DIABETES PREDICTION SYSTEM

WHOA!

For More Imformation

- Visit : GitHub Repository
- Contact Us : **se.abdelrahman968@gmail.com**

TABLE OF CONTENTS

T O

01

PROBLEM STATEMENT

Understanding the Importance of Early Diabetes Detection

T 1

INTRODUCTION

Diabetes is a chronic disease that occurs when the body is unable to produce or effectively use insulin, leading to high blood sugar levels. The World Health Organization (WHO) reports that over 422 million people suffer from diabetes, and the number continues to rise due to lifestyle changes, obesity, and genetic factors.

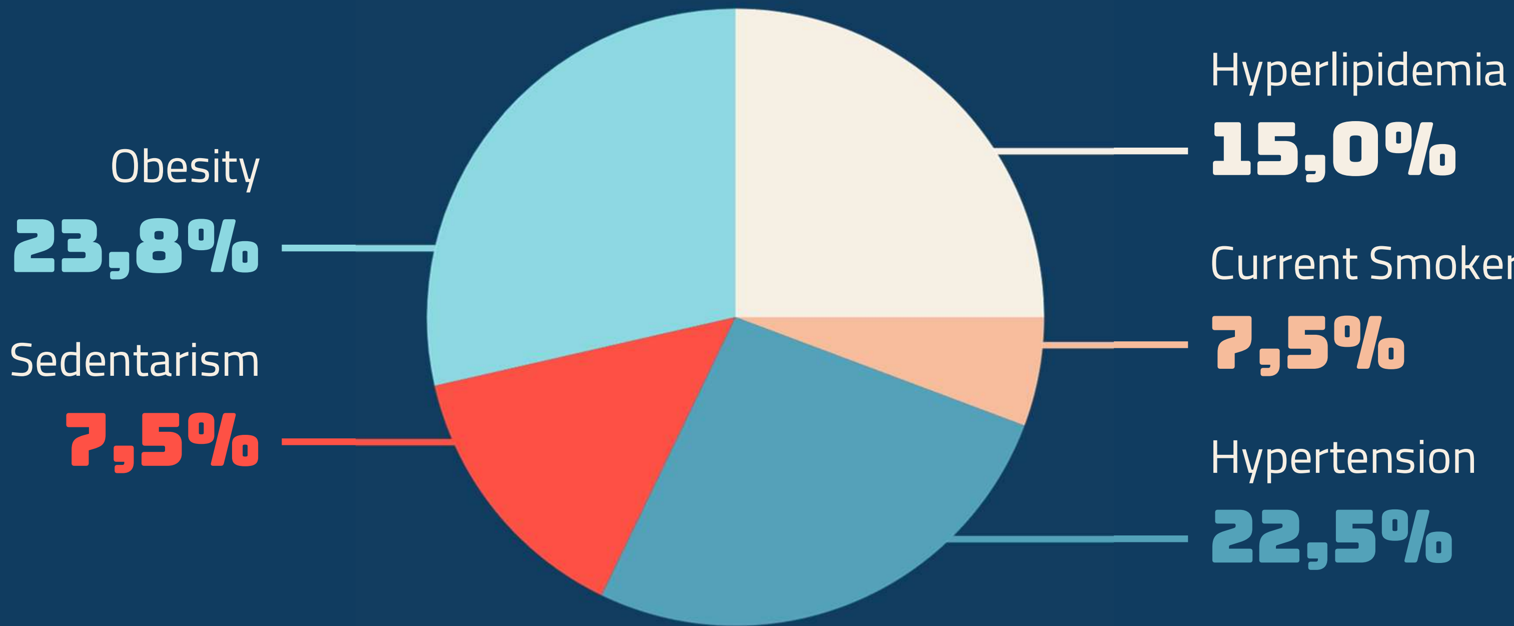
The condition is often asymptomatic in its early stages, making it difficult for individuals to recognize the risk until significant damage has already occurred. Late detection can result in severe complications such as cardiovascular diseases, kidney failure, vision loss, and even amputations.

PROBLEM

The challenge lies in the lack of early and accurate screening tools that can predict the risk of diabetes in individuals using readily available health data. Traditional diagnostic methods can be time-consuming and may not be feasible for large-scale health screenings.

Late detection can lead to severe complications and high treatment costs, placing a burden on healthcare systems and reducing the quality of life for affected individuals.

RISK FACTORS



For more info, click here

T 1

THE EFFECTS OF DIABETES

EYES

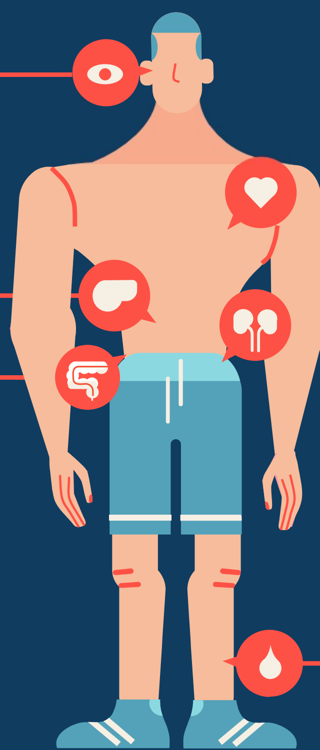
Risk of Diabetic Retinopathy and Vision Loss

LIVER

Impact on Liver Function and Metabolism

INTESTINES

Effects on Digestion and Nutrient Absorption



HEART

Impact on Cardiovascular Health

KIDNEYS

Risk of Kidney Disease and Failure

BLOOD

Effects on Blood Circulation and Glucose Levels

FIRST SYMPTOMS OF DIABETES

1 HEADACHE

A common early sign due to high blood sugar levels.

2 BLURRY VISION

Caused by fluid changes in the eyes, affecting focus.

3 TIREDNESS

Fatigue results from the body's inefficiency in using glucose for energy.

T 1

RISK AGE TYPE 1 DIABETES

60%

4-7 YEARS

Despite being red, Mars is actually a cold place

40%

10-14 YEARS

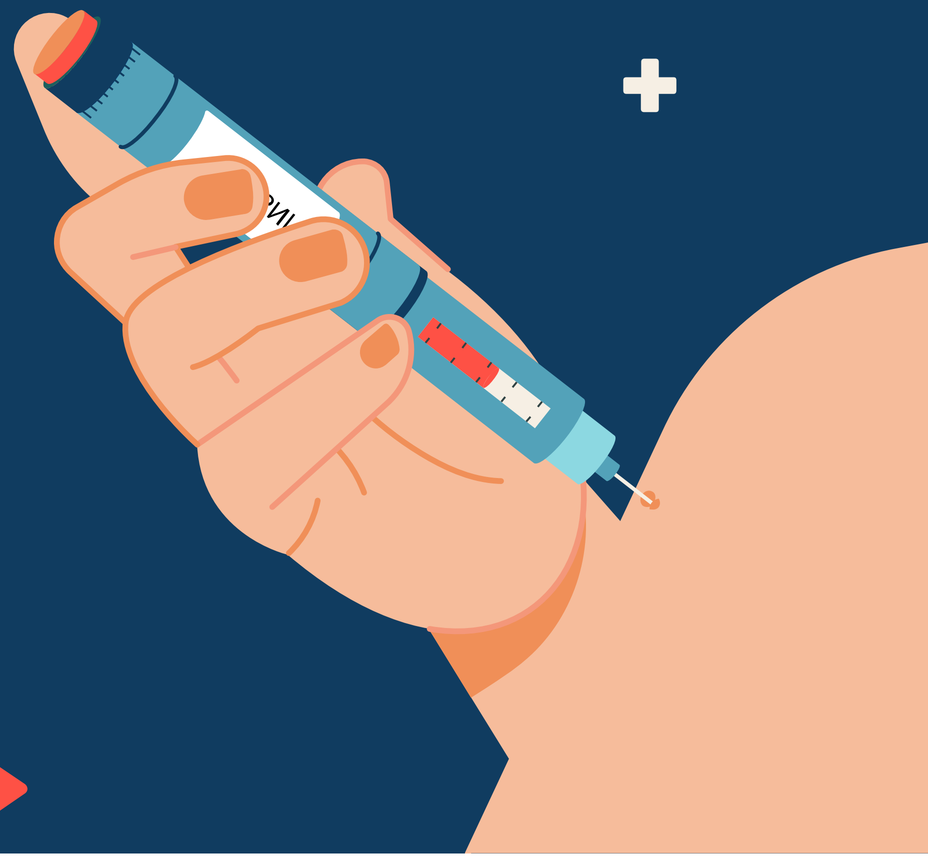
Mercury is the closest to the Sun of them all

For more info, click here

02

RELATED WORK

Overview of current research
and methods



RELATED WORK

STUDY	YEAR	METHOD USED	RESULTS	LIMITATIONS
Smith et al.	2020	Logistic Regression	Accuracy of 85%	Limited data diversity
Chen et al.	2021	Random Forest Classifier	Accuracy of 90%	High computational cost
Lee & Kim	2019	Decision Tree	Precision of 80%	Prone to overfitting
Patel et al.	2022	Support Vector Machine	Recall of 87%	Not robust to noisy data

TIMELINE OF AI-POWERED TREATMENT DEVELOPMENTS

2010
Early AI Integration

2012
AI in Medical Imaging

2015
AI for Diabetes Management

2017
Machine Learning for Personalized Medicine

2018
AI-Driven Drug Discovery

T 2

TIMELINE OF AI-POWERED TREATMENT DEVELOPMENTS

2020
Advances in Predictive Analytics

2021
AI for Treatment Optimization

2022
AI in Continuous Glucose Monitoring (CGM)

2023
AI-Powered Robotic Surgery

2024
Next-Generation AI Models

03

PROPOSED METHODOLOGY

Approach for Developing an Accurate Diabetes Prediction Model

PROPOSED METHODOLOGY

01 DATA COLLECTION

02 DATA PREPROCESSING

03 FEATURE SELECTION AND ENGINEERING

T 3

PROPOSED METHODOLOGY

04

MODEL SELECTION AND TRAINING

05

MODEL EVALUATION

06

VALIDATION

T 3

DATA COLLECTION

- Utilized a publicly available dataset containing various health-related attributes.
- Ensured data includes relevant features like age, BMI, blood glucose levels, and more.

DATA PREPROCESSING

- Handled missing data and performed feature scaling using **MinMaxScaler** for normalization.
- Coded categorical features with **LabelEncoder** to convert non-numeric labels into numerical values.

T 3

FEATURE SELECTION AND ENGINEERING

- Selected features based on domain knowledge and correlation analysis.
- Created new features if necessary for better prediction.

MODEL SELECTION AND TRAINING

- Applied multiple machine learning algorithms: Logistic Regression, Decision Trees, Random Forests, Gradient Boosting, and Support Vector Classifier (SVC).
- Used **GridSearchCV** for hyperparameter tuning to optimize model performance.

MODEL EVALUATION

- Evaluated using metrics like accuracy, precision, recall, F1-score, and confusion matrix.
- Ensured data includes relevant features like age, BMI, blood glucose levels, and more.

VALIDATION

- Split data into training and test sets for robust validation.
- Performed cross-validation to ensure consistent results.

IMPLEMENTATION

Finalized the best model and created a user-friendly interface to input data and obtain predictions.

04

RESULTS

Evaluation and Performance Metrics of the Model

THE MODEL BEING WORKED ON

LOGISTIC REGRESSION

SVM MODEL

DECISION TREE

GRADIENT BOOSTING

RANDOM FOREST

ACCURACY RESULTS

95.93% — SVM MODEL — A reliable baseline model that performed well for binary classification.

95.91% — SVM MODEL — Highly effective in finding the optimal hyperplane for classification, with comparable accuracy to logistic regression.

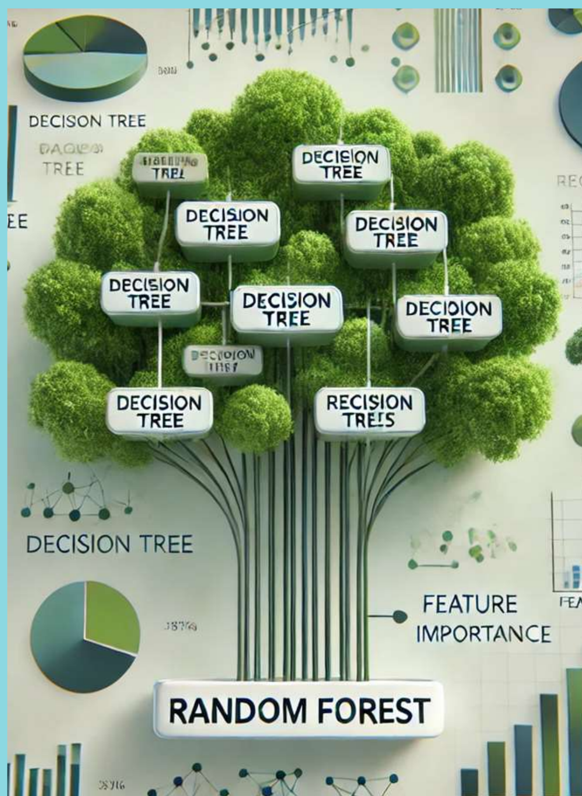
99.93% — RANDOM FOREST — Outperformed other models significantly, showing exceptional performance due to its ensemble approach.

ACCURACY RESULTS

97.13% — **GRADIENT BOOSTING** — Demonstrated strong predictive capabilities with boosting techniques to improve accuracy.

97.08% — **DECISION TREE** — Performed well but slightly lower than ensemble models, providing clear and interpretable decision boundaries.

RANDOM FOREST CLASSIFIER



Key Points:

- Model: Random Forest Classifier
- Accuracy: 99.93%
- **Why It Stood Out**:
  - **Ensemble Method**: Combines multiple decision trees to improve prediction accuracy.
  - **Robust Performance**: High accuracy indicates strong generalization to unseen data.
  - **Feature Importance**: Capable of providing insight into the importance of each feature.
  - **Resilience**: Handles overfitting better than individual decision trees.

Conclusion: The Random Forest Classifier outperformed all other models in terms of accuracy, making it the most suitable choice for this diabetes prediction task.

T 4

05

CONCLUSIONS

Key Takeaways and Future Directions

CONCLUSIONS

- **Model Performance**:
  - The Random Forest Classifier demonstrated the highest accuracy at **99.93%**, showcasing its effectiveness in diabetes prediction.

- **Comparison with Other Models**:
  - Logistic Regression, SVM, and Gradient Boosting Classifier also showed strong results with accuracies of **95.93%**, **95.91%**, and **97.13%**, respectively.
  - Decision Tree achieved **97.08%**, proving its usefulness for simpler models.

CONCLUSIONS

- **Key Insights**:
  - Ensemble models like Random Forest provide significant performance improvements due to their ability to combine multiple decision trees to enhance predictions.
  - The results suggest that tree-based algorithms are highly suitable for complex classification tasks like diabetes prediction.

FUTURE WORK

FUTURE WORK

- Implementing additional data preprocessing and feature engineering.
- Exploring the integration of more advanced models such as neural networks.
- Real-time deployment of the model for practical use.

OUR TEAM

Abdelrahman Ayman Elsayed

Abdulrahman Ashraf Mokhtar

Abdulbari Sabry Abdulbari

Sami Mohammed Sami

THANKS!

Do you have any questions?

se.abdelrahman968@gmail.com
+20 1275502636