

Prediction of fasting plasma glucose and glycated haemoglobin using machine learning based on tongue features

Jun Li (✉ ljzy05@126.com)

Shanghai University of Traditional Chinese Medicine <https://orcid.org/0000-0001-5624-2146>

Jingbin Huang

Shanghai University of Traditional Chinese Medicine

Liping Tu

Shanghai University of Traditional Chinese Medicine

Longtao Cui

Shanghai University of Traditional Chinese Medicine

Ji Cui

Shanghai University of Traditional Chinese Medicine

Xuxiang Ma

Shanghai University of Traditional Chinese Medicine

Tao Jiang

Shanghai University of Traditional Chinese Medicine

Xinghua Yao

Shanghai University of Traditional Chinese Medicine

Jiacai Li

Shanghai University of Traditional Chinese Medicine

Yulin Shi

Shanghai University of Traditional Chinese Medicine

Zijuan Bi

Shanghai University of Traditional Chinese Medicine

Yu Wang

Shanghai University of Traditional Chinese Medicine

Hongyuan Fu

Shanghai University of Traditional Chinese Medicine

Jue Wang

Shanghai University of Traditional Chinese Medicine

Yenting Lin

Shanghai University of Traditional Chinese Medicine

ChingHsuan Pai

Shanghai University of Traditional Chinese Medicine

Xiaojing Guo

Shanghai University of Traditional Chinese Medicine

Changle Zhou

Xiamen University

Xiaojuan Hu

Shanghai University of Traditional Chinese Medicine

Jiatuo Xu

Shanghai University of Traditional Chinese Medicine

Research

Keywords: Chinese medicine, Diabetes, Fasting plasma glucose, Glycated haemoglobin, Non-invasive, Machine learning, Risk prediction, Tongue diagnosis

Posted Date: August 18th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-58084/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background

Given tongue features and basic features, this study aimed to develop and assess a non-invasive machine learning model to perform regression prediction on fasting plasma glucose and glycated haemoglobin which will help optimize diabetes risk warning.

Methods

We collected the basic features, tongue features and blood features of the subjects. Using machine learning algorithms to analyze these data, we built models to predict fasting plasma glucose and glycated haemoglobin. Then the performance of the models was evaluated through 5-fold cross-validation results and test set results.

Results

The results of cross validation on the training set showed that given non-invasive input features, the minimum average mean square error of fasting plasma glucose and glycated haemoglobin prediction was 1.227 and 0.438. Our non-invasive fasting plasma glucose prediction model with tongue features and basic features combined achieved a minimum mean square error of 0.601 and a maximum coefficient of determination of 0.606 on the test set. The glycated haemoglobin prediction model product a minimum mean square error of 0.272 and a maximum coefficient of determination of 0.539 on the test set. The Clarke's Error Grid Analysis showed that the non-invasive blending model had 90.83% of points in zone A and 8.49% of points in zone B on the test set.

Conclusions

We developed an effective non-invasive method for estimating fasting plasma glucose and glycated haemoglobin from tongue features and basic features combined, which may help identify individuals at high risk for diabetes.

Background

Diabetes takes a significant toll on health care systems and public health[1]. The prevalence rate is currently 9.3%, and there are 463 million people aged 20–79 with DM globally. It is estimated that the number will reach 578.4 million by 2030. The number of diabetic patients in China is as high as 116.4 million, ranking the first in the world. Only in 2019, the medical cost of diabetic patients in China is as high as \$109 billion, second only to the United States[2]. China is the country with the most severe diabetes prevalence. Diabetes affects multiple organ systems and is associated with multiple vascular

and non-vascular complications[3]. Excess morbidity and premature death place a heavy burden and far-reaching impact on individuals, families and societies[4].

Effective interventions to prevent or delay diabetes and its complications have the potential to save significant health care costs and improve the health of a population in the long run[5, 6]. However, for those people at a low risk of diabetes, early intervention should not be undertaken, and hence it is vital to identify high-risk individuals. In order to reduce the influence of diabetes and lower the morbidity, non-invasive risk models are needed to detect individuals at a high risk. Moreover, non-invasive risk models are more economical and more suitable for large-scale screening than invasive risk models[7, 8].

There is growing evidence that diabetes is associated with changes in tongue image.

The yellow tongue coating is associated with a high prevalence of diabetes, and at the same time, it is also associated with prediabetes among Asian people[9]. As a non-invasive and readily available feature, purple tongue, thick tongue coating, and yellow tongue coating can be used for early screening of diabetes[10]. At present, the researches that use machine learning to analyze the features of tongue image and establish a diabetes risk prediction model are mostly qualitative and two-category studies[11, 12].

We hypothesize that we could train a machine learning model to identify the subtle features acquired from a standard tongue image that are due to glucose metabolism level changes associated with a history of hyperglycemia. To the best of our knowledge, this is the first attempt to use tongue features to quantitatively predict fasting plasma glucose (FPG) and glycated haemoglobin(HbA_{1c}).

Methods

We use machine learning to establish the mapping relation between various types of features and the glucose metabolism index, make predictions for FPG and HbA_{1c} in undiagnosed subjects based on the mapping relation, and finally achieve the predictions of diabetes risk(Fig. 1).

Data Collection

We included all subjects aged 18 years or older with at least one digital, normal tongue capture acquired at Shu Guang Hospital Affiliated to Shanghai University of Traditional Chinese Medicine (SHUTCM) between Apr 13, 2015, and Nov 15, 2019. A total of 2,500 subjects participated in the study. Due to incomplete data or incorrect data recording, 320 subjects were excluded, and eventually 2,180 subjects were enrolled in the study. Basic information and laboratory test results were provided by Shu Guang Hospital. Our study has been approved by the ethics committee of SHUTCM. All subjects were willing to participate in the study and signed the informed consent.

Tongue Features Extraction

Tongue images are captured by the TFDA-1 Tongue Diagnosis Instrument(Fig. 2). The key component of TFDA-1 is a stable light source, which has color temperature of 5000K and color rendering index of 97. Given the stable light source and white diffuse reflection coating, TFDA-1 can ensure a standard and stable environment for acquiring tongue images[13].

After tongue images acquired by TFDA-1, features are extracted by the Tongue Diagnosis Analysis System V2.0. TDAS first segments the tongue area from the original image, separates the tongue body (TB) and tongue coating (TC) by the "chrominance-threshold method"[14], and automatically calculates the color features and texture features(Fig. 3). R, G, B represent the three components of the RGB color space; L, a, b represent the three components of the Lab color space; H, S, I represent the three components of the HSI color space; Y, Cr, Cb represent the three components of YCrCb color space. perAll is the ratio of tongue coating to the entire tongue surface, perPart is the ratio of tongue coating to the tongue surface without the tongue coating. Texture features include contrast (CON), angle second moments (ASM), entropy (ENT), and mean[15, 16].

Both TFDA-1 and TDAS V2.0 are developed by the intelligent diagnostic laboratory of SHUTCM.

Handling Missing and Abnormal Values

Missing data may weaken the representativeness of the samples and complicate the research analysis[17]. In order to ensure the reliability of the experiment, samples with missing values were deleted directly. According to the Tukey Method[18], if the sample contains two or more outliers which are higher than upper whisker or lower than lower whisker, then it is deleted directly. A few abnormal values of data were handled by replacing them with the mean value of feature. Where X denotes the original value of the feature, $Q3$ denotes upper quartile, $Q1$ denotes lower quartile, IQR denotes the difference between $Q3$ and $Q1$, the upper whisker denotes $Q3$ plus 1.5 times IQR , and the lower whisker denotes the difference between $Q1$ and 1.5 times IQR .

$$X > Q3 + 1.5 \times IQR \cup X < Q1 - 1.5 \times IQR \quad (1)$$

Data Scaling and Normalization

The appropriate conditions for the machine learning model to work require that the eigenvalues are in a similar scale and approximate to a normal distribution. Normalization is applied to each observation so that the values in a row have a unit norm.

$$X_{scaled} = \frac{X - \bar{X}}{Std} \quad (2)$$

where X denotes the original value of the feature, \bar{X} denotes the mean value of the feature, and Std denotes the standard deviation of the feature.

Variable Definition

Hyperglycemia group was defined as FPG > 6.0mmol/L(108mg/dl)[19]; HbA_{1c} >= 38.801mmol/mol(5.7%) [20].

Basic features included in this study were Age, Weight, systolic blood pressure (SBP) and diastolic blood pressure (DBP). Age was calculated as Year, Weight was calculated as Kg, and SBP and DBP were calculated as mmHg.

Blood features included in this study were white blood cell(WBC), red blood cell(RBC), hemoglobin(HGB), total cholesterol(TCHO), triglyceride(TG), low density lipoprotein(LDL), high density lipoprotein(HDL), alanine aminotransferase(ALT). WBC was calculated as 10⁹/L, RBC was calculated as 10¹²/L, HGB was calculated as g/L, TCHO, TG, LDL and HDL were calculated as mmol/L, ALT was calculated as U/L.

Tongue features included in this study were perAll, perPart, TC-CON, TC-ASM, TC-ENT, TC-MEAN, TC-R, TC-G, TC-B, TC-L, TC-a, TC-b, TC-H, TC-I, TC-S, TC-Y, TC-Cr, TC-Cb, TB-CON, TB-ASM, TB-ENT, TB-MEAN, TB-R, TB-G, TB-B, TB-L, TB-a, TB-b, TB-H, TB-I, TB-S, TB-Y, TB-Cr, TB-Cb.

Tongue features included in machine learning model were TC-L, TC-a, TC-b, TB-L, TB-a, TB-b.

Non-invasive features were the fusion of tongue features and basic features.

Full features were the fusion of tongue features, basic features and blood features.

Statistical analysis

Normally distributed variables were presented as mean \pm standard deviation. Skewed variables were presented as the median, 25% quartile and 75% quartile. Continuous data were tested for normality using the D'Agostino and Pearson's tests. Continuous data were tested for homogeneity of variance using the Bartlett's test. The two-sample t-test, separate variance estimation t-test and Wilcoxon rank-sum test were used to compare continuous variables between normal group and **hyperglycemia** group. The correlations were executed using Pearson's and Spearman's methods. All tests were two-sided and statistical significance was assumed at $P < 0.05$. All analyses were performed with Python software, version 3.7.4.

Table 1 Statistical analysis was used to compare the differences between the basic and blood features of the normal and **hyperglycemia** group, and the correlations between the features and glucose metabolism index were analyzed.

Feature	Normal (N=1546)	Hyperglycemia (N=634)	<i>p</i> value ^a	CC ^a FPG	<i>p</i> value ^b	CC ^b HbA _{1c}	<i>p</i> value ^c
Age(Year)	38.0(30.0-47.0)	55.0(46.0-63.75)	<0.001	0.373	<0.001	0.536	<0.001
Weight(Kg)	64.0(55.0-73.0)	71.0(63.0-78.0)	<0.001	0.257	<0.001	0.25	<0.001
SBP(mmHg)	119.0(109.0-131.0)	133.0(121.25-145.0)	<0.001	0.397	<0.001	0.354	<0.001
DBP(mmHg)	75.0(68.0-83.0)	80.0(73.0-88.0)	<0.001	0.264	<0.001	0.224	<0.001
WBC(109/L)	5.8(5.1-6.8)	6.3(5.4-7.4)	<0.001	0.1	<0.001	0.158	<0.001
RBC(1012/L)	4.76(4.45-5.13)	4.86(4.57-5.14)	<0.001	0.11	<0.001	0.078	<0.001
HGB(g/L)	145.0(134.0-156.0)	150.0(140.0-157.0)	<0.001	0.163	<0.001	0.075	<0.001
TCHO(mmol/L)	4.875(4.33-5.42)	5.2(4.62-5.868)	<0.001	0.113	<0.001	0.239	<0.001
TG(mmol/L)	1.04(0.75-1.487)	1.46(1.07-2.09)	<0.001	0.293	<0.001	0.333	<0.001
HDL(mmol/L)	1.37(1.17-1.578)	1.2(1.06-1.42)	<0.001	-0.267	<0.001	-0.228	<0.001
LDL(mmol/L)	2.82(2.41-3.32)	3.14(2.63-3.695)	<0.001	0.134	<0.001	0.246	<0.001
ALT(U/L)	17.0(12.0-24.0)	21.0(16.0-29.0)	<0.001	0.212	<0.001	0.211	<0.001

p value^a Significance level for the difference between the normal and the hyperglycemia group.

CC^a Correlation coefficient between features and FPG.

p value^b Significance level of correlation coefficient between features and FPG.

CC^b Correlation coefficient between features and HbA_{1c}.

p value^c Significance level of correlation coefficient between features and HbA_{1c}.

Table 2 Statistical analysis was used to compare the differences between the tongue features of the normal and hyperglycemia group, and the correlations between the tongue features and the glucose

metabolism index were analyzed.

Feature	Normal (N=1546)	Hyperglycemia (N=634)	<i>p</i> value ^a	CC ^a FPG	<i>p</i> value ^b	CC ^b HbA _{1c}	<i>p</i> value ^c
perAll	0.414(0.314-0.563)	0.454(0.354-0.792)	<0.001	0.302	<0.001	0.137	<0.001
perPart	1.152(1.067-1.289)	1.1(1.023-1.252)	<0.001	-0.208	<0.001	-0.144	<0.001
TB-CON	60.946(43.702-83.004)	62.968(43.366-86.508)	0.675	-0.074	0.001	0.027	0.214
TB-ASM	0.085(0.072-0.101)	0.084(0.071-0.102)	0.606	0.068	0.002	-0.028	0.198
TB-ENT	1.173(1.098-1.244)	1.179(1.092-1.254)	0.759	-0.078	<0.001	0.024	0.271
TB-MEAN	0.024(0.02-0.028)	0.024(0.02-0.028)	0.655	-0.07	0.001	0.028	0.195
TC-CON	75.281(49.555-102.545)	85.722(59.51-114.987)	<0.001	0.058	0.007	0.138	<0.001
TC-ASM	0.073(0.062-0.092)	0.068(0.058-0.084)	<0.001	-0.056	0.009	-0.133	<0.001
TC-ENT	1.226(1.131-1.297)	1.256(1.172-1.324)	<0.001	0.06	0.005	0.138	<0.001
TC-MEAN	0.027(0.022-0.031)	0.028(0.024-0.033)	<0.001	0.056	0.009	0.135	<0.001
TB-R	162.0(154.0-168.0)	157.0(149.0-164.0)	<0.001	-0.153	<0.001	-0.191	<0.001
TB-G	100.0(93.0-108.0)	97.0(89.25-106.0)	<0.001	-0.014	0.517	-0.127	<0.001
TB-B	105.0(98.0-115.0)	105.0(97.0-116.0)	0.86	0.145	<0.001	-0.018	0.394
TC-R	152.102±15.38	150.278±16.516	0.017	-0.051	0.016	-0.071	0.001
TC-G	110.0(101.0-120.0)	111.0(101.0-123.0)	0.141	0.101	<0.001	0.019	0.381
TC-B	114.0(104.0-125.0)	117.0(106.0-134.0)	<0.001	0.2	<0.001	0.088	<0.001
TB-L	104.988(102.774-107.829)	104.07(101.362-106.988)	<0.001	-0.038	0.076	-0.138	<0.001
TB-a	21.372(19.351-23.381)	21.824(19.66-23.626)	0.024	-0.069	0.001	0.037	0.087
TB-b	5.182(2.674-6.671)	4.46(-2.886-6.318)	<0.001	-0.276	<0.001	-0.135	<0.001

TC-L	107.233±5.185	107.52±5.421	0.247	0.077	<0.001	0.006	0.795
TC-a	14.393±2.651	13.917±2.762	<0.001	-0.198	<0.001	-0.081	<0.001
TC-b	3.527(1.179-5.206)	2.925(-3.833-4.75)	<0.001	-0.251	<0.001	-0.13	<0.001
TB-H	356.037(352.221-358.425)	354.791(339.386-357.969)	<0.001	-0.24	<0.001	-0.132	<0.001
TB-I	121.0(115.0-130.0)	119.0(112.0-127.0)	<0.001	-0.014	0.504	-0.122	<0.001
TB-S	0.181±0.025	0.189±0.031	<0.001	0.023	0.287	0.113	<0.001
TC-H	356.62(350.777-360.0)	354.924(330.0-360.0)	<0.001	-0.212	<0.001	-0.125	<0.001
TC-I	125.0(116.0-135.0)	126.0(117.0-138.0)	0.088	0.096	<0.001	0.021	0.331
TC-S	0.125±0.024	0.124±0.025	0.698	-0.104	<0.001	-0.015	0.492
TB-Y	117.672(112.442-124.673)	115.605(109.439-122.641)	<0.001	-0.043	0.043	-0.14	<0.001
TB-Cr	154.292(151.371-156.631)	153.123(149.379-155.681)	<0.001	-0.278	<0.001	-0.147	<0.001
TB-Cb	120.853(119.551-123.211)	121.596(119.981-128.45)	<0.001	0.282	<0.001	0.147	<0.001
TC-Y	121.972(113.952-130.15)	121.851(113.796-132.073)	0.409	0.071	0.001	0.002	0.908
TC-Cr	146.131(143.219-148.418)	145.058(140.772-147.693)	<0.001	-0.286	<0.001	-0.145	<0.001
TC-Cb	122.801(121.473-125.023)	123.394(121.764-130.292)	<0.001	0.252	<0.001	0.132	<0.001

p value^a Significance level for the difference between the normal and hyperglycemia group.

CC^a Correlation coefficient between features and FPG.

p value^b Significance level of correlation coefficient between features and FPG.

CC^b Correlation coefficient between features and HbA_{1c}.

p value^c Significance level of correlation coefficient between features and HbA_{1c}.

Data Partition

Our research divides the data set into a training set and a test set according to the ratio of 8: 2. The training set consisted of 80% of the original database (1744 individuals). The test set consisted of 20% of the original database (436 individuals). We performed a 5-fold cross-validation on the training set and selected the model through cross-validation, i.e., we used 80% of the training set, leaving 20% as the validation set, and then rotated five times. The test set didn't participate in the training process of the model, and the actual detection performance of the model was checked on the test set.

Feature Selection

In this paper, we selected Age, Weight, BP, RBC, HGB, WBC, TCHO, TG, HDL LDL, ALT, TB-L, TB-a, TB-b, TC-L, TC-a, and TC-b according to research practice. Age is an important risk factor for the diagnosis, development and prognosis of type 2 diabetes; however, HbA_{1c} has a tendency to decline with age, and only increases when it is older than 90 years of age[21]. Weight is highly correlated with the occurrence of diabetes. After the diagnosis of diabetes, weight loss of more than 5% can improve the level of HbA_{1c} and reduce the risk of cardiovascular disease within 10 years[22]. There is no causal relationship between hypertension and diabetes, but diabetes often leads to increased blood pressure[23]. Hyperglycemia can affect blood indicators, which are recognized a risk factor for complications. Due to the chronic inflammatory state of diabetes, pro-inflammatory factors can differentiate and mature white blood cells. Diabetes changes the surface charge and aggregation of red blood cells to reduce the number of red blood cells and the content of hemoglobin[24]. Fat accumulation in the body leads to excessive saturated fatty acids remaining in the cells. These saturated fatty acids can be toxic to liver and pancreatic islet cells. Therefore, it is very important to evaluate lipid toxicity by detecting TCHO, TG, HDL LDL and ALT [25].

Tongue diagnosis has been widely used in the diagnosis of diabetes by traditional Chinese medicine. In clinical practice, 80% of diabetic patients do not show typical symptoms, and tongue signs have changed accordingly. For example, the tongue of diabetic patients is often red and yellow[26]. Studies have shown that yellow tongue coating is associated with a high incidence of diabetes[9]. Given the above research, we choose the related features to train the machine learning regression model.

Model Construction

In our study, we utilize multiple supervised learning models for regression of FPG and HbA_{1c} of subjects. We not only used the popular artificial neural network(ANN) model, but also used three ensemble models including Gradient Boosting Decision Tree(GBDT), Random Forest(RF) and eXtreme Gradient Boosting Tree(XGBT).

In a simple neural network, it generally includes three layers, an input layer, a hidden layer, and an output layer. The neurons are connected to each other. The data enters the neural network at the input layer and propagates forward to the output layer to get the solution. There is a corresponding weight at the input of each neuron to control the strength of the input[27].

RF[28] is an integrated tree algorithm based on the idea of bagging. Using a different bootstrap sample of the data, RF builds a foundational decision tree. Each node is split using the best among a subset of predictors randomly chosen at that node. The double randomness of feature selection and sample selection can not only improve the generalization ability of the model, but also avoid overfitting the model.

GBDT and XGBT[29] have been widely used in a number of data mining and machine learning challenges. They are tree ensemble models which use K additive functions to predict the output. However, XGBT controls the complexity of the model by adding regular terms to the objective function to avoid overfitting the model.

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i) \quad f_k \in F \quad (3)$$

where F is the space of regression trees.

$$Obj = \sum_{n=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (4)$$

where l is the loss of model, and Ω is the complexity of model.

We constructed models for prediction of FPG and HbA_{1c} with machine learning algorithms(Fig. 5). In the current experiment, to study the contribution of various types of features in modeling, we combined different types of features, including basic features, blood features and tongue features.

Fusion Strategy

The machine learning based model blending approach is a very common fusion technique[30]. The idea of model fusion is to use many independent models to calculate the initial prediction, and then mix the initial prediction to achieve a better final prediction result. Four independent models were used to fit the training set, the prediction verification set and the test set. The prediction results of the verification set were combined into a new training matrix, and the prediction results of the test set were combined into a new test matrix. The second layer of linear regression was used to fit the new training matrix, predict the new test matrix, and get the final test result(Fig. 6). The stacking and generalization operation of the blending approach uses different data to avoid the leakage of learning information.

Evaluation Criteria

Mean Squared Error(MSE) and coefficient of determination(R-squared) were used to evaluate the performance of our prediction model. The calculation of MSE was simple and the meaning was clear. It was a commonly used evaluation index in statistical analysis[31]. We should not only consider the actual

working accuracy of the prediction model, but also consider the influence of the prediction model on the clinical decision. Consequently, the Clark's Error Grid Analysis(EGA)[32, 33] was used to determine the acceptable error for the accuracy of predictive value of FPG in comparison with the actual value. The more values that appear in Zones A and B, the more accurate the model is in terms of clinical utility[34].The scatter-plot was used to access the performance of the HbA_{1c} prediction model[35]. The closer the slope of the fitted line is to 1, the closer the intercept is to 0, the better the model performance.

$$M_{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (5)$$

$$M_{R-squared} = 1 - \frac{\sum_{i=0}^n (y_i - \hat{y}_i)^2}{\sum_{i=0}^n (y_i - \bar{y}_i)^2} \quad (6)$$

where y_i is actual value, \hat{y}_i is predictive value.

Results

Cross Validation

In order to estimate the effect of model, we performed five-fold cross validation on training set including 1744 samples. Average MSE is used to evaluate the performance of the model. When the fusion features model predicted FPG and HbA_{1c}, the mean MSE of the model with the full input features was the lowest, followed by the model with tongue features in combination with blood features, and then the model with the non-invasive features. When the single-features model predicted FPG and HbA_{1c}, the average MSE of the model with the tongue features alone was the lowest, followed by the model with the blood features alone, and then the model with the basic features alone.

The FPG and HbA_{1c} prediction results showed tongue features improved the performance of the models. Combining tongue features with basic features would decrease the average MSE compared with FPG and HbA_{1c} prediction with basic features alone. Combining tongue features with blood features would decrease the average MSE compared with FPG and HbA_{1c} prediction with blood features alone. Combining basic features, blood features and tongue features would decrease the average MSE compared with FPG and HbA_{1c} prediction using the model with basic features and blood features combined (Fig. 7 and Table 3; Fig. 8 and Table 4).

Table 3 Summary average MSE of various FPG prediction models with different types of features using cross validation.

Feature	ANN	XGBT	RF	GBDT
Basic	1.722	1.736	1.745	1.737
Blood	1.534	1.575	1.567	1.588
Tongue	1.394	1.354	1.353	1.388
Basic+Blood	1.457	1.499	1.516	1.526
Basic+Tongue	1.308	1.232	1.227	1.249
Blood+Tongue	1.289	1.237	1.158	1.198
Basic+Blood+Tongue	1.244	1.149	1.124	1.151

Table 4 Summary average MSE of various HbA_{1c} prediction algorithms with different types of features using cross validation.

Feature	ANN	XGBT	RF	GBDT
Basic	0.593	0.621	0.595	0.598
Blood	0.574	0.586	0.57	0.561
Tongue	0.521	0.534	0.505	0.513
Basic+Blood	0.497	0.502	0.504	0.51
Basic+Tongue	0.467	0.439	0.438	0.439
Blood+Tongue	0.48	0.42	0.416	0.405
Basic+Blood+Tongue	0.426	0.393	0.391	0.384

Model performance evaluation on the test set

Indicators including MSE and R-squared were used to evaluate the generalization performance of the model on the test set. In addition, EGA was used to evaluate the clinical value of the FPG prediction model and scatter-plot was used to assess the performance of the HbA_{1c} prediction model.

The lowest MSE and highest R-squared were obtained by the machine learning models that were applied for prediction of FPG and HbA_{1c} using non-invasive features included basic features and tongue features. Combining tongue features with basic features would decrease MSE and increase R-squared compared with FPG and HbA_{1c} prediction with basic features alone. Combining tongue features with blood features

would decrease MSE and increase R-squared compared with FPG and HbA_{1c} prediction with blood features alone. Combining basic features, blood features and tongue features would decrease MSE and increase R-squared compared with FPG and HbA_{1c} prediction using the model combined basic features and blood features(Fig. 9 and Table 5; Fig. 10 and Table 6; Fig. 12 and Table 7; Fig. 13 and Table 8).

Table 5 Summary MSE of various FPG prediction models with different types of features on the test set.

Feature	ANN	XGBT	RF	GBDT	Blending
Basic	1.267	1.269	1.278	1.277	1.264
Blood	1.332	1.486	1.436	1.426	1.444
Tongue	0.876	0.761	0.764	0.793	0.807
Basic+Blood	1.131	1.24	1.125	1.179	1.196
Basic+Tongue	0.725	0.649	0.601	0.644	0.683
Blood+Tongue	0.893	0.828	0.781	0.841	0.769
Basic+Blood+Tongue	0.823	0.727	0.668	0.75	0.654

Table 6 Summary R-squared of various FPG prediction models with different types of features on the test set.

Feature	ANN	XGBT	RF	GBDT	Blending
Basic	0.169	0.168	0.162	0.162	0.171
Blood	0.126	0.026	0.058	0.065	0.053
Tongue	0.425	0.501	0.499	0.48	0.47
Basic+Blood	0.258	0.186	0.262	0.226	0.216
Basic+Tongue	0.525	0.574	0.606	0.578	0.552
Blood+Tongue	0.414	0.457	0.488	0.448	0.496
Basic+Blood+Tongue	0.46	0.523	0.562	0.508	0.571

The EGA results of four models on the test set were presented, including the RF model with non-invasive features, the XGBT model with non-invasive features, the blending model with non-invasive features and blending model with full features. Because non-invasive RF model achieve highest R-squared and lowest MSE; non-invasive XGBT and blending models got better EGA results than non-invasive RF model; The best EGA results were obtained by the blending model with full features.

Given non-invasive features, 89.68% of the results predicted by the RF model appear in zone A, 9.63% of the results appear in zone B; 90.14% of the results predicted by the XGBT model appear in zone A, 9.17% of the results appear in zone B; 90.83% of the results predicted by the blending model appear in zone A, and 8.49% of the results predicted in zone B. The EGA results obtained on the blending model with non-invasive features input closely agreed with the values using the best model with full features(Fig. 11).

Table 7 Summary MSE of various HbA_{1c} prediction models with different types of features on the test set.

Feature	ANN	XGBT	RF	GBDT	Blending
Basic	0.477	0.469	0.484	0.469	0.477
Blood	0.547	0.588	0.541	0.55	0.557
Tongue	0.414	0.425	0.38	0.393	0.38
Basic+Blood	0.451	0.468	0.427	0.433	0.436
Basic+Tongue	0.327	0.302	0.273	0.272	0.293
Blood+Tongue	0.391	0.388	0.366	0.381	0.365
Basic+Blood+Tongue	0.344	0.353	0.282	0.299	0.291

Table 8 Summary R-squared of various HbA_{1c} prediction models with different types of features on the test set.

Feature	ANN	XGBT	RF	GBDT	Blending
Basic	0.193	0.205	0.18	0.205	0.192
Blood	0.073	0.005	0.083	0.068	0.057
Tongue	0.299	0.281	0.357	0.334	0.356
Basic+Blood	0.235	0.208	0.277	0.266	0.261
Basic+Tongue	0.447	0.488	0.538	0.539	0.504
Blood+Tongue	0.337	0.342	0.381	0.354	0.381
Basic+Blood+Tongue	0.417	0.402	0.522	0.494	0.507

Scatter-plots of HbA_{1c} from five models were shown. The GBDT model with the best MSE and R-squared achieved lower slope and intercept than the XGBT model and RF model. The blending model with non-invasive input features didn't produce better results. Based on non-invasive features, the XGBT model with obtained highest slope and lowest intercept, which was better than the blending model with full features(Fig. 14).

Discussion

Obesity, physical inactivity, and smoking have a tremendous harmful effect on the health of patients with diabetes[36]. Recent research shows that timely lifestyle-modifying interventions can prevent and delay the onset of diabetes[37]. By identifying high-risk subjects early and taking active intervention measures as early as possible, it can prevent people who are in prediabetes from developing diabetes and other diseases[38]. Furthermore, people who already have diabetes can be diagnosed as early as possible, receive treatment, and avoid complications occur[39–41]. Therefore, in order to reduce the morbidity and mortality of diabetes, we have strong reasons to use diabetes risk prediction models for population screening. Using machine learning methodologies, we established the diabetes risk prediction regression model based on non-invasive features. These features include 6 tongue features and 4 basic features: TB-L, TB-a, TB-b, TC-L, TC-a, TC-b, Age, Weight, SBP, and DBP, all of which are easy to detect and the risk prediction model we developed is simple to use. The model can be deployed in mobile phones or other portable health testing equipment to predict FBG and HbA_{1c}. The model can improve the risk prediction of diabetes and prediabetes.

The experimental results showed that adding tongue features can improve the predictive ability of the model. Given non-invasive features, the model was basically equivalent to the full-feature model. Although cross validation results showed that the performance of the model with the full features is better than non-invasive model, but the model with the full features required blood tests, which limits the scope of use of the model. We build a full feature model to explore performance boundaries and provide performance references for non-invasive model. In a study, a non-invasive machine learning model for prediction of blood glucose based on photoplethysmography achieved 87.7% of points in zone A and 10.3% of points in zone B on EGA[35]. Then our best non-invasive model obtained 90.83% of points in zone A and 8.49% of points in zone B. Through 30-day self-monitoring of blood glucose data, a regression equation model for the prediction of glycosylated hemoglobin could be established, with a multiple correlation coefficient of 0.778[42]. In our research, a non-invasive machine learning model based on tongue features and basic features combined predicts HbA_{1c} with a multiple correlation coefficient of 0.734. Through research and practice, it has been suggested that the tongue features as a basic physiological feature of humans have a predictive value for FBG and HbA_{1c}.

Machine learning is a branch of artificial intelligence, which can extract the inherent relationships of data such as decision rules and patterns[43]. Several diabetes risk prediction models in previous studies were based on machine learning algorithms[44]. Choi et al.[45] has developed two effective machine learning models for predicting pre-diabetes. The input features included age, gender, family history of diabetes,

hypertension, alcohol intake, BMI, smoking status, waist circumference, and physical activity. A recent cross-sectional study conducted in USA showed that machine learning models based on survey questionnaires are able to identify individuals at high risk of diabetes[46]. In another study[32], a deep learning model for dynamically predicting blood glucose was developed, which is conducive to self-management of diabetes. Machine learning algorithms have become a key process for mining the internal relationship between clinical data and diabetes or pre-diabetes[47]. In our study, classical machine learning models with different structures are used to explore the linear relationship between tongue features and glucose metabolism indicators, moreover, we attempted to enhance this relationship through model fusion.

The current research is mainly to use statistical methods to explore the relationship between tongue features and diabetes and pre-diabetes. Given the tongue features, the diabetes risk prediction model established by machine learning method mainly performs qualitative classification prediction[11, 12]. In our study, the prediction accuracy of the model is very high, MSE of FPG and HbA_{1c} prediction was 0.601 and 0.272 respectively, and the interpretability of the model is also very strong. R-squared of FPG and HbA_{1c} prediction are 0.606 and 0.539, respectively. These results indicate that there is a linear relationship between the tongue features and glucose metabolism index, and that our model can capture this relationship. We systematically reviewed the literature, and no similar published research work has been found. The variables required for model work are non-invasive and easy to measure, and are very suitable for large-scale diabetes screening. The prediction result of the model is the specific values of FPG and HbA_{1c}, through which the risk of pre-diabetes and diabetes can be further evaluated.

However, several potential limitations of this study should be mentioned. Despite our best efforts, the sample size we included in the study was still relatively small. In addition, our study requires independent evaluations from data collected in other sources to further validate the performance of the model.

Conclusion

We discovered a linear relationship between tongue features and glucose metabolism index by means of machine learning algorithms. Tongue features made an important contribution to develop diabetes risk prediction model. Based on this, we established a non-invasive diabetes risk prediction model to predict FPG and HbA_{1c}. The model can provide timely risk warning to prevent or delay the onset of diabetes.

Abbreviations

FPG

Fasting plasma glucose

TB

Tongue body

TC

Tongue coating

SBP
Systolic blood pressure
DBP
Diastolic blood pressure
WBC
White blood cell
RBC
Red blood cell
HGB
Hemoglobin
TCHO
Total cholesterol
TG
Triglyceride
ALT
Alanine aminotransferase
CC
Correlation coefficient
MSE
Mean squared error
R-squared
Coefficient of determination
EGA
Error Grid Analysis

Declarations

Ethics approval and consent to participate

This study was reviewed and approved by the ethics committee of Shanghai University of Traditional Chinese Medicine.

Consent for publication

Not applicable

Availability of data and materials

Not applicable

Competing interests

The authors declare that there is no duality of interest associated with this manuscript.

Funding

This study was supported by the National Key Research and Development Program of China (Grant Number: 2017YFC1703300 and 2017YFC1703301) and National Natural Science Foundation of China (Grant Numbers: 81373556, 81102558, and 81873235). The study sponsors were not involved in the design of the study, the collection, analysis, and interpretation of data, writing the report or the decision to submit the report for publication.

Acknowledgements

We thank all investigators, study teams and subjects for participating in these studies. We thank the doctors at the Shuguang Hospital for their assistance in this study. We are thankful to Dong Xu (University of Missouri, Columbia, MO, USA) for writing support.

Authors' information

¹ School of Basic Medicine, Shanghai University of Traditional Chinese Medicine, Shanghai, China. ² Shanghai Collaborative Innovation Center of Health Service in Traditional Chinese Medicine, Shanghai University of Traditional Chinese Medicine, Shanghai, China. ³ Department of Intelligent Science and Technology, Xiamen University, Xiamen, Fujian, China

Authors' contributions

JL (Jun Li) wrote the whole manuscript text, analyzing the data, coding the program and plotting the results. These drafts were revised for important scientific content by JH, XH, CZ and JX. JH, LT, LC, JC, XM, TJ, XY, CZ, XH and JX helped in the ideas of study. JL (Jun Li), JL (Jiacai Li), YS, ZB, YW, HF, JW, YL, CP and XG acquired and analyzed the ADVANCE trial data. All authors gave final approval of the version to be published. JX is the guarantor of this work.

References

[1] Ely EK, Gruss SM, Luman ET, et al. (2017) A national effort to prevent type 2 diabetes: participant-level evaluation of CDC's National Diabetes Prevention Program. *Diabetes care* 40(10): 1331-1341. <https://doi.org/10.2337/dc16-2099>

[2] International Diabetes Federation (2019) IDF DIABETES ATLAS 9th ed. In: Chapter 3 Global picture. International Diabetes Federation, Brussels, Belgium, pp 34-57

- [3] Wang L, Gao P, Zhang M, et al. (2017) Prevalence and Ethnic Pattern of Diabetes and Prediabetes in China in 2013. *JAMA* 317(24): 2515-2523. <https://doi.org/10.1001/jama.2017.7596>
- [4] Fagg J, Valabhji J (2019) How do we identify people at high risk of Type 2 diabetes and help prevent the condition from developing? *Diabetic medicine : a journal of the British Diabetic Association* 36(3): 316-325. <https://doi.org/10.1111/dme.13867>
- [5] Zhuo X, Zhang P, Hoerger TJ (2013) Lifetime direct medical costs of treating type 2 diabetes and diabetic complications. *American journal of preventive medicine* 45(3): 253-261. <https://doi.org/10.1016/j.amepre.2013.04.017>
- [6] Gillies CL, Abrams KR, Lambert PC, et al. (2007) Pharmacological and lifestyle interventions to prevent or delay type 2 diabetes in people with impaired glucose tolerance: systematic review and meta-analysis. *Bmj* 334(7588): 1-9. <https://doi.org/10.1136/bmj.39063.689375.55>
- [7] Kengne AP, Beulens JWJ, Peelen LM, et al. (2014) Non-invasive risk scores for prediction of type 2 diabetes (EPIC-InterAct): a validation of existing models. *The Lancet Diabetes & Endocrinology* 2(1): 19-29. [https://doi.org/10.1016/S2213-8587\(13\)70103-7](https://doi.org/10.1016/S2213-8587(13)70103-7)
- [8] Wu H, Yang S, Huang Z, He J, Wang X (2018) Type 2 diabetes mellitus prediction model based on data mining. *Informatics in Medicine Unlocked* 10(2018): 100-107. <https://doi.org/10.1016/j.imu.2017.12.006>
- [9] Tomooka K, Saito I, Furukawa S, et al. (2018) Yellow Tongue Coating is Associated With Diabetes Mellitus Among Japanese Non-smoking Men and Women: The Toon Health Study. *Journal of epidemiology* 28(6): 287-291. <https://doi.org/10.2188/jea.JE20160169>
- [10] Hsu P-C, Wu H-K, Huang Y-C, et al. (2019) The tongue features associated with type 2 diabetes mellitus. *Medicine* 98(19): 1-6. <https://doi.org/10.1097/md.00000000000015567>
- [11] Zhang J, Xu J, Hu X, et al. (2017) Diagnostic Method of Diabetes Based on Support Vector Machine and Tongue Images. *BioMed research international* 2017: 7961494-7961494. <https://doi.org/10.1155/2017/7961494>
- [12] Li J, Zhang B, Lu G, You J, Zhang D (2019) Body surface feature-based multi-modal Learning for Diabetes Mellitus detection. *Information Sciences* 472(2019): 1-14. <https://doi.org/10.1016/j.ins.2018.09.010>
- [13] Jiao W, Hu XJ, Tu LP, et al. (2020) Tongue color clustering and visual application based on 2D information. *International journal of computer assisted radiology and surgery* 15(2): 203-212. <https://doi.org/10.1007/s11548-019-02076-z>
- [14] Xu J, Tu L, Zhang Z, Zhang L, Zhou C (2008) The region partition of quality and coating for tongue image based on color image segmentation method. In: 2008 IEEE International Symposium on IT in

Medicine and Education. IEEE, pp 817-821. <https://doi.org/10.1109/ITME.2008.4743981>

- [15] Xu J, Sun Y, Zhang Z (2003) Analysis and discrimination of tongue texture characteristics by difference statistics. *Journal of Shanghai University of Traditional Chinese Medicine* 17(3): 55-58. <https://doi.org/10.16306/j.1008-861x.2003.03.017>
- [16] Xu J, Zhou C, Zhang Z, et al. (2004) Computerized analysis and recognition of tongue and its coating color in tongue diagnosis. *Journal of Shanghai University of Traditional Chinese Medicine* 18(3): 43-47. <https://doi.org/10.16306/j.1008-861x.2004.03.015>
- [17] Kang H (2013) The prevention and handling of the missing data. *Korean J Anesthesiol* 64(5): 402-406. <https://doi.org/10.4097/kjae.2013.64.5.402>
- [18] Schwertman NC, Owens MA, Adnan R (2004) A simple more general boxplot method for identifying outliers. *Computational Statistics & Data Analysis* 47(1): 165-174. <https://doi.org/10.1016/j.csda.2003.10.012>
- [19] Organization WH (2016) Definition and diagnosis of diabetes mellitus and intermediate hyperglycemia. Geneva: World Health Organization. 2006. Google Scholar
- [20] Association AD (2018) Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes-2018. *Diabetes care* 41(Suppl 1): S13-s27. <https://doi.org/10.2337/dc18-S002>
- [21] Sattar N, Rawshani A, Franzén S, et al. (2019) Age at Diagnosis of Type 2 Diabetes Mellitus and Associations With Cardiovascular and Mortality Risks. *Circulation* 139(19): 2228-2237. <https://doi.org/10.1161/CIRCULATIONAHA.118.037885>
- [22] Strelitz J, Ahern AL, Long GH, et al. (2019) Moderate weight change following diabetes diagnosis and 10 year incidence of cardiovascular disease and mortality. *Diabetologia* 62(8): 1391-1402. <https://doi.org/10.1007/s00125-019-4886-1>
- [23] Sun D, Zhou T, Heianza Y, et al. (2019) Type 2 Diabetes and Hypertension. *Circulation Research* 124(6): 930-937. <https://doi.org/10.1161/CIRCRESAHA.118.314487>
- [24] Milosevic D, Panin VL (2019) Relationship Between Hematological Parameters and Glycemic Control in Type 2 Diabetes Mellitus Patients. *J Med Biochem* 38(2): 164-171. <https://doi.org/10.2478/jomb-2018-0021>
- [25] Piccolis M, Bond LM, Kampmann M, et al. (2019) Probing the Global Cellular Responses to Lipotoxicity Caused by Saturated Fatty Acids. *Molecular Cell* 74(1): 32-44. <https://doi.org/10.1016/j.molcel.2019.01.036>
- [26] Guo J, Chen H, Song J, Wang J, Zhao L, Tong X (2014) Syndrome Differentiation of Diabetes by the Traditional Chinese Medicine according to Evidence-Based Medicine and Expert Consensus Opinion.

Evidence-Based Complementary and Alternative Medicine 2014(2014): 492193.

<https://doi.org/10.1155/2014/492193>

[27] Heiat A (2002) Comparison of artificial neural network and regression models for estimating software development effort. *Information and Software Technology* 44(15): 911-922.

[https://doi.org/10.1016/S0950-5849\(02\)00128-3](https://doi.org/10.1016/S0950-5849(02)00128-3)

[28] Liaw A, Wiener M (2002) Classification and regression by randomForest. *R news* 2(3): 18-22.

Available from <https://www.r-project.org>. Accessed December 2002

[29] Chen T, Guestrin C (2016) XGBoost: A Scalable Tree Boosting System. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Association for Computing Machinery, San Francisco, California, USA, pp 785–794.

<https://doi.org/10.1145/2939672.2939785>

[30] Lu S, Hwang Y, Khabibrakhmanov I, et al. (2015) Machine learning based multi-physical-model blending for enhancing renewable energy forecast-improvement via situation dependent error correction. In: *2015 European control conference (ECC)*. IEEE, pp 283-290.

<https://doi.org/10.1109/ECC.2015.7330558>

[31] Wang Z, Bovik AC (2009) Mean squared error: Love it or leave it? A new look at signal fidelity measures. *IEEE signal processing magazine* 26(1): 98-117. <https://doi.org/10.1109/MSP.2008.930649>

[32] Faruqui SHA, Du Y, Meka R, et al. (2019) Development of a Deep Learning Model for Dynamic Forecasting of Blood Glucose Level for Type 2 Diabetes Mellitus: Secondary Analysis of a Randomized Controlled Trial. *JMIR mHealth and uHealth* 7(11): 1-14. <https://doi.org/10.2196/14452>

[33] Clarke WL, Cox D, Gonder-Frederick LA, Carter W, Pohl SL (1987) Evaluating clinical accuracy of systems for self-monitoring of blood glucose. *Diabetes care* 10(5): 622-628.

<https://doi.org/10.2337/diacare.10.5.622>

[34] (2005) The Original Clarke Error Grid Analysis (EGA). *Diabetes Technology & Therapeutics* 7(5): 776-779. <https://doi.org/10.1089/dia.2005.7.776>

[35] Monte-Moreno E (2011) Non-invasive estimate of blood glucose and blood pressure from a photoplethysmograph by means of machine learning techniques. *Artificial intelligence in medicine* 53(2): 127-138. <https://doi.org/10.1016/j.artmed.2011.05.001>

[36] Association AD (2016) 1. Strategies for Improving Care. *Diabetes care* 39(Supplement 1): S6.

<https://doi.org/10.2337/dc16-S004>

[37] Association AD (2016) 4. Prevention or delay of type 2 diabetes. *Diabetes care* 39(Supplement 1): S36-S38.

<https://doi.org/10.2337/dc16-S007>

- [38] Zand A, Ibrahim K, Patham B (2018) Prediabetes: Why Should We Care? *Methodist DeBakey cardiovascular journal* 14(4): 289-297. <https://doi.org/10.14797/mdcj-14-4-289>
- [39] Lindström J, Tuomilehto J (2003) The diabetes risk score: a practical tool to predict type 2 diabetes risk. *Diabetes care* 26(3): 725-731. <https://doi.org/10.2337/diacare.26.3.725>
- [40] Waugh N, Scotland G, McNamee P, et al. (2007) Screening for type 2 diabetes: literature review and economic modelling. *Health technology assessment (Winchester, England)* 11(17): iii-iv, ix-xi, 1-125. <https://doi.org/10.3310/hta11170>
- [41] Han X, Wang J, Li Y, et al. (2018) Development of a new scoring system to predict 5-year incident diabetes risk in middle-aged and older Chinese. *Acta diabetologica* 55(1): 13-19. <https://doi.org/10.1007/s00592-017-1047-1>
- [42] Cho JH, Park RW, Shin JA, et al. (2009) HbA1c Prediction Model Using SMBG Data in Real Practice for Subjects with Type 2 Diabetes: The Catholic-Ajou HbA1c Predictor. *DIABETES*. 58(2009). A108(Abstract)
- [43] Dagliati A, Marini S, Sacchi L, et al. (2018) Machine Learning Methods to Predict Diabetes Complications. *Journal of diabetes science and technology* 12(2): 295-302. <https://doi.org/10.1177/1932296817706375>
- [44] Zheng T, Xie W, Xu L, et al. (2017) A machine learning-based framework to identify type 2 diabetes through electronic health records. *International journal of medical informatics* 97(2017): 120-127. <https://doi.org/10.1016/j.ijmedinf.2016.09.014>
- [45] Choi SB, Kim WJ, Yoo TK, et al. (2014) Screening for Prediabetes Using Machine Learning Models. *Computational and Mathematical Methods in Medicine* 2014(2014): 618976. <https://doi.org/10.1155/2014/618976>
- [46] Dinh A, Miertschin S, Young A, Mohanty SD (2019) A data-driven approach to predicting diabetes and cardiovascular disease with machine learning. *BMC medical informatics and decision making* 19(1): 1-15. <https://doi.org/10.1186/s12911-019-0918-5>
- [47] Kavakiotis I, Tsave O, Salifoglou A, Maglaveras N, Vlahavas I, Chouvarda I (2017) Machine Learning and Data Mining Methods in Diabetes Research. *Comput Struct Biotechnol J* 15(2017): 104-116. <https://doi.org/10.1016/j.csbj.2016.12.005>

Figures

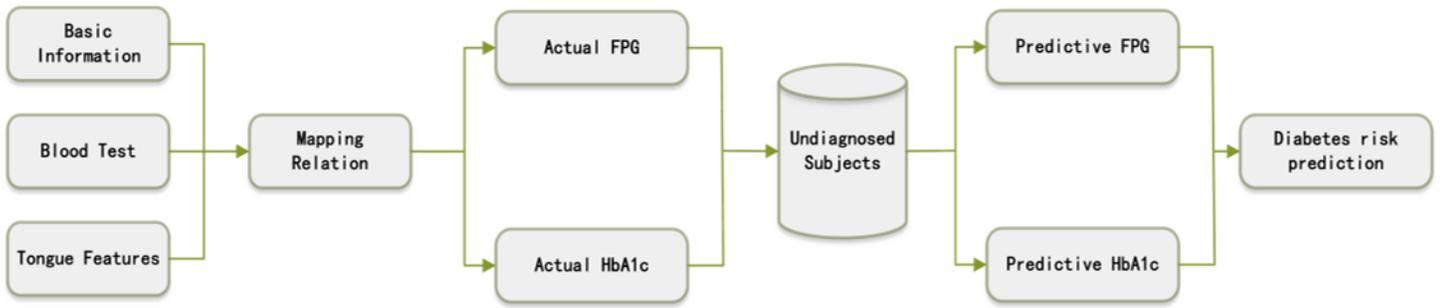


Figure 1

Study Flow Diagram



(a) Front View

(b) Side View

Figure 2

TFDA-1 Tongue Diagnosis Instrument. 1: ring light source, 2: camera lens, 3: mandible support plate.

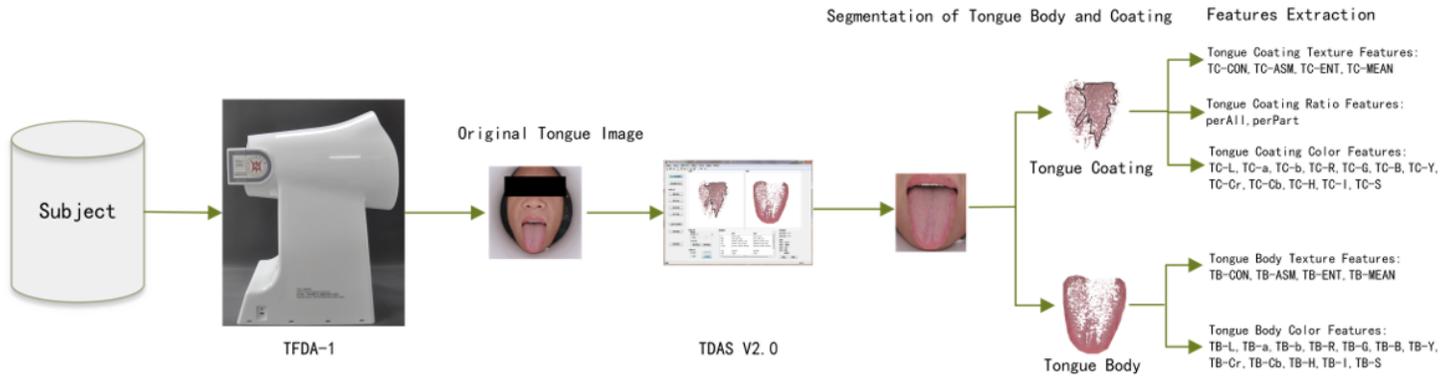


Figure 3

Tongue Feature Extraction Process

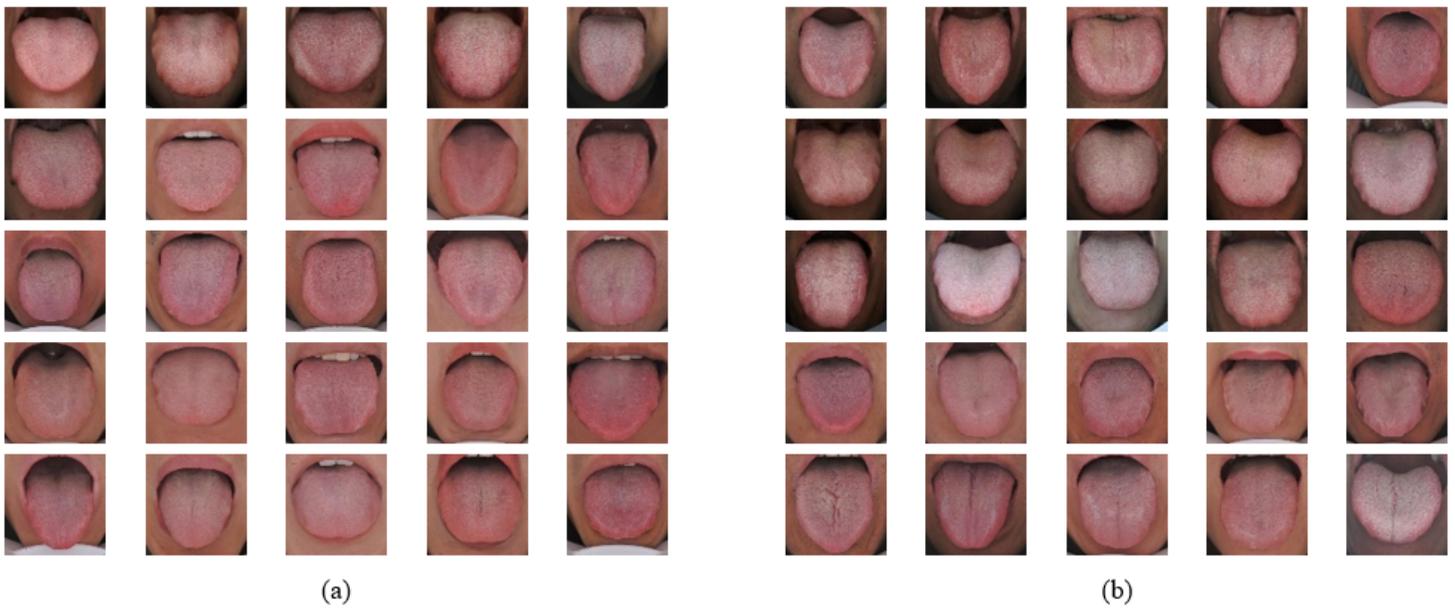


Figure 4

Typical Tongue Images. (a) The tongues of 25 individuals with normal blood glucose. (b) The tongues of 25 individuals with hyperglycemia

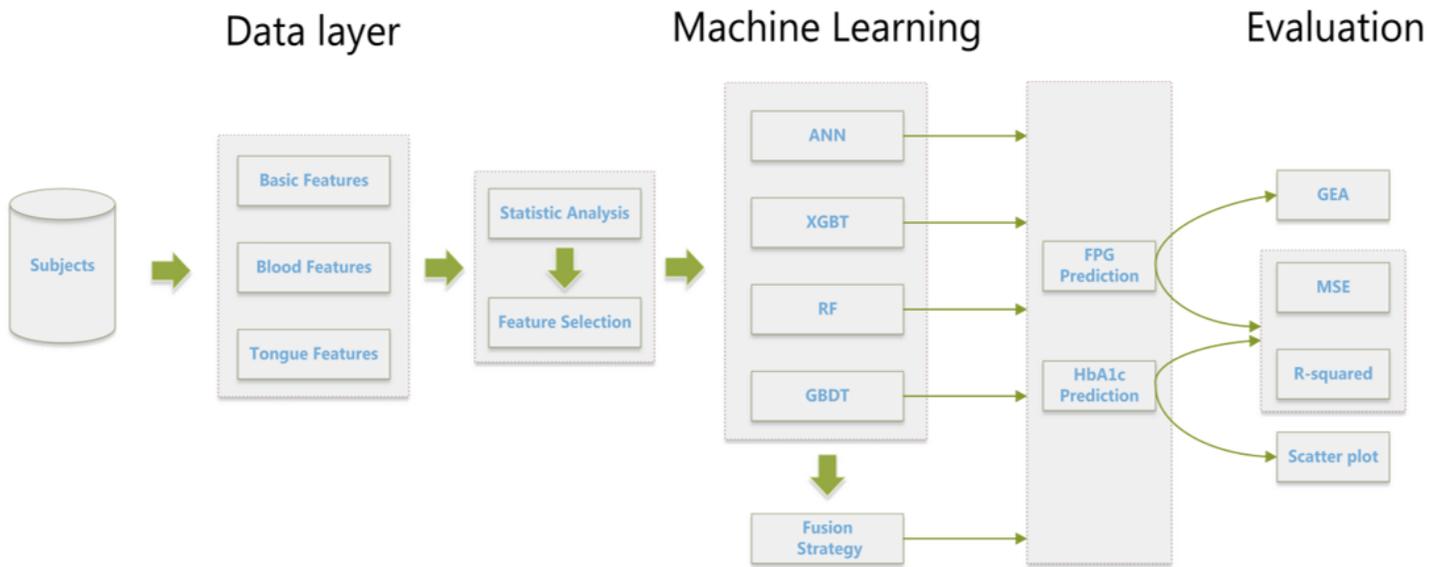


Figure 5

Machine Learning Flow Diagram

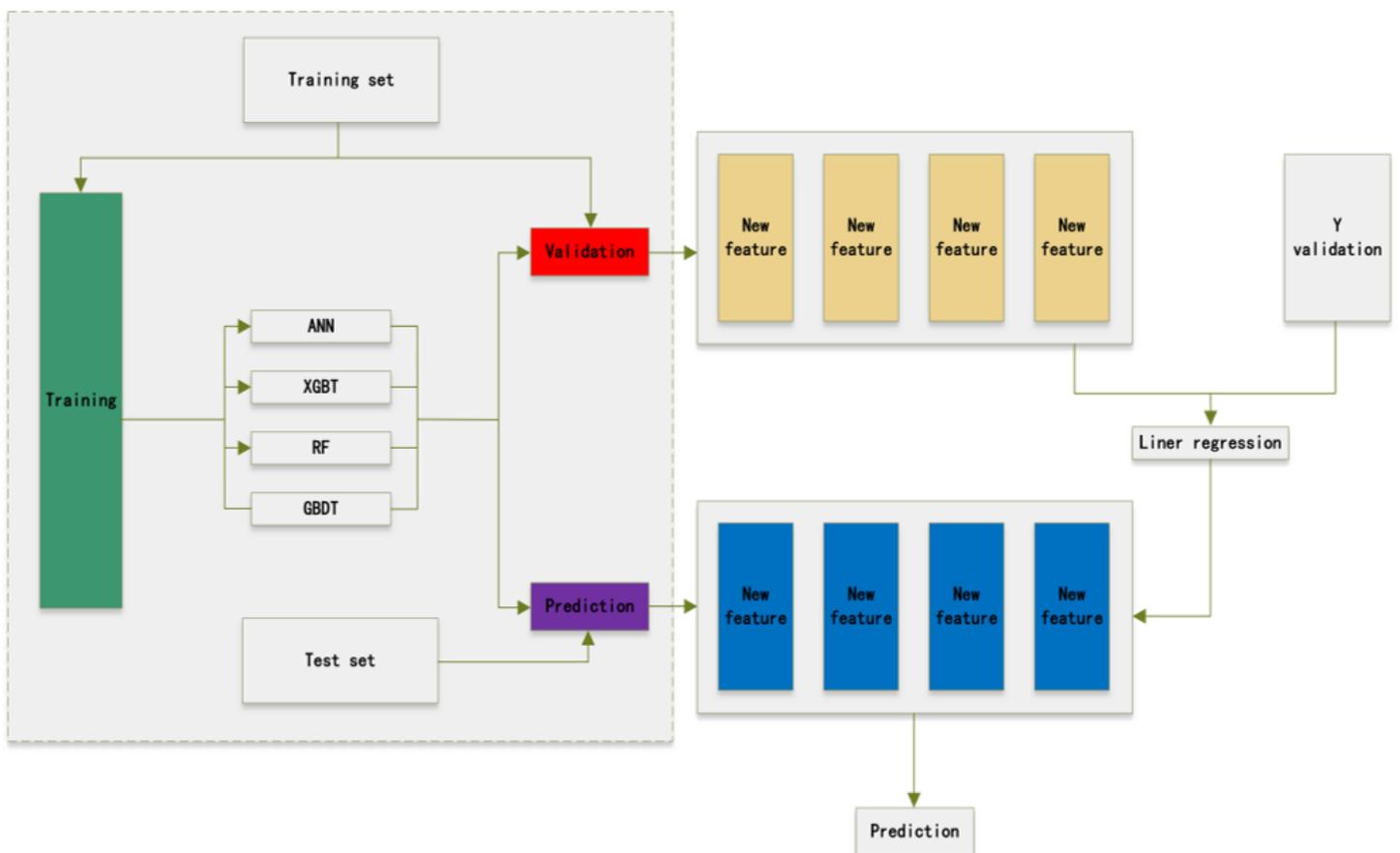


Figure 6

Fusion strategy of blending approach

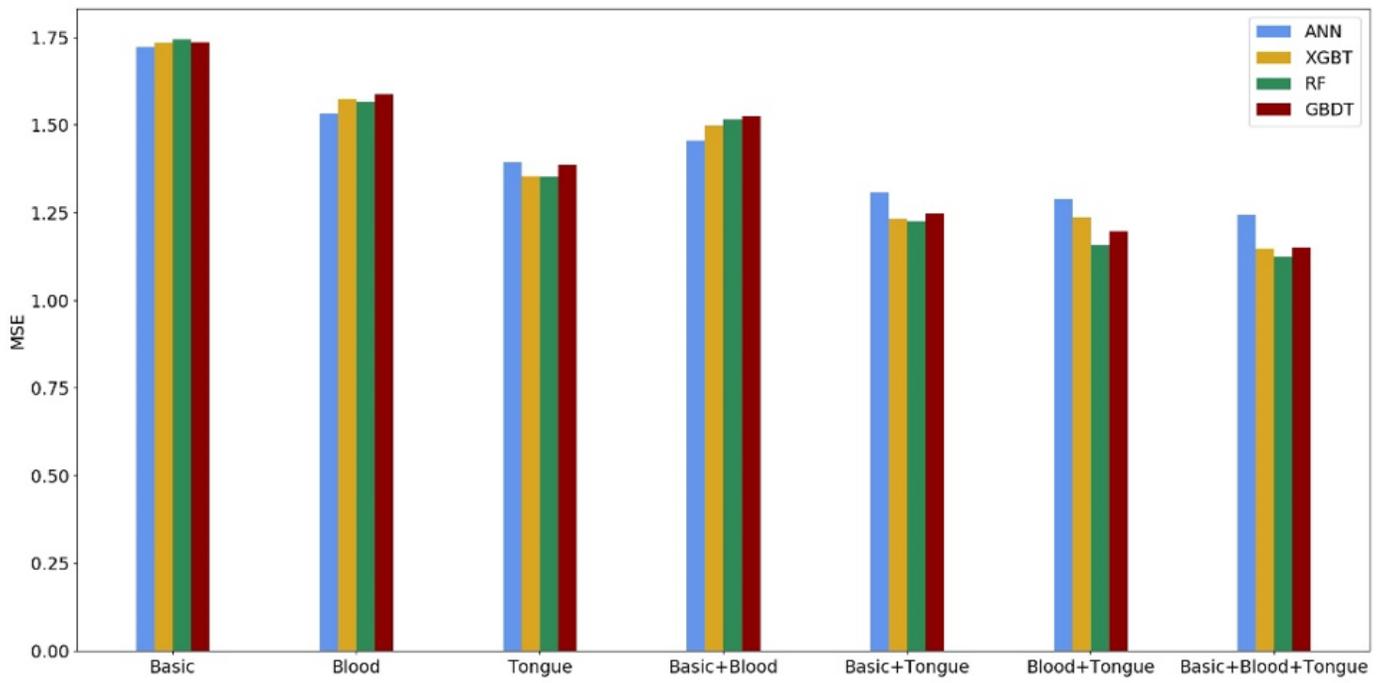


Figure 7

Average MSE of various FPG prediction models with different types of features using cross validation.

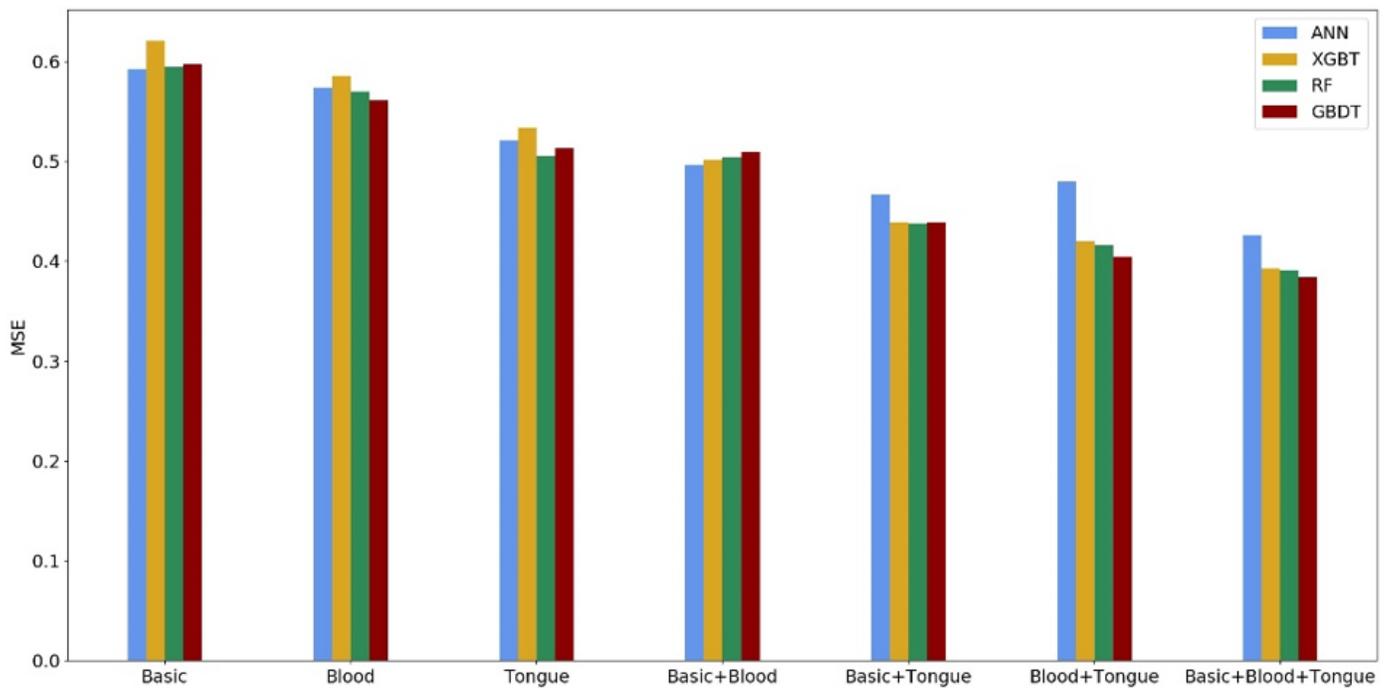


Figure 8

Average MSE of various HbA1c prediction models with different types of features using cross validation.

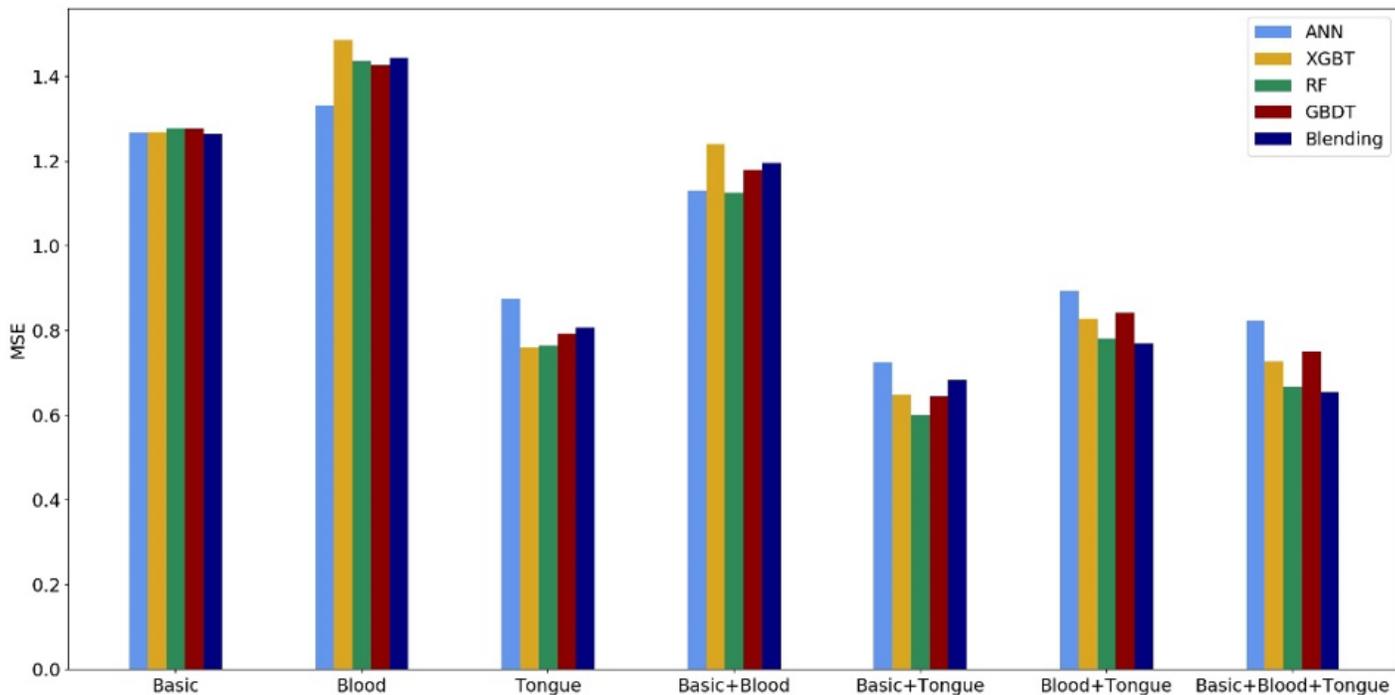


Figure 9

MSE of various FPG prediction models with different types of features on the test set.

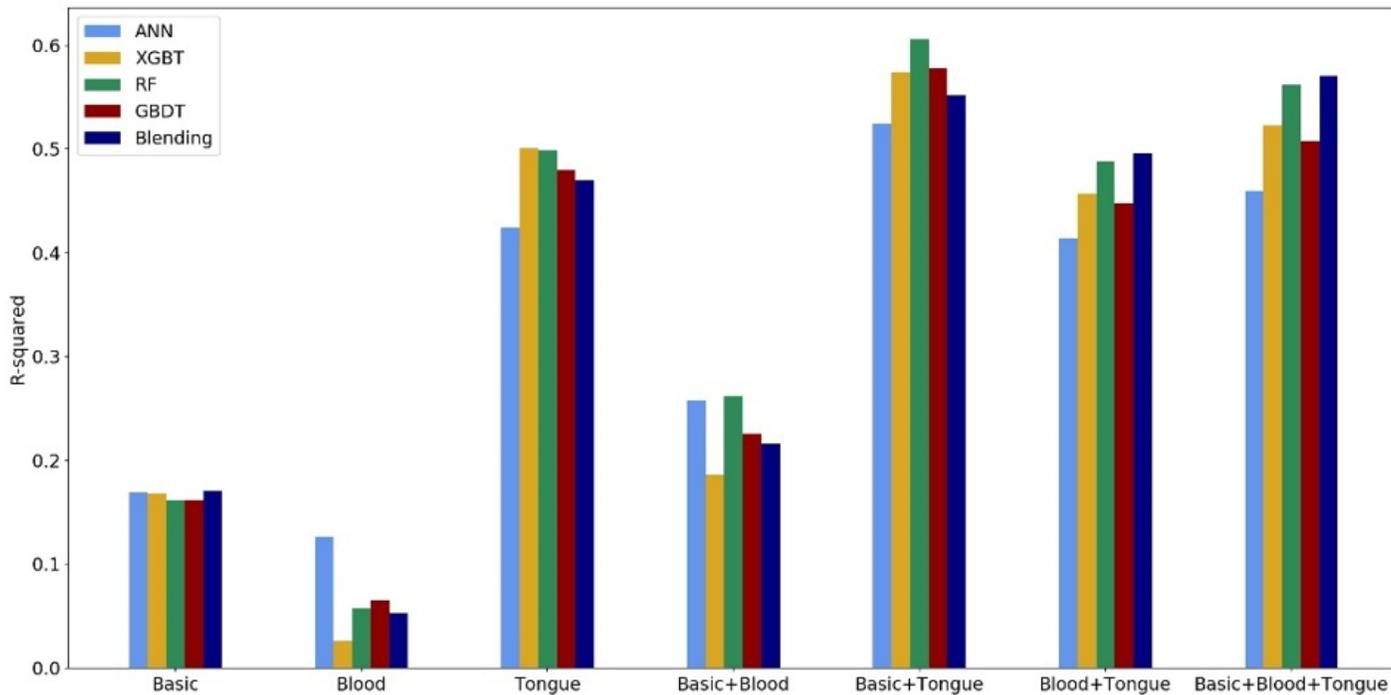


Figure 10

R-squared of various FPG prediction models with different types of features on the test set.

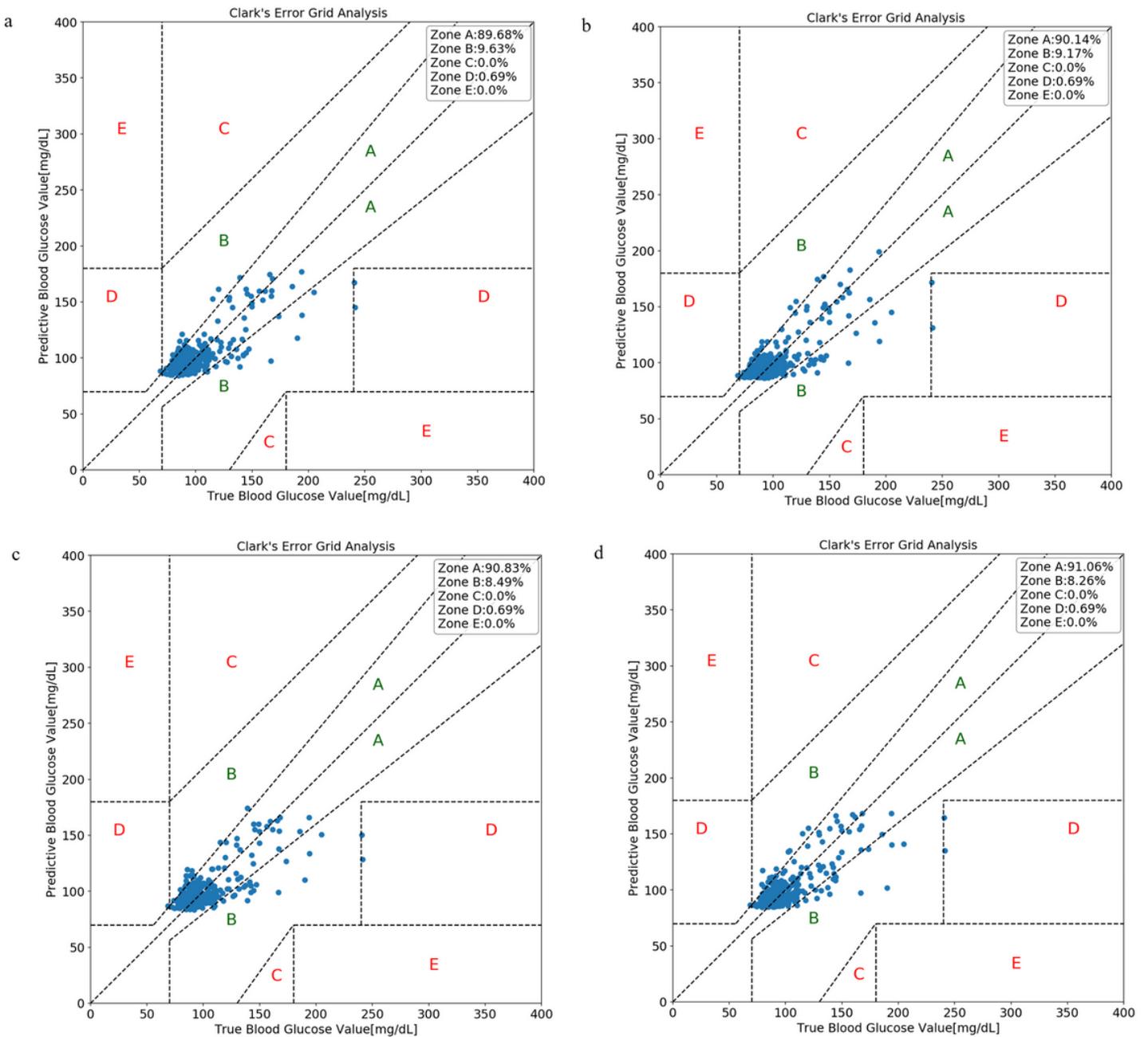


Figure 11

The EGA results of (a)RF model with non-invasive features, (b)XGBT model with non-invasive features, (c)blending model with non-invasive features and (d)blending model with full features on the test set. Non-invasive features refer to tongue features and basic features combined.

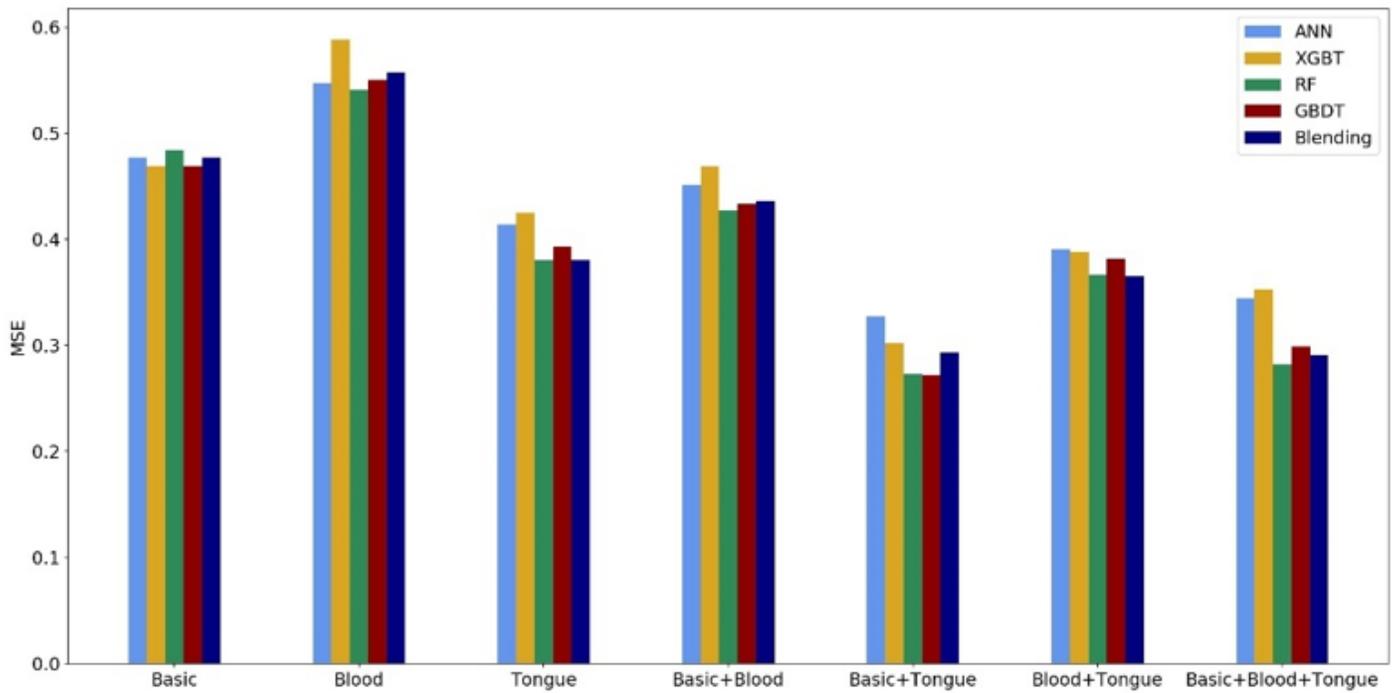


Figure 12

MSE of various HbA1c prediction models with different types of features on the test set.

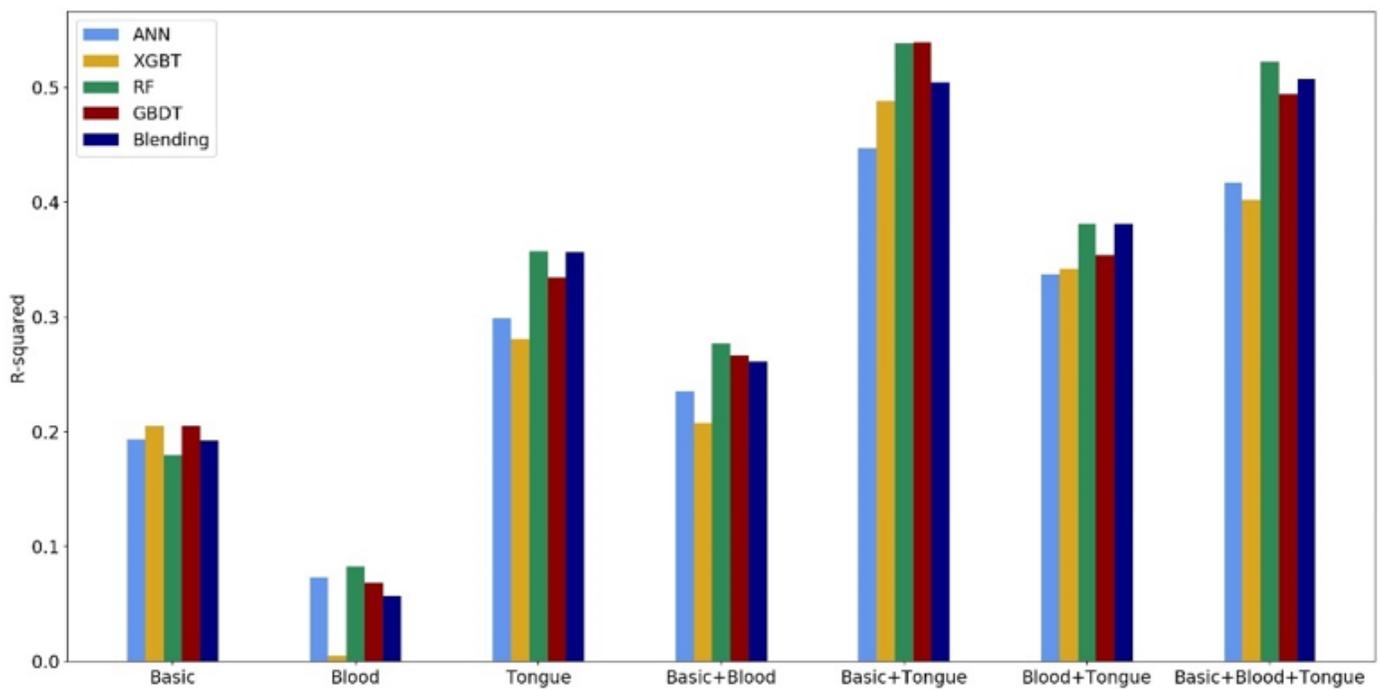


Figure 13

R-squared of various HbA1c prediction models with different types of features on the test set.

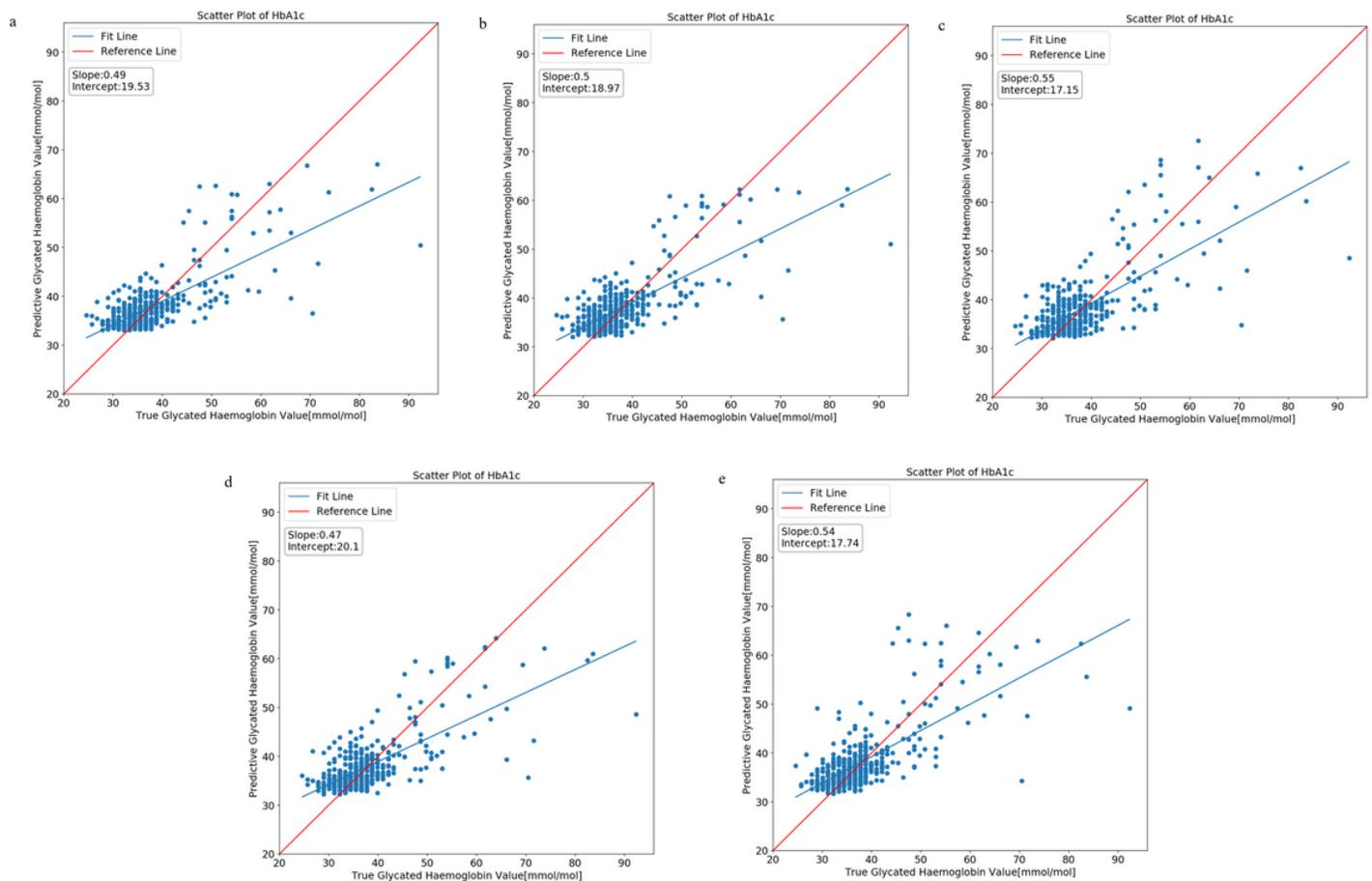


Figure 14

Scatter-plot of HbA1c predicted by (a) GBDT model with non-invasive features, (b) RF model with non-invasive features, (c) XGBT model with non-invasive features, (d) blending model with non-invasive features and (e) blending model with full features on the test set.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [graghabstract.jpg](#)