Predicting Diabetes Mellitus in ICU Patients

Hena Ghonia Shanaya Mehta

Introduction

Diabetes is a serious chronic health condition affecting millions of people across the planet. In the ICU, patients needing urgent care might not be able to provide all information beforehand to care providers. Giving the wrong medicine might affect patient outcomes negatively. We present here a comparative study of tree-based methods used to predict diabetes mellitus in ICU patients.

The dataset used in this study was from the WiDS datathon hosted on Kaggle, provided by the GOSSIS Consortium [1]. The data contains electronic patient health records from various countries such as the USA, Australia and the UK. From the 130,157 encounters in the dataset, the percent of positive samples is very low, approximately 21.62%. There are 59,573 females in the dataset and of these, 12,726 females have diabetes.

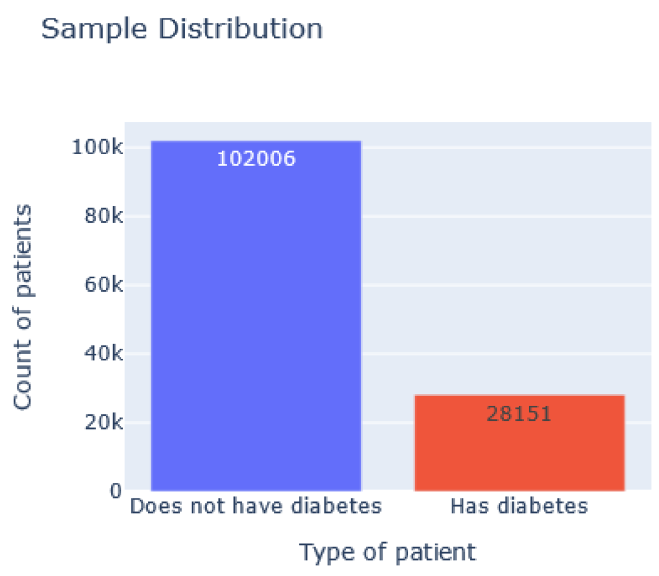

Figure 1. Dataset Imbalance distribution plot

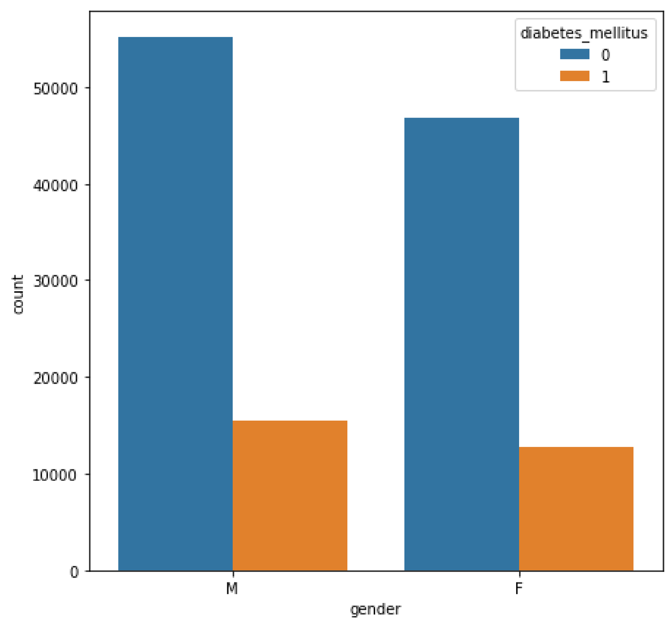

Figure 2. Gender distribution

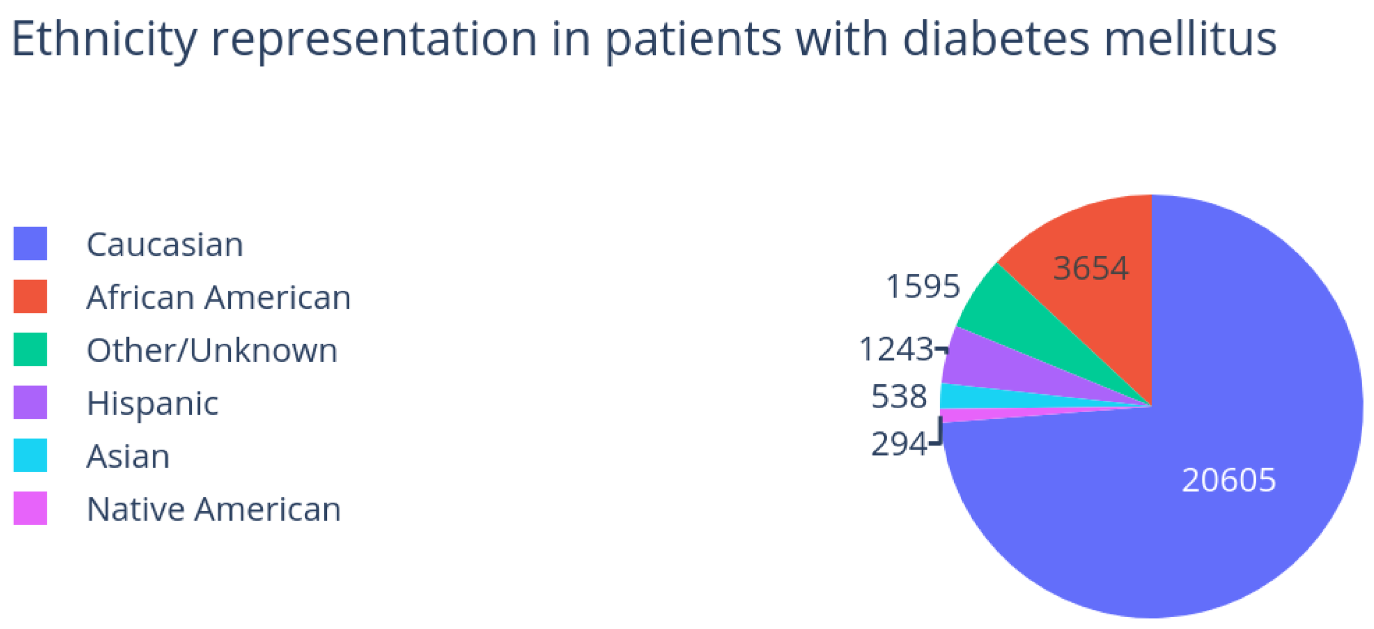

Figure 3. Ethnicity representation

Methodology

For feature engineering, target encoding was done for the categorical columns like ethnicity, gender, hospital admit source, icu admit source, icu stay type and icu type and highly correlated features were dropped.

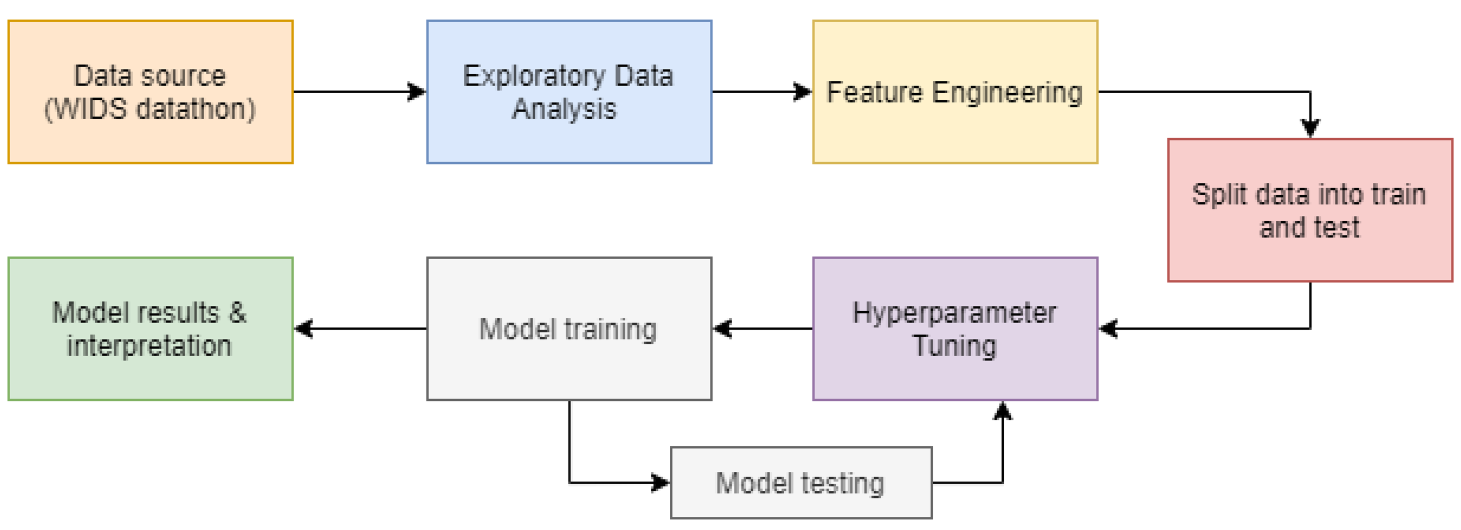

Figure 4. Methodology

We performed 5-fold cross validation, hyperparameter tuning and model training on 147 total features. The metric used for scoring the model was Area Under the Receiver Operator Characteristics Curve (AUCROC).

Models Used

Table 1 shows the models used for baseline setting. The best accuracy was achieved for Random Forest classifier with default parameters and so we set it as our baseline.

Classifier	Accuracy
Dummy Classifier	78.37%
Logistic Regression	80%
Random Forest	81.7%

Table 1. Accuracy comparison of various Baseline models

Table 2 shows all approaches used in this study which include weighted ensembles, stacked ensembles, autoML approaches and standalone boost-ing algorithms.

Algorithm	Model Details
XGBoost Ensemble	Ensemble of XGBoost, SVM, Random Forest
LightGBM	Ensemble of 2 LightGBMs
PyStackNet 1	2 layers, 2 classifiers: LightGBM & XGBoost
PyStackNet 2	2 Layers, 3 classifiers: CatBoost, LightGBM & XGBoost

Table 2. All algorithms used in this study

Results

As seen from the table, it is evident that the ensemble of LightGBM models outperformed the rest (95% confidence interval 0.8632-0.8902). The StackNets' performance is also similar; the addition of CatBoost to the ensemble only creates a slight difference.

Model	Valid AUC	Test AUC
Ensemble of LightGBMs	0.9860	0.8767
XGBoost with Ensemble of SVM and KNN	0.8450	0.8572
StackNet with XGBoost and LightGBM	0.8766	0.8680
StackNet with XGBoost, LightGBM and Catboost	0.8790	0.8683

Table 3. AUC Scores of all models

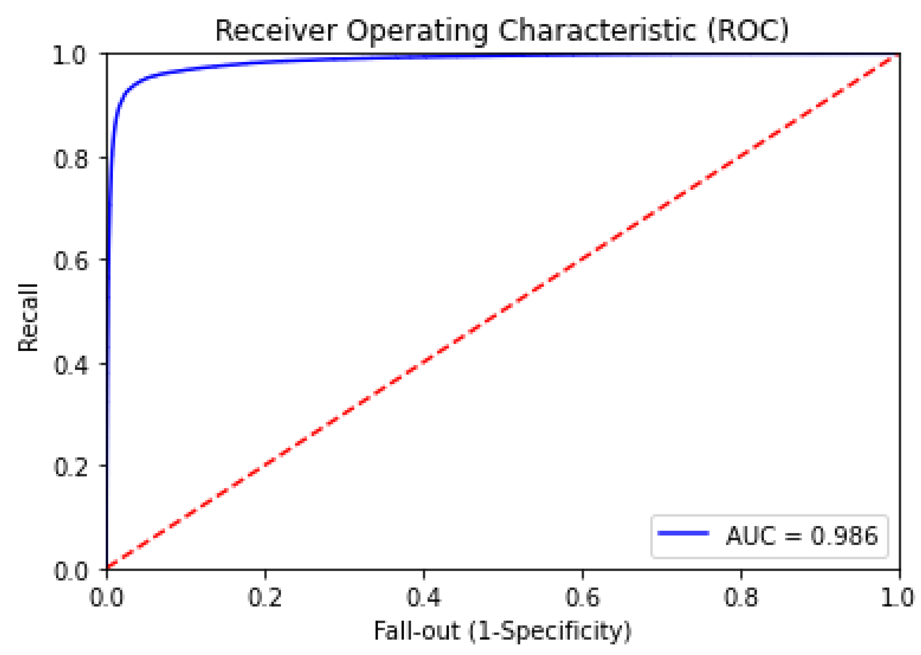

Figure 5. AUROC Curve for the validation set - LightGBM

Model Explainability

For medical data, being able to understand the results of the model training is more important than getting good accuracy or AUC score. In our study, we found that age, weight and BMI are important indicators of diabetes. These findings are in line with other research done in the field of diabetes mellitus.

From figure 6, it can be seen that higher values of glucose readings, age, weight and BMI have a high impact on the model's predicting diabetes mellitus whereas lower values of these variables has a lesser negative impact on the prediction. The feature icu_id shows up in the top-10 SHAP summary plot since 91% of icu_ids overlap between train data and test data.

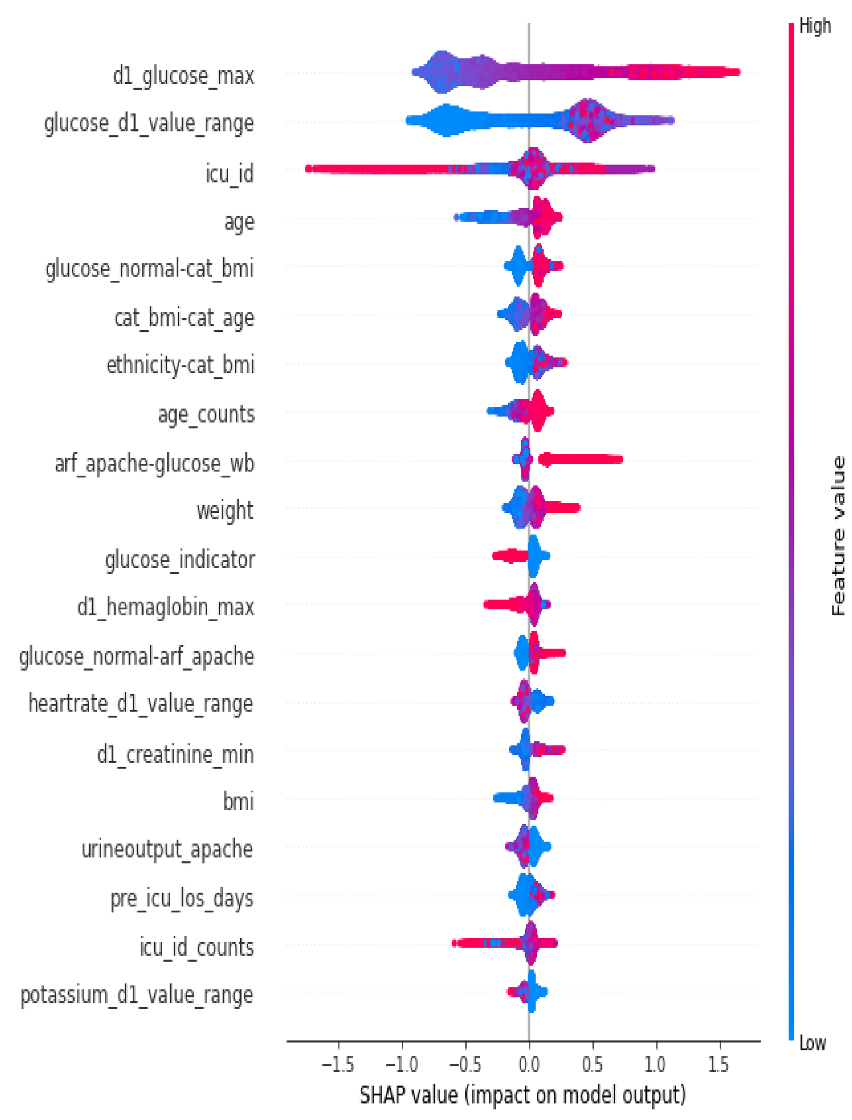

Figure 6. SHAP Summary Plot for LightGBM ensemble

References

[1] Jesse Raffa, Alistair Johnson, Leo Anthony Celi, Tom Pollard, David Pilcher, and Omar Badawi. 33: The global open source severity of illness score (gossis). *Critical Care Medicine*, 47(1):17, 2019.