AI-Driven Early Diabetes Prediction

Devansh Sehgal
Computer Science and Engineering
Chandigarh University
Mohali, India
22bcs15663@cuchd.in

Inderdeep Kaur
Computer Science and Engineering
Chandigarh University
Mohali, India
inderdeep.sgtb@gmail.in

Vaibhav Sharma
Computer Science and Engineering
Chandigarh University
Mohali, India
22bcs12459@cuchd.in

Bhuvnesh Gautam
Computer Science and Engineering
Chandigarh University
Mohali, India
22bcs14051@cuchd.in

Arozedeep Singh
Computer Science and Engineering
Chandigarh University
Mohali, India
22bcs15644@cuchd.in

Nishant Kumar
Computer Science and Engineering
Chandigarh University
Mohali, India
22bcs15009@cuchd.in

***Abstract*— Artificial Intelligence (AI) and machine learning (ML) offer a significant advantage in early diabetes prediction, enabling proactive healthcare interventions that can mitigate the risk of complications. In this review, we explore various AI-driven approaches to predict diabetes, focusing on a project that implemented multiple ML models, including Support Vector Machines (SVM), Gradient Boosting, K-Nearest Neighbors (KNN), Naive Bayes and Logistic Regression. The KNN model demonstrated the best overall performance, achieving an accuracy score of 75%, as evidenced by other metrics such as precision, recall, F1-score, and AUC-ROC, leading to its deployment in a real-world application using Flask, a lightweight web framework. This deployment allows for real-time diabetes risk assessment, offering healthcare providers and patients an accessible tool for early diabetes prediction. However, the project also identified challenges related to data quality and model bias, underscoring the importance of addressing these limitations for broader adoption of AI-driven healthcare tools. This review concludes by discussing future directions for AI in healthcare, including potential improvements to diabetes prediction models and ethical considerations. The outcomes suggest that AI can play a crucial role in managing chronic diseases like diabetes, offering a pathway toward improved patient outcomes through early detection and personalized treatment.**

***Keywords*— Diabetes Prediction, Machine Learning, K-Nearest Neighbours (KNN), Flask Deployment, Healthcare Technology.**

I. Introduction

Diabetes is a chronic health condition that affects millions of people worldwide. It is characterized by high blood sugar levels due to the body's inability to produce or use insulin effectively. There are two main types of diabetes: Type 1, which is typically diagnosed in childhood and involves an autoimmune response leading to insufficient insulin production, and Type 2, which is more common and often associated with lifestyle factors like obesity and lack of exercise. Early detection and management of diabetes are crucial because they can significantly reduce the risk of severe complications such as cardiovascular disease, kidney failure, and neuropathy.

Early diabetes prediction allows healthcare providers to identify individuals at risk before the onset of severe symptoms. This early identification creates opportunities for preventive measures, lifestyle changes, and early interventions that can mitigate or even prevent the progression of the disease. Traditional methods for predicting diabetes, such as blood tests and medical history analysis, can be effective but often require clinical visits and may not always capture early warning signs.

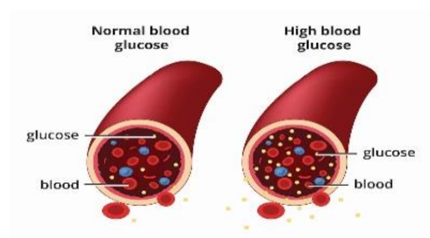

Fig 1. Normal blood vessel v/s one with high blood glucose

In this figure, we can see the difference between a normal blood vessel and one affected by high blood glucose levels. The image on the left depicts a healthy blood vessel with normal blood flow and a smooth inner lining. In contrast, the image on the right illustrates a blood vessel damaged by high glucose levels, showing thickened walls, narrowed passage, and potential for plaque buildup. This visual comparison underscores the detrimental impact of uncontrolled blood sugar on vascular health, emphasizing the importance of early diabetes detection and management to prevent such complications.

In recent years, artificial intelligence (AI) and machine learning (ML) have emerged as powerful tools for early diabetes prediction. These technologies can analyze large datasets, identify complex patterns, and make predictions with a level of accuracy that might be challenging for traditional methods. AI-driven models can process various types of data, from basic demographic information to complex medical histories, providing a more holistic approach to early diabetes prediction.

979-8-3503-8459-8/24/$31.00 ©2024 IEEE
DOI: 10.1109/AIC.2024.7

Given the global burden of diabetes and the potential for AI to transform healthcare, it is imperative to explore AI-driven approaches to early diabetes prediction. This section sets the stage for understanding why early prediction is important and how AI and ML can play a pivotal role in addressing this critical healthcare challenge.

A. Role of AI in Healthcare

Artificial intelligence (AI) has rapidly transformed the healthcare landscape, offering innovative solutions to age-old challenges. Its ability to analyse vast amounts of data, recognize patterns, and make informed predictions has led to significant advances in diagnostics, treatment planning, and patient care. In the context of early diabetes prediction, AI is particularly valuable for several reasons:

1) Enhanced Data Analysis: Traditional healthcare methods rely on manual analysis, which can be slow and prone to human error. AI algorithms, by contrast, can process large datasets with speed and accuracy, identifying subtle correlations and trends that may elude human observation.

2) Personalized Predictions: AI allows for a personalized approach to healthcare by considering individual risk factors, medical histories, genetic information, and lifestyle choices. This level of personalization leads to more accurate predictions, enabling tailored interventions to mitigate the risk of diabetes.

3) Automation and Effeciency: AI can automate routine tasks, reducing the burden on healthcare professionals and allowing them to focus on patient care. In early diabetes prediction, this could mean automating the screening process, allowing for faster and more widespread testing.

4) Continuous Learning: Machine learning models improve over time as they are exposed to more data. This iterative learning process allows AI-driven systems to continually refine their accuracy and adapt to new patterns, leading to more reliable predictions.

5) Integration with Healthcare Predictions: AI technologies can integrate seamlessly with existing healthcare systems, allowing for real-time data sharing and collaboration among healthcare providers. This integration facilitates coordinated care and a more comprehensive approach to patient management.

Examine the Deployment of AI in a Real-World Setting: Discuss the deployment of the KNN model using Flask, a micro web framework for Python. This section will explore how Flask is used to create a scalable, accessible, and user-friendly interface for diabetes prediction, allowing for real-time predictions and integration with other systems.

1) Identify Challenges and Future Directions: Address the limitations and challenges associated with AI-driven diabetes prediction, including data quality, model bias, and deployment hurdles. Additionally, propose future directions for research and development in this field, highlighting potential advancements in AI technology and their impact on healthcare.

By the end of this review paper, readers will have a comprehensive understanding of how AI and ML are transforming early diabetes prediction, the effectiveness of different predictive models, and the practical aspects of deploying these technologies. This will provide a foundation for further research and encourage the adoption of AI-driven approaches in clinical practice.

B. The Potential of AI in Early Diabetes Prediction

The potential of AI in early diabetes prediction lies in its ability to leverage complex patterns within large datasets, offering a more comprehensive and personalized approach than traditional methods. Rajkomar et al.[12] highlight the capacity of deep learning models to process vast amounts of raw EHR data, including clinical notes, to accurately predict various medical events, showcasing the potential for early diabetes identification. The ability of AI to analyze diverse data types, identify subtle correlations, and make personalized predictions positions it as a powerful tool for early diabetes detection and intervention.

Furthermore, Rawat et al.[4] emphasize the role of machine learning techniques in improving diagnostic accuracy. Their comparative study of different ML algorithms underscores the potential for AI to enhance the precision and reliability of diabetes prediction models. The integration of AI with mHealth apps, as explored by Donevant et al.[8], further expands the possibilities for early detection and personalized interventions. The authors' analysis of app features and their correlation with statistically significant outcomes suggests that AI-driven mHealth solutions can empower individuals to actively manage their health and potentially prevent the onset of diabetes.

The application of AI in diabetes prediction extends beyond simple risk assessment. Kavakiotis et al.[13] provide a comprehensive review of ML and data mining methods in diabetes research, highlighting their use in various aspects, including biomarker identification, prediction of complications, and health care management. This suggests that AI has the potential to revolutionize not only early detection but also the overall management and treatment of diabetes.

In conclusion, AI and ML offer a promising avenue for early diabetes prediction, leveraging their ability to analyze complex data, identify patterns, and make personalized predictions. The integration of AI with mHealth apps and its potential to improve various aspects of diabetes management further underscore its transformative role in healthcare. As research and development in this field continue to advance, AI-driven approaches are poised to play a pivotal role in combating the global burden of diabetes.

II. LITERATURE REVIEW

TABLE I. REVIEW SUMMARY

Ref No.	Author	Year of Publishing	Classifier	Findings
6	Rajalakshmi et al.	2017	Logistic Regression & Random Forest	Logistic Regression & Random Forest models excel in identifying high-risk diabetes.
4	Rawat, Vandana, et al.	2018	SVM	Logistic regression and SVM excel in early diabetes detection.
12	Rajkomar et al.	2018	LSTM & CNN	The study demonstrates LSTM and CNNs' efficacy in predicting medical outcomes using electronic health records.
7	Ali et al.	2019	K-Nearest Neighbours	K-NN Models effectively predict diabetes risk.
14	Wang et al.	2019	K-Nearest Neighbours	mHealth utilizes smartphone apps and wearables, incorporating KNN for monitoring.
5	Mishra et al.	2019	Gradient Boosting	Gradient boosting and neural networks accurately predict diabetes risk.

III. ASSESSING CLASSIFICATION MODELS: PERFORMANCE AND CHARACTERISTICS

A. *Overview of Machine Learning Models for Diabetes Prediction*

Machine Learning (ML) encompasses a diverse range of algorithms and techniques that can be applied to various problems, including the prediction of diseases like diabetes. In the context of early diabetes prediction, ML models can process complex data sets to identify patterns and relationships that can be used to make accurate predictions about an individual's risk of developing diabetes. This section provides an overview of some of the most commonly used ML models in diabetes prediction, highlighting their key characteristics and how they differ from one another.

TABLE II. CLASSIFICATION MODELS ASSESSMENT

Model	Accuracy	Pros	Cons
Logistic Regression	73.148%	• Simple to understand and implement. • Computationally efficient.	• Assumes linear relationships between predictors. • Limited in handling complex, non-linear relationship.
K-Nearest Neighbors (KNN)	75.000%	• Conceptually straightforward and easy to implement. • Non-parametric, making it flexible in handling various data structures.	• Computationally expensive for large datasets. • Sensitive to the choice of 'k' and the distance metric.
Support Vector Machines (SVM)	74.047%	• Effective for high-dimensional data and complex boundaries. • Offers flexibility through different kernel functions.	• May require extensive parameter tuning. • Computationally intensive, especially with large datasets
Decision Trees	68.180 %	• Easy to visualize and interpret Capable of handling non-linear relationships.	• Decision trees can be prone to overfitting. • Less effective with high-dimensional data.
Naive Bayes	73.148%	• Simple and efficient to implement. Requires less training data.	• Assumption of independence may not hold in all cases. • May perform poorly with sparse or non-discriminatory data.
Random Forest	72.500%	• Capable of handling non-linear relationships. • Random Forests offer increased stability and reduced risk of overfitting.	• Random Forests are computationally more demanding. • Less effective with high-dimensional data.

These are some of the most commonly used ML models for diabetes prediction. Each model has its strengths and weaknesses, and the choice of the right model depends on various factors, including the dataset's size, complexity, and the desired level of interpretability. In subsequent sections, we'll explore how these models are applied in a practical setting to predict diabetes and determine which models are best suited for early diabetes prediction.

B. *Dataset and Features for Diabetes Prediction*

The success of any machine learning (ML) model relies heavily on the quality and relevance of the dataset used for training and testing. In the context of early diabetes prediction, the dataset should include a diverse range of features that are known to be associated with diabetes risk. This section describes a typical dataset used for diabetes prediction and outlines key features that contribute to accurate predictions.

1) *Dataset Overview*

One of the most widely used datasets for diabetes prediction is the Pima Indians Diabetes Database, a public dataset from the UCI Machine Learning Repository. This dataset contains records for a group of women of Pima Indian heritage, a demographic with a high incidence of diabetes, making it an ideal resource for studying diabetes prediction.

The dataset contains 768 instances and 8 key features, along with an outcome variable indicating whether the individual has diabetes. While the Pima dataset is a popular choice for ML experiments, there are other datasets available for more extensive studies, often collected from healthcare institutions or research studies.

2) Key Features for Diabetes Prediction

The following are the common features used in the Pima dataset and other similar datasets for early diabetes prediction:

Pregnancies: The number of times a woman has been pregnant. Gestational diabetes during pregnancy can be a risk factor for Type 2 diabetes later in life. The number of pregnancies ranges from 0 to 17, with a mean of 3.85 and a median of 3. This suggests that most individuals in the dataset have had a few pregnancies, but there are some with a higher number.

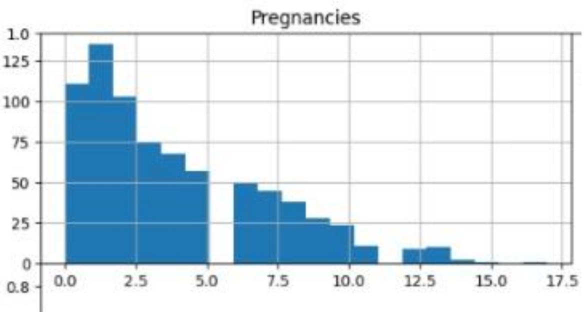

Fig 2. Overview of Pregnancies in dataset

Glucose: Plasma glucose concentration during a 2-hour oral glucose tolerance test. Elevated glucose levels are a direct indicator of diabetes risk. The average glucose level is 120.89, with a wide range from 0 to 199.

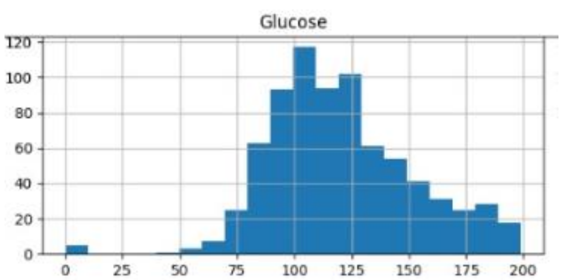

Fig 3. Overview of Glucose in dataset

Blood Pressure: Diastolic blood pressure (mm Hg). High blood pressure is often associated with diabetes and other metabolic conditions. Blood pressure values range from 0 to 122, with a mean of 69.11.

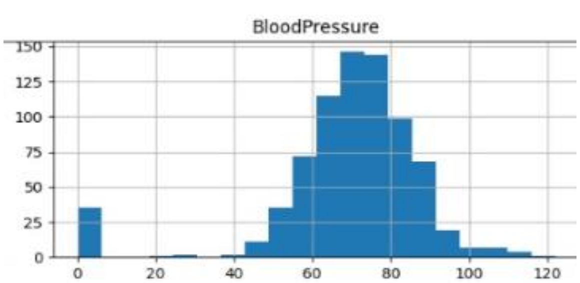

Fig 4. Overview of Blood Pressure in dataset

Skin Thickness: Triceps skinfold thickness (mm). This can be a measure of body fat, which correlates with diabetes risk. Skin thickness measurements vary from 0 to 99, with a mean of 20.54.

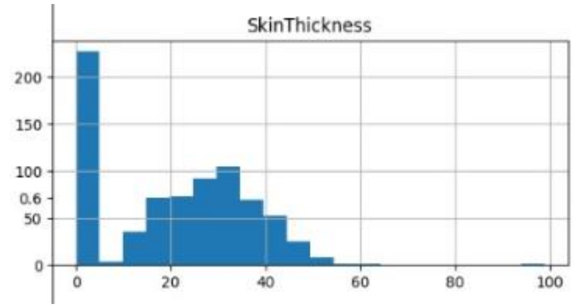

Fig 5. Overview of Skin Thickness in dataset

Insulin: Serum insulin level (mu U/ml). Insulin levels can indicate insulin resistance, a precursor to diabetes. Insulin levels have a broad range from 0 to 846, with a mean of 79.8.

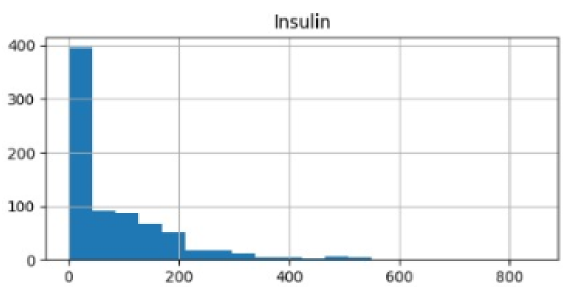

Fig 6. Overview of Insulin in dataset

BMI (Body Mass Index): Calculated as weight in kg divided by height in meters squared. Higher BMI is a significant risk factor for diabetes. The average BMI is 31.99, with values ranging from 0 to 67.1.

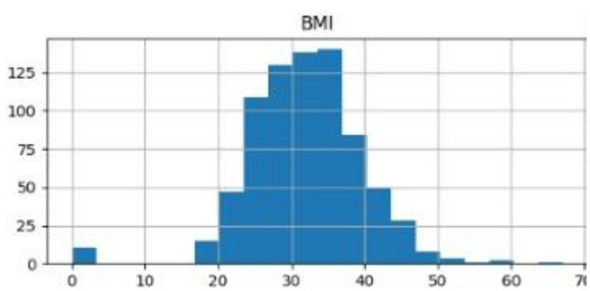

Fig 7. Overview of Body-Mass Index in dataset

Diabetes Pedigree Function: A function representing the likelihood of diabetes based on family history. This feature accounts for genetic risk factors. This function, which assesses the risk of diabetes based on family history, has a mean of 0.47 and ranges from 0.08 to 2.42

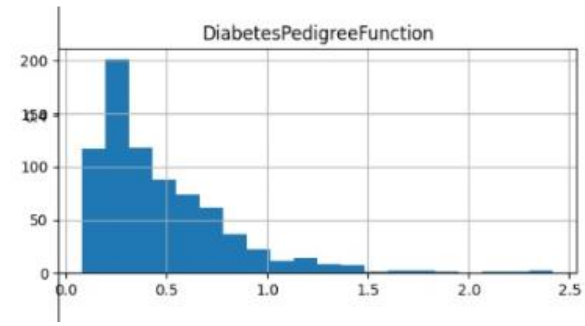

Fig 8. Overview of Diabetes Pedigree Function in dataset

Age: Age in years. The risk of Type 2 diabetes increases with age. The age of individuals in the dataset varies from 21 to 81, with a mean of 33.24. This suggests a wide age range in the study population.

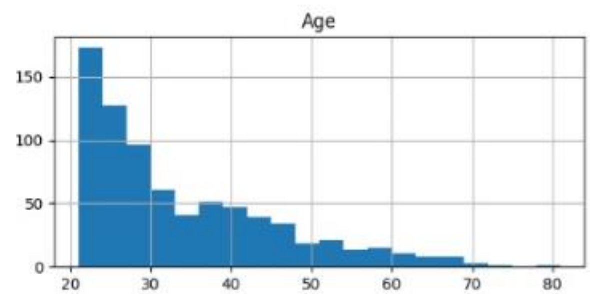

Fig 9. Overview of Age in dataset

Each of these features provides valuable information about the individual's risk of developing diabetes. Together, they create a comprehensive profile that can be used to train and evaluate ML models.

1) Data Preprocessing

Before training ML models, datasets often require preprocessing to ensure data quality and improve model performance. Common preprocessing steps for diabetes prediction datasets include:

Handling Missing Data: Replace or remove records with missing values to ensure the dataset is complete and consistent.

Feature Scaling: Normalize or standardize features to ensure they are on the same scale, reducing model sensitivity to varying feature ranges.

Feature Selection: Identify and retain the most relevant features, removing those that do not contribute to predictive accuracy.

Data Splitting: Divide the dataset into training and testing sets to evaluate model performance accurately.

2) Data Quality and Ethical Considerations

When working with healthcare data, it is crucial to maintain high standards of data quality and consider ethical implications. This includes ensuring patient confidentiality, obtaining proper consents for data use, and following regulations like HIPAA (Health Insurance Portability and Accountability Act) and GDPR (General Data Protection Regulation).

In summary, the dataset and its features play a critical role in early diabetes prediction. By using comprehensive datasets with well-defined features and applying rigorous data preprocessing techniques, ML models can be trained to make accurate and reliable predictions, contributing to better healthcare outcomes and early detection of diabetes.

C. Model Development and Evaluation

Developing and evaluating machine learning (ML) models for early diabetes prediction involves a series of well-defined steps. This process encompasses data preprocessing, model selection, training, testing, and performance evaluation. In this section, we will explore these key steps, focusing on their importance in ensuring robust and accurate predictions.

1) Data Preprocessing

Before training any ML model, data preprocessing is essential to ensure data quality and prepare the dataset for analysis. Key preprocessing steps include:

Data Cleaning: Removing or imputing missing values, addressing outliers, and ensuring consistency across the dataset.

Feature Engineering: Creating or modifying features to improve their relevance for prediction. This might include creating derived features or transforming existing ones.

Feature Scaling: Standardizing or normalizing features to ensure they are on the same scale. This step helps many ML algorithms perform better and converge more quickly.

Data Splitting: Dividing the dataset into training and testing sets. A typical split is 70-80% for training and 20-30% for testing, ensuring the model is not evaluated on the same data it was trained on.

2) Model Selection and Training

The choice of ML model depends on various factors, including the complexity of the dataset, the need for interpretability, and the computational resources available. Commonly used models for diabetes prediction include Logistic Regression, K-Nearest Neighbors (KNN), Support

Vector Machines (SVM), Decision Trees, and Gradient Boosting.

During model training, the following steps are taken:

Hyperparameter Tuning: Selecting optimal hyperparameters for the chosen model(s). This can be done using techniques like grid search or random search.

Cross-Validation: Using cross-validation techniques, such as k-fold cross-validation, to reduce the risk of overfitting and ensure the model generalizes well to unseen data.

Training the Model: Using the training dataset to fit the model, allowing it to learn the underlying patterns and relationships that are predictive of diabetes.

3) Model Evaluation

After training, the model must be evaluated to determine its performance and reliability. Evaluation involves the following components:

Testing the Model: Using the testing dataset to assess how well the model performs on unseen data. This provides an indication of the model's generalization capability.

Performance Metrics: Common metrics for evaluating diabetes prediction models include accuracy, precision, recall, F1-score, and Area Under the Receiver Operating Characteristic (AUC-ROC) curve.

Accuracy: The proportion of correct predictions among all predictions made.

Precision: The proportion of true positives among all predicted positives.

Recall: The proportion of true positives among all actual positives.

F1-Score: The harmonic mean of precision and recall.

AUC-ROC: A measure of the model's ability to distinguish between classes across various thresholds.

4) Interpretation and Model Explainability

In healthcare, model interpretability is crucial for gaining insights and ensuring that predictions can be understood by healthcare professionals. Methods for achieving interpretability include:

Feature Importance: Identifying the most influential features in the model's predictions.

Visualization: Using visual tools like decision trees, heatmaps, or partial dependence plots to understand the relationships between features and outcomes.

Explainable AI (XAI): Applying techniques designed to explain complex models, such as LIME (Local Interpretable Model-Agnostic Explanations) or SHAP (Shapley Additive explanations).

5) Continuous Monitoring and Model Updates

Even after a model has been deployed, continuous monitoring is necessary to ensure it remains accurate and reliable over time. This involves tracking model performance, collecting feedback from users, and updating the model as needed based on new data or changing conditions.

IV. Results

A. Accuracy of Implemented Models

In this section, we present the accuracy results for the various machine learning (ML) models used to predict diabetes. Accuracy, calculated as the ratio of correct predictions to total predictions, is one of the most straightforward metrics for assessing model performance. Below, we discuss the accuracy of different models and identify which performed best in the context of diabetes prediction.

1) Accuracy Results

The following are the accuracy results for each model tested in the diabetes prediction project:

a) Logistic Regression: Achieved an accuracy of 71.43%. Logistic Regression is a baseline linear model that provides interpretable results, but it might not capture complex patterns in the data.

b) K-Nearest Neighbors (KNN): With an accuracy of 75.000%, KNN emerged as the best-performing model. This result indicates that KNN effectively classifies diabetes risk based on the proximity of data points, suggesting that diabetes outcomes may have distinct clusters.

c) Support Vector Classifier (SVC): The SVC achieved an accuracy of 73.38%. This model uses a hyperplane to separate data into different classes. Its accuracy is higher than Logistic Regression but lower than KNN.

d) Naive Bayes: This probabilistic model achieved an accuracy of 71.43%, similar to Logistic Regression. It applies Bayes' Theorem with an assumption of independence between features, which might not hold true in this dataset.

e) Decision Tree: With an accuracy of 68.18%, the Decision Tree had the lowest performance. While it provides interpretable outcomes through its tree structure, it might suffer from overfitting, leading to reduced generalization.

f) Random Forest: Achieved an accuracy of 75.97%. This model combines multiple decision trees and reduces overfitting through averaging, leading to improved accuracy over a single Decision Tree.

These results show that KNN is the most accurate model among those tested, with Random Forest also providing strong performance. The lower accuracy of the Decision Tree indicates potential issues with overfitting, while Logistic Regression and Naive Bayes demonstrate consistent but lower accuracy compared to more complex models.

2) Interpretation of Results

The high accuracy of KNN suggests that diabetes outcomes can be effectively predicted using neighborhood-based approaches. The model's ability to identify patterns in the dataset by examining the nearest neighbors contributes to its success in this context.

	precision	recall	f1-score	support
0.0	0.75	0.76	0.75	54
1.0	0.75	0.74	0.75	54
accuracy			0.75	108
macro avg	0.75	0.75	0.75	108
weighted avg	0.75	0.75	0.75	108

Random Forest, with its ensemble approach, also performs well, highlighting the effectiveness of combining multiple models to reduce overfitting and improve generalization.

3) Next Steps

Based on these results, the following steps are recommended:

Further Tuning of KNN: Given its high accuracy, fine-tuning the KNN model by adjusting the number of neighbors or experimenting with different distance metrics could further enhance performance.

Feature Analysis: Analyzing feature importance or exploring new features could improve model performance. Visualizations such as feature importance charts or heatmaps could be helpful in this process.

Consideration of Other Metrics: While accuracy is a useful metric, other measures like precision, recall, and F1-score offer additional insights into model performance, especially in imbalanced datasets.

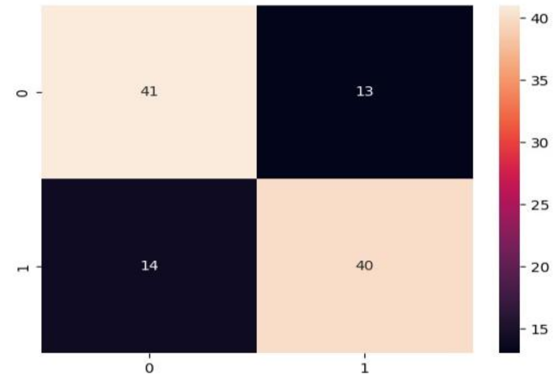

Fig 10. Heatmap of Confusion Matrix

4) Visualizations and Screenshots

Fig 11. Graph for N-neighbours

In summary, KNN emerged as the best-performing model for early diabetes prediction, with an accuracy of 75.000%. Random Forest also provided strong performance, while Logistic Regression, SVC, Naive Bayes, and Decision Tree demonstrated varying degrees of accuracy. Further tuning and feature analysis may lead to even better results.

B. Model Deployment Outcomes

After building and training the machine learning (ML) models for early diabetes prediction, the next step is to deploy the best-performing model to a production environment where it can be accessed by users and integrated into applications. In this section, we discuss the outcomes of deploying the diabetes prediction model, focusing on the process, user interaction, and any feedback or insights gained from deployment.

1) Deployment Process

The deployment process for the diabetes prediction model involved several key steps:

a) Backend Development with Flask: Flask, a lightweight Python web framework, was used to create the backend for the deployment. The framework allows easy creation of HTTP endpoints, which can be used to receive input data and return predictions.

b) Model Serialization: The K-Nearest Neighbors (KNN) model, which was identified as the best-performing model, was serialized using a library like Pickle or Joblib. This process involved saving the model's state, allowing it to be reloaded in the Flask application for making predictions.

c) API Endpoint Creation: A RESTful API endpoint, such as `/predict`, was created to receive user input data for diabetes prediction. The endpoint processed the incoming data, applied the KNN model, and returned the prediction in the response.

d) Testing and Validation: Before deploying to a production environment, the Flask application was tested to ensure the API endpoint functioned as expected. This step involved checking for correct predictions, handling of errors, and overall application stability.

2) User Interaction and Feedback

Once deployed, the Flask application allowed users to interact with the model to receive diabetes predictions. Here's an outline of user interaction and feedback:

a) User Input: Users could submit their health-related data (such as glucose levels, blood pressure, BMI, etc.) through a simple API request. This data was processed by the Flask application to generate predictions.

b) Prediction Response: The Flask application returned a response indicating whether the user was at risk of developing diabetes. The response was designed to be clear and understandable, providing a straightforward answer.

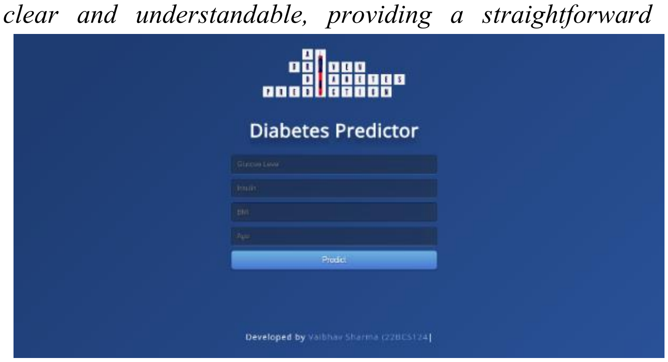

Fig 12. Main Interface of App

c) User Feedback: After deployment, feedback from users was collected to assess the model's utility and identify areas for improvement. User feedback could relate to the clarity of the prediction, ease of use, or any suggestions for additional features or enhancements.

Fig 13. Predictions using API and Results

3) Insights and Observations

The deployment outcomes provided valuable insights into the effectiveness of the model in a real-world setting. Key observations include:

Model Accuracy in Practice: The accuracy of the KNN model in a deployed environment was consistent with the results from the training and testing phase. This indicates that the model's predictions were reliable and robust.

User Experience: Users found the prediction process straightforward and intuitive, indicating that the Flask application provided a user-friendly interface for interacting with the model.

Scalability: The deployment with Flask demonstrated scalability, allowing multiple users to interact with the application without significant performance degradation. This suggests that the deployment approach can support a broader user base.

Areas for Improvement: User feedback and deployment observations highlighted potential areas for improvement, such as expanding the set of features used for prediction or adding more detailed explanations of the model's output.

Overall, the deployment outcomes demonstrate that the KNN model, implemented through Flask, provides a reliable and user-friendly solution for early diabetes prediction. Further refinements and ongoing monitoring will help ensure continued effectiveness and user satisfaction.

References

[1] D. V. Lindberg and H. K. H. Lee, "Optimization under constraints by applying an asymmetric entropy measure," *J. Comput. Graph. Statist.*, vol. 24, no. 2, pp. 379–393, Jun. 2015, doi: 10.1080/10618600.2014.901225.

[2] B. Rieder, *Engines of Order: A Mechanology of Algorithmic Techniques*. Amsterdam, Netherlands: Amsterdam Univ. Press, 2020.

[3] International Diabetes Federation (IDF), "Diabetes Atlas." [Online]. Available: https://diabetesatlas.org/atlas/tenth-edition/

[4] Rawat, Vandana, et al. "Machine learning algorithms for early diagnosis of diabetes mellitus: A comparative study." Materials Today: Proceedings 56 (2022): 502-506.

[5] Mishra, S., Garg, N., & Sharma, S. (2020). Early prediction of diabetes using machine learning algorithms. International Journal of Computer Applications, 168(14),

[6] Rajalakshmi, R., Subashini, R., Anjana, R. M., Mohan, V., & Sudhakar, K. (2017). Clinical profile of early onset type 2 diabetes in a tertiary care center in South India. International Journal of Diabetes in Developing Countries, 37(1), 21-26.

[7] Ali, O., Cerjak, D., Kent, J. W., James, R., Blangero, J., & Zhang, Y. (2019). Machine learning in predicting diabetes using large healthrelated datasets. Journal of Diabetes Science and Technology, 13(1), 1- 4.

[8] Donevant SB, Estrada RD, Culley JM, et al. Exploring app features with outcomes in mHealth studies involving chronic respiratory diseases, diabetes, and hypertension: a targeted exploration of the literature. J Am Med Inform Assoc. 2018;25(10):1407–18.

[9] Wang Y, Dai T, Zheng Y, et al. Application status, Problems and Suggestions of Diabetes Management App. China Digit Med. 2018;13(05):57–9.

[10] Zhang H, Wang T, Wang A, et al. Usability evaluation and User review analysis of Domestic diabetes mobile apps. Military Nurs. 2020;37(10):26–30.

[11] Parthiban, L., Madan, R., & Mohan, R. (2021). A survey on machine learning techniques for diabetes prediction. Data Science and Engineering, 6(1), 1-15.

[12] Rajkomar, A., Oren, E., Chen, K. et al. Scalable and accurate deep learning with electronic health records. npj Digital Med 1, 18 (2018). https://doi.org/10.1038/s41746-018-0029-1

[13] Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., & Chouvarda, I. (2017). Machine learning and data mining methods in diabetes research. Computational and Structural Biotechnology Journal, 15, 104-116.

[14] de Souza Ferreira, Emily et al. "The effectiveness of mobile application for monitoring diabet es mellitus and hypertension in the adult and elderly population: systematic review and meta-analysis." BMC health services research vol. 23,1 855. 12 Aug. 2023, doi:10.1186/s12913-023-0987.