



A Comparative Study on Machine Learning Based Type 2 Diabetes Mellitus Prediction

Weiye Zhan *

School of Computer Science, Guangdong Polytechnic Normal University, Guangzhou, China

* Corresponding author: wcx@bbc.edu.cn

Abstract. As is well known, one of the biggest health issues of our time is type 2 diabetes mellitus (T2DM). And it has been predicted that by 2045, its prevalence will have increased by more than 50% worldwide. Machine learning has emerged as a promising option for the prediction of diabetes after years of study in computational diagnosis of diabetes. However, the accuracy rate to date suggests that there is still a room for improvement. Using the PIMA Indian dataset, four machine learning methods (Hybrid Random Forests, Random Forests, XGBoost, and LGBM Classifiers) are examined for diabetes diagnosis and prediction. The data are prepared using Exploratory Data Analysis (EDA) and data standardization. The author also carried out auto-parameter tweaking using the optuna machine learning model. This project aims to create a data science model to Predict T2DM. For each of the methods proposed, the accuracy level is calculated as a percentage, and it is shown that Hybrid Random Forests with a Linear Model generate the highest degree of accuracy (86.4%).

Keywords: Hybrid Random Forests Linear Model (HRFLM), type 2 diabetes mellitus(T2DM), Optuna, Exploratory Data Analysis (EDA)

1 Introduction

A sign of the metabolic disorder diabetes mellitus is hyperglycemia. It typically manifests itself in several ways, including Prediabetes signifies having a glycemia that is greater than normal. And overt diabetes refers to gestational diabetes, also known as type I and type II. The World Health Organization (WHO) estimates that diabetes claims the lives of 1.6 million people per year. One condition that develops when blood glucose/blood sugar levels are extremely high in the body is diabetes. According to medical experts, diabetes is a condition that develops when the pancreas, a gland found in the human body, is unable to create enough insulin (Type 1 diabetes in this case), or when the body's cells are unable to use the insulin that is generated (which is Type 2 diabetes) [2]. One of the biggest health issues of our day is type 2 diabetes mellitus (T2DM). And it has been predicted that by 2045, its prevalence would have increased by more than 50% worldwide. [3]. Medical research has con-

nected diabetes to long-term damage to important organs such as the heart, blood vessels, kidneys, nerves, eyes, and renal function. Diabetes causes several problems for vital body parts, including the heart, kidneys, blood vessels, and nerves, if it is not properly controlled or diagnosed in a timely manner, which can be fatal [1],[4]. For the clinical diagnosis of diabetes, a plasma glucose level of more than 126 mg/dl during a fasting period or a plasma glucose result of more than 200 mg/dl following a two- or three-hour oral glucose tolerance test are both necessary [5]. And Diabetes is a growing threat nowadays, one of the reasons being that there is no perfect cure for it. Type 2 diabetes should be paid more attention to by human beings. However, the glycemic risks of various ethnic groups vary. As a result, regardless of the patient's ethnic origin, clinicians must choose a glycemic threshold for diabetes diagnosis and must take into account whether such a threshold may be reliable without a battery of follow-up tests to confirm the diagnosis [4]. It takes a lot of blood sugar testing to make a valid diagnosis during a single clinical visit since tests must be done both before and after meals. However, the diagnostic procedure may be made simpler computationally.

Machine learning has been regarded as a prominent option for diabetes prediction after studying in the computational diagnosis of diabetes. In the past few years, there have been some conventional machine learning models (CML) in computational diagnosis of diabetes [2]. The PIMA Indian diabetes dataset (PIDD) was utilized by several studies to predict diabetes using machine learning (ML) techniques. Alam, T.M., et al. demonstrated 75.7 percent accuracy on PIDD using the ANN algorithm [6]. Additionally, Tigga et al. predicted diabetes using Logistic Regression model on PIDD. [7]. They found that, of all the PIDD features, the number of pregnancies, blood glucose levels, and BMI degree are the most crucial indicators for predicting diabetes. Application of Random Forest, Decision Tree, and ANN for classification algorithms on PIDD was done after feature reduction using Principal Component Analysis (PCA) and Minimum Redundancy Maximum Relevance (mRMR) techniques in the work of Zou et al. [8]. They discovered that the random forest with the mRMR feature reduction technique produced the maximum accuracy (77.21%) for PIDD.

However, the accuracy rate to date suggests that there are still some rooms for development. The author of this article compared four machine learning algorithms including Random Forest, XGBoost, LGBM, and HRFLM for predicting and diagnosing T2DM using the PIDD. This project aims to make a data science model to predict T2DM. Over 60 million individuals (7.8% of the population) in India have diabetes, more than 30 million of whom are undiagnosed or untreated, increasing their risk of complications and early death. As a result, the PIDD is ideal for this research.

2 Method

2.1 Exploratory Data Analysis (EDA)

2.1.1 Descriptive Statistics.

Descriptive statistics are a type of summary statistic that, first and foremost, quantitatively explains or summarizes characteristics from a group of data. After a simple data exploration in the early stage, it is simple to determine that this dataset has no missing values. Table 1 then displays the dataset's descriptive data.

Table 1. Descriptive statistical results for various aspects of the properties of type 2 diabetes data.

	Preg-nan-cies	Glucose	Blood Pres-sure	Skin-Thick-ness	Insu-lin	BMI	Diabe-tesPedi-greeFunc-tion	Age
count	768.00	768.00	768.00	768.00	768.00	768.00	768.00	768.00
mean	3.85	120.89	69.11	20.54	79.80	31.99	0.47	33.24
std	3.37	31.97	19.36	15.95	115.24	7.88	0.33	11.76
min	0.00	0.00	0.00	0.00	0.00	0.00	0.08	21.00
25%	1.00	99.00	62.00	0.00	0.00	27.30	0.243	24.000
50%	3.00	117.00	72.00	23.00	30.50	32.00	0.37	29.00
75%	6.00	140.25	80.00	32.00	127.25	36.60	0.63	41.00
max	17.00	199.00	122.00	99.00	846.00	67.10	2.42	81.00

2.1.2 Correlation

Correlation explains how one or more variables are related to each other. The author used heatmap from seaborn library to visualize the Correlation which is shown in Figure 1(a). The magnitude of the connection between the variables may be calculated using the size of the correlation coefficient that corresponds to the color of the various blocks in the heat map. The following equation (1) is used to compute the correlation coefficient between two variables:

$$\rho_{x_1x_2} = \frac{Cov(X_1,X_2)}{\sqrt{DX_1}DX_2} = \frac{EX_1X_2 - EX_1 * EX_2}{\sqrt{DX_1 * DX_2}} \tag{1}$$

In equation (1), the correlation coefficient is represented by ρ . Cov stands for covariance and E for expectation or mean in mathematics. This correlation coefficient can only assess the linear correlation between variables. Small correlation coefficients cannot point to the absence of any other correlation relationship between the variables. Therefore, according to Figure 1 (b), the outcomes were the most relevant to glucose. Moreover, the second and third most relevant factors were BMI and age, respectively. And the age was the most relevant to pregnancies. Based on this correlation, it can be estimated that the diagnosis of type 2 diabetes is closely related to factors affecting blood glucose, BMI, and pregnancy.

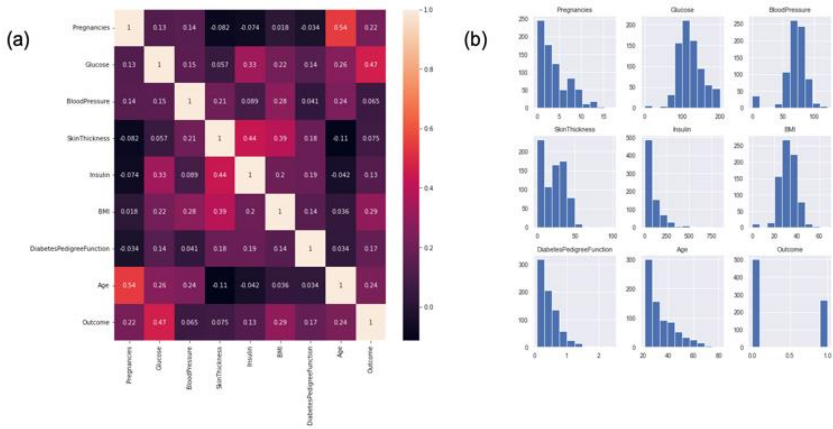


Fig. 1. (a) A heat map displays the degree of association between the distribution of their data and (b) the characteristics of type 2 diabetes data.

The scatter plot matrices are a great way to identify trends for subsequent analysis. It can be found from Figure 2 that the classification of positive (which means outcome is equal to 1) and negative (which means outcome is equal to 0) samples and their overall trend. Except for glucose, blood pressure, and BMI sample data, most of which are concentrated in the middle, the sample data of other features is mostly biased to the left. Also, the number of samples with an outcome of 0 is about twice the number of samples with an outcome of 1.



Fig. 2. This plot shows that positive and negative samples are roughly distributed. And negative samples are way more than positive samples.

2.1.3 Outlier Visualization.

The box plot's data shape would not change by outliers. Up to 25% of the data may be moved arbitrarily far away without significantly disrupting the quartiles, since quartiles have a certain resilience. The author used box plots to visualize data features

like Figure 3. If there are outliers in the data, which means outside the upper and lower edge areas, represented as dots. The outliers of the insulin data were the highest of all 9 characteristic values. This most likely indicates that the diagnosis of type 2 diabetes has a reference to insulin abnormalities. When the box plot is short, it means that many data points are spread over a small range, such as Pregnancies, Blood Pressure, BMI, Diabetes Pedigree Function and Age.

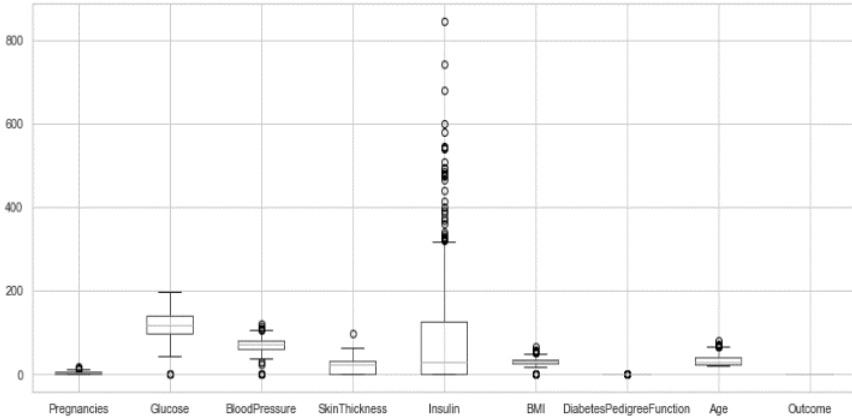


Fig. 3. Visually and clearly identify outliers for each attribute of type 2 diabetes data. Among them, insulin has the most outliers.

2.1.4 Distributions for Some Features.

As for pregnancies, the author used a point plot in Figure 4. From the chart, the positive and negative samples with more than 7 pregnancies were generally between 40 and 55 years old. And the largest number of samples was found with two pregnancies.

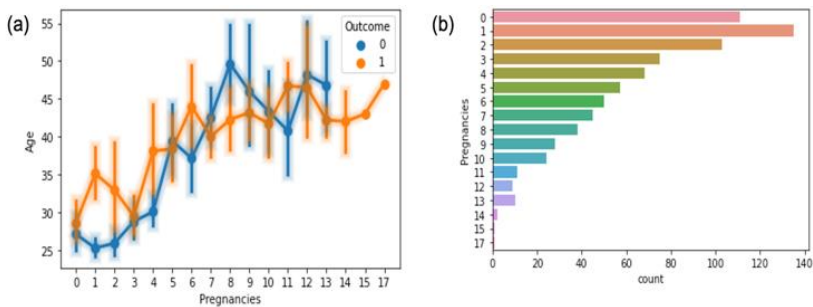


Fig. 4. (a) The positive and negative samples with more pregnancies were relatively older than the samples with fewer pregnancies. (b) The sample size shows a downward trend.

As for glucose from the diagrams in Figure 5 (a), negative samples with glucose above 150 diagnosed with type 2 diabetes were significantly increased. When blood glucose is above 75, the trend of negative samples diagnosed with type 2 diabetes

increases significantly. As for skin thickness with insulin from the diagram in Figure 5 (b), the positive and negative samples for skin thickness are mainly distributed between 0 and 60. And the samples of insulin are mainly distributed between 0 and 400. Among them, the negative sample can be approximated as more than twice as much as the positive sample.

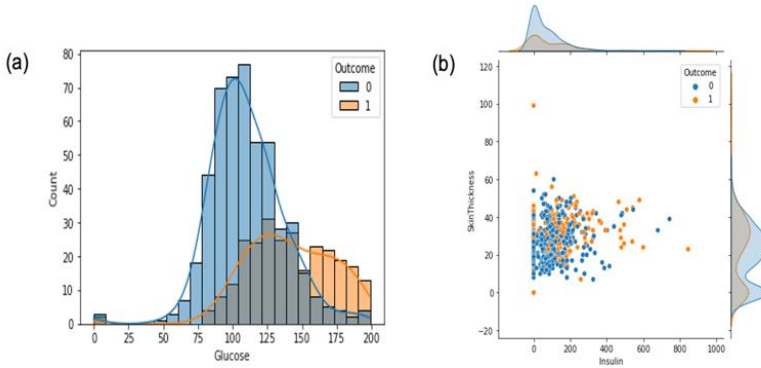


Fig. 5. (a) As the value of glucose changes, the number of positive and negative samples changes. (b) Skin thickness versus the distribution of insulin for positive and negative samples.

2.2 Data Processing

It is easy to know through data exploration that PIDD is clean, but the data is imbalanced. To deal with that, the author used oversampling and data standardization. The National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) first acquired the PIDD, which was used in this research [9]. The dataset includes details about 768 patients and the nine distinctive traits that correspond to them. The description of the characteristics in this dataset is shown in Table 2. Nine parameters are used to predict the occurrence of diabetes: pregnancies, glucose, BMI, skin thickness, insulin, age, function of the diabetes lineage, blood pressure, and outcome. The "outcome" attribute is regarded as a goal variable, whilst the other eight qualities are handled as feature variables. The study's author employed machine learning algorithms to more accurately determine if a patient had diabetes or not. Type 2 diabetes is far more likely to occur in obese adults. The average body mass index of the 768 patients, shown in Table 2, is 32. Due to the definition of obesity as having a BMI of 30 or more, the dataset is for people with T2DM [10].

Table 2. Statistical description of each attribute

Attribute	Description	Type	Average/Mean
Pregnancies	number of pregnancies.	Numeric	3.85
Glucose	Plasma glucose concentration was 2 hours in an oral glucose tolerance test.	Numeric	120.89
BloodPressure	Diastolic blood pressure (mm Hg).	Numeric	69.11

SkinThickness	Triceps skinfold thickness (mm).	Numeric	20.54
Insulin	2-hour serum insulin ($\mu\text{IU}/\text{mL}$).	Numeric	79.80
BMI	Body mass index (kg/m^2)	Numeric	32
DPF	Diabetes pedigree function.	Numeric	0.47
Age	Age (years).	Numeric	33
Outcome	Diabetes diagnosis results	Nominal	-

SMOTE is one of the oversampling methods to fix imbalanced datasets by increasing the minority data. The author chose this method because it is better compared with custom weighting by trial and error. And multiple feature variables of the same magnitude can influence the target variable, and each feature variable must have the same scale for the model to be trained with the same level of influence from different features on the parameters. For this reason, normalizing the dataset's data is necessary before modelling.

2.3 Machine Learning Classification Models

2.3.1 Random Forest Model.

To choose the most crucial significant property, the RF algorithm employs a straightforward predefined probability. Breiman created the RF approach by creating numerous decision trees from a sample of feature subspaces that were randomly mapped to sample data subsets [11]. The following is a description of the RF algorithm: (1) Initial: Replacement sampling with an initial probability of D_1, \dots, D_K . (2) D_K builds an own decision tree model for every document which are chosen at random from the available features using the m-try dimension's subspace. The best data split is produced by the leaf node. Up till the saturation criteria is reached, the procedure will be continued. (3) Create a random forests ensemble using the K trees $h_1(X_1)$, $h_2(X_2)$, etc., which haven't been trimmed. And apply the high probability value for classification decisions [13].

2.3.2 Xgboost Classifier.

The XGBoost method systematically instructs decision trees using training data [12]. To raise the goal function's value, the method adds a new decision tree to the prior decision trees at each iteration. The regularization term (Ω) in equation (2) and the loss term from equation (2) make up the objective function that is intended to be minimized. Equation (2) provides the objective function for the t-th iteration (L^t), where n is the training set's number of examples, y_i is the actual class label for example i, \hat{y}_i is the class label of instance i for prediction, f_k is the function of the tree, and Ω is regularization term.

$$L^t = \sum_{i=1}^n \left[l(y_i, \hat{y}_i^{t-1} + f_t(x_i)) \right] + \Omega(f_t) \quad (2)$$

To penalize the complexity of the model and prevent overfitting, where γ and λ are the hyperparameters, w is the weight of each leaf, and T is the number of leaves in the tree in equation (3).

$$\Omega(f_t) = \gamma T_t + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \tag{3}$$

The training set's number of examples is n in equation (4), where g_i stands for the first order gradient statistics on the loss in equation (5), and h_i stands for the second order gradient statistics on the loss in equation (6).

$$L^t \cong \sum_{i=1}^n \left[l(y_i, \hat{y}_i^{t-1}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t) \tag{4}$$

$$g_i = \partial_{\hat{y}_i^{t-1}} l(y_i, \hat{y}_i^{t-1}) \tag{5}$$

$$h_i = \partial_{\hat{y}_i^{t-1}}^2 l(y_i, \hat{y}_i^{t-1}) \tag{6}$$

2.3.3 Lightgbm Classifier.

Microsoft introduced LGBM, a new gradient boosting library, in 2017. The objective was to speed up gradient boosting for decision trees. When making new leaves, it should just examine a subset of the splits. The author just iterated over all of the buckets. The creators of this implementation refer to it as a histogram implementation. In contrast to previous gradient boosting techniques, the trees were developed depth first (or leaf wise) while maintaining the presorted state, in Figure 6. The algorithm was chosen to grow the leaf with the greatest delta loss.

When compared to the other algorithms used, LGBM provides the dataset evaluated with the highest level of accuracy. The algorithm will decide which data to include in the dataset and train itself based on the results once the user has divided the training and testing split. Attempts are made to forecast using the remaining data. The final findings are derived from the training and testing data and checked against the model predicting, which will output the outcome.

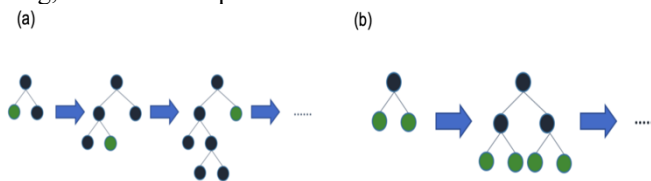


Fig. 6. (a) Leaf wise growth deeply (b) Level wise growth widely

2.3.4 Hybrid Random Forests Linear Model

In this paper, the author used the Logistic Regression model as HRFLM's linear model. The Hybrid Random Forest with Logistic Regression Model is a relatively novel approach in diabetes prediction. This HRFLM infers three random forests models and a Logistic Regression model. In this model, the author used a voting classifier to ensemble models. First, a plurality of base models based on different algorithms are obtained through training. And the training model predicts the test set following optu-

na tweaking. Finally, the output results of other models are fused and voted, using the rule that indicates the minority obeys the majority to finally decide the prediction result [14]. For example, the author now chooses logistic regression, XGBoost, LightGBM, and random forest as the base models for training, and then uses the outputs of the base models to determine the final true classification result.

3 Results

3.1 Classification Accuracy on Different Testing Sample Ratios And Different Models

The classification accuracy with 10%, 20%, and 40% of samples as test sets in each dataset is displayed in Figure 7 to analyze how well more clearly these models work when the training samples are insufficient. It can be observed that the HRFLM model's accuracy grows from 10% to 40% of the test set. And when the test set accounts for 10%, the accuracy of the HRFLM model reaches its best, about 86.4%. It is easy to know that, in general, the smaller the proportion of test samples, the better the results. The larger the training samples in proportion, the more data can be contained. Therefore, a test set of 10% is selected for subsequent performance evaluation. In addition, regardless of how the proportions of the test set change, the accuracy of the HRFLM model is the best.

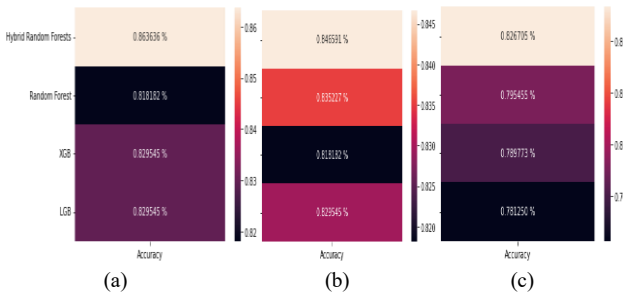


Fig. 7. Classification accuracy with different ratios of dataset when training samples are insufficient. (a) 10% samples as test set, (b) 20% samples as test set, (c) 40% samples as test set.

3.2 Comparison of HRFLM, RF, XGB and LGB

To further illustrate the performance of the HRFLM as a powerful predictor, the author used a voting classifier to ensemble the Logistic Regression Model and 4 Random Forests Models, which together form a Hybrid Random Forest with Logistic regression Model. In this paper, the parameters of these classifiers are shown in Table 3. The ratio of testing dataset is 10%. From Table 3, HRFLM has the greatest AUC rating, which is closest to 1. As is well known, the application performs better overall the closer the AUC score is to value 1 [15]. In addition, the recall score of HRFLM is equal to the recall score of RF. And the AUC scores between HRFLM and RF are so

close. The author should utilize the F1 score to evaluate the accuracy of the binary classification models more accurately as it takes into consideration both the accuracy and recall of the classifiers and could be thought of as a kind of harmonic average of the two. According to Table 3, the F1 score of HRFLM is closest to 1. Therefore, the HRFLM classification effect is the best.

Table 3. Statistical description of each attribute

Methods	AUC Score	Sensitivity	Recall score	F1 Score
HRFLM	0.949126	0.686	0.973	0.86
RF	0.942236	0.706	0.973	0.85
LGB	0.936407	0.765	0.919	0.83

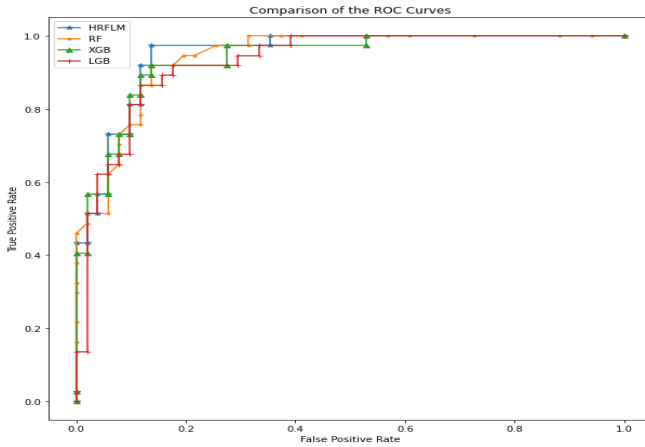


Fig. 8. The analysis of ROC curve comparison between HRFLM, RF, XGB and LGB.

In Table 4, the HRFLM model has much longer testing and training times than other models across all datasets. According to the model itself, the test time of HRFLM inflates when more complex integrated strategies are used. For the RF and XGB models, especially LGB, the total time consumed declines. The total time of LightGBM is the fastest of these models. LightGBM accelerates training by using a histogram-based approach that encodes continuous eigenvalues into discrete bins. This algorithm separates the leaves of the tree into an ideal fit. Therefore, the leaf-wise approach will minimize the loss more quickly than the level-wise strategy in LightGBM when grown on the same leaf.

Table 4. Comparison of the training and testing time about HRFLM, RF, XGBM and LGBM.

Methods	Training time	Testing time	Total
HRFLM	0.441	0.036	0.477
RF	0.334	0.020	0.354
XGB	0.360	0.003	0.363
LGBM	0.055	0.0	0.055

4 Discussion

In this paper, type 2 diabetes prediction by HRFLM was being studied. When predicting the T2DM, the author looked at how well machine learning-based prediction models performed, such as HRFLM, RF, LightGBM, and XGBoost. Machine learning algorithms can help healthcare clinicians, service users, and laypeople with improved T2DM treatment and management when combined with other ideas proposed in the learning healthcare systems approach [6]. In addition, when using HRFLM to predict the T2DM, HRFLM performed with higher accuracy compared to the other three models of RF, XGB, and LGB. And it can be observed clearly that the model effects of XGB and LGB are often very close. This might be because both XGB and LGB are optimising algorithms based on decision trees. And the model we used, which is called HRFLM, is brand new for type 2 diabetes prediction. Similar techniques have been applied to the prognosis and forecasting of various illnesses, including heart disease [16],[17]. When introducing novel prediction models, it is important to evaluate both the advantages and disadvantages of machine learning techniques in addition to the predictive performance.

However, there are still some places to improve HRFLM performance. According to the AUC results, the performance of HRFLM and the Random Forests model on the classification results is very close. But RF is faster than HRFLM's training speed. The time cost of HRFLM training is higher than the original random forests model, which might lead to inconvenience for new dataset fitting. In addition, the sensitivity of HRFLM is lower than other models, which indicates that this model is slightly more likely to miss other patients with type 2 diabetes than other models. This is a large part of the optimization that needs to be improved in the later model. To overcome these flaws in this model, much more delicate optimizing tricks will be introduced in future work. To make the necessary medical advances and treat Diabetes Mellitus Disease, the Hybrid Random Forests algorithm may also be further conducted utilizing the stacking method.

5 Conclusions

One of the biggest health issues of our day is type 2 diabetes mellitus (T2DM). The author used HRFLM for T2DM in this paper, which consists of Logistic Regression and 4 random forests as the classifier. Also, nowadays a very popular automatic hyperparameter optimization software framework, optuna is used in this paper. The HRFLM model is fully evaluated with a bunch of well-designed experiments. In addition, HRFLM is the first application for type 2 diabetes prediction for now, which is also highly creative. When the testing set was set by 10%, the results demonstrate that the Hybrid Random Forest with Linear Model produced the best accuracy with a percentage of 86.4% when using the PIMA Indian Dataset, which indicates the effectiveness of HRFLM. Additionally, it may be applied to the future development of a method for identifying and forecasting diabetes mellitus.

References

1. C. Olisah Chollette, L. Smith, M. Smith, "Diabetes mellitus prediction and diagnosis from a data preprocessing and machine learning perspective, " *Computer Methods and Programs in Biomedicine*, Volume 220, 106773, ISSN 0169-2607, 2022.
2. J. K. Jobeda, Y. Foo Simon, "A comparison of machine learning algorithms for diabetes prediction, " *ICT Express*, Volume 7, Issue 4, ISSN 2405-9595, 2021, pp. 432-439.
3. L. Perreault, J.S. Skyler & J. Rosenstock, "Novel therapies with precision mechanisms for type 2 diabetes mellitus, " *Nat Rev Endocrinol* 17, 2021, pp. 364–377.
4. Akiba, Takuya, et al. "Optuna: A next-generation hyperparameter optimization framework," *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2019.
5. R.M.M. Khan; Z.J.Y. Chua; J.C. Tan; Y. Yang; Z. Liao; Y. Zhao, "From Pre-Diabetes to Diabetes: Diagnosis, Treatments and Translational Research, " *Medicina* 55, 2019, p. 546.
6. L. Kopitar, P. Kocbek, L. Cilar, et al. "Early detection of type 2 diabetes mellitus using machine learning-based prediction models,". *Sci Rep* 10, 11981, 2020.
7. Malone, John I., and Barbara C. Hansen. "Does obesity cause type 2 diabetes mellitus (T2DM)? Or is it the opposite?," *Pediatric diabetes* 20.1, 2019,pp. 5-9.
8. Alam, Talha Mahboob, et al. "A model for early prediction of diabetes," *Informatics in Medicine Unlocked* 16: 100204, 2019.
9. Tigga, Neha Prerna, and Shruti Garg. "Predicting type 2 diabetes using logistic regression," *Proceedings of the fourth international conference on microelectronics, computing and communication systems*. Springer, Singapore, 2021.
10. Zou Q, Qu K, Luo Y, et al. "Predicting diabetes mellitus with machine learning techniques, " *Frontiers in genetics*, 2018, p. 515.
11. V. Jackins, et al. "AI-based smart prediction of clinical disease using random forest classifier and Naive Bayes," *The Journal of Supercomputing* 77.5, 2021, pp. 5198-5219.
12. H. Kumar Sharma, and Amrit Lal Sangal. "Hybrid Model for Heart Disease Prediction Using Random Forest and Logistic Regression," *Innovations in Computer Science and Engineering*. Springer, Singapore, 2021, pp. 657-662.
13. De Silva, Kushan, et al. "Use and performance of machine learning models for type 2 diabetes prediction in community settings: A systematic review and meta-analysis," *International journal of medical informatics* 143: 104268, 2020.
14. U. Kumar, Karthik, MB Sai Nikhil, and K. Sumangali. "Prediction of breast cancer using voting classifier technique," *2017 IEEE international conference on smart technologies and management for computing, communication, controls, energy and materials (ICSTM)*. IEEE, 2017.
15. Sisodia et al. D.Sisodia, "Prediction of Diabetes using Classification Algorithms, " *a.Procedia Comput. Sci*,132, 2018, pp. 1578–1585
16. A. Kondababu, V. Siddhartha, BHK. Bhagath Kumar, Bujjibabu Penumutchi, "A comparative study on machine learning based heart disease prediction, " *Materials Today: Proceedings*, ISSN 2214-7853, 2021.
17. U. S. N. V. K. R. C. S. T. S. Mamatha Singh, "A Hybrid Random Forest Linear Model approach to predict the Heart Disease," *Annals of the Romanian Society for Cell Biology*, 25(6), 2021, pp. 7810–7814.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

