



OPEN ACCESS

EDITED BY

Muhammad Azeem Ashraf,
Hunan University, China

REVIEWED BY

Melanie Stephens,
University of Salford, United Kingdom
Nadia Shabnam,
National University of Medical Sciences
Rawalpindi, Pakistan
Rita Chhikara,
The NorthCap University, India

*CORRESPONDENCE

Yimin Zhang
✉ zhang_yimin@163.com

RECEIVED 29 April 2024

ACCEPTED 11 July 2024

PUBLISHED 07 August 2024

CITATION

Du Q, Wang D and Zhang Y (2024) The role of artificial intelligence in disease prediction: using ensemble model to predict disease mellitus.

Front. Med. 11:1425305.

doi: 10.3389/fmed.2024.1425305

COPYRIGHT

© 2024 Du, Wang and Zhang. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

The role of artificial intelligence in disease prediction: using ensemble model to predict disease mellitus

Qinyuan Du, Dongli Wang and Yimin Zhang*

Key Laboratory of Traditional Chinese Medicine Classical Theory, Ministry of Education, Shandong University of Traditional Chinese Medicine, Jinan, China

The traditional complications of diabetes are well known and continue to pose a considerable burden to millions of people with diabetes mellitus (DM). With the continuous accumulation of medical data and technological advances, artificial intelligence has shown great potential and advantages in the prediction, diagnosis, and treatment of DM. When DM is diagnosed, some subjective factors and diagnostic methods of doctors will have an impact on the diagnostic results, so the use of artificial intelligence for fast and effective early prediction of DM patients can provide decision-making support to doctors and give more accurate treatment services to patients in time, which is of great clinical medical significance and practical significance. In this paper, an adaptive Stacking ensemble model is proposed based on the theory of "error-ambiguity decomposition," which can adaptively select the base classifiers from the pre-selected models. The adaptive Stacking ensemble model proposed in this paper is compared with KNN, SVM, RF, LR, DT, GBDT, XGBoost, LightGBM, CatBoost, MLP and traditional Stacking ensemble models. The results showed that the adaptive Stacking ensemble model achieved the best performance in five evaluation metrics: accuracy, precision, recall, F1 value and AUC value, which were 0.7559, 0.7286, 0.8132, 0.7686 and 0.8436. The model can effectively predict DM patients and provide a reference value for the screening and diagnosis of clinical DM.

KEYWORDS

diabetes mellitus, disease prediction, machine learning, artificial intelligence, Stacking ensemble model

1 Introduction

Diabetes mellitus (DM) is a metabolic disease clinically characterized by chronic hyperglycemia, dyslipidemia and protein abnormalities, and other symptoms that increase the risk of morbidity and mortality, of which type 2 diabetes mellitus (T2DM) is a major public health challenge globally, and the assessment and management of this chronic disease carries a heavy economic burden (1–3). Worldwide, 537 million adults (aged 20–79) have diabetes (10%), and this number is expected to rise to 643 million by 2030 and 783 million by 2045 (4, 5).

Along with the rapid development of the intersection of artificial intelligence and medical diagnostics, machine learning (ML) has once become the most concerned topic among researchers, which can provide accurate predictive analysis of diseases, effectively identify high-risk factors as well as patients with high morbidity, and then provide accurate decision support for hospital administrators (6). By mining potential healthcare data through machine learning and constructing a novel DM prediction model, early warning of high-risk groups can be performed, and appropriate healthcare management can be taken to patients in advance, which can also provide certain decision support to doctors and reduce the rate of missed diagnosis and misdiagnosis (7).

Initially, researchers predicted DM through traditional machine learning and verified that random forest (RF) based on tree model has better prediction effect (8–12). Some researchers focused on the preliminary data processing to get a better DM prediction model using feature selection and data imbalance processing (13–15). Meanwhile, considering the influence of different factors on diabetes, researchers began to study the three aspects of age, gender and geography, and obtained a better prediction effect of the targeted population prediction model (16–18).

In recent years researchers began to consider the use of ensemble learning to predict diabetes and obtained diabetes prediction models that are superior to traditional machine learning (19, 20). In DM prediction, although machine learning is superior to traditional statistical methods, most of the research has focused on a single prediction model. Each prediction model has its advantages, disadvantages and limitations, and researchers use ensemble learning to combine the advantages of a single prediction model to build a more powerful ensemble model, of which the most effective Stacking ensemble model is gradually applied (21–23). From the perspective of single classification model, using Stacking ensemble model can solve the limitations of single classification model, but the selection of base classifiers and meta-learners of Stacking ensemble model has randomness (24).

To solve the above-mentioned problem, this paper follows the theory of “error-ambiguity decomposition” (25) and designs an adaptive Stacking ensemble model, which can further improve the prediction performance of DM model.

2 Materials and methods

In this paper, an adaptive Stacking ensemble model is designed and implemented on the DM dataset to predict DM patients, and the overall process framework is shown in Figure 1. Firstly, the DM dataset was normalized, followed by feature screening using the gradient boosting decision tree (GBDT) (26) feature selection method, and next the adaptive Stacking ensemble model was constructed by first adaptively selecting the base classifiers from n pre-selected models, and then traversing the models in the selection of meta-learners. The performance of the adaptive Stacking ensemble model, proposed in this paper was evaluated by five evaluation metrics: accuracy, precision, recall, F1 value and AUC value.

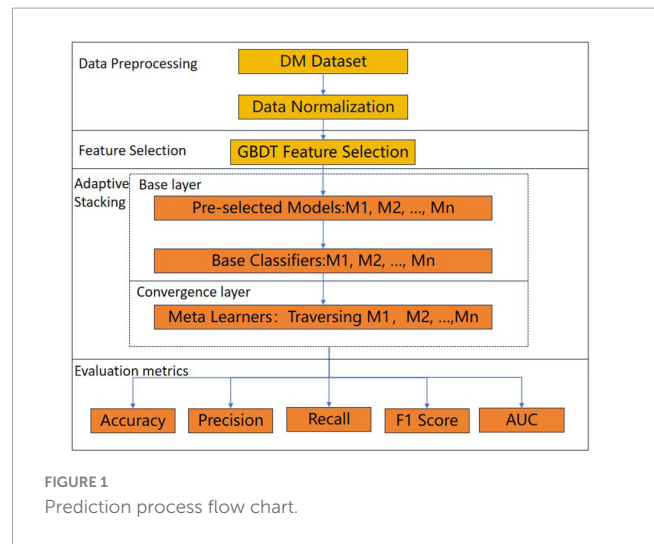


TABLE 1 CDC’s BRFSS2015.

ID	Feature	Detailed description
1	HighBP	High blood pressure
2	HighChol	High cholesterol
3	CholCheck	Had a cholesterol test within 5 years
4	BMI	Body mass index
5	Smoker	Smoker
6	Stroke	Stroke
7	HeartDiseaseor	Coronary heart disease or myocardial infarction
8	PhysActivity	Physical activity in the last 30 days
9	Fruits	Fruit 1 or more times per day
10	Veggies	Veggie 1 or more times per day
11	Alcoholic	Alcoholic
12	AnyHealthcare	Have any type of health insurance
13	NoDocbcCost	In the past 12 months, have you needed to see a doctor but were unable to do so due to cost?
14	GenHlth	Health status
15	MentHlth	Mental health status
16	PhysHlth	Health status
17	DiffWalk	Have severe difficulty walking or climbing stairs
18	Sex	Sex
19	Age	Age
20	Education	Educational level
21	Income	Income situation

2.1 Dataset

The dataset selected for this paper is a balanced dataset processed on CDC’s BRFSS2015, which has the same proportion of

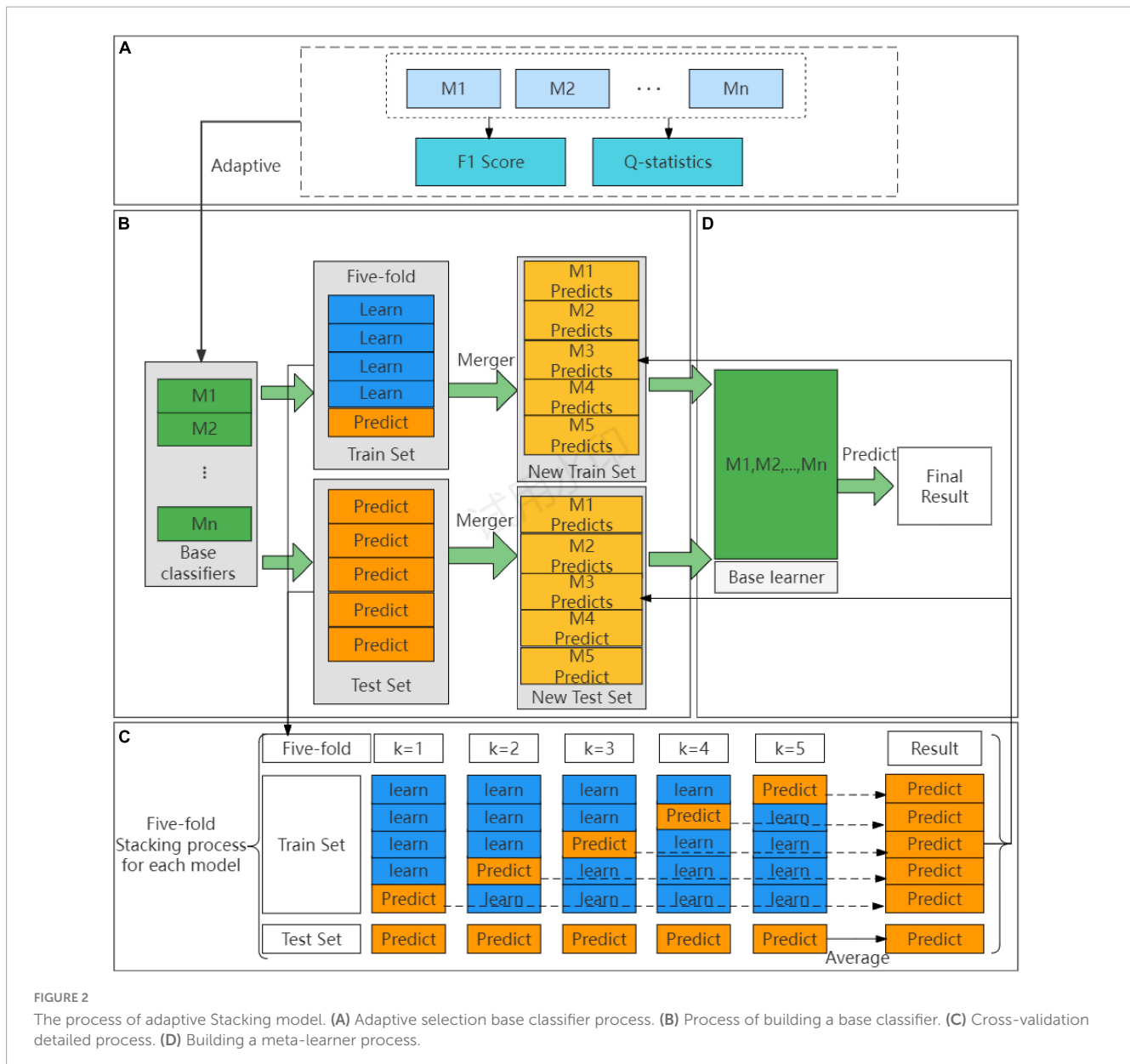


FIGURE 2 The process of adaptive Stacking model. (A) Adaptive selection base classifier process. (B) Process of building a base classifier. (C) Cross-validation detailed process. (D) Building a meta-learner process.

TABLE 2 Relationship between two pre-selected model.

	M_k correct (1)	M_k wrong (0)
M_i correct (1)	N_{11}	N_{10}
M_i wrong (0)	N_{01}	N_{00}

diabetic and non-diabetic interviews, totaling 70,691 samples. This dataset contains 21 characteristic variables as shown in Table 1.

2.2 Feature selection

Feature selection is an important technique in machine learning to filter out the most valuable and relevant sample features in the data for use in building machine learning models (27, 28). The purpose of feature selection is to reduce the number of sample features in the data, improve the accuracy and operational

efficiency of the model, reduce the risk of overfitting, and improve the interpretability of the model (29). Not all sample features have a significant impact on the prediction results, which contains many sample features with low or irrelevant contribution to the prediction results, and too many sample features will cause computational resource consumption and reduce the training speed of the model, and may also reduce the accuracy of the model, so this chapter is to eliminate the sample features with low or irrelevant contribution to the prediction results. There are three common feature selection methods: filter, wrapper, and embedding (30). Since embedding methods have better predictive performance than filter methods and run much faster than wrapper methods (31), our study uses the embedding method GBDT to select feature variables.

GBDT is an ensemble learning method that improves the predictive performance of a model by constructing a series of weak learners (usually decision trees). The basic idea of GBDT is to

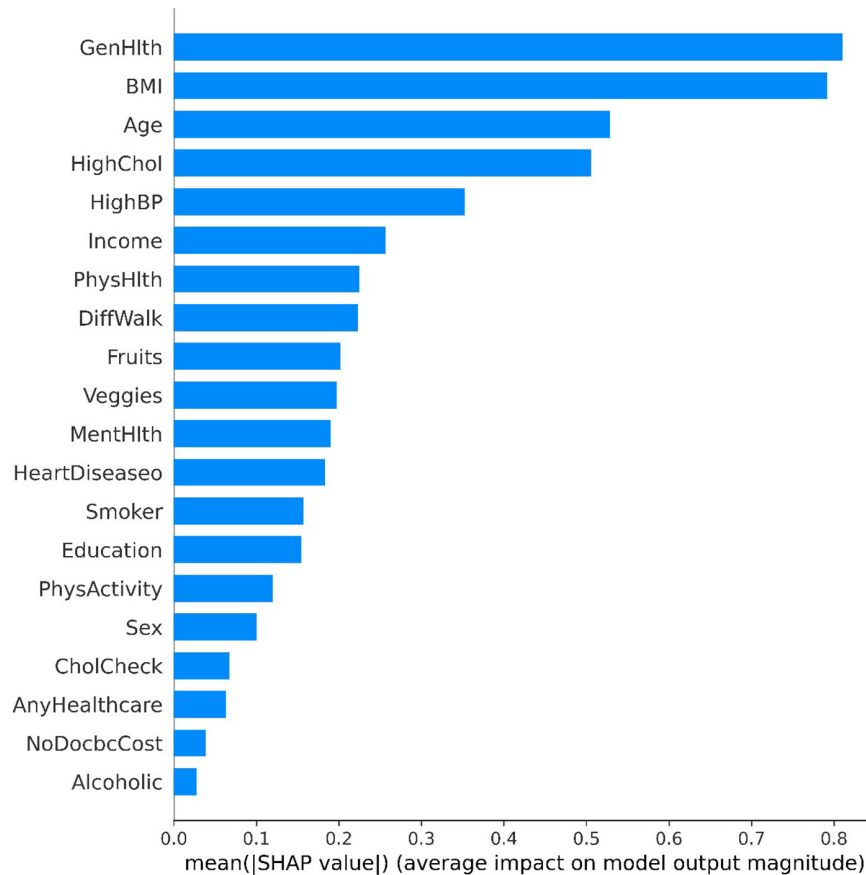


FIGURE 3 Global feature contributions.

combine multiple simple models (weak learners) so that each new model corrects as much as possible the errors of the previous model. Since the decision tree splits based on the importance of the features and features with high importance are more frequently selected as split points, GBDT can rank the importance of the features and handle high dimensional data. Secondly, GBDT can handle different types of features, including numerical and categorical. Finally, GBDT focuses on the residuals of the current model rather than modeling the target values directly, which makes the model more tolerant to noisy data.

When we use GBDT with embedding method to obtain feature importance ranking, one of the problems we face is the inability to accurately interpret the impact of individual features on the final prediction results. To solve this problem, we used a technique called feature interpolation method (32). This method represents the explanatory model as a linear function of the feature interpolation values, thus providing a clearer understanding of the model behavior. By this method we can reduce the repetition rate and better understand the contribution of individual features to the prediction results. The method is formulated as follows.

$$l(z') = \vartheta_0 + \sum_{i=1}^N \vartheta_i z'_i \tag{1}$$

Where N is the number of features, ϑ_i is the value of the feature attribute of the feature, $z'_i = 0$ or 1 to indicate whether the feature

is observed or not, where the feature attribute can be regarded as the “feature contribution.”

In order to compute the ϑ_i values in Equation 1, a tree-value estimation method based on game theoretic ideas (33), the SHAP method, is introduced as feature attribute values. In this method, the model f and the set S contain non-zero indexes in z' and each feature has the classical Shapley value attribute ϑ_i , which is formulated in Equation 2.

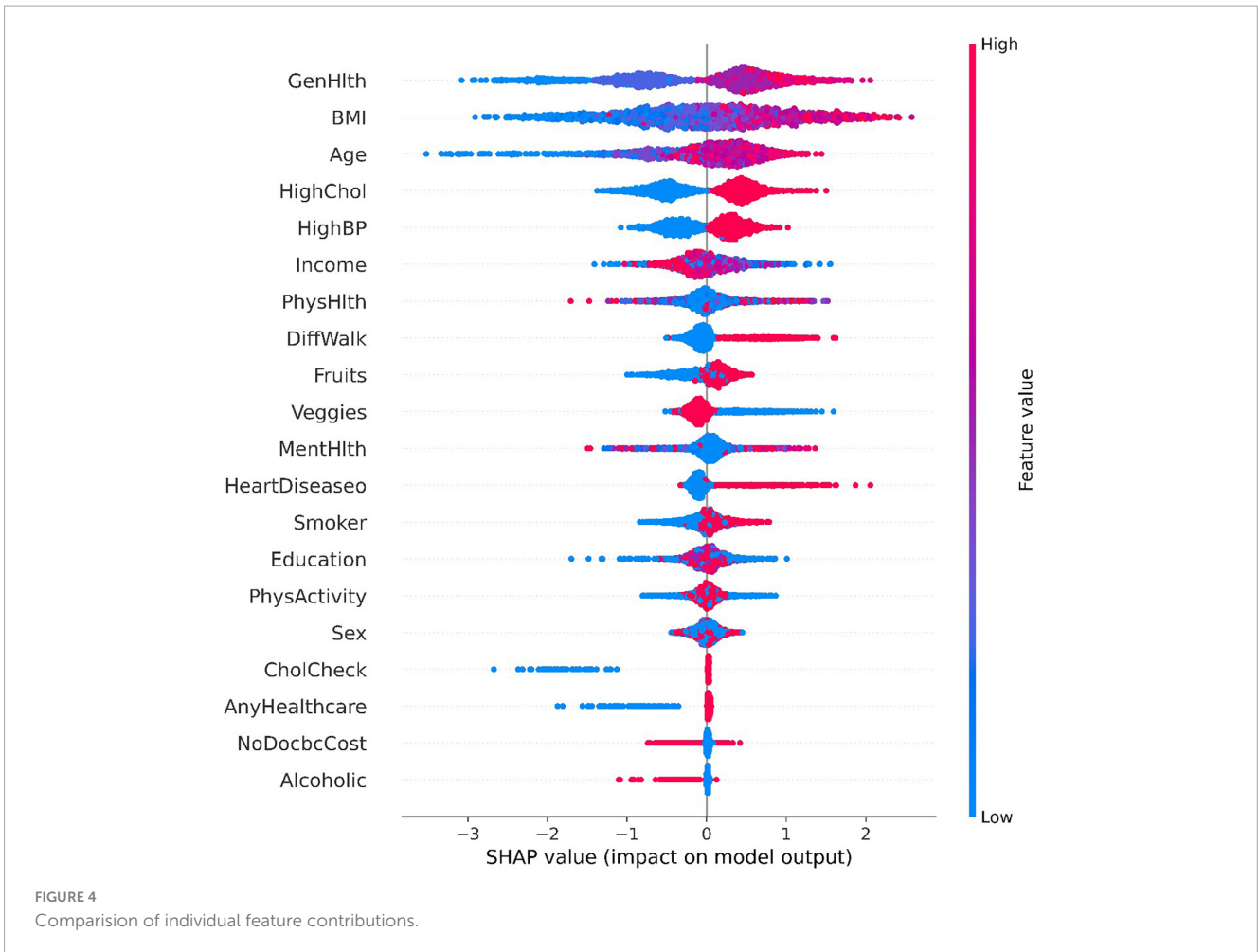
$$\vartheta_i = \sum_{S \in M\{i\}} \frac{|S|!(N - |S| - 1)!}{N!} [f(S \cup \{i\}) - f(S)] \tag{2}$$

where M is the set of all input features.

The SHAP method is a locally accurate, personalized feature attribution method that, unlike tree model gain, provides consistent global feature attribute results (34). In our study, we use the SHAP method for feature filtering and interpretation of individual feature attributes, which helps to reduce the repetition rate and provide a clearer understanding of the impact of each feature on the results.

2.3 Model building

When the base classifiers predict accurately, the greater the variability of the base classifiers, the better the integration of the model will be, which is the famous “error- ambiguity



decomposition” theory. This implies that the variability of the base classifiers should be taken into account while guaranteeing the prediction effect of the base classifiers.

In principle, as long as the base classifier of the Stacking ensemble model predicts well (35), the number of layers of the Stacking ensemble model can be stacked infinitely, but this increases the complexity of the model. Therefore, we try to reduce the model complexity as much as possible while ensuring the prediction effect of the model, and only select the Stacking ensemble model, with the two-layer structure of the base layer and the convergence layer. To solve the problem of randomness of the traditional Stacking ensemble model, in selecting base classifiers and meta-learners, this paper proposes an adaptive Stacking ensemble model, and the process is shown in Figure 2.

The first step is to adaptively construct the base classifier of the Stacking base layer, as shown in Figure 2A. When choosing the base classifiers, the traditional Stacking ensemble model usually select the classifiers with good prediction effect, ignoring the principle of “error- ambiguity decomposition.” In this paper, we design a method to adaptively construct the base classifiers of Stacking, from the pre-selected models of M_1, M_2, \dots, M_n , according to the F1 value of the comprehensive evaluation metrics, we set a threshold to adaptively select the pre-selected models from the high to the low to construct the base classifiers, which is a step to ensure the prediction effect of the models. To ensure the variability of the

models, this paper chooses the Q-statistics method (36) to compare the variability between the pre-selected models.

The detailed steps of the Q-statistics method are as follows, labeling the DM dataset as $Z = z_1, z_2, \dots, z_n$, the pre-selected model as M_i , after using n pre-selected models for classification prediction, if the pre-selected model M_i predicts the ensemble correctly it will be 1, and the prediction is wrong it will be 0, and the relationship between the two pre-selected models is shown in Table 2.

The differential Q-statistics for the two pre-selected model M_i, M_k are shown in Equation 3 as follows.

$$Q_{i,k} = \frac{N_{11}N_{00} - N_{01}N_{10}}{N_{11}N_{00} + N_{01}N_{10}} \quad (3)$$

Statistically, the expected value of two completely independent pre-selection models $Q_{i,k}$ is 0. The range of Q is between $[-1, 1]$, and the smaller the absolute value, the greater the variability between the pre-selection models. In the L pre-selection models of this paper, the average value of Q is shown in Equation 4.

$$Q_{avg} = \frac{2}{L(L-1)} \sum_{L-1}^{i=1} \sum_{L}^{k=i+1} Q_{i,k} \quad (4)$$

The steps for adaptively constructing a base classifier for the Stacking ensemble model, are shown below:

Algorithm 1 Adaptive Stacking algorithm.

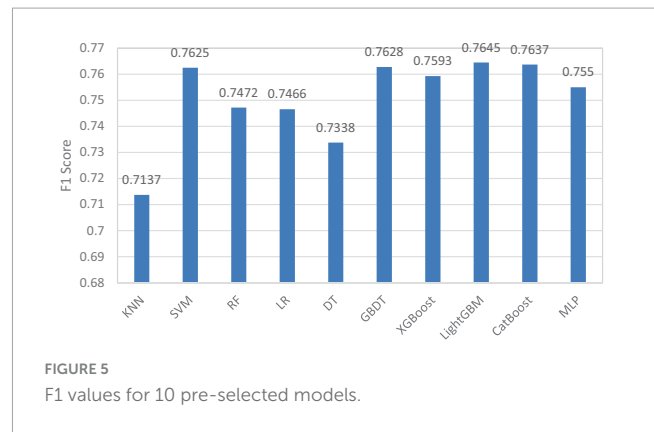
Symbols are defined: threshold λ , F_i^N is the F1 value of M_i , \odot is the model satisfying the discreteness condition, M is the set of base classifiers, the train set T and the test set S , $T_{f_i}^N$ is the training set for the five-fold cross-validation, $S_{f_i}^N$ is the test set for the five-fold cross validation, ϵl is the base classifier, LR_{meta} is the meta learners, T^l is the train set for ϵl , and S^l is the test set for ϵl

1. For $N = [M_1, M_2, \dots, M_n]$ // Preselection model
2. If $F_i^N > \lambda$ // Determine if the F1 value of model M_i is greater than a threshold value
3. If $M = \emptyset$ // Determine if M is empty
4. $M = M + M_i$ // Model M_i is added to M
5. If $M_i \odot M_j$ // Determine whether two models satisfy the difference condition
6. $M = M + M_j$ // Model M_j is added to M
7. For $l = M = [M_1, M_2, \dots, M_n]$ // Base classifiers
8. For $k = 1, 2, \dots, 5$
9. $T_{f_i}^l \rightarrow \odot l$ // Use $T_{f_i}^l$ to train ϵl
10. $S_{f_i}^l \rightarrow \epsilon l \rightarrow train_f, S^l \rightarrow \epsilon l \rightarrow test_f$ // Predict $S_{f_i}^l, S^l$ by ϵl to get $train_f, test_f$
11. $train_1 = (train_1 + train_2 + \dots + train_5)$ // Vertical stack
12. $test_1 = (test_1 + test_2 + \dots + test_5)/5$ // Level average
13. $train_{new} = [train_1, train_2, \dots, train_5]$ and $train_{new} = [train_1, train_2, \dots, train_5]$
14. $train_{new} \rightarrow LR_{meta}$ //Train LR_{meta} with $train_{new}$
15. $test_{new} \rightarrow LR_{meta} \rightarrow result_{pre}$ //Predict $test_{new}$ with LR_{meta} to get the final result
16. Return $result_{pre}$ // Returns the final prediction

- (1) Train n pre-selected models M_1, M_2, \dots, M_n .
- (2) Calculate the comprehensive evaluation metrics F1 value of the n pre-selected models, and set the threshold as the average F1 value.
- (3) Eliminate the models with F1 values smaller than the threshold and retain the models with F1 values larger than the threshold.
- (4) Models with small variance are eliminated and models with large variance are retained based on Q-statistics.
- (5) Select M_1, M_2, \dots, M_n as the final base classification.

Figure 2B shows the cross-validation part of the adaptive Stacking ensemble model. After adaptive selection as base classifiers, the train set and test set are divided according to an 8:2 ratio. In the train set, each base classifier using five-fold cross-validation. Taking five base classifiers M_1, M_2, \dots, M_5 as an example, the specific operation is shown in Figure 2D. A base classifier can get five predictions, which are vertically stacked into a one-dimensional matrix. Five base classifiers can be combined into a five-dimensional matrix as a new train set for the convergence layer. In the test set, again each base classifier using five-fold cross-validation and again five predictions are obtained. To ensure the division ratio between the train set and test set, the predictions of the test set are horizontally averaged to obtain a one-dimensional matrix. The predictions of the five base classifiers are combined into a five-dimensional matrix that serves as a new test set for the convergence layer.

The second step adaptively constructs the meta-learner of the Stacking ensemble model. As shown in Figure 2C, this paper



traverses the whole base classifier model to select the meta-learner, and obtains the final prediction result through the meta-learner.

Finally, to better understand the implementation process of the Adaptive Stacking algorithm, this paper gives the pseudo-code of the Adaptive Stacking algorithm, as shown in Algorithm 1.

Typically, machine learning evaluates the performance of a model using True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) metrics. The commonly used accuracy, precision, recall and F1 score are calculated from these metrics, which can be calculated by referring to Equations 5–8.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{8}$$

3 Results and discussion

3.1 Results of feature selection and analysis

The tree SHAP method calculates the individual contribution value of each feature in the sample dataset. Figure 3 demonstrates the global feature contribution in the GBDT model, the horizontal coordinate represents the sample feature contribution, the larger the value, the more important the sample feature is, and the vertical coordinate is the sample feature based on the feature importance from the largest to the smallest. From the figure, it can be seen that the features “GenHlth” and “BMI” have significant contribution degrees, which indicates a strong correlation with diabetes. While the inverse features “CHolCheck,” “AnyHealthcare,” “NoDocbcCost,” and “Alcoholic” features have less than 0.1 contribution, this paper directly excludes these four features and finally retains 16 sample features.

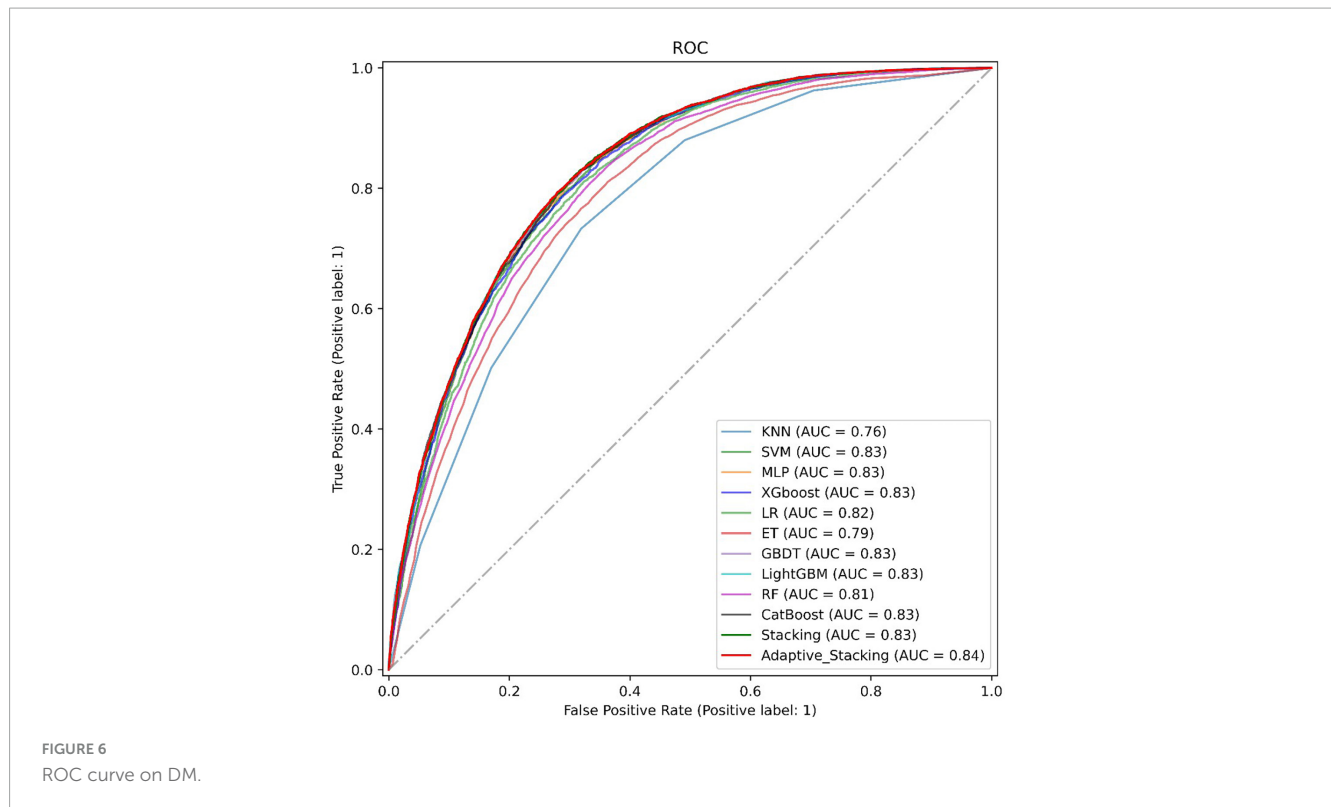


TABLE 3 Results of classifiers.

Model	Accuracy	Precision	Recall	F1 score
KNN	0.7068	0.6954	0.733	0.7137
SVM	0.74.9	0.7217	0.8081	0.7625
RF	0.7371	0.7176	0.7794	0.7472
LR	0.741	0.7289	0.7652	0.7466
DT	0.7232	0.7048	0.7652	0.7338
GBDT	0.7536	0.73.3	0.7949	0.7628
XGBoost	0.7493	0.7283	0.7932	0.7593
LightGBM	0.7547	0.7331	0.7987	0.7645
CatBoost	0.7557	0.7336	0.8008	0.7637
MLP	0.7493	0.7362	0.7746	0.755
Stacking	0.7554	0.7307	0.8007	0.7668
Adaptive Stacking	0.7559	0.7286	0.8132	0.7686

The results with the best predictions are highlighted in bold.

The Tree SHAP values depend on how the features are interpreted, so we can obtain the feature interpretation for each sample from the model (37). Figure 4 presents some information about the contribution of individual features to the model output and details how their values affect the model. the x-axis indicates the magnitude of the feature contribution, while the magnitude of the feature value is indicated by the color of the different points. The highest contribution of characteristics, “GenHlth” indicates that the poorer the physical condition the more likely to get diabetes, and similarly from the second and third ranked “BMI” and “Age,” it can be seen that the higher the weight coefficient the more likely to get diabetes, and the higher the age the more likely to get diabetes.

3.2 Results of the proposed adaptive Stacking ensemble model

In the adaptive Stacking ensemble model, the 10 pre-selected models are sorted according to the size of the F1 value, and the sorting results are shown in Figure 5. The horizontal coordinates in the figure indicate the 10 pre-selected models and the vertical coordinates indicate the F1 values of the pre-selected models.

In this paper, we set the average of F1 values as the threshold and first excluded four pre-selected models, K-nearest neighbor (KNN), RF, logistic regression (LR) and decision tree (DT). Four models, support vector machine (SVM), multilayer perceptron (MLP) and XGBoost and CatBoost, are selected as base classifiers from the remaining six pre-selected models adaptively based on Q-statistics. The top-ranked GBDT model and LightGBM model have better prediction results, but the difference between them and CatBoost model is proved to be small by Q-statistics method, so these two models are eliminated in adaptive way. In the selection of meta-learner, this paper traverses all the base classifier models and selects the optimal meta-learner according to the evaluation metrics.

In this paper, the prediction results of 10 pre-selected models, traditional Stacking ensemble model, and the adaptive Stacking ensemble model proposed in this paper are compared by four evaluation metrics: Accuracy, Precision, Recall and F1 value, and the comparison results are shown in Table 3. It can be seen that compared with the traditional machine learning models KNN and SVM, the traditional Stacking ensemble model outperforms the single classification model in predicting DM. The adaptive Stacking ensemble model proposed in this paper has very high accuracy, recall and F1 values of 0.7559, 0.8132 and 0.7668, which are

higher than the traditional Stacking ensemble model. The adaptive Stacking ensemble model proposed in this paper is completely better than the traditional Stacking ensemble model, and can adaptively make the best adjustment. This shows that the adaptive Stacking ensemble model has obvious advantages in predicting DM.

Finally, this paper plots the ROC curves of 11 representative models with the adaptive Stacking ensemble model, as shown in Figure 6. In the dataset of this paper, the ensemble learning model predicts better than traditional machine learning, and the AUC value of the ensemble learning model reaches 0.83. Among the Stacking ensemble model is better than the Bagging model and the Boosting model. The AUC value of the adaptive Stacking ensemble model reaches 0.84, which is higher than the traditional Stacking ensemble model. This shows that the adaptive superposition ensemble model proposed in this paper has excellent prediction effect in predicting DM.

4 Conclusion

In this paper, we propose a DM prediction model based on adaptive Stacking and analyze it in comparison with 10 pre-selected models and traditional Stacking ensemble model by five evaluation metrics: accuracy, precision, recall, F1 value and AUC value. The results show that the adaptive Stacking ensemble model proposed in this paper outperforms other models in several evaluation metrics, with accuracy, precision, recall, F1 value, and AUC value of 0.7559, 0.7268, 0.8132, 0.7686, and 0.8436, which suggests that the adaptive Stacking ensemble model, proposed in this paper is able to integrate the advantages of a single model and adaptive selection of pre-selected models to obtain better prediction results, which can provide clinical diagnostic advice and decision support for doctors and provide patients with appropriate medical and health management as early as possible. Although the adaptive Stacking model proposed in this study has a better prediction effect compared to a single model, the model complexity is high, and the complexity of the model needs to be further optimized according to the actual application scenarios. In addition, this study is limited to the study of machine learning models, and will collect more datasets and try to use deep learning models for the study. Finally, we stay on top of recently released healthcare policies, continually communicating with local hospitals and healthcare professionals to collect data on patient metrics.

References

- Choudhury AA, Rajeswari VD. Gestational diabetes mellitus-A metabolic and reproductive disorder. *Biomed Pharmacother.* (2021) 143:112183.
- Onuigwe F, Ambi H, Uchechukwu N, Obeagu E. Platelet dysfunction in diabetes mellitus. *Elite J Med.* (2024) 2:1–7.
- Vlacho B, Rossell-Rusiñol J, Granado-Casas M, Mauricio D, Julve J. *Overview on chronic complications of diabetes mellitus. Chronic complications of diabetes mellitus.* Amsterdam: Elsevier (2024).
- Alam MM. *Prevalence of type 2 diabetes mellitus complications in human.* Chittagong: Chattogram Veterinary & Animal Sciences University (2021).
- Ikwuka AO, Omoju DI, Mahanera OK. Profiling of clinical dynamics of type 2 diabetes mellitus in patients: A perspective review. *World J Curr Med Pharm Res.* (2023) 5:210–8.
- Picca A, Ronconi D, Coelho-Junior HJ, Calvani R, Marini F, Biancolillo A, et al. The “development of metabolic and functional markers of dementia in older people” (Odino) study: Rationale, design and methods. *J Pers Med.* (2020) 10:22.
- Alanazi HO, Abdullah AH, Qureshi KN. A critical review for developing accurate and dynamic predictive models using machine learning methods in medicine and health care. *J Med Syst.* (2017) 41:1–10.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: <https://www.kaggle.com/datasets/alexteboul/diabetes-health-indicators-dataset>.

Author contributions

QD: Data curation, Software, Writing – original draft, Writing – review & editing. DW: Data curation, Supervision, Formal analysis, Project administration, Writing – review & editing. YZ: Writing – original draft, Writing – review & editing.

Funding

The authors declare that financial support was received for the research, authorship, and/or publication of this article. This study was supported by the National Natural Science Foundation of China (No. 81173183), the Natural Science Foundation of Shandong Province (No. ZR2020KH003), the Shandong Province Higher Educational Institutions' Youth Creative Talent Inducement Programmer Project (Year.2021), and the Open Project of the Key Laboratory of Classical Theory of Traditional Chinese Medicine of the Ministry of Education.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

8. Kopitar L, Kocbek P, Cilar L, Sheikh A, Stiglic G. Early detection of type 2 diabetes mellitus using machine learning-based prediction models. *Sci Rep.* (2020) 10:11981.
9. Ganie SM, Malik MB. Comparative analysis of various supervised machine learning algorithms for the early prediction of type-II diabetes mellitus. *Int J Med Eng Inform.* (2022) 14:473–83.
10. Laila UE, Mahboob K, Khan AW, Khan F, Taekeun W. An ensemble approach to predict early-stage diabetes risk using machine learning: An empirical study. *Sensors.* (2022) 22:5247.
11. Muhammad L, Algehyne EA, Usman SS. Predictive supervised machine learning models for diabetes mellitus. *SN Comput Sci.* (2020) 1:240.
12. Malik S, Harous S, El-Sayed H. Comparative analysis of machine learning algorithms for early prediction of diabetes mellitus in women. *Proceedings of the international symposium on modelling and implementation of complex systems.* Cham: Springer (2020). p. 95–106.
13. Rodríguez-Rodríguez I, Rodríguez J-V, Woo WL, Wei B, Pardo-Quiles D-J. A comparison of feature selection and forecasting machine learning algorithms for predicting glycaemia in type 1 diabetes mellitus. *Appl Sci.* (2021) 11:1742.
14. Ali, MS, Islam MK, Das AA, Duranta D, Haque M, Rahman MH. A novel approach for best parameters selection and feature engineering to analyze and detect diabetes: Machine learning insights. *BioMed Res Int.* (2023) 2023.
15. Tasin I, Nabil TU, Islam S, Khan R. Diabetes prediction using machine learning and explainable AI techniques. *Healthc Technol Lett.* (2023) 10:1–10.
16. Chou C-Y, Hsu D-Y, Chou C-H. Predicting the onset of diabetes with machine learning methods. *J Pers Med.* (2023) 13:406.
17. Abegaz TM, Ahmed M, Sherbeny F, Diaby V, Chi H, Ali A. Application of machine learning algorithms to predict uncontrolled diabetes using the all of us research program data. *Healthcare.* (2023) 11:1138.
18. Su Y, Huang C, Yin W, Lyu X, Ma L, Tao Z. Diabetes Mellitus risk prediction using age adaptation models. *Biomed Signal Process Control.* (2023) 80:104381.
19. Zohair M, Chandra R, Tiwari S, Agarwal S. A model fusion approach for severity prediction of diabetes with respect to binary and multiclass classification. *Int J Inform Technol.* (2024) 16:1955–65.
20. Doğru A, Buyrukoglu S, Arı M. A hybrid super ensemble learning model for the early-stage prediction of diabetes risk. *Med Biol Eng Comput.* (2023) 61:785–97.
21. Mienye ID, Sun Y. Improved heart disease prediction using particle swarm optimization based stacked sparse autoencoder. *Electronics.* (2021) 10:2347.
22. Cui S, Yin Y, Wang D, Li Z, Wang Y. A stacking-based ensemble learning method for earthquake casualty prediction. *Appl Soft Comput.* (2021) 101:107038.
23. Liang M, Chang T, An B, Duan X, Du L, Wang X, et al. A stacking ensemble learning framework for genomic prediction. *Front Genet.* (2021) 12:600040. doi: 10.3389/fgene.2021.600040
24. Liu J, Dong X, Zhao H, Tian Y. Predictive classifier for cardiovascular disease based on stacking model fusion. *Processes.* (2022) 10:749.
25. Zhou Z-H, Wu J, Tang W. Ensembling neural networks: Many could be better than all. *Artif Intell.* (2002) 137:239–63.
26. Liang W, Luo S, Zhao G, Wu H. Predicting hard rock pillar stability using GBDT, XGBoost, and LightGBM algorithms. *Mathematics.* (2020) 8:765.
27. Khalid S, Khalil T, Nasreen S. A survey of feature selection and feature extraction techniques in machine learning. *Proceedings of the 2014 science and information conference.* New York, NY: IEEE (2014). p. 372–8.
28. Solorio-Fernández S, Carrasco-Ochoa JA, Martínez-Trinidad JF. A review of unsupervised feature selection methods. *Artif Intell Rev.* (2020) 53:907–48.
29. Dougherty ER. Feature-selection overfitting with small-sample classifier design. *Proceedings of the IEEE computer society 10662 LOS VAQUEROS CIRCLE, PO BOX 3014.* Los Alamitos, CA: (2005).
30. Yang P, Huang H, Liu C. Feature selection revisited in the single-cell era. *Genome Biol.* (2021) 22:1–17.
31. Senliol B, Gulgezen G, Yu L, Cataltepe Z. Fast correlation based filter (FCBF) with a different search strategy. *Proceedings of the 2008 23rd international symposium on computer and information sciences.* Berlin: IEEE (2008). p. 1–4.
32. Parsa AB, Movahedi A, Taghipour H, Derrible S, Mohammadian AK. Toward safer highways, application of XGBoost and Shap for real-time accident detection and feature analysis. *Accident Anal Prev.* (2020) 136:105405.
33. Lundberg SM, Erion GG, Lee S-I. Consistent individualized feature attribution for tree ensembles. *arXiv [Preprint].* (2018). arXiv:1802.03888.
34. Holzinger A, Saranti A, Molnar C, Biecek P, Samek W. Explainable AI methods—a brief overview. *Proceedings of the international workshop on extending explainable AI beyond deep models and classifiers.* Springer (2022). p. 13–38.
35. Soleymanzadeh R, Aljasim M, Qadeer MW, Kashaf R. Cyberattack and fraud detection using ensemble stacking. *AI.* (2022) 3:22–36.
36. Barbosa JC, Willoughby P, Rosenberg CA, Mrtek RG. Statistical methodology: VII. Q-methodology, a structural analytic approach to medical subjectivity. *Acad Emerg Med.* (1998) 5:1032–40.
37. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. *Adv Neural Inform Process Syst.* (2017) 30:4768–77.