DIABETES PREDICTION USING MACHINE LEARNING

Project work submitted in partial fulfillment of the

requirements for the degree of

MASTER OF SCIENCE

IN

COMPUTER SCIENCE

to the

Annamalai University, Chidambaram

By

SRINIVAS. K

(Register Number: 23131056504321022)

Under the supervision of

Dr. A. MURUGAN, M.C.A., M. Phil., Ph.D.,

(P.G Department of Computer Science)

MARCH- 2025

PERIYAR ARTS COLLEGE, CUDDALORE – 607001

(Affiliated to Annamalai University, Chidambaram)

PROJECT WORK

DIABETES PREDICTION USING MACHINE LEARNING

is a Bonafide Work

Done by

SRINIVAS. K

(Register Number: 23131056504321022)

Submitted to Annamalai University, Chidambaram

In partial fulfillment of the requirements for the award of the

Degree of

Master of Science

In

Computer Science

Under my guidance and supervision

INTERNAL GUIDE HEAD OF THE DEPARTMENT

Submitted for the Viva-Voice Examination held on ______________________

Internal Examiner External Examiner

ACKNOWLEGEMENT

I take this privilege to express few words of gratitude and respect to all those who helped me in the successful completion of this project.

First and foremost, I thank the Almighty for showering his blessing upon me to complete this project.

I would like to express my thanks to our Dr. R. Rajendiran M. A., M. Phil., Ph. D, Principal, Periyar Arts College, Cuddalore, who gave me the golden opportunity.

I would like to express my sincere thanks and gratitude to **Dr. K. GEETHA, M. Sc., M. Phil., Ph.D.,** Head of PG Department Of Computer Science, Periyar Arts College, for her constant support.

I would like to express my sincere thanks to my Internal Guide **Dr. A. Murugan M. C. A., M. Phil., Ph. D.**, Assistant Professor in PG Department Of Computer Science, Periyar Arts College, for his valuable guidance provided during the development of my project.

I would like to express my special gratitude and thanks to all the staff members of Computer Science Department for supporting towards my project.

Finally, yet importantly, I would like to express my heartfelt thanks to my beloved parents, my friends/classmates for their encouragement and co-operation for the successful completion of my project.

SRINIVAS. K

(23131056504321022)

ABSTRACT

Diabetes Mellitus is among critical diseases and lots of people are suffering from this disease. Age, obesity, lack of exercise, hereditary diabetes, living style, bad diet, high blood pressure, etc. can cause Diabetes Mellitus. People having diabetes have high risk of diseases like heart disease, kidney disease, stroke, eye problem, nerve damage, etc. Current practice in hospital is to collect required information for diabetes diagnosis through various tests and appropriate treatment is provided based on diagnosis. Big Data Analytics plays an significant role in healthcare industries. Healthcare industries have large volume databases. Using big data analytics one can study huge datasets and find hidden information, hidden patterns to discover knowledge from the data and predict outcomes accordingly. In existing method, the classification and prediction accuracy is not so high. In this paper, we have proposed a diabetes prediction model for better classification of diabetes which includes few external factors responsible for diabetes along with regular factors like Glucose, BMI, Age, Insulin, etc. Classification accuracy is boosted with new dataset compared to existing dataset. Further with imposed a pipeline model for diabetes prediction intended towards improving the accuracy of classification.

TABLE OF CONTENTS

1. INTRODUCTION TO THE PROJECT

The field of medicine has been quick to embrace the idea of “machine learning,” which stands for “machine learning.” The predictions made by the research community and the analysis of medical data sets contribute to the prevention of disease by indicating the most effective treatments and preventative measures. In the field of machine learning, algorithms that are able to assist in decision making and prediction. In addition, we investigate a wide variety of applications of machine learning in the area of medicine, with a special emphasis on the use of machine learning for diabetes prediction. Diabetes is one of the illnesses that are spreading at the quickest rate throughout the globe, and it is imperative that it be constantly managed at all times. In order to investigate this, we investigate a wide variety of machine learning algorithms that have the potential to assist in early sickness prediction. This article delves into a wide range of subjects pertaining to machine learning, including the several classifications of algorithms that provide assistance with decision making and prediction. People are able to prevent becoming ill by taking the appropriate care and precautions thanks to the forecasts and analyses provided by the scientific community for medical datasets. Explore the many different uses of machine learning in the medical field, with a particular emphasis on utilizing machine learning to predict diabetes. This article delves into a wide range of subjects pertaining to machine learning, including the several classifications of algorithms that provide assistance with decision making and prediction. People are able to prevent becoming ill by taking the appropriate care and precautions thanks to the forecasts and analyses provided by the scientific community for medical datasets. Explore the many uses of machine learning in medicine, with a particular focus on making predictions about diabetes using machine learning. This article examines a variety of concerns pertaining to machine learning, such as the several classifications of algorithms that might be of assistance in decision making and prediction. Patients are given assistance in adopting the appropriate preventative measures and treatment thanks to the scientific community’s forecasts and analyses of medical datasets, which help patients, avoid becoming unwell. Explore the many uses of machine learning in medicine, with a particular focus on making predictions about diabetes using machine learning. Because of this, there is a higher concentration of glucose in the urine. Diabetes may lead to organ failure, cardiovascular disease, and disturbances in a variety of physiological systems if it is not well managed. The World Health Organization (WHO) identifies diabetes as one of the four most significant non-communicable diseases (NCDs) that are now having an effect on the global population (World Health Day, 2016). The most recent results from the WHO are quite concerning. As was mentioned before, diabetes may result in a wide range of major problems with the cardiovascular system. According to the World Health Organization (WHO), diabetes and cardiovascular illnesses would be responsible for the deaths of 3.7 million people before they reach the age of 70. A blood glucose level that is not under control is the primary contributor to diabetes. Diabetes is a serious illness that affects a significant portion of the world's population. Diabetes may be brought on by a variety of factors, including advancing age, being overweight, experiencing a rapid loss of weight, overeating (polyphagia), irritability, and muscular stiffness, amongst others. We proposed a model for the prediction of diabetes as a means of improving the classification of diabetes. This model combines a number of diabetes-specific exterior characteristics with regular aspects such as delayed healing, partial paresis, irritation, itching, and visual blurring, among other things. People would be able to take better care of themselves if diabetes could be foreseen. Diabetes develops when the body is unable to produce an adequate amount of insulin. Diabetes affects around 422 million people all over the globe, most significantly in low- and middle-income nations, and it is anticipated that this number will rise to 490 billion by the year 2030. The completion of this task has the potential to save a significant number of lives.

1.2 SYSTEM SPECIFICATION

1.2.1 HARDWARE SPECIFICATION

Processor	:	AMD PRO A4-4350B R4, 5 COMPUTE CORES 2C+3G
Speed	:	2.50 GHz
RAM	:	4.00 GB (3.84 GB usable)
Device Name	:	DESKTOP-D3D34DG
System Type	:	64-bit operating system, x64-based processor
Device ID	:	A346BBF4-0854-4B13-B78D-98B5965B8A83

1.2.1 SOFTWARE SPECIFICATION

Operating System	**:** Windows (64-bit) (Likely Windows 10 or 11)
Product ID	**:** 00379-20000-00001-AAOEM (OEM Windows license)
Python Version	**:** Python 3.12
Flask Version	**:** Flask 3.1.0
Database	**:** MySQL

2. SYSTEM STUDY

2.1 EXISTING SYSTEM

The existing system of diabetes prediction primarily depends on traditional diagnostic approaches, such as blood tests, patient history analysis, and clinical observations. These methods, while effective to some extent, often lack the ability to provide early detection or predict potential risks before symptoms become apparent. Additionally, they rely heavily on manual assessments by healthcare professionals, which can lead to delays in diagnosis and increased chances of human error. Without integrating advanced technologies like machine learning and artificial intelligence, these conventional methods may fail to identify complex patterns in patient data that could enhance prediction accuracy. As a result, there is a growing need to adopt intelligent systems that can analyze vast amounts of medical data, provide real-time insights, and improve early diagnosis, ultimately leading to better disease management and prevention strategies.

2.1.1. DRAWBACKS

Machine learning models require large and high-quality datasets, but medical data often contains missing, inconsistent, or biased information, which can reduce prediction accuracy. The performance of the model depends on the chosen features (e.g., glucose levels, BMI, age). Incorrect or irrelevant feature selection can reduce accuracy. Training and deploying ML models require computational power. Hosting a real-time prediction system on a web application may require cloud-based solutions, increasing costs. The model requires continuous updates, retraining with new data, and performance monitoring to maintain its accuracy and relevance in changing healthcare scenarios.

By addressing these challenges through robust data handling, optimized feature selection, scalable cloud infrastructure, and regular model updates, the proposed diabetes prediction system can significantly enhance prediction accuracy, efficiency, and overall healthcare impact.

2.2 PROPOSED SYSTEM

In earlier technologies, diabetes prediction systems using machine learning were limited to displaying only the prediction result. These systems analyzed patient data and provided a simple output indicating whether the patient was at risk of diabetes. However, they lacked advanced functionalities such as patient record management and automated alerts, making them less effective for real-world healthcare applications.

In my project, I aim to enhance the system by introducing additional features that improve both functionality and patient engagement. Instead of only showing the prediction result, the system will now store and display complete patient details, including name, medical history, glucose levels, BMI, age, and other relevant health metrics. This will allow for better tracking and analysis of patient data over time. Additionally, if a patient is predicted to have diabetes, an automated email alert will be sent to their registered email address.

By implementing these improvements, my project will go beyond just predicting diabetes. It will provide a more comprehensive and proactive healthcare solution, ensuring timely alerts, better patient management, and improved medical intervention to help in early detection and treatment of diabetes.

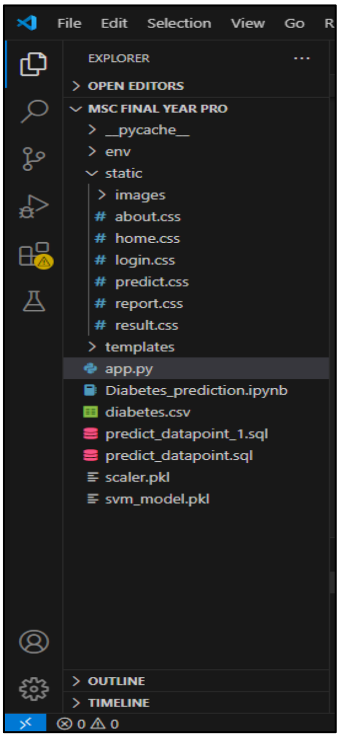
2.2.1. FEATURES

The proposed system offers several advanced features to enhance diabetes prediction and patient management. Unlike traditional machine learning-based prediction systems that only display the result, this system stores and displays complete patient details, including name, age, medical history, glucose levels, BMI, blood pressure, and other relevant health metrics. This allows for better tracking and analysis of patient data over time.

One of the key features of this system is the automatic email alert mechanism. If a patient is predicted to have diabetes, an email notification is instantly sent to their registered email address. This email includes the prediction result, precautionary health measures, and a recommendation to consult a doctor for further evaluation. Additionally, the system ensures data storage and retrieval, allowing healthcare professionals and patients to access previous records for monitoring and comparison.

3. SYSTEM DESIGN AND DEVELOPMENT

3.1 FILE DESIGN



```
MSC FINAL YEAR PRO
  > __pycache__
  > env
  v static
    > images
    # about.css
    # home.css
    # login.css
    # predict.css
    # report.css
    # result.css
  > templates
  app.py
  Diabetes_prediction.ipynb
  diabetes.csv
  predict_datapoint_1.sql
  predict_datapoint.sql
  scaler.pkl
  svm_model.pkl
```

1. Main Files

- **app.py** → This is the main backend script that handles the web application's logic. It likely includes:
  - ➢ Loading the machine learning model (svm_model.pkl)
  - ➢ Connecting to the database
  - ➢ Handling user input for diabetes prediction
  - ➢ Sending email alerts if a patient tests positive
- **Diabetes_prediction.ipynb** → This is a Jupyter Notebook file, probably used for:
  - ➢ Developing and training the ML model

- ➢ Exploratory Data Analysis (EDA)
- ➢ Evaluating the model performance before deployment

- **diabetes.csv** → This is the dataset file, which contains patient health data such as glucose levels, BMI, insulin levels, and other features used for training the machine learning model.
- **predict_datapoint.sql** → These are SQL files that might store queries related to:
  - ➢ Fetching patient details from the database
  - ➢ Saving predicted results
  - ➢ Managing patient records for future analysis
- **scaler.pkl** → A pre-trained scaler (likely StandardScaler or MinMaxScaler) used for normalizing patient data before making predictions.
- **svm_model.pkl** → A saved machine learning model (probably Support Vector Machine) that is used for diabetes prediction in the web application.

2. Static Folder (static/)

This folder contains static files (CSS, images) used for the web interface.

- **images/** → Likely contains logo, background images, or other UI elements.
- **CSS files(about.css, home.css, login.css, predict.css, report.css, result.css)** → Define the styling of different pages in the frontend.

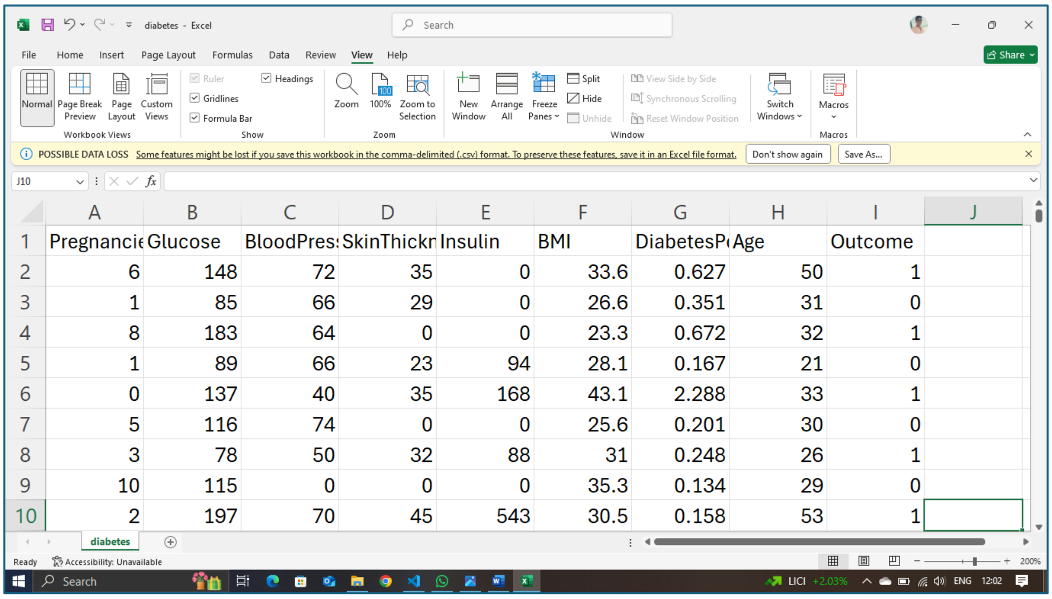
3. Templates Folder (templates/)

Although not expanded in the image, the templates/ folder likely contains HTML templates for the web application's UI. It could have:

- index.html (Homepage)
- login.html
- predict.html (Form for patient input)
- report.html (Shows patient test results)
- result.html (Displays diabetes prediction result)

3.2. INPUT DESIGN

A. Patient Data (CSV File Input)

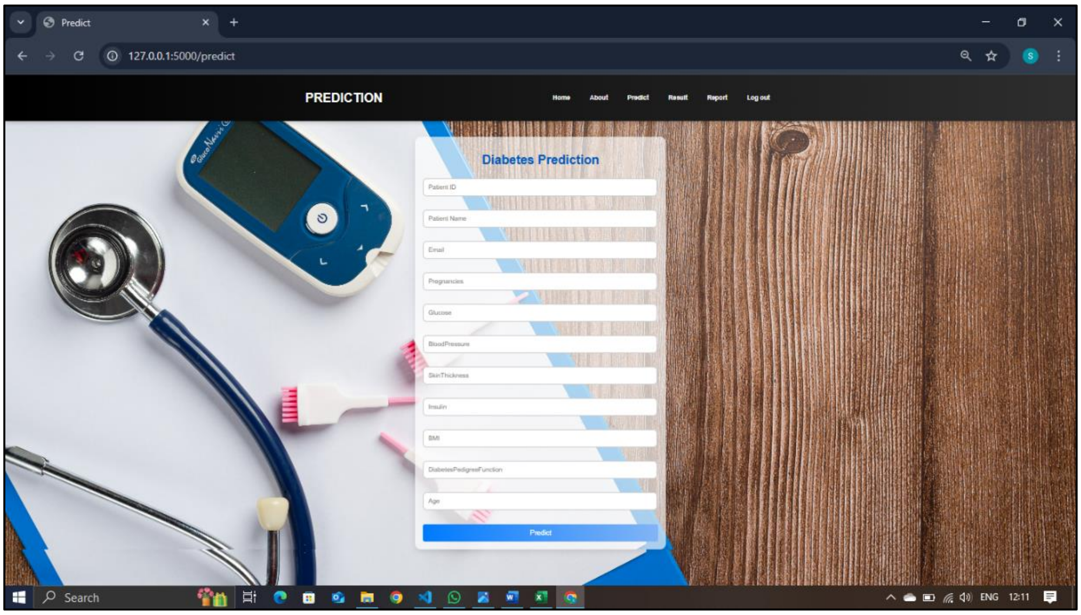


ⓘ POSSIBLE DATA LOSS Some features might be lost if you save this workbook in the comma-delimited (.csv) format. To preserve these features, save it in an Excel file format.

	A	B	C	D	E	F	G	H	I	J
1	Pregnancie	Glucose	BloodPres	SkinThickr	Insulin	BMI	DiabetesP	Age	Outcome	
2	6	148	72	35	0	33.6	0.627	50	1	
3	1	85	66	29	0	26.6	0.351	31	0	
4	8	183	64	0	0	23.3	0.672	32	1	
5	1	89	66	23	94	28.1	0.167	21	0	
6	0	137	40	35	168	43.1	2.288	33	1	
7	5	116	74	0	0	25.6	0.201	30	0	
8	3	78	50	32	88	31	0.248	26	1	
9	10	115	0	0	0	35.3	0.134	29	0	
10	2	197	70	45	543	30.5	0.158	53	1	

- The system processes patient health records stored in diabetes.csv.
- This dataset includes key medical parameters like:
  - ➢ Glucose Level
  - ➢ Blood Pressure
  - ➢ BMI (Body Mass Index)
  - ➢ Age
  - ➢ Insulin Levels
  - ➢ Diabetes Pedigree Function
  - ➢ Pregnancies
  - ➢ Skin Thickness

B. User Inputs via Web Application



- Users enter their health information manually through a web-based interface.
- The form fields may include:
  - Patient ID, Name, Email
  - Pregnancies (The number of time patient has been pregnant).
  - Glucose Level (Blood glucose concentration measured in mg/dl).
  - Blood Pressure (Diastolic blood pressure measured in mm hg).
  - Skin Thickness (Triceps skinfold thickness measured in mm).
  - Insulin (Serum insulin levels measured in µu/ml).
  - Body Mass Index (weight)
  - Diabetes Pedigree Function(DPF) A function that scores the likelihood of diabetes based on family history.
  - Age (Patient's age)

C. SQL Database Input

- SQL files (predict_datapoint.sql, predict_datapoint_1.sql) handle structured data storage.
- The system retrieves previous patient records for analysis and comparison.

D. Model Input

- The system loads pre-trained machine learning models (svm_model.pkl) for prediction.
- Feature scaling is performed using the scaler.pkl file.

2. Input Validation and Processing

To ensure the quality and reliability of inputs, validation is applied:

1. **Data Type Validation**
   - ➢ Ensures numerical fields (Glucose, BMI, Age) contain valid numbers.
   - ➢ Prevents empty fields or incorrect entries.
2. **Range Validation**
   - ➢ Ensures health metrics fall within medically accepted limits.
   - ➢ Example: Glucose levels should be within a normal human range (e.g., 70–200 mg/dL).
3. **Missing Value Handling**
   - ➢ If an input is missing, the system either prompts the user or fills it using mean/mode imputation.
4. **File Format Validation**
   - ➢ Ensures CSV files are properly formatted before processing.

3. How the Input is Processed

- The user uploads a CSV file or manually enters details.
- The system validates and preprocesses the data.
- The machine learning model (svm_model.pkl) analyzes the data.

- The system generates a diabetes prediction result.
- If the result is positive, an alert email is sent to the patient.

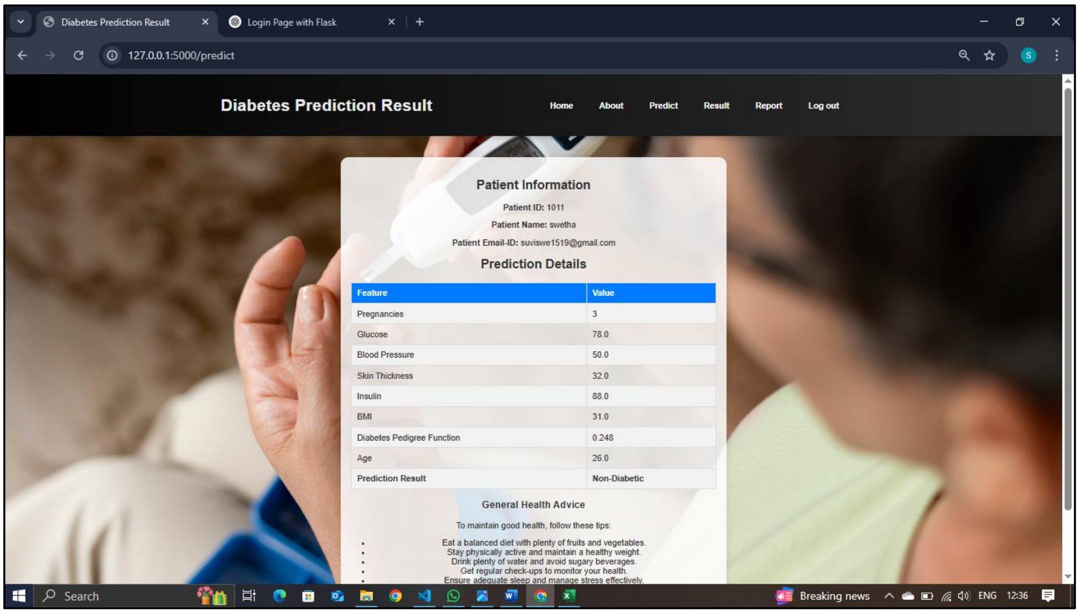
4. User Experience Considerations

- A clean and simple input form for easy data entry.
- Dropdown menus for selecting medical history instead of free text input.
- File upload feature to allow bulk processing of patient records.

3.3. OUTPUT DESIGN

1. Types of Outputs

A. Prediction Result Display (Web Interface)



Diabetes Prediction Result

Patient Information

Patient ID: 1011

Patient Name: swetha

Patient Email-ID: suviswe1519@gmail.com

Prediction Details

Feature	Value
Pregnancies	3
Glucose	78.0
Blood Pressure	50.0
Skin Thickness	32.0
Insulin	88.0
BMI	31.0
Diabetes Pedigree Function	0.248
Age	26.0
Prediction Result	**Non-Diabetic**

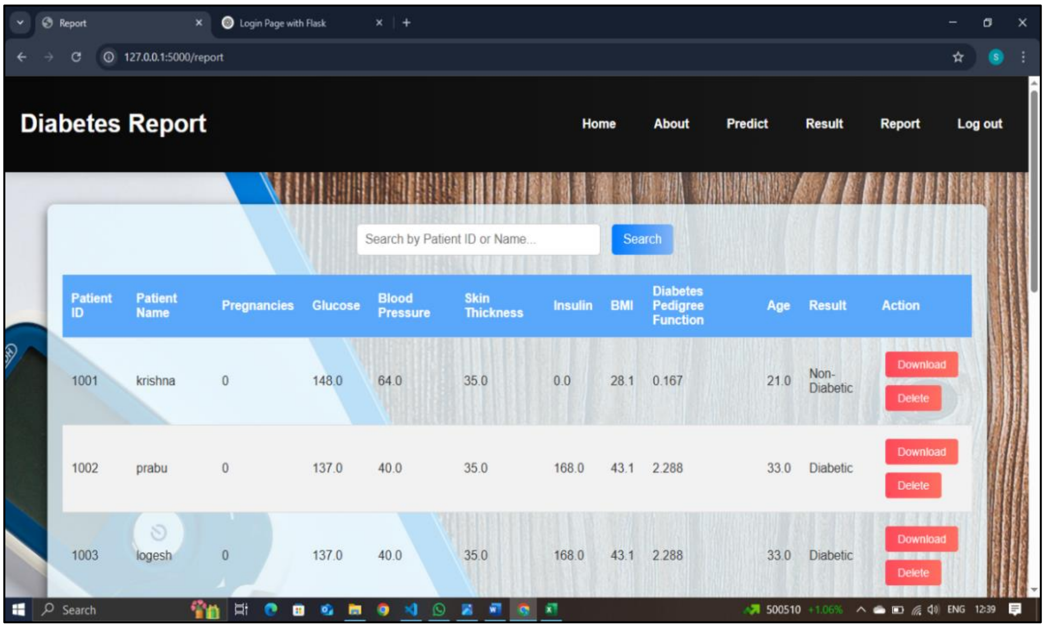
General Health Advice

To maintain good health, follow these tips:

- Eat a balanced diet with plenty of fruits and vegetables.
- Stay physically active and maintain a healthy weight.
- Drink plenty of water and avoid sugary beverages.
- Get regular check-ups to monitor your health.
- Ensure adequate sleep and manage stress effectively.

- After entering health data, the system processes it through the machine learning model (svm_model.pkl) and displays the result:
  - ➢ Positive Prediction: If the patient is at risk of diabetes.
  - ➢ Negative Prediction: If the patient is not at risk.
- The output page will also show confidence scores to indicate prediction accuracy.
- The interface uses result.css for styling to ensure clarity and readability.

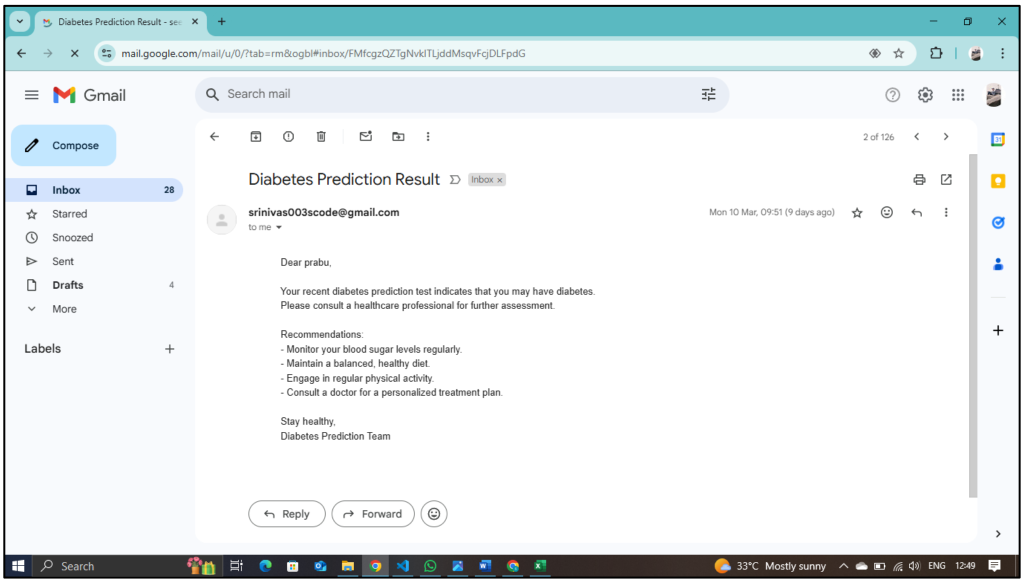
B. Patient Report



Patient ID	Patient Name	Pregnancies	Glucose	Blood Pressure	Skin Thickness	Insulin	BMI	Diabetes Pedigree Function	Age	Result	Action
1001	krishna	0	148.0	64.0	35.0	0.0	28.1	0.167	21.0	Non-Diabetic	Download Delete
1002	prabu	0	137.0	40.0	35.0	168.0	43.1	2.288	33.0	Diabetic	Download Delete
1003	logesh	0	137.0	40.0	35.0	168.0	43.1	2.288	33.0	Diabetic	Download Delete

- The system generates a detailed report that includes:
  - ➢ Patient Name & ID
  - ➢ Entered Health Metrics
  - ➢ Diabetes Prediction Result
- The report is displayed in the web app and can be downloaded as a csv file for future reference.

C. Alert System (Email Notification)


Diabetes Prediction Result - sec
mail.google.com/mail/u/0/?tab=rm&ogbl#inbox/FMfcgzQZTgNvklTLjddMsqvFcjDLFpdG
Gmail
Search mail
Compose
Inbox 28
Starred
Snoozed
Sent
Drafts 4
More
Labels
2 of 126

Diabetes Prediction Result Inbox ×

srinivas003scode@gmail.com
to me

Mon 10 Mar, 09:51 (9 days ago)

Dear prabu,

Your recent diabetes prediction test indicates that you may have diabetes.
Please consult a healthcare professional for further assessment.

Recommendations:
- Monitor your blood sugar levels regularly.
- Maintain a balanced, healthy diet.
- Engage in regular physical activity.
- Consult a doctor for a personalized treatment plan.

Stay healthy,
Diabetes Prediction Team

- If the prediction result is positive, an automated email alert is sent to the patient's registered email.
- The email contains:
    - ➢ Diagnosis Result
    - ➢ Advice on Next Steps (e.g., Consulting a Doctor)
    - ➢ Health Tips to Manage Diabetes

D. Database Output (SQL Storage)

- The system stores prediction results in the SQL database (predict_datapoint.sql).
- The stored data helps in:
    - ➢ Tracking patient history
    - ➢ Comparing previous and current results
    - ➢ Enhancing model performance through retraining

2. How the Output is Displayed

1. **Web Application Output:**
    - ➢ Displayed in a structured HTML page styled with report.css and result.css.
2. **Email Alert:**
    - ➢ Sent using a backend email service (e.g., SMTP).
3. **Data Logging:**
    - ➢ Saved in the SQL database for future reference.

3.4. DATABASE DESIGN

Database Design for Diabetes Prediction System

The database design ensures efficient storage, retrieval, and management of patient health data and prediction results. The system stores user details, health parameters, prediction outcomes, and email alerts.

The database for this project is structured using **MySQL** and consists of a schema named **prediction_data**, which contains a table named **predictions**. This table is designed to store patient details and their respective diabetes prediction results.

Table: predictions

This table holds critical patient information and their medical parameters required for diabetes prediction.

Columns:

1. **id** *(Primary Key)* – Unique identifier for each record.
2. **patient_Id** – Unique ID assigned to each patient.
3. **p_name** – Name of the patient.
4. **email** – Patient's email address (used for sending alerts).
5. **Pregnancies** – Number of times the patient has been pregnant.
6. **Glucose** – Blood glucose level.
7. **BloodPressure** – Blood pressure measurement.
8. **SkinThickness** – Skin fold thickness measurement.
9. **Insulin** – Insulin level in the body.
10. **BMI** – Body Mass Index (BMI) value.
11. **DiabetesPedigree** – Diabetes pedigree function (genetic risk factor).
12. **Age** – Age of the patient.
13. **result** – The diabetes prediction result (Positive/Negative).

3.5. SYSTEM DEVELOPMENT

1. Software Development Life Cycle (SDLC) Approach

The system is developed using the Waterfall Model, ensuring a step-by-step process from requirement gathering to deployment. The stages include:

1. Requirement Analysis – Understanding the need for diabetes prediction and additional features.
2. System Design – Creating a structured framework for the system.
3. Implementation – Developing the frontend, backend, and integrating the ML model.
4. Testing – Ensuring accuracy, performance, and security.
5. Deployment – Hosting the system for real-time usage.
6. Maintenance & Updates – Continuous monitoring and improvements.

2. Technology Stack

The project utilizes the following technologies:

- Frontend: HTML, CSS, JavaScript, Bootstrap
- Backend: Python (Flask)
- Database: MySQL
- Machine Learning: Scikit-learn (SVM Model)
- Cloud Services: SMTP for email alerts

3. System Architecture

The system follows a 3-Tier Architecture:

1. **Presentation Layer (Frontend)**
    - ➢ Provides user interaction interface.
    - ➢ Collects patient health details.
    - ➢ Displays prediction results.

2. **Application Layer (Backend & ML Model)**

   - Handles data processing and ML predictions.
   - Stores patient records.
   - Triggers email alerts.

3. **Database Layer (MySQL)**

   - Stores user details, health records, predictions, and emails.

4. Development Modules

A. User Management Module

- Allows users (doctors) to log in.
- Stores user details securely.

B. Health Data Input Module

- Doctor enter health parameters such as glucose level, BMI, blood pressure, insulin, etc.
- Data is validated before processing.

C. Machine Learning Prediction Module

- Uses an SVM model trained on diabetes datasets.
- Takes user input and predicts the probability of diabetes.
- Displays results with a confidence score.

D. Email Alert Module

- If a patient is predicted to be diabetic, an automatic alert is sent.
- Email logs are maintained for tracking.

E. Report & Visualization Module

- Displays previous predictions with graphical insights.
- Allows patients to track health trends.

5. Implementation Process

1. **Frontend Development**

   - ➢ Designed using HTML, CSS, JavaScript, and Bootstrap.
   - ➢ Ensures responsive UI for a better user experience.

2. **Backend Development**

   - ➢ Developed using Flask (Python framework).
   - ➢ Handles user requests, processes health data, and interacts with the ML model.

3. **Machine Learning Model**

   - ➢ Trained using a dataset containing patient health records.
   - ➢ Uses an SVM classifier for high accuracy.

4. **Database Integration**

   - ➢ MySQL database manages user and prediction records.
   - ➢ Ensures data integrity and security.

5. **Email Automation**

   - ➢ Configured using Python’s smtplib.
   - ➢ Sends email alerts to patients if diabetes is detected.

6. Testing & Debugging

- Unit Testing: Ensures individual components (login, data input, prediction) function correctly.
- Integration Testing: Verifies smooth interaction between frontend, backend, and database.
- Performance Testing: Tests response time and load handling.
- Security Testing: Protects user data using encryption and secure authentication.

7. Deployment & Maintenance

- The system is deployed on a local server.

- Periodic updates are done to improve the model's accuracy and add new features.
- User feedback is incorporated for continuous enhancement.

3.5.1. DESCRIPTION OF THE MODULES

1. User Management Module

Description:

This module is responsible for handling authentication and access control. In this system, only the admin has access to the web application to manage the diabetes prediction process.

Functionalities:

- Admin Login & Authentication: Secure login using a username and password.
- Role-Based Access: Only the admin can access the system to input patient data and generate predictions.
- Profile Management: The admin can update account details.
- System Control: The admin oversees data entry, prediction results, and report generation.

2. Health Data Input Module

Description:

This module collects patient health data, which is then used for diabetes prediction.

Functionalities:

- Collects important health parameters such as:
    - ➢ Blood glucose level
    - ➢ Blood pressure
    - ➢ BMI (Body Mass Index)
    - ➢ Insulin level
    - ➢ Age, gender, etc.

- Input Validation: Ensures correct and complete data entry.
- Data Preprocessing: Handles missing or incorrect values before prediction.

3. Machine Learning Prediction Module

Description:

This is the core module where the machine learning model analyzes patient data and predicts whether they are diabetic or not.

Functionalities:

- Feature Extraction: Extracts relevant features from input data.
- ML Model Integration: Uses a trained Support Vector Machine (SVM) model for prediction.
- Prediction Output: Classifies a patient as Diabetic or Non-Diabetic.
- Confidence Score: Displays the probability of the prediction.

4. Email Alert Module

Description:

This module is responsible for notifying patients via email if they are at risk of diabetes.

Functionalities:

- Automatic Email Alerts: If diabetes is detected, the system sends an email to the patient.
- SMTP Configuration: Uses Python’s smtplib to send emails.
- Customizable Messages: Emails include the prediction result and health advice.

5. Report & Visualization Module

Description:

This module provides insights into patient health trends using graphical reports.

Functionalities:

- Graphical Representation: Displays charts for glucose levels, BMI, etc.
- Historical Data Tracking: Users can view previous test results.

- Downloadable Reports: Generates PDF or CSV reports for medical consultation.

6. Database Management Module

Description:

This module is responsible for storing and managing patient records, predictions, and system logs.

Functionalities:

- MySQL Database Integration: Stores user data and prediction history.
- Data Security & Encryption: Ensures patient data confidentiality.
- Data Retrieval: Doctors can access and analyze patient records.

7. Web Interface Module

Description:

This module provides a user-friendly interface for patients and doctors to interact with the system.

Functionalities:

- Responsive Design: Works on both mobile and desktop.
- User-Friendly Dashboard: Simple navigation for patients and doctors.
- Interactive Forms: Easy data entry and real-time prediction results.

4. TESTING AND IMPLEMENTATION

Testing

Testing is a crucial phase in software development that ensures the system functions as expected, meets requirements, and is free of major bugs. The following testing strategies are applied:

1. Unit Testing

- Individual components, such as the diabetes prediction model, database connections, and email alert system, are tested separately.
- Ensures each function operates correctly before integrating it into the full system.

2. Integration Testing

- Verifies that different modules, such as data input, prediction, and email notifications, work together without issues.
- Ensures seamless interaction between the machine learning model, database, and web application.

3. Functional Testing

- Ensures that the application meets functional requirements:
  - Accepting user input
  - Predicting diabetes based on patient data
  - Sending alert emails if the result is positive
- Checks that all features operate correctly under expected conditions.

4. Performance Testing

- Evaluates the speed and responsiveness of the system.
- Ensures that the model prediction and email alert functionalities work efficiently even with large datasets.

5. Security Testing

- Ensures that the admin login system is secure against unauthorized access.
- Protects sensitive patient data by preventing SQL injection, cross-site scripting (XSS), and other attacks.

Implementation

1. System Deployment

- The web application is hosted on a local server or cloud platform (e.g., AWS, Azure, or Firebase).
- The database is set up for storing patient records and predictions.

2. Model Integration

- The trained machine learning model (SVM or another algorithm) is integrated into the web application for real-time predictions.
- The model is loaded from a saved .pkl file to avoid retraining every time.

3. Email Alert System

- When a patient is predicted as diabetic, an automated email notification is sent to the registered email ID.
- Uses SMTP protocol to send emails securely.

4. Admin Dashboard Setup

- The admin interface is designed to input patient data, view prediction results, and manage records.
- The interface includes graphs, reports, and prediction history for easy monitoring.

5. System Maintenance and Updates

- Regular updates to improve model accuracy by retraining with new data.
- Fixing bugs and ensuring smooth performance by monitoring logs and user feedback.

5. FUTURE ENHANCEMENTS

To enhance its functionality, the improved system will include features for storing and displaying all patient details, allowing the admin to manage records efficiently. Additionally, an automatic email alert system will notify patients of their health status, ensuring timely medical attention. To further improve accuracy, deep learning algorithms will be integrated, offering more precise predictions compared to traditional machine learning models. Moreover, the system will introduce user and admin profiles, enabling personalized access—patients can view their test results, while the admin can manage records and oversee the prediction process. These enhancements will make the system more efficient, accurate, and user-friendly, supporting better diabetes monitoring and management.

6. CONCLUSION

This project successfully enhances traditional diabetes prediction by incorporating patient record management, automated email alerts, and a role-based access system. By storing detailed patient information, the system enables better tracking and analysis of health conditions over time. The email alert system ensures timely communication, allowing at-risk patients to take necessary precautions and seek medical consultation. These improvements, combined with the integration of machine learning for accurate predictions, make the system a comprehensive, proactive, and user-friendly healthcare solution. Ultimately, this project contributes to early detection, improved patient care, and effective diabetes management, supporting better healthcare outcomes through technology.

7. BIBLIOGRAPHY

APPENDICES

A. DATA FLOW DIAGRAM

B. TABLE STRUCTURE

Column name	**Data type**	**Constraints**
Patient ID	INT	PRIMARY KEY
Patient Name	VARCHAR(100)	NOT NULL
Email	VARCHAR(100)	NOT NULL
Pregnancies	INT	NOT NULL
Glucose Level	FLOAT	NOT NULL
Blood Pressure	FLOAT	NOT NULL
Skin Thickness	FLOAT	NOT NULL
Insulin	FLOAT	NULL
Body Mass Index	FLOAT	NOT NULL
Diabetes Pedigree Function(DPF)	FLOAT	NULL
Age	INT	NOT NULL
Result	VARCHAR(100)	NOT NULL

C. SAMPLE CODING

D. SAMPLE INPUT

E. SAMPLE OUTPUT