Rate, No. of Filters and Filter size for CNN are correctly selected.	Validated that optimized hyperparameters lead to improved performance on unseen data.
Testing (Diabetes Prediction Model)	Evaluate Diabetes Prediction model on test dataset	Verified testing results by computing accuracy.	Validated that the model achieved higher accuracy (96.19%) and reliably predicts diabetes outcomes using real-world data from our team members, friends, family and relatives.
Testing (Foot Ulcer Detection Model)	Evaluate Ulcer Detection model on test dataset	Verified the model's ability to classify ulcer vs. no-ulcer images accurately.	Validated that the CNN achieves high sensitivity and specificity in detecting foot ulcers even on images sourced from various public online datasets.
Testing (Retinopathy Detection Model)	Evaluate Retinopathy Detection model on test dataset	Verified the model's ability to classify No Dr to Proliferative DR images accurately.	Validated that CNN correctly detects and stages diabetic retinopathy even on images sourced from various public online datasets.
Classification Outcome	Generate diabetic/non-diabetic, ulcer/no ulcer, No DR to Proliferative DR results	Verified that predictions align with expected results based on the test dataset.	Validated that predictions align with real-world clinical interpretations and diagnosis even on unseen and real data.

Table 3: Verification and Validation Table

CHAPTER 4: EPILOGUE

4.1 RESULTS AND CONCLUSION

RESULT

The implementation and evaluation of Our System yielded encouraging results across multiple functionalities, demonstrating its potential as a holistic solution for diabetes management and early complication detection. Key outcomes are as follows:

Diabetes Prediction using RBF Kernel SVM

The RBF Kernel Support Vector Machine (SVM) model achieved an **96.19% Test Accuracy**, providing reliable diabetes risk assessments based on various clinical parameters, Age, Gender, Hypertension, Heart Disease, Smoking, BMI, HbA1c level and Blood Glucose Level. Effective data preprocessing (normalization, missing value handling, and feature selection) with Hyperparameter Tunning enhanced prediction accuracy and improved the model's performance.

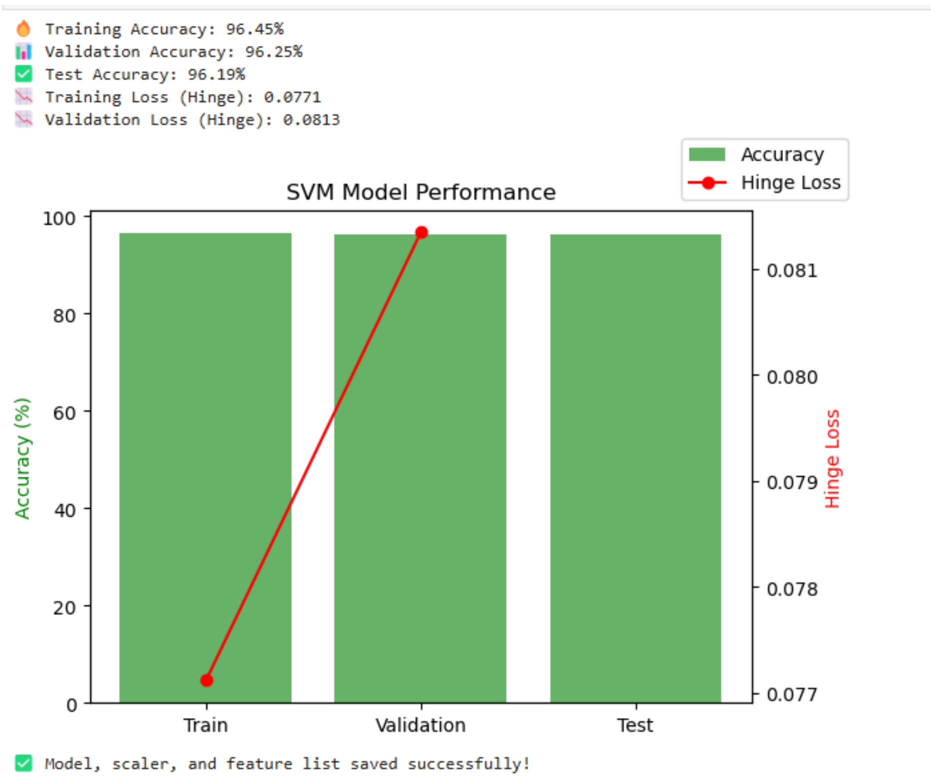


Figure 16: SVM Model Accuracy

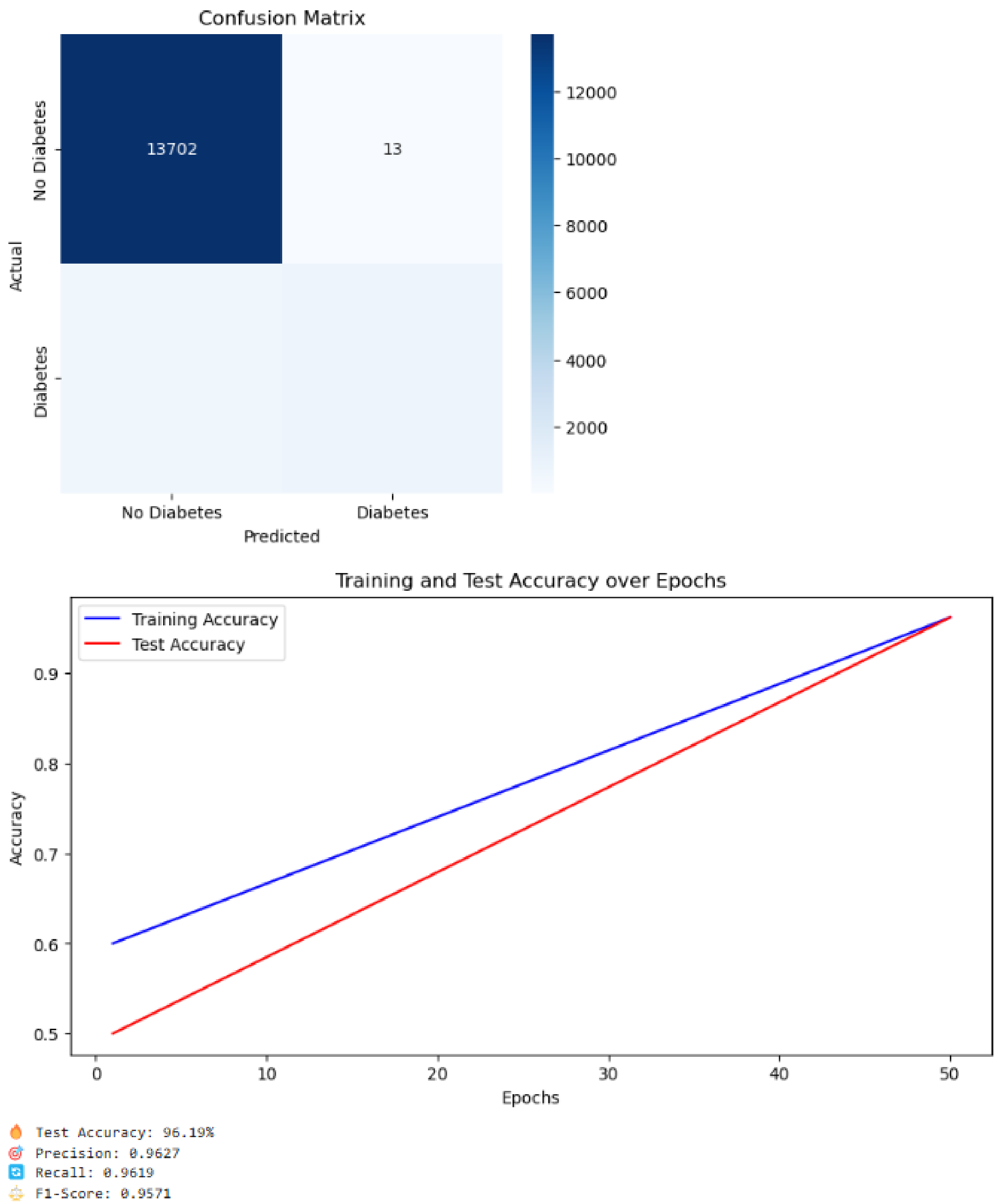


Figure 17: Performance Evaluation of Diabetes Classification Model: Confusion Matrix & Accuracy Trend Over Epochs

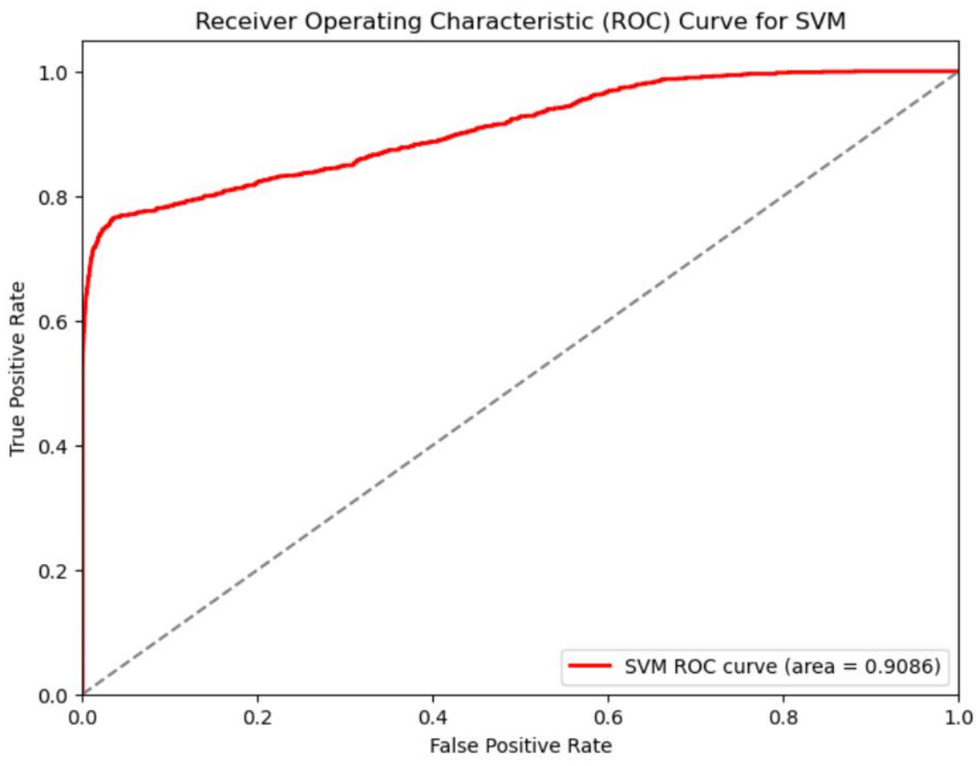


Figure 18: ROC Curve of model for Diabetes Prediction

Analysis:

The RBF Kernel based SVM diabetes prediction system demonstrates excellent performance, achieving high accuracy across training (96.45%), validation (96.25%), and test (96.19%) datasets, with minimal hinge loss, ensuring effective classification. The confusion matrix shows a very low false positive rate (only 13 cases misclassified), and strong performance metrics with precision (0.9627), recall (0.9619), and F1-score (0.9571), confirming the model's reliability. The accuracy trend over epochs indicates consistent learning without overfitting. The ROC curve (0.9086) further highlights the model's strong ability to distinguish diabetic and non-diabetic cases. Overall, this SVM model is highly effective and well-suited for real-world diabetes prediction.

Diabetic Foot Ulcer Detection using Custom CNN

The Custom Convolutional Neural Network (CNN) model successfully detected diabetic foot ulcers with **94% Test Accuracy**. Image preprocessing techniques, such as resizing, augmentation, and normalization, contributed to consistent and accurate detection. Feature extraction from high-level CNN architecture enhanced the model's capability to identify foot ulcers.

Figure 19: Foot Ulcer CNN Model Accuracy

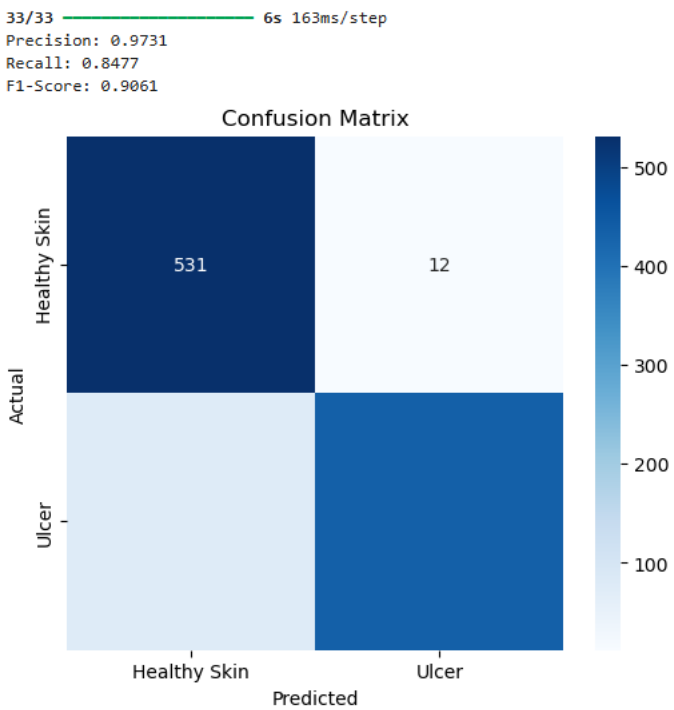


Figure 20: Confusion Matrix of Model for Foot Ulcer Detection

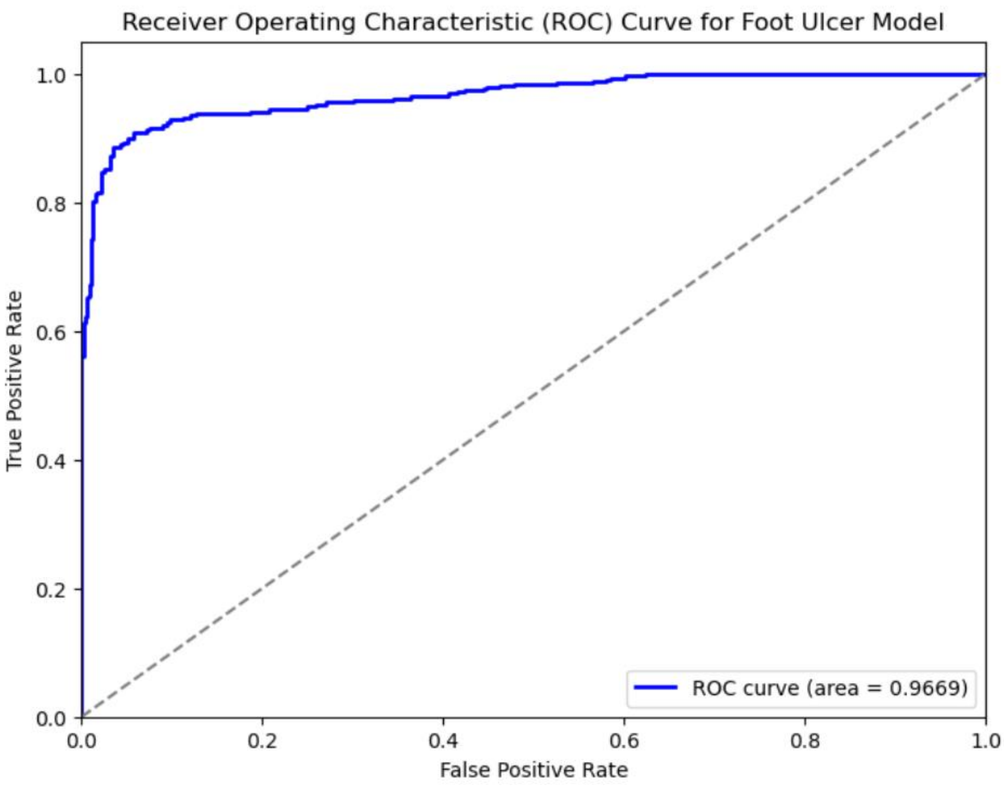


Figure 21: ROC Curve of model for Diabetic Foot Ulcer Detection

Analysis:

The custom CNN-based diabetic foot ulcer detection model demonstrates strong performance, achieving high accuracy across training (92%), validation (89%), and test datasets, with stable loss trends indicating effective learning. The confusion matrix reveals a high precision (97.3%) and an F1-score of 90.6%, with minimal misclassifications, confirming the model's reliability. The accuracy trend shows consistent learning without significant overfitting. Additionally, the ROC score of 0.967 highlights excellent discrimination between ulcer and healthy skin images. Overall, this model proves to be highly effective and well-suited for real-world diabetic foot ulcer detection.

Diabetes Retinopathy Detection Using ResNet50 CNN

The system also incorporated diabetes retinopathy detection using a ResNet50 Convolutional Neural Network (CNN) model, achieving **80% Test Accuracy** in identifying different stages of retinopathy. Image preprocessing, including contrast enhancement, noise reduction, and feature extraction, significantly improved detection reliability. The model effectively classified retinal images into categories such as mild, moderate, severe, and proliferative diabetic retinopathy, enabling early diagnosis and timely medical intervention.

Figure 22: Retinopathy CNN Model Accuracy

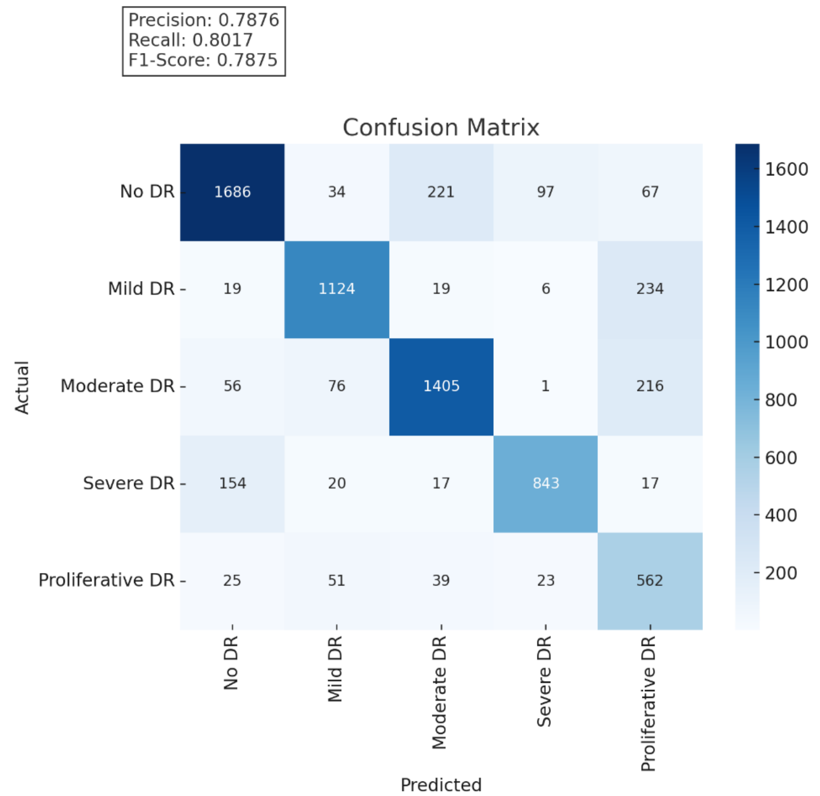


Figure 23: Confusion Matrix of Model for Retinopathy Detection

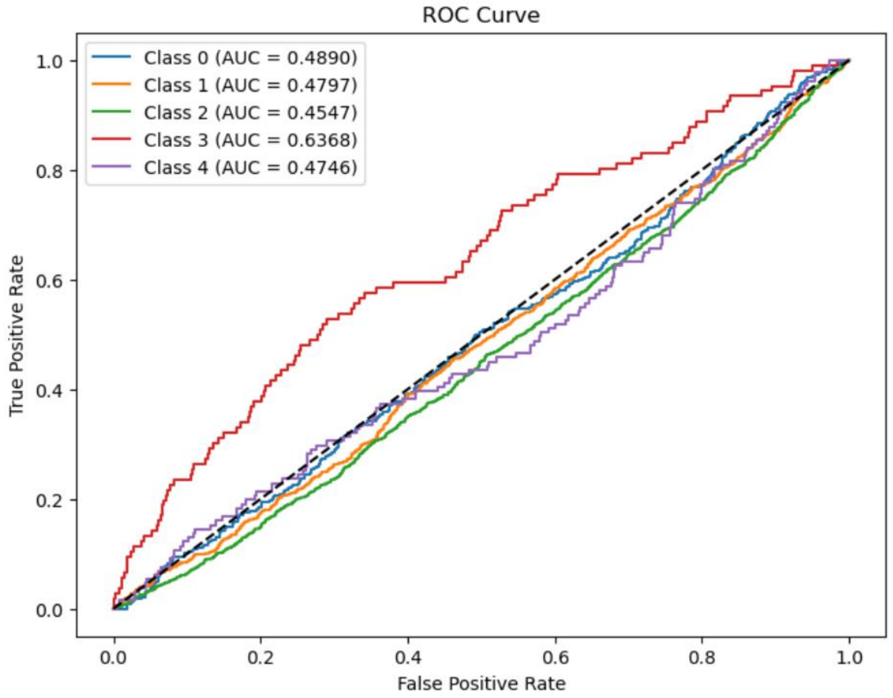


Figure 24: ROC Curve of Model for Diabetic Retinopathy Detection

Analysis:

The ResNet50-based diabetic retinopathy detection model demonstrates moderate performance, achieving an accuracy of 88% on the training set, 79% on validation, and 80% on the test dataset. However, fluctuating validation loss suggests potential overfitting, indicating room for improvement in generalization. The confusion matrix reveals a precision of 78.76% and an F1-score of 78.75%, with notable misclassifications, especially in Severe and Proliferative DR cases, affecting overall reliability. The ROC scores are relatively low, with the highest being 0.6368 for Severe DR, suggesting poor class-wise discrimination. Overall, while the model shows promise, improvements such as regularization, data augmentation, and fine-tuning are necessary to enhance real-world diabetic retinopathy detection.

Integrated Features:

- **Medication Suggestions:** Personalized medication recommendations based on user health profiles.
- **Lifestyle Recommendations:** Proper Diets and Lifestyle Recommendations dedicated to Diabetes.
- **Yoga Animations:** Interactive yoga tutorials to encourage physical activity as a supportive lifestyle intervention.
- **Fitness Tracker:** A comprehensive tool to monitor physical activity and track health metrics.
- **Risk Calculator:** A tool for calculating the risk of Diabetes.
- **Food Nutrition and Exercise Suggestions:** A feature that calculates potential nutrition of a food and recommend corresponding exercises to burn calories.
- **Blogs:** Informative articles on diabetes management, healthy diets, personal experiences and wellness tips.
- **Chatbot:** An AI-powered conversational assistant to answer user queries and provide instant guidance.
- **Chat-with-Doctor:** Direct consultation feature allowing users to connect with healthcare professionals for personalized medical advice.

CONCLUSION

The development of Our System successfully addressed critical challenges in diabetes prediction and complication management. By integrating machine learning techniques such as SVM for diabetes prediction, CNN for ulcer detection, and CNN for diabetic retinopathy detection, the platform provided a reliable and user-friendly solution for healthcare providers and individuals alike. The inclusion of lifestyle-oriented features, such as yoga animations, fitness tracking, and blog content, ensures a holistic approach to diabetes care that extends beyond traditional medical diagnostics.

The platform is scalable for deployment in hospitals, pharmacies, health camps, and community health programs, making it especially valuable in underserved regions. Its modular design allows for future enhancements, such as the adoption of ensemble machine learning methods, multilingual support, and deeper neural network architectures.

4.2 FUTURE ENHANCEMENT

To further improve the functionality and scalability of Our System, several future enhancements are planned:

1. **Dataset Expansion**: Collect more diverse and larger medical datasets to improve prediction, ulcer and retinopathy detection accuracy.
2. **Accuracy Improvement:** Use advanced machine learning techniques and ensemble models to enhance both prediction and detection accuracy.
3. **Language Support:** Add multiple languages to make the platform accessible to a wider audience.
4. **Real Doctor for Chat-with-Doctor:** Replace Dummy Doctor with Real Doctor.

4.3 GANTT CHART

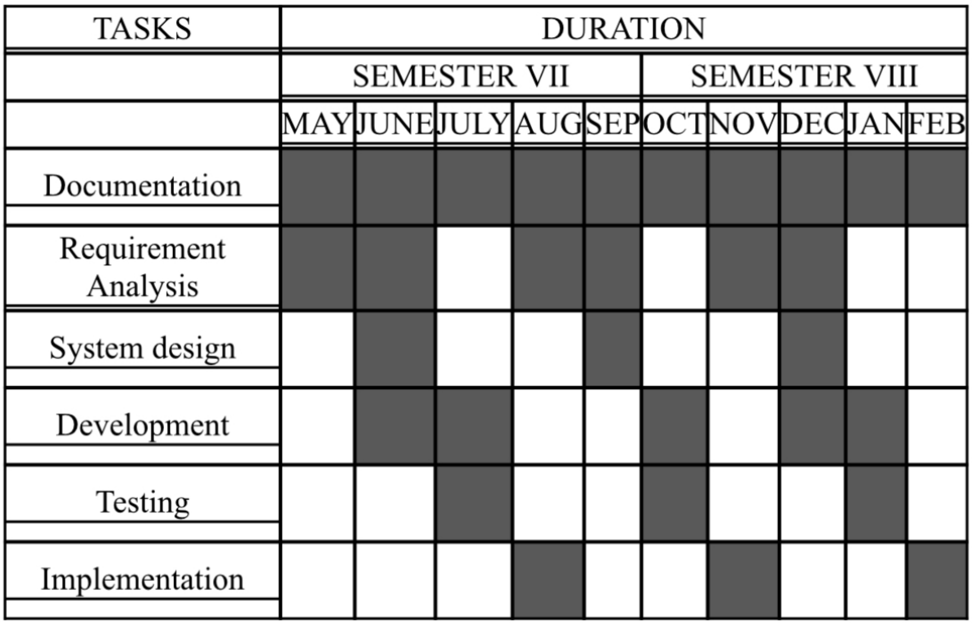


Figure 25: Gantt Chart

REFERENCES

[1] S. Shrestha, R. Khatri, and R. Paudel, "Burden of diabetes and prediabetes in Nepal: A systematic review and meta-analysis," J. Diabetes Metab. Disord., vol. 19, no. 4, pp., 2020. [Online]. Available: https://link.springer.com/article/10.1007/s13300-020-00884-0/figures/4

[2] A. Fleck, "Diabetes cases are climbing," Statista, 2024. [Online]. Available: https://www.statista.com/chart/23491/share-of-adults-with-diabetes-world-region/

[3] Healthline, "Diabetes," 2023.

[4] E. Feldman, "New Horizons in Diabetic Neurpathy:Mechanism,Bioener-getics, and Pain Science," vol. 93, no. 6, pp. 1296-1313, 2017.

[5] K. E. R. O. a. M. Orlando I., "Application of Machine Learning Models for Early Prediction," 2023.

[6] U. I. U. I. Essien V., "A Bio-Informatics System for Intelligent Classification of Severity Index of Hypertension. International Journal of Innovative Research in Sciences and Engineering Studies.," pp. 1-8, 2021.

[7] G. Y. K. H. Z. C. &. G. Q. Pei D., "Accurate and rapid screening model for potential diabetes mellitus," in *BMC Medical Informatics and Decision making*, 2019, p. 41.

[8] A. Ekong, "Evaluation of Machine Learning Techniques towards Early Prediction of Cardiovascular Diseases.," *American journal of artificial intelligence,* vol. 7, no. 1, pp. 6-16, 2023.

[9] M. R. M. J. A. B. &. A. M. M. Maniruzzaman, "Classification and prediction of diabetes diseases using machine learning paradigm," in *Health information science and system*, 2020, pp. 1-14.

[10] A. E. B. &. E. A. Ekong, "Supervised Machine Learning Model For Effective Classification of Patients With Covid-19 Symptoms Based On Bayesian Belief Network,," *Researchers Journal of Science and Technology, 2,* pp. 27-33, 2022.

[11] L. G. X. A. H. T. Q. K. A. &. A. Q. H. Miao, "Using Machine learning to predict the future development of disease," in *International Conferenceon UK-China Emerging Technologies(UCET)*, 2020.

[12] H. &. O. A. Alshari, "Machine Learning Model to Diagnose Diabetes Type 2 Based on Health Behavior," *Gazi University Journal of Science35 ,* pp. 834-852, 2022.

[13] U. R. V. J. Prasannavenkatesan T., "Diagnosis and Classification of the Diabetes Using machine learning algorithm," *computer science(4).*

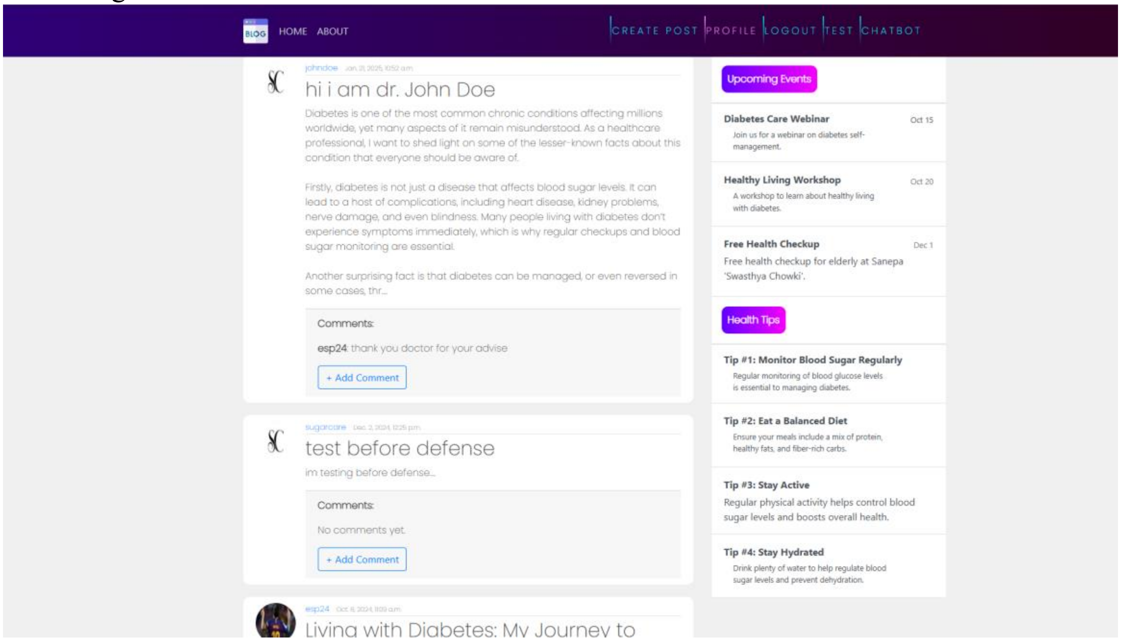
[14] N. U. a. R. K., "Classification of Diabetes using machine learning," in *International conference on computational performance evaluation* , North-Eastern Hill University, Shillong, Meghalaya, India. , 2021.

[14] A. Mahmood, V. Pandey, R. Raj, and G. S. Mishra, "Detection of phishing sites using machine learning techniques," ResearchGate, 2024. [Online].

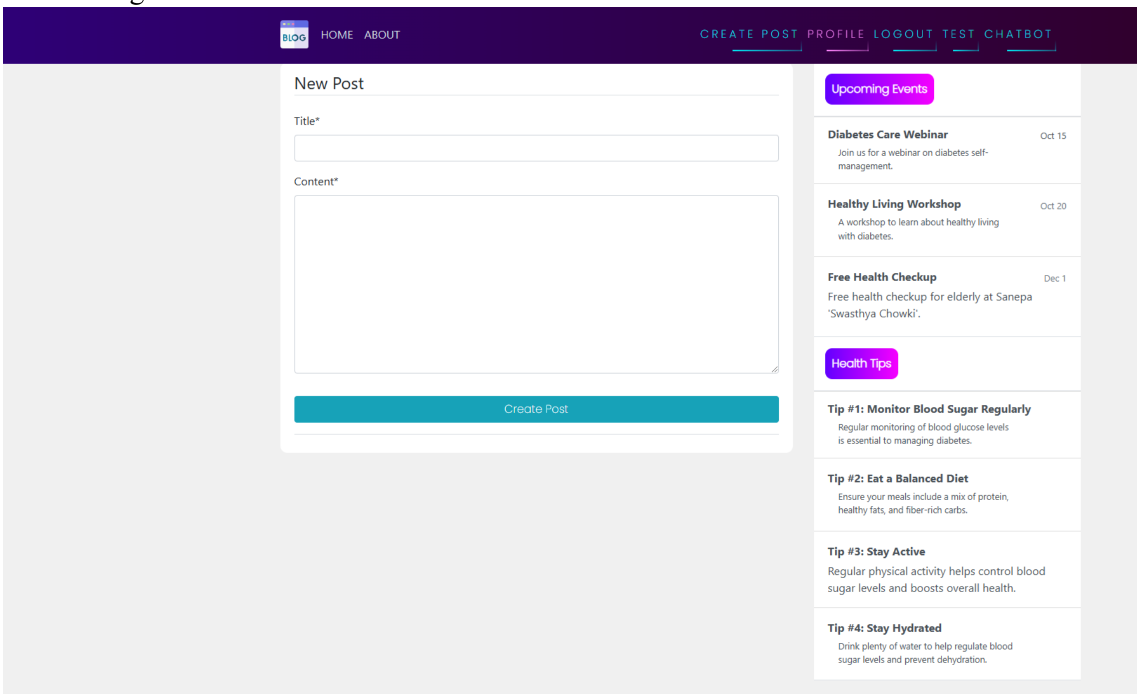
[15] D. Phung and J. Rhee, "High-accuracy ensemble CNN model for cloud classification," ResearchGate, 2019. [Online].

SCREENSHOT

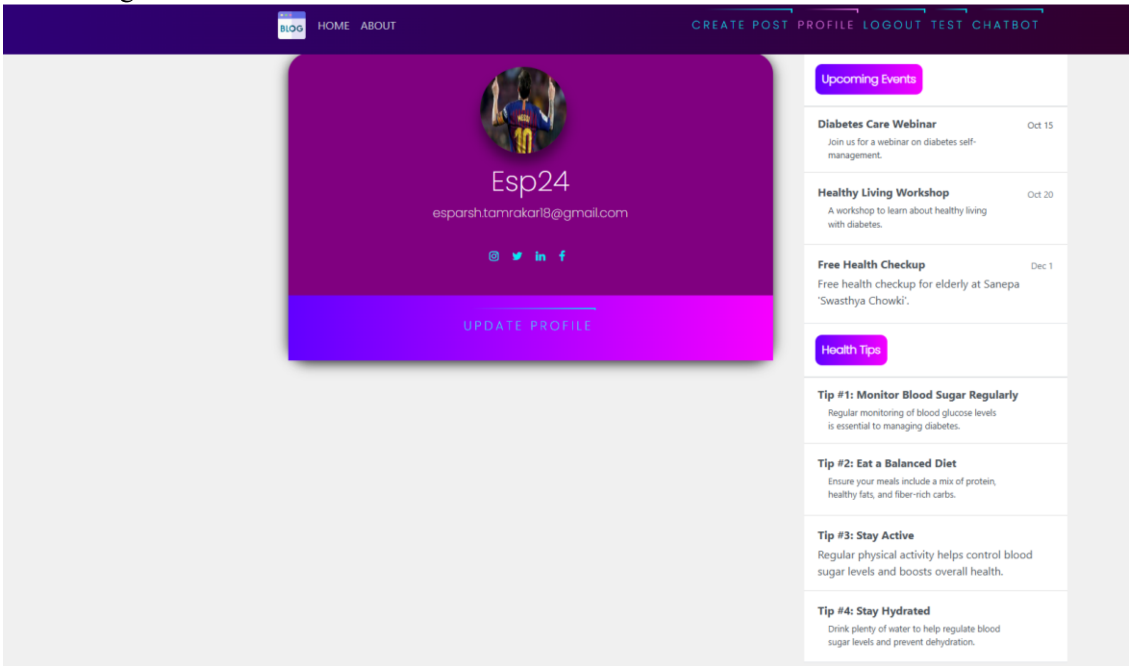
Home Page:



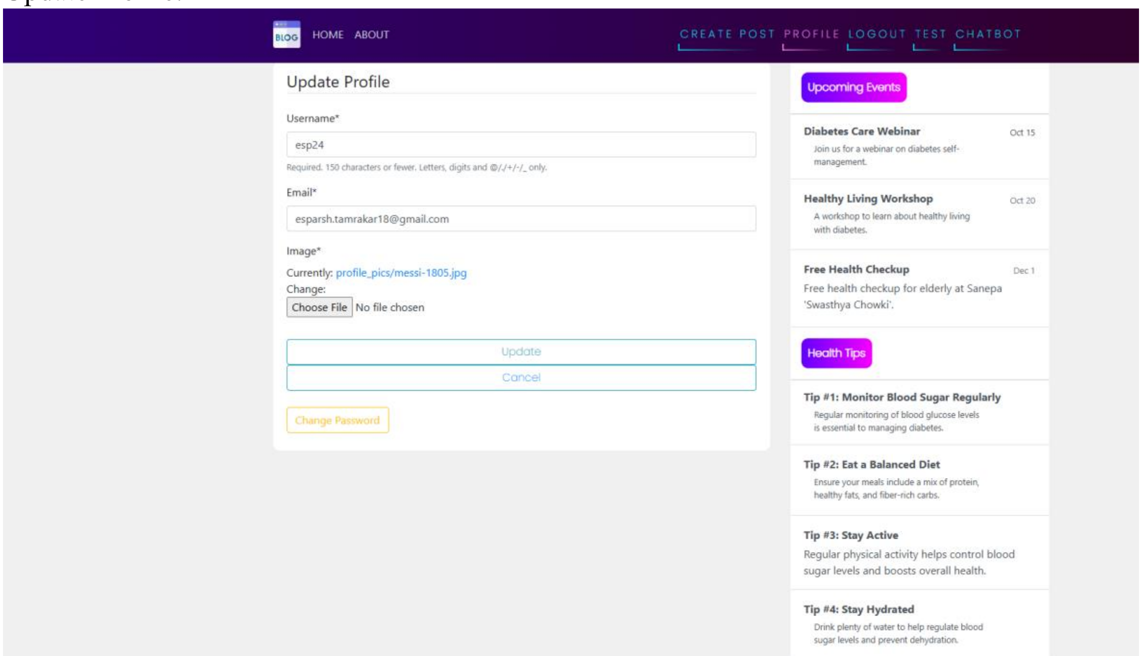
Create Blog:



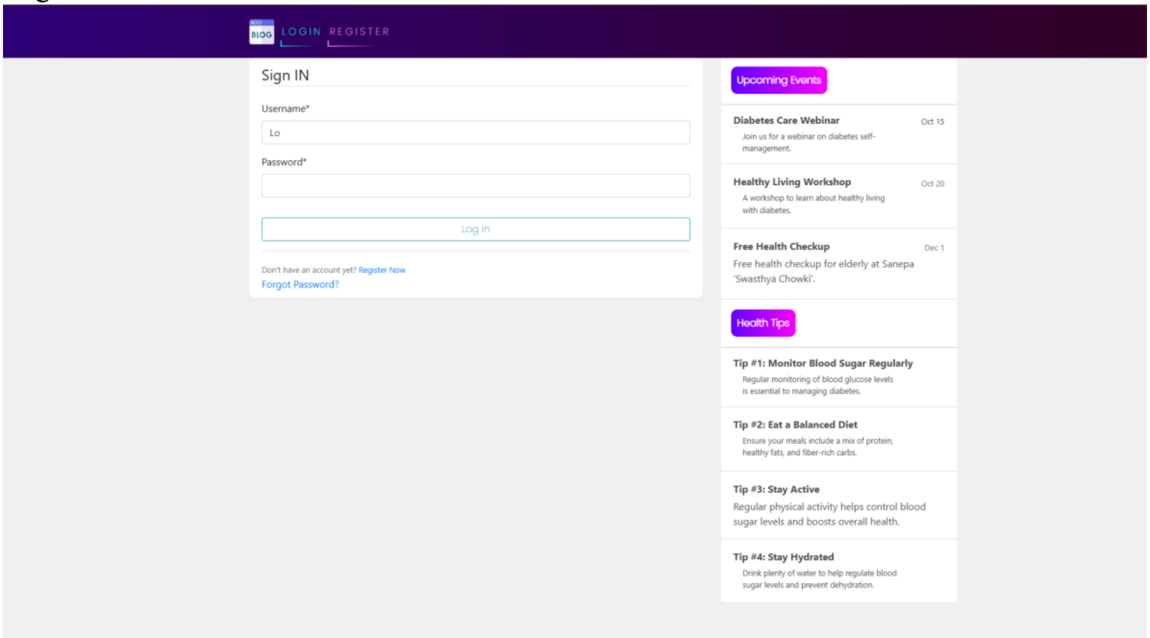
Profile Page:



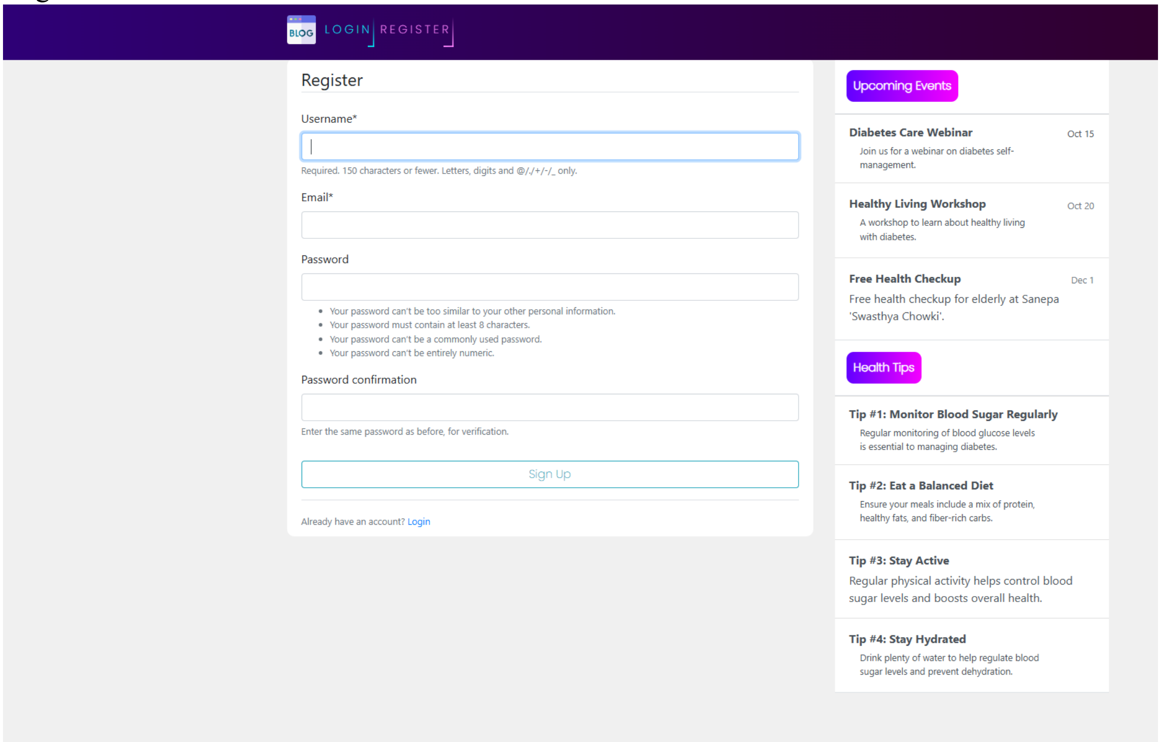
Update Profile:



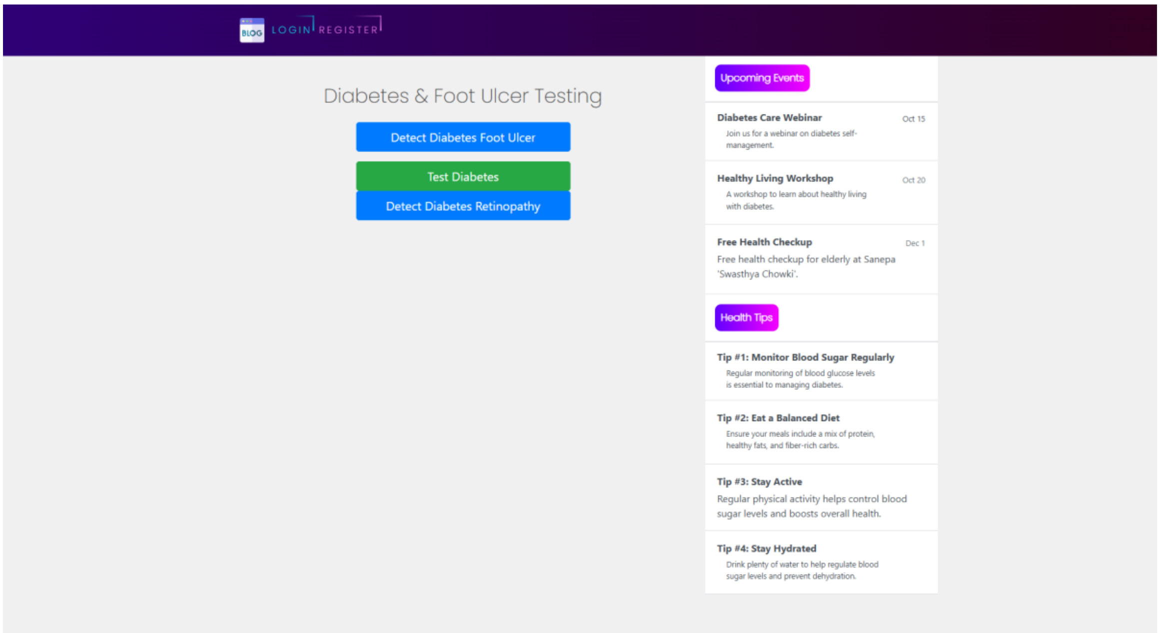
Login:



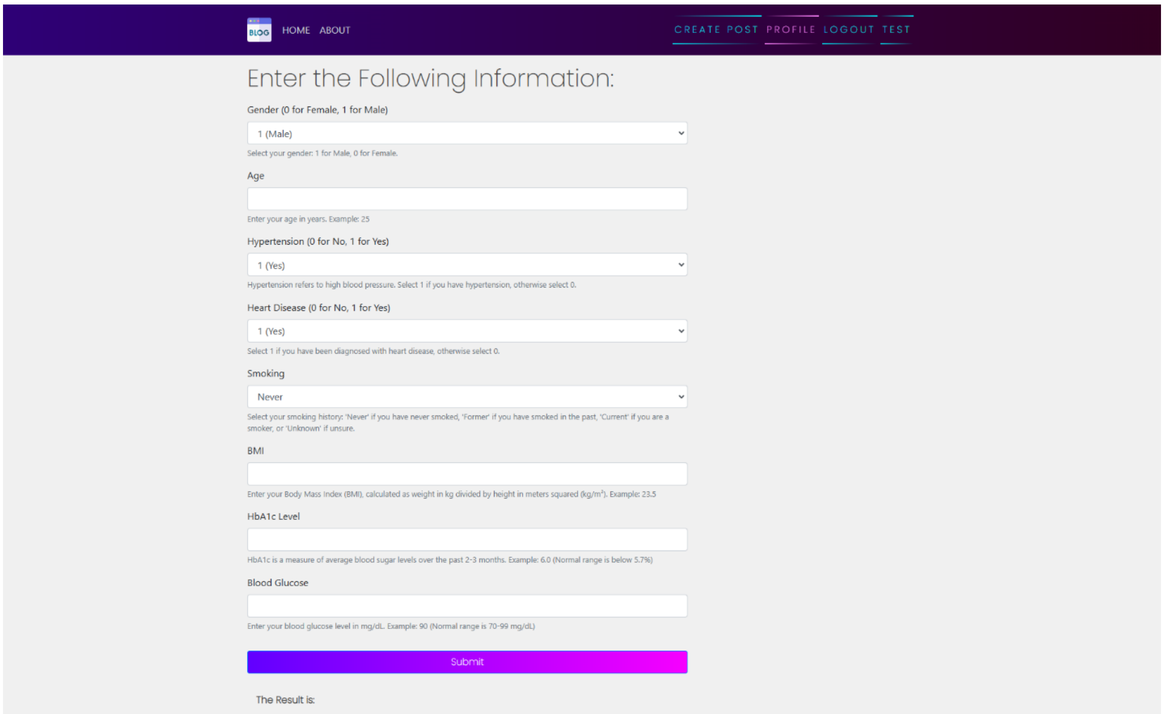
Register:



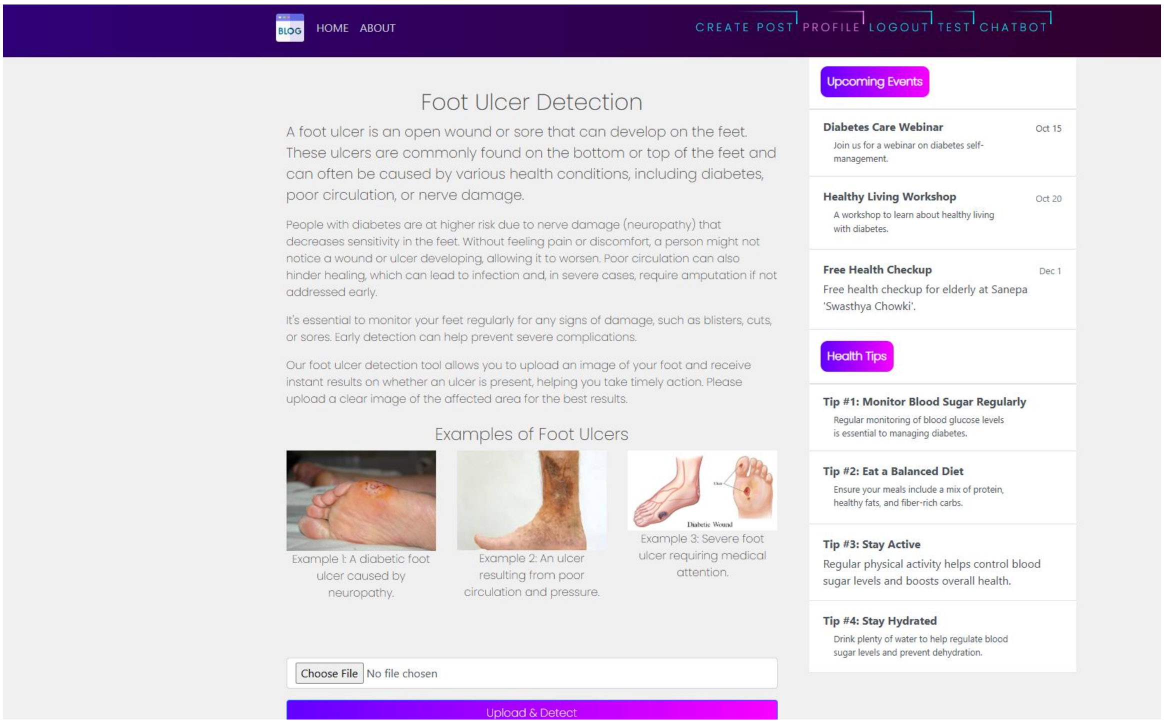
Test Selection:



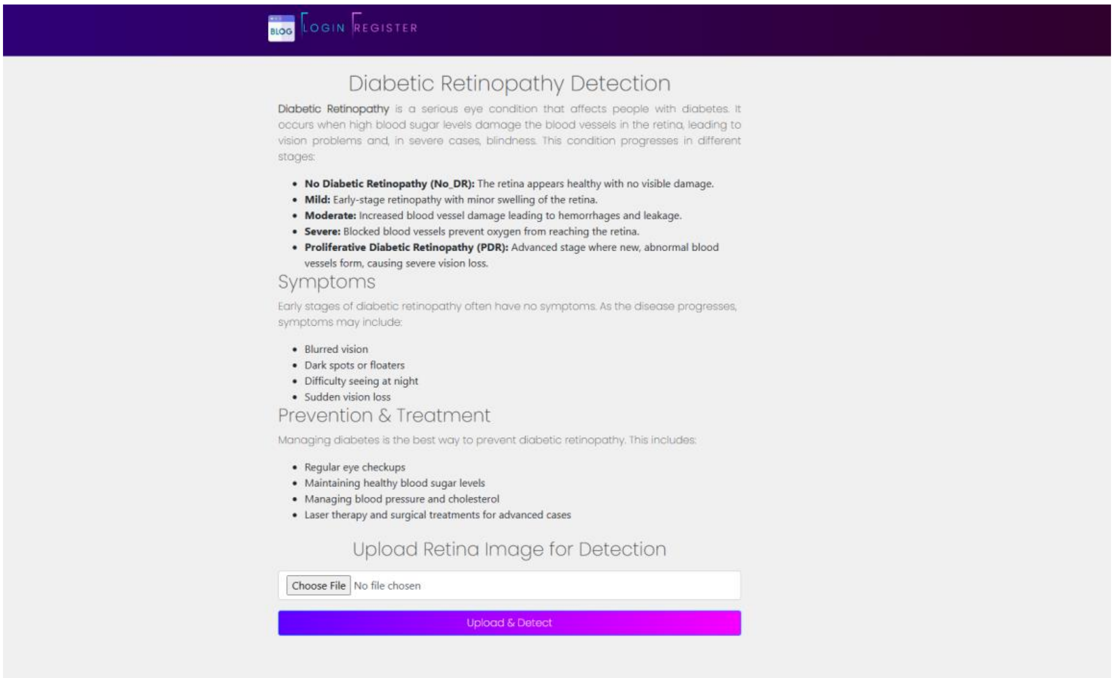
Diabetes Test:



Foot Ulcer Test:

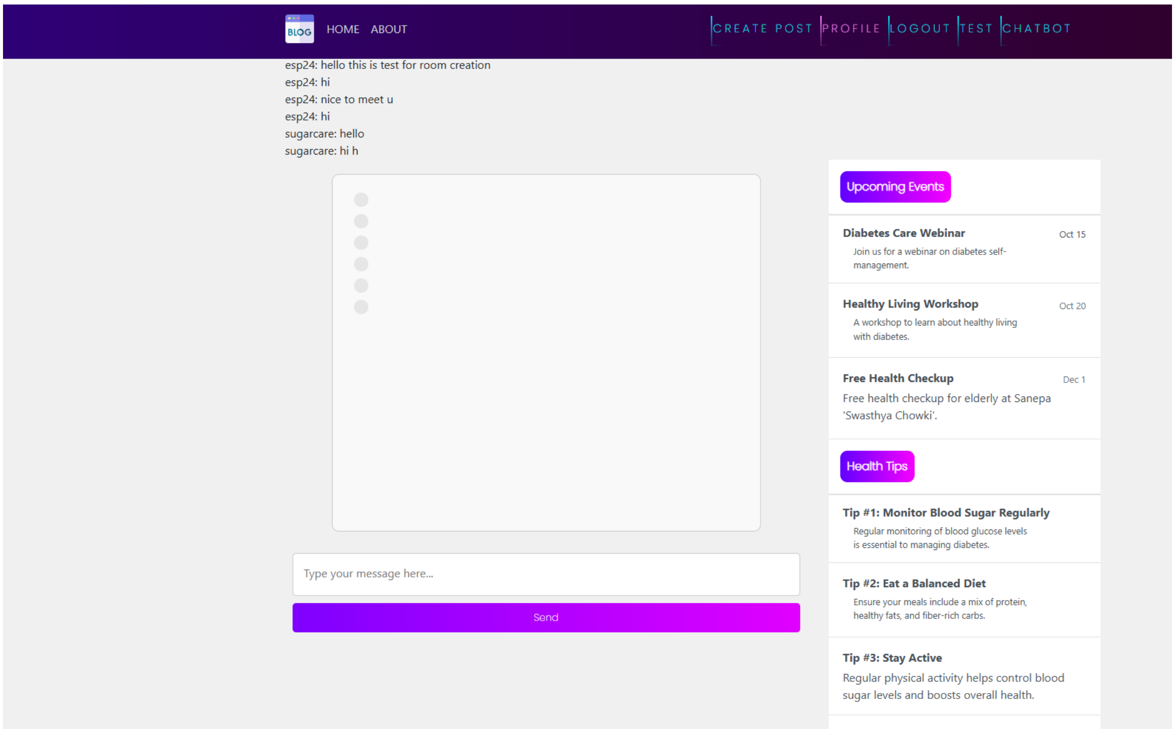


Retinopathy Test:

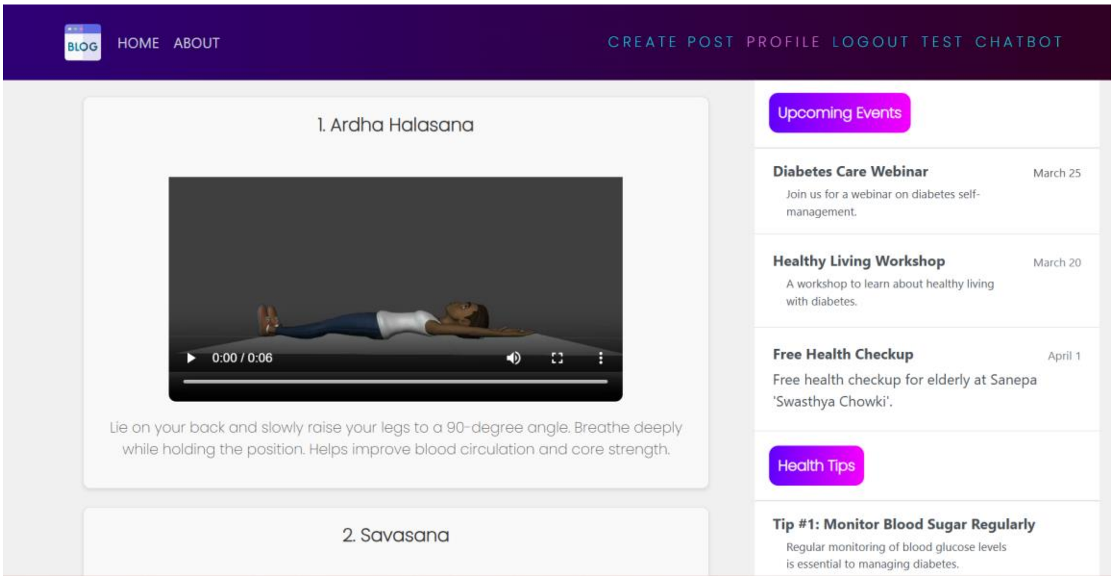


Chatbot:

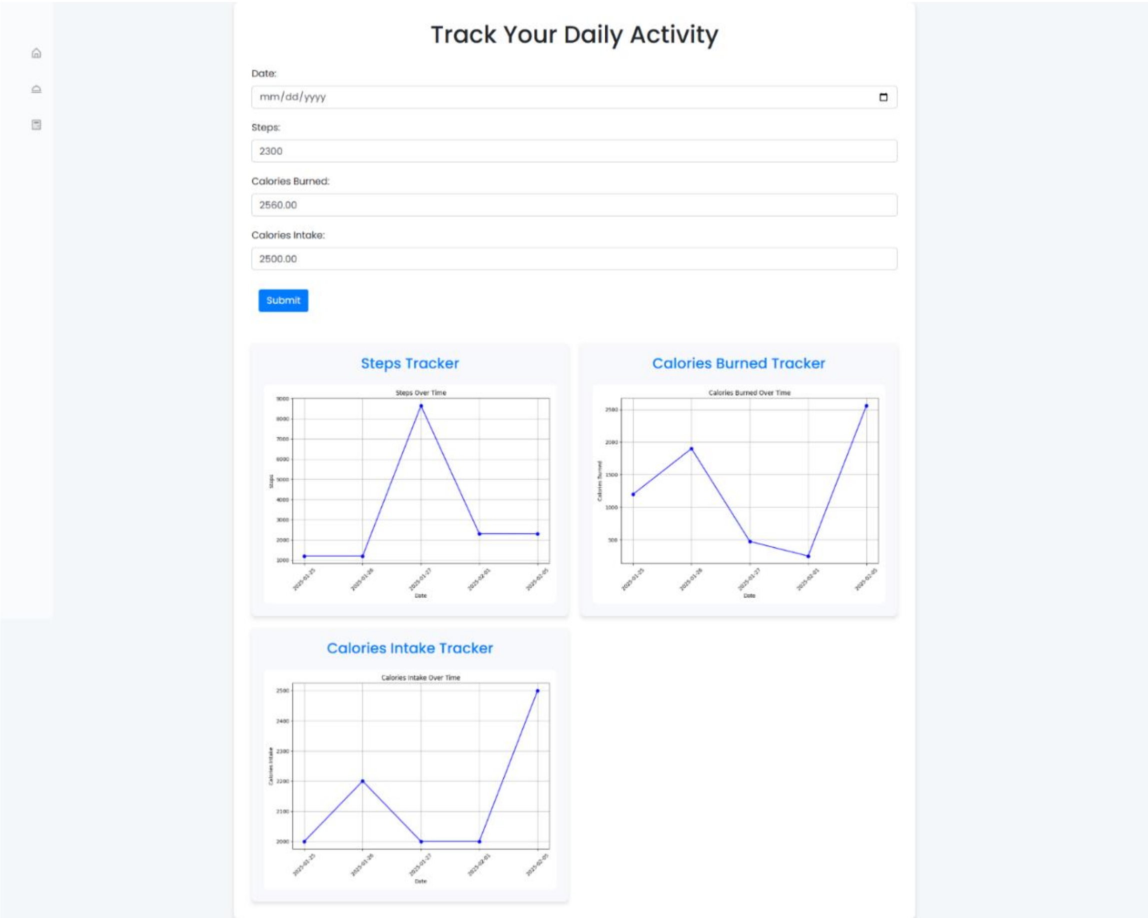
Chat-with-Doctor:



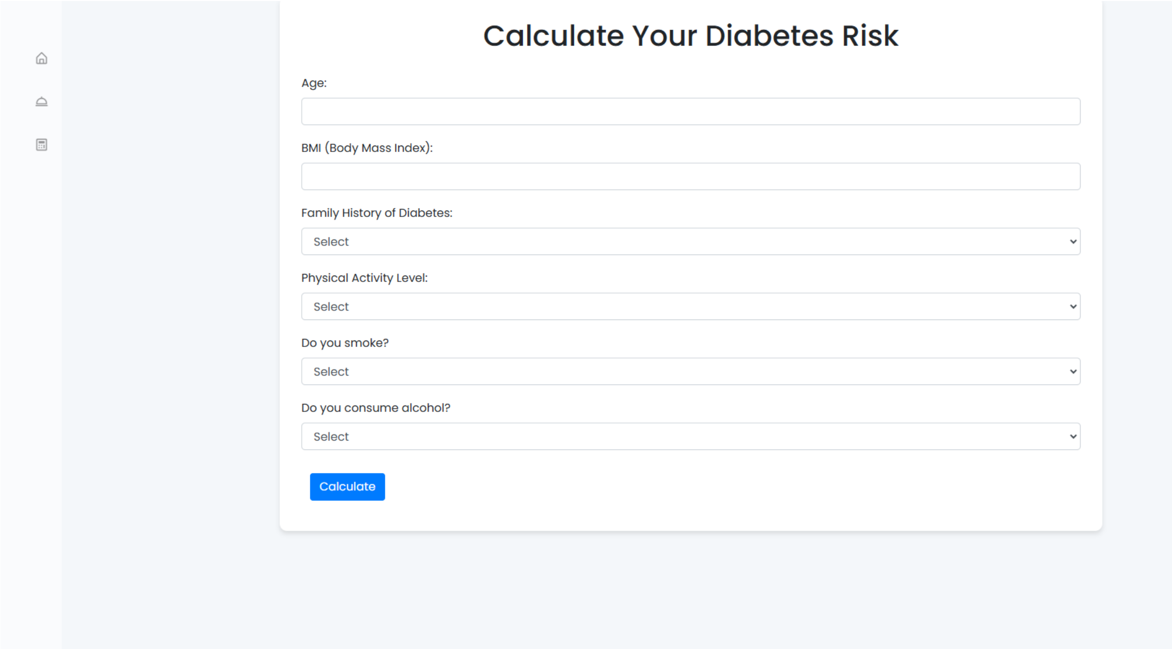
Yoga Animations:



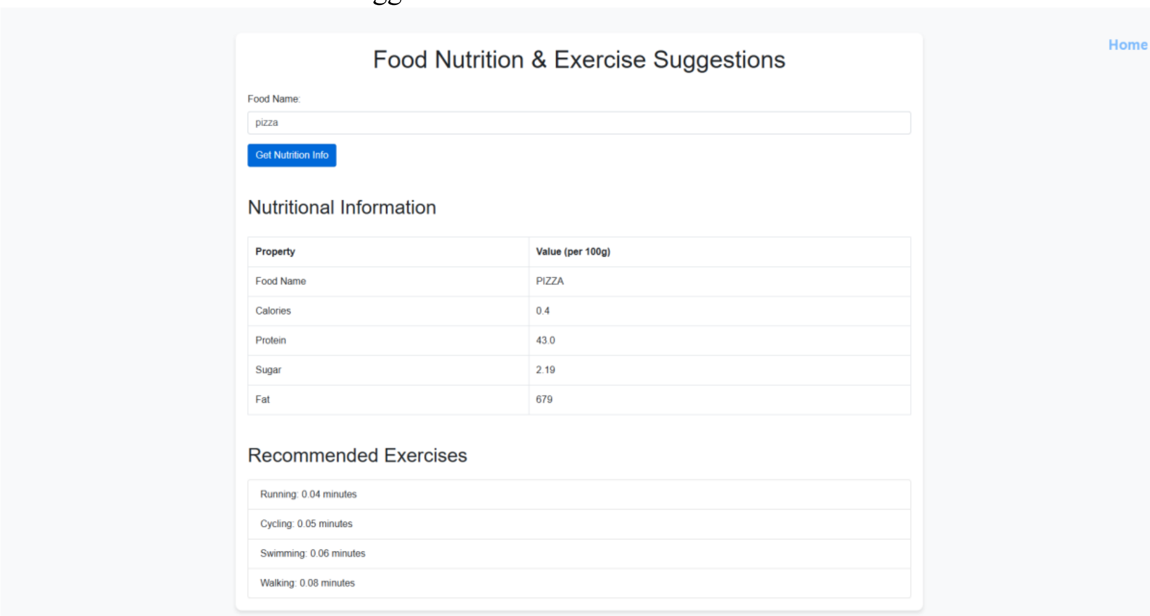
Fitness Tracker:



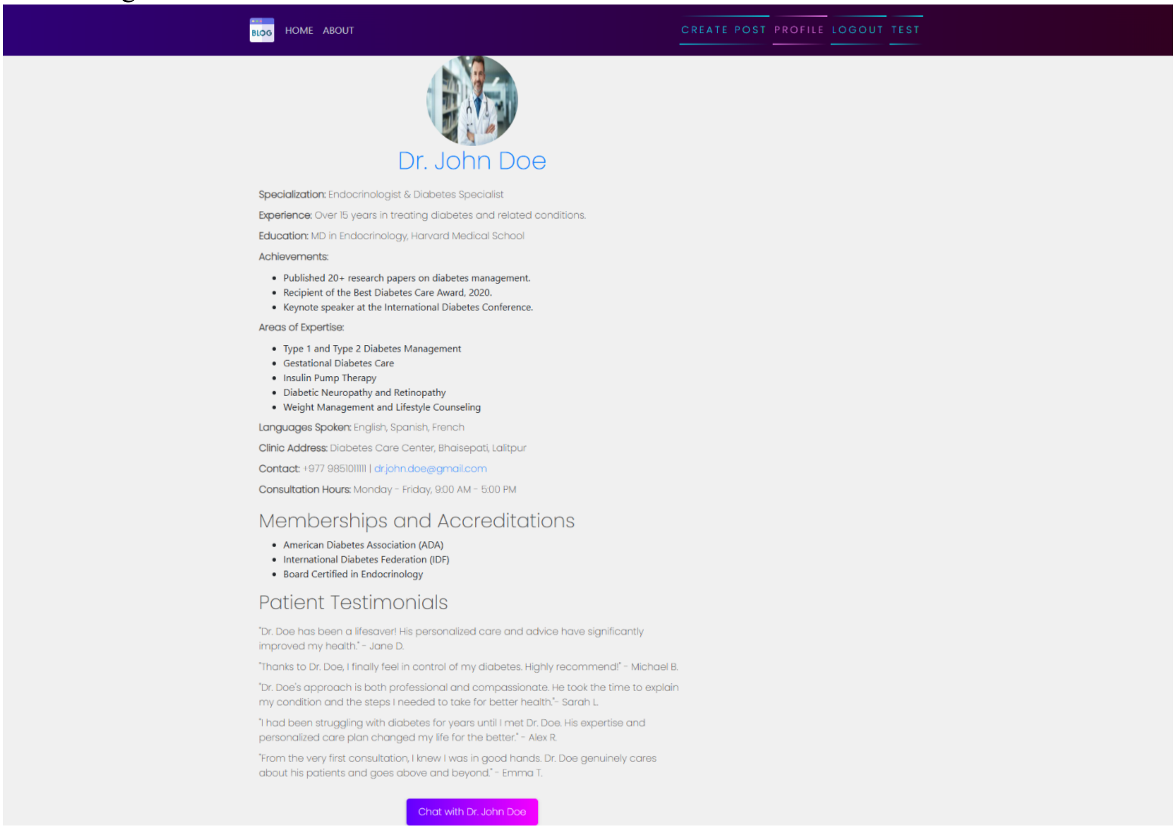
Risk Calculator:



Food Nutrition and Exercise Suggestions:



Doctor Page:


BLOG
HOME
ABOUT
CREATE POST
PROFILE
LOGOUT
TEST
Dr. John Doe
Specialization: Endocrinologist & Diabetes Specialist
Experience: Over 15 years in treating diabetes and related conditions.
Education: MD in Endocrinology, Harvard Medical School
Achievements:
Published 20+ research papers on diabetes management.
Recipient of the Best Diabetes Care Award, 2020.
Keynote speaker at the International Diabetes Conference.
Areas of Expertise:
Type 1 and Type 2 Diabetes Management
Gestational Diabetes Care
Insulin Pump Therapy
Diabetic Neuropathy and Retinopathy
Weight Management and Lifestyle Counseling
Languages Spoken: English, Spanish, French
Clinic Address: Diabetes Care Center, Bhaisepati, Lalitpur
Contact: +977 9851011111 | dr.john.doe@gmail.com
Consultation Hours: Monday - Friday, 9:00 AM - 5:00 PM
Memberships and Accreditations
American Diabetes Association (ADA)
International Diabetes Federation (IDF)
Board Certified in Endocrinology
Patient Testimonials
"Dr. Doe has been a lifesaver! His personalized care and advice have significantly improved my health." - Jane D.
"Thanks to Dr. Doe, I finally feel in control of my diabetes. Highly recommend!" - Michael B.
"Dr. Doe's approach is both professional and compassionate. He took the time to explain my condition and the steps I needed to take for better health."- Sarah L.
"I had been struggling with diabetes for years until I met Dr. Doe. His expertise and personalized care plan changed my life for the better." - Alex R.
"From the very first consultation, I knew I was in good hands. Dr. Doe genuinely cares about his patients and goes above and beyond." - Emma T.
Chat with Dr. John Doe

Medicine Suggestions: