AI - BASED DIABETES PREDICTION SYSTEM

- S.Vasanthi
- BE(ECE) III YEAR
- EMAIL ID: vasanthilithiya812@gmail.com
- NM ID: au513521106037

PROBLEM STATEMENT

• A system is used to predict whether a patient has diabetes based on some of its health related details such as BMI(Body Mass Index),blood pressure,Insulin,etc.

• This system is only for females as the dataset used to make this system exclusively belongs to the females

PROJECT OVERVIEW

• This work is used four types of kernels,Linear, polynomial,RBF, and sigmoid,to predict diabetes in the machine learning platform.

- The authors obtained diverse accuracies in different kernals, ranging between 0.69 and 0.82.
  - The SVM technique with radial basis kernel function obtained the highest accuracy of 0.82.

CURRENT DIABETES REPORT

PROBLEM DEFINITION

•Logistic regression (LR) is used to identify the risk factors for diabetes disease based on p value and odds radio(OR).

•We have adopted four classifiers like naive Bayes(NB), Decision tree(DT),Adaboost(AB),and random forest (RF) to predict the diabetic patient.

DATA SOURCE

•The diabetes prediction dataset is a collection of medical and domographic data from patients, along with their diabetes status (positive or negative).

•The data includes features such as age,gender,Body Mass Index(BMI), hypertension,heart disease, smoking history,HbA1c level ,and blood glucose level

DATA PREPROCESSING

- Clinicians seek a reliable one time diagnostic system for diabetes diagnosis without several blood sugar tests

- We framed working data preprocessing steps for features selection and missing value imputation.
- We developed a deep neural network model for accuratly predicting diabetes mellitus and determining the disease severity during diagnosis.
- A performance difference that ranges from 8.68% to 21.99% is achieved in comparison to State of the art.
- This work can be reproduced with the publicly available source code.

FEATURES SELECTION

- The model uses the Boruta feature selection algorithm,Kmeans++ unsupervised cluster learning algorithm and stacking ensemble learning method.

- There are three main contributions in the paper.

- Feature selection was used to extract valuable features from the dataset. Then tha dataset was trained using SVM,fuzzy rules were generated, and finally, the output was classified using the fuzzy inference method.

- The method achieves an accuracy of 89.02% on the PIMA Indian Diabetes dataset.