



Smart Diabetes Prediction System Using Machine Learning Algorithms

IEEE BASE PAPER TITLE:

Improving Healthcare Prediction of Diabetic Patients Using KNN Imputed Features and Tri-Ensemble Model

IEEE BASE PAPER ABSTRACT:

Objective: Diabetes ranks as the most prevalent ailment in developing nations. Vital steps to mitigate the consequences of diabetes include early detection and expert medical intervention. A highly effective approach for identifying diabetes involves assessing the specific indicators associated with this condition. When it comes to automated diabetes detection, frequently encountered datasets frequently exhibit gaps in data, which can markedly impact the effectiveness of machine learning models. Methods: The aim of this study is to propose an automated method for predicting diabetes, with a focus on appropriately dealing with missing data and improving accuracy. The proposed framework makes use of K-Nearest Neighbour (KNN) imputed features along with a Tri-ensemble voting classifier model. Results: By incorporating the KNN imputer, the presented model demonstrates impressive performance metrics, including an accuracy of 97.49%, precision of 98.16%, recall of 99.35%, and an F1 score of 98.84%. The study conducted a thorough comparison of this proposed model against seven alternative machine learning algorithms, assessing them under two conditions: one with omitted missing values and another with the KNN imputer applied. These findings



support the proposed model's efficacy, highlighting its superiority over currently established state-of-the-art techniques. Conclusion: This research explores the problem of missing data in diabetes diagnosis and highlights the efficacy of the KNN-imputed technique. The results are promising for healthcare practitioners as they could facilitate early detection and improve the quality of diabetic patient care.

OUR PROPOSED PROJECT ABSTRACT:

Diabetes is a chronic medical condition that occurs when the body is unable to properly process and regulate blood glucose levels. It is a major global health concern, with both genetic and lifestyle factors contributing to its development. Early detection and prediction of diabetes can help in the management and prevention of complications, making predictive systems essential in healthcare today.

The "Smart Diabetes Prediction System Using Machine Learning Algorithms" is designed to predict the likelihood of diabetes in individuals based on various health parameters. This system is developed using Python for the backend and Flask as the web framework, with the frontend built using HTML, CSS, and JavaScript for a seamless user experience. The system utilizes four powerful machine learning models: Stacking Classifier, ExtraTree Classifier, LGBM Classifier, and CatBoost Classifier, which are trained and evaluated using the enhanced Pima Indians Diabetes Database. The dataset comprises 1,382 records with features such as Pregnancies, Glucose levels, Blood Pressure, Skin Thickness, Insulin, BMI, Diabetes Pedigree Function, Age, and Outcome.



Among the models, the Stacking Classifier delivers a Train Accuracy of 99% and a Test Accuracy of 92%. The ExtraTree Classifier achieves perfect Train Accuracy of 100% and a Test Accuracy of 93%. The LGBM Classifier, known for its efficiency in handling large datasets, also performs remarkably, with 100% Train Accuracy and a Test Accuracy of 95%. Finally, the CatBoost Classifier, a gradient boosting model, outperforms the others, achieving 100% accuracy on both the Train and Test datasets.

This system offers a high-accuracy, robust solution for early diabetes detection using advanced machine learning techniques with an intuitive web interface. The user inputs health data through the web interface and the model predicts the likelihood of diabetes, aiding healthcare providers in early diagnosis and preventive care. By integrating machine learning with accessible web technology, the system not only demonstrates the power of predictive analytics but also offers a practical tool for improving diabetes care and management.

SYSTEM REQUIREMENTS:

HARDWARE REQUIREMENTS:

- System : Pentium i3 Processor.
- Hard Disk : 500 GB.
- Monitor : 15'' LED.
- Input Devices : Keyboard, Mouse.
- Ram : 8 GB.



SOFTWARE REQUIREMENTS:

- Operating System : Windows 10 / 11.
- Coding Language : Python 3.12.0.
- Web Framework : Flask.
- Frontend : HTML, CSS, JavaScript.

REFERENCE:

Khaled Alnowaiser, “Improving Healthcare Prediction of Diabetic Patients Using KNN Imputed Features and Tri-Ensemble Model”, IEEE Access, vol. 12, pp. 16783-16793, 2024.