DIABETES PREDICTION USING AI

Overview:

Diabetes prediction is the estimation of the future trends and risk factors of diabetes, a chronic disease that affects the body's ability to use glucose. According to the CDC, the number of diagnosed cases of diabetes in the US increased by nearly half in the past decade and is expected to rise by 165 percent by 2050. The best clinical predictor of diabetes is adiposity (body fat), and the best biological predictor is baseline glucose (blood sugar). Genetic polymorphisms (variations in DNA) have little impact on the prediction of diabetes.

Project Title: AI Based Diabetes Prediction System

Software Component: The software used in our project is Google Colab or Jupiter Notebook.

Problem Statement:

Develop an AI-powered diabetes prediction system that leverages machine learning algorithms to analyze medical data and predict the likelihood of an individual developing diabetes, providing early risk assessment and personalized preventive measures.

Dataset Link: **https://www.kaggle.com/datasets/mathchi/diabetes-data-set**

Problem Definition:

The problem is to build an AI-powered diabetes prediction system that uses machine learning algorithms to analyze medical data and predict the likelihood of an individual developing diabetes. The system aims to provide early risk assessment and personalized preventive measures, allowing individuals to take proactive actions to manage their health.

Project steps:

The Project is divided into 5 phases.

Phase 1: Problem Definition and Design Thinking

In this phase, we'll define the problem of developing an AI-powered diabetes prediction system, understand the requirements, and plan the design of the system.

Phase 2: Innovation

In this phase, we can explore innovative techniques such as ensemble methods and deep learning architectures to improve the prediction system's accuracy and robustness.

Phase 3: Development Part

In this phase, we'll begin developing the diabetes prediction system by preparing the data and selecting relevant features.

Phase 4: Development Part 2

In this phase, we'll continue building the diabetes prediction system by selecting a machine learning algorithm, training the model, and evaluating its performance.

Phase 5: Project Documentation & Submission

In this phase, we'll document the entire diabetes prediction system and prepare it for submission.

Methodology:

Dataset Description:

The diabetes data set was originated from **https://www.kaggle.com/datasets/mathchi/diabetes-data-set** . Diabetes dataset containing 768 instances with 9 features. The objective is to predict if the patient is diabetic or not. The “Outcome” is the feature we are going to predict ,0 means No diabetes, 1 means diabetes.

Data Collection:

We need a dataset containing medical features such as glucose levels, blood pressure, BMI, etc., along with information about whether the individual has diabetes or not.

Data Preprocessing:

The medical data needs to be cleaned, normalized, and prepared for training machine learning models.

Feature Selection:

We will select relevant features that can impact diabetes risk prediction.

Model Selection:

We can experiment with various machine learning algorithms like Logistic Regression, Random Forest, and Gradient Boosting.

Evaluation:

We will evaluate the model's performance using metrics like accuracy, precision, recall, F1-score, and ROC-AUC.

Iterative Improvement:

We will fine-tune the model parameters and explore techniques like feature engineering to enhance prediction accuracy.

Conclusion and Future work:

One of the important real-world medical problem is the detection of diabetes at its early stage. In this study, systematic efforts are made in designing a system which results in the prediction of diabetes.During this work , the machine learning classification algorithms are studied and evaluated on various measures.

In future, the designed system with the used machine learning classification algorithms can be used to predict or diagnose other diseases. The work can be extended and improved for the automation of diabetes analysis including some other machine learning algorithms.