**1.ABSTRACT**

* The primary objectives of this project involve the preparation of a comprehensive dataset employing diverse methods and the development of a highly accurate predictive model for diabetes.
* To achieve these objectives, the project encompasses two key tasks. The first task involves the meticulous preparation of a robust dataset utilizing various methodologies. This dataset will serve as the foundation for training the predictive model.
* The second task revolves around the construction of a sophisticated model capable of delivering high accuracy in predicting the occurrence of diabetes.

**2.Introduction**

Diabetes is a chronic disease that occurs either when the pancreas does not produce enough insulin or when the body cannot effectively use the insulin it produces. Insulin is a hormone that regulates blood glucose. Hyperglycaemia, also called raised blood glucose or raised blood sugar, is a common effect of uncontrolled diabetes and over time leads to serious damage to many of the body's systems, especially the nerves and blood vessels.

**3. Existing Method**

* **Data Collection**: Utilization of various types of data, such as patient demographics, medical history, lifestyle factors, and possibly genetic information. Integration with wearable devices or sensors to collect real-time health data, such as glucose levels, physical activity, and sleep patterns. Feature Selection:
* **User Interface:** Development of user-friendly interfaces for healthcare professionals or end-users to input data, view predictions, and interpret results. Integration with mobile applications or web platforms for easy accessibility.
* **Risk Stratification:** Categorization of individuals into different risk groups based on their likelihood of developing diabetes. Providing personalized recommendations for lifestyle modifications.

**4.Implementation**

* This Data set Comprises of 769 rows of data
* Here I Used RandomForestClassifier to test and train the data
* The accuracy of the model is 0.75 out of 1

**Reason to choose Random Forest Classifier**:- It gives the best accuracy for predicating the outcome

Through some of the graph representation from which we can conclude data easily .

**5.Conclusion**

successful implementation of these tasks holds the potential to significantly contribute to early diagnosis and intervention strategies, thereby promoting healthier lifestyles and improving the overall management of diabetes. Through some of the graph representation from which we can conclude that Male are more affected than Female