A new model for predicting the occurrence of polycystic ovary syndrome: Based on data of tongue and pulse

Weiying Wang[1,*] 🆔, Weiwei Zeng[1,*], Shunli He[1], Yulin Shi[2], Xinmin Chen[1], Liping Tu[2], Bingyi Yang[1], Jiatuo Xu[2] and Xiuqi Yin[1]

Abstract

Background and objective: Polycystic ovary syndrome is one of the most common types of endocrine and metabolic diseases in women of reproductive age that needs to be screened early and assessed non-invasively. The objective of the current study was to develop prediction models for polycystic ovary syndrome based on data of tongue and pulse using machine learning techniques.

Methods: A dataset of 285 polycystic ovary syndrome patients and 201 healthy women were investigated to identify the significant tongue and pulse parameters for predicting polycystic ovary syndrome. In this study, feature selection was performed using least absolute shrinkage and selection operator regression. Several machine learning algorithms (multilayer perceptron classifier, eXtreme gradient boosting classifier, and support vector machine) were used to construct the classification models to predict the presence of polycystic ovary syndrome.

Results: TB-L, TB-a, TB-b, TC-L, TC-a, h_3, and h_4/h_1 in tongue and pulse parameters were statistically associated with polycystic ovary syndrome presence. Among the several machine learning techniques, the support vector machine model was optimal for the comprehensive evaluation of this dataset and deduced the area under the receiver operating characteristic curve, DeLong test, calibration curve, and decision curve analysis.

Conclusion: The machine learning model with tongue and pulse factors can predict the existence of polycystic ovary syndrome precisely.

Keywords

Polycystic ovarian syndrome, tongue diagnosis, pulse diagnosis, machine learning

Submission date: 8 August 2022; Acceptance date: 12 February 2023

Introduction

Polycystic ovary syndrome (PCOS) is the most common endocrine disease in women of childbearing age, affecting 6%–13% of females worldwide.[1–4] It has diverse clinical manifestations as it is a heterogeneous syndrome of systemic neuro-endocrine-metabolic network disorders.[5] The main symptoms are ovulatory and menstrual with high androgen levels, depression, and other mood disturbances.[6,7] Some studies have shown that PCOS is a lifelong metabolic disorder with evidence of the increased prevalence of insulin resistance, metabolic syndrome,

[1]Department of Gynecology and Obstetrics, Shuguang Hospital Affiliated to Shanghai University of Chinese Medicine, Shanghai, P.R. China
[2]Basic Medical College, Shanghai University of Traditional Chinese Medicine, Shanghai, P.R. China
*These authors contributed equally to this work.

Corresponding authors:
Xiuqi Yin, Department of Gynecology and Obstetrics, Shuguang Hospital Affiliated to Shanghai University of Chinese Medicine, Shanghai 201203, P.R. China.
Email: yxq0402@shutcm.edu.cn

Jiatuo Xu, Basic Medical College, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, P.R. China.
Email: xjt@fudan.edu.cn

cardiometabolic disease, and endometrial cancer.[8–10] Therefore, early screening of PCOS is essential to treat the disease and prevent its long-term complications. Currently, the diagnosis of the syndrome relies on ultrasonography and serology. Since the results of these methods are related to the women's menstrual cycle, and there are strict requirements on the detection time, the degree of patient cooperation is low. Tongue and pulse diagnoses in Traditional Chinese Medicine (TCM) are convenient and noninvasive detection methods that provide objective reference indicators for the diagnosis and treatment of the disease. Therefore, it is necessary to establish a PCOS risk prediction model based on tongue and pulse diagnosis data.

Tongue diagnosis and pulse diagnosis are essential components of TCM diagnosis. Doctors infer the occurrence and development of the disease by observing the tongue, tongue coating, and pulse. With the development of computer technology, many studies have been carried out on the acquisition and simulation of TCM pulse and tongue diagnoses. Current research determined that tongue and pulse are closely related to digestive, endocrine, and circulatory diseases.[11–13] There is also progress in female tongue and pulse research. Hsieh et al.[14] found differences in tongue color during the menstrual cycle in women. Another study suggested that there was a close relationship between the performance of the tongue and pulse in women and the incidence of coronary heart disease.[15] Thus, tongue and pulse conditions could be used to assess the physical condition of women.

Machine learning, a broad branch of artificial intelligence, is a multifield interdisciplinary subject involving probability theory, statistics, and computer science, and is widely used in various fields. The application of machine learning in clinical medicine simplifies the decision-making for doctors. Rubio et al.[16] used the logic model as a decision support system in the clinic, where the algorithm provided a decision analysis of the care trajectories of patients. Artificial intelligence and statistical analysis can be applied to identify and predict medical patterns from the extensive data of patients, and machine learning algorithms can be used to predict certain diseases based on tongue and pulse data. Li et al.[17] used the Stacking and ResNet50 models to predict type 2 diabetes based on tongue images. Shi et al.[18] used the neural network to establish a model for fatigue classification based on tongue and pulse data. Zhang et al.[19] proposed a Bayesian network-based classification system to diagnose cor pulmonale, appendicitis, gastritis, pancreatitis, and bronchitis.

The present study aimed to establish a noninvasive and convenient PCOS prediction model and evaluate the ability of tongue and pulse parameters combined to predict PCOS risk. This study used different machine learning to determine the accurate estimated scores for PCOS predictions by the proposed independent risk factors.

Material and methods

Study subjects

This retrospective study used a dataset of 285 patients diagnosed with PCOS. Data were collected from The Shuguang Hospital Affiliated with the Shanghai University of TCM, China. All patients were evaluated according to the 2003 Rotterdam diagnostic criteria.[20] We also enrolled 201 healthy female volunteers who comprised the control group in the Physical Examination Center of the hospital. The study followed the ethical principles of the Helsinki Declaration of Good Clinical Practice Guidelines. All participants signed informed consent, and the study was approved by the Medical Ethics Committee of Shuguang Hospital Affiliated with Shanghai University of TCM (authorization no. 2018-626-55-01).

Inclusion criteria and exclusion criteria

Inclusion criteria were as follows (2/3 should be fulfilled): The 2003 Rotterdam diagnostic criteria[20] were as follows:

1. Oligo- and/or anovulation;
2. Clinical and/or biochemical signs of hyperandrogenism;
3. Polycystic ovaries.

Exclusion criteria were as follows:

1. Cases that did not meet the inclusion criteria of this project;
2. Hyperandrogenism is caused by other causes, such as congenital adrenal hyperplasia, Cushing's syndrome, and androgen-secreting tumors;
3. Other causes of ovulation disorders, such as hyperprolactinemia, premature ovarian failure, pituitary or hypothalamic amenorrhea, and abnormal thyroid function;
4. Patients had taken hormonal medications (such as contraceptives, ovulation stimulants, and glucocorticoids) within the past month;
5. Patients with underlying severe diseases in internal medicine and surgery;
6. Patients who were unable to cooperate with completing the study plan, including those with a history of infectious diseases, mental disorders, and other medical records.

Clinical tongue and pulse data collection and analysis method

Tongue and pulse data collection. Tongue and pulse diagnosis data were collected using the TDA-1 tongue diagnosis instrument and PDA-1 pulse diagnosis instrument developed by the Shanghai University of TCM Intelligent Diagnosis Technology Research Team. The collection process was operated by Professional researchers from

Shuguang Hospital Affiliated with Shanghai University of TCM. Before tongue data collection, the staff used alcohol cotton balls to wipe the instrument for disinfection and confirmed that the examinee had an empty stomach, a clean mouth, and no foreign matter or stained coating on the tongue. The subjects took the sitting position, put the mandible on the mandible rest, opened the mouth, and stretched the tongue to fully expose the tongue surface. The tongue body was relaxed, and the tongue surface was flat and stretched.

Before pulse data collection, the subjects were required to ensure that they did not stay up late, and did not drink strong tea or coffee. The subjects were breathing calmly, sitting upright, with the upper arms relaxed, and the elbows bent about 120°. The forearms were naturally flat, the wrists were placed on the pulse pillow, and the palms were facing up. The PDA-1 single-segment pulse detector was placed in the sampler where the subject's pulse was most obvious, and it could automatically collect a single-part pulse diagram with a stable waveform of 30 s.

Tongue and pulse data analysis method. Tongue data were analyzed using the Tongue Diagnosis Analysis System (TDAS) developed by the same team. The data of tongue images were selected from the indexes of the CIE-Lab color space. Tongue body L value (TB-L) and tongue coating L value (TC-L) represented the lightness of 0–100, indicating the range from pure black to pure white. Tongue body a value (TB-a) and tongue coating a value (TC-a) represented the green-red axis, while the positive a represented red, and the negative a represented green. Tongue body b value (TB-b) and tongue coating b value (TC-b) represented the blue-yellow axis, while the positive b represented yellow, and the negative b represented blue.[18] Pulse data were analyzed by the pulse classification standard in "Modern TCM Pulse Diagnosis."[21] The specific parameters are defined in Supplemental Table 1.

Variable selection

Variable selection plays a critical role in machine learning and hence is used in model-building for simplifying, facilitating interpretation, and eliminating redundant variables.[22] The present study used Propensity Score Matching (PSM)[23] to accurately match the experimental and control groups' age variables and eliminate selection bias. The minor absolute shrinkage and selection operator Least Absolute Shrinkage and Selection Operator (LASSO) regression[24] was used for variable screening in this study; nonetheless, the sample size was small.

Statistical analysis

SPSS 26.0 software was used for statistical analysis. Continuous data with normal distribution were presented as mean $\pm$ standard deviation (SD), and those with abnormal distribution were represented as the median and interquartile range (IQR). IQR was equal to the third quartile minus the first quartile. The comparisons between groups were conducted using a t-test or Mann–Whitney *U*-test for continuous variables. In this study, the accuracy of the proposed prediction models was calculated using the area under the receiver operating characteristic curve (AUROC). Then, the accuracy, sensitivity, and specificity were calculated. The DeLong test was performed to examine the differences between the ROC curves of various machine learning classifiers. The area under the curve (AUC) was directly compared in a pairwise approach using paired DeLong test.[25] $P < 0.05$ (two-tailed) indicated a statistically significant difference in the comparisons when the inspection level was $\alpha = 0.05$. A calibration curve was used to evaluate the model performance. In addition, decision curve analysis (DCA) was used to assess the clinical benefit of the models. It was a suitable method for evaluating alternative diagnostic and prognostic strategies that exhibited advantages over other commonly used measures and techniques.[26]

Machine learning approach

In this study, three machine learning classification algorithms were used to establish a PCOS risk prediction model: Multilayer Perceptron (MLP) classifier, eXtreme Gradient Boosting (XGBoost) classifier, and Support Vector Machine (SVM). Each algorithm incorporates tongue and pulse parameters to build models and make two-class predictions, respectively. MLP is a feedforward artificial neural network consisting of three layers: an input layer, one or more hidden layers, and an output layer.[27] XGBoost is an implementation of gradient boosting decision trees focused on computational speed and model performance. It achieves the best results by adding multiple decision trees.[28] SVM is a supervised machine learning algorithm which proves simpler in structure and offers global optimality and good generalization; hence, it has been widely applied in classification.[29–31] The selection of patients' tongue and pulse feature parameters and model classification are shown in Figure 1. All machine learning processes were carried out using Python 3.7.

Results

Baseline statistics

The dataset included 486 subjects (285 PCOS patients and 201 healthy women) aged 20–44 years. The height, weight, BMI, and age of women in the healthy group and the PCOS group do not conform to the normal distribution, so the relevant data are represented by median (IQR). The height of the subjects in the healthy group was 1.640 m (0.080), the
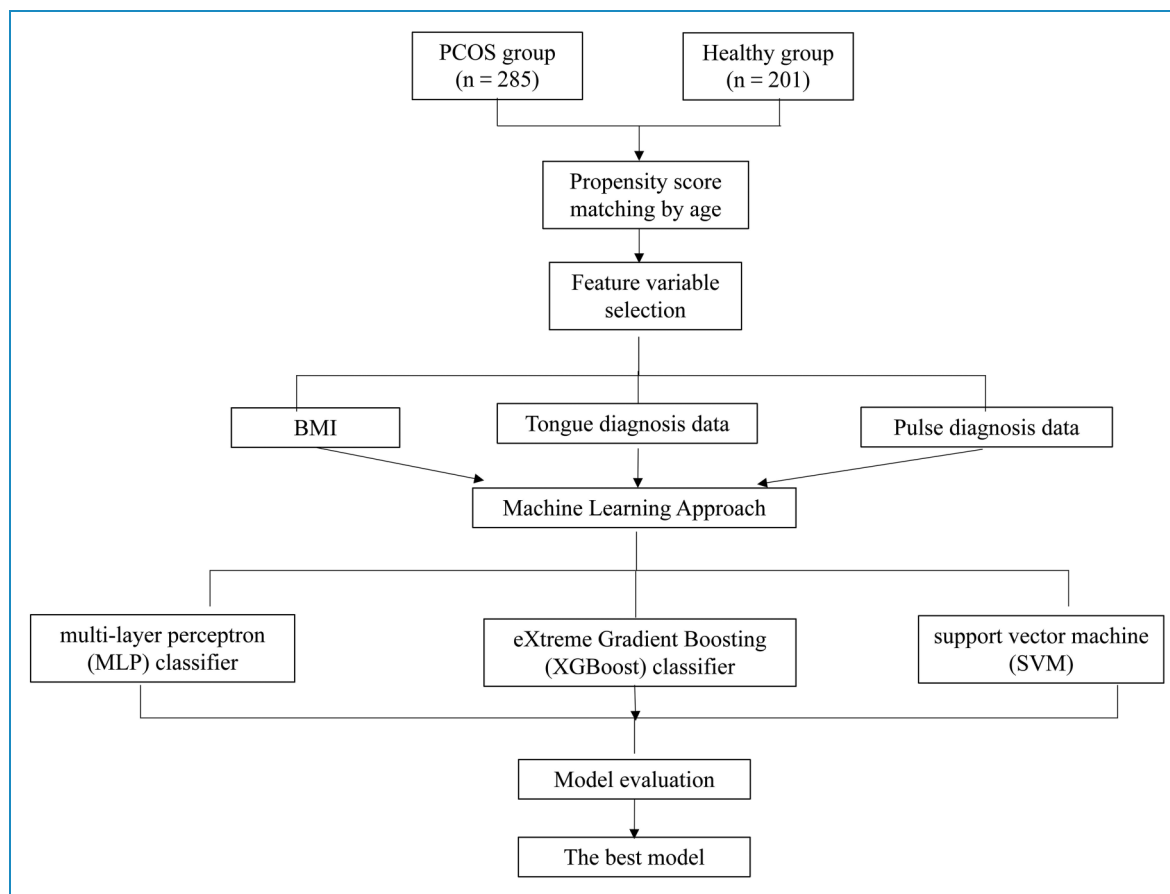
Figure 1. Overall flowchart.

Table 1. Original baseline characteristics in the dataset.

Baseline characteristics	Healthy group (n = 201)	PCOS group (n = 285)	Statistical magnitude	P-value
Height (m, median (IQR))	1.640 (0.080)	1.620 (0.030)	4.085	<0.001
Weight (kg, median (IQR))	61.000 (9.200)	62.100 (8.800)	−2.617	0.009
BMI (kg/m², median (IQR))	22.684 (2.534)	24.221 (2.384)	−6.507	< 0.001
Age (years, median (IQR))	28.000 (5.000)	31.000 (6.000)	−4.900	< 0.001

PCOS: polycystic ovary syndrome; IQR: interquartile range; BMI: body mass index.

height of the subjects in the PCOS group was 1.620 (0.030) m; the weight of the subjects in the healthy group was 61.000 (9.200) kg, the weight of the subjects in the PCOS group was 62.100 (8.800) kg; the BMI of the subjects in the healthy group was 22 .684 (2.534) kg/m², PCOS group subjects BMI was 24.221 (2.384) kg/m². There were differences in height, weight, and BMI between the two groups of subjects ($P < 0.05$). In addition, the age of subjects in the healthy group was 28.000 (5.000) years old, and the age of subjects in the PCOS group was 31.000 (6.000) years old. There was a statistical difference in age between the two groups ($P < 0.05$). The baseline characteristics of the included subjects are listed in Table 1.

In order to exclude the interference of age differences in this study, we used PSM to take the occurrence of PCOS as the outcome, together with the age of the two groups of subjects as the scoring item in a ratio of 1:1. The corrected baseline data was obtained after accurate matching. After PSM matching, the age of the subjects in the healthy group was 28.000 (5.000) years old, and the age of the

subjects in the PCOS group was 28.000 (5.000) years old. There was no statistical difference in the ages of the two groups of subjects ($P > 0.05$), which met the research requirements. After matching, the height of the subjects in the healthy group was 1.640 (0.080) m, the height of the subjects in the PCOS group was 1.620 (0.030) m; the weight of the subjects in the healthy group was 61.000 (9.200) kg, and the weight of the subjects in the PCOS group was 62.400 (7.900) kg; The BMI of the subjects in the PCOS group was 22.684 (2.534) kg/m^2, and the BMI of the subjects in the PCOS group was 24.221 (2.360) kg/m^2. There were still differences in height, weight, and BMI between the two groups ($P < 0.05$). The baseline data of the two groups after matching are shown in Table 2.

Data analysis of tongue and pulse

The tongue and pulse data of women in the healthy group and the PCOS group do not conform to the normal distribution, so the relevant data are represented by the median (IQR). The clinical tongue and pulse data of the subjects are shown in Table 3. It can be seen from Table 3 that in terms of tongue coating in the PCOS group: the TC-L value and TC-a value were significantly lower than those of the healthy group. PCOS group's TC-b value was significantly higher than that of the healthy group. In terms of tongue quality in the PCOS group: the TB-L value and The TB-a value were significantly lower than those of the healthy group. The TB-b value of the PCOS group was significantly higher than that of the healthy group. The differences in the above tongue data were statistically significant ($P < 0.05$). In addition, in terms of pulse data, t_1/t, h_4/h_1 in the PCOS group were significantly higher than those in the healthy group, and h_1/t_1, h_3/h_1, t, t_4, t_5, h_1, h_3, h_4, and h_5 were significantly lower than those in the healthy group. Also, significant differences were detected in the above pulse data ($P < 0.05$).

Variable filter

The "Glmnet" package (version 4.1.2) was applied to fit the logistic LASSO regression. Feature selection was performed using Lasso, and the Lasso model's tuning parameter (λ) was selected by a minimum criterion using 10-fold cross-validation. BMI, tongue data, and pulse data were included in the variable selection dataset. AUROC versus log(λ) was plotted. Vertical dashed lines were drawn at the optimal values with the logarithm λ of the minimum mean squared error and the logarithm λ of the standard error of the minimum distance (Figure 2A and B). According to 10-fold cross-validation, the λ of the minimum mean squared error was 0.006, and the variable selection of the corresponding model was TB-L + TB-a + TB-b + TC-L + TC-a + h_3 + h_4/h_1. Thus, a PCOS risk prediction model was established using these seven variables as independent factors.

Machine learning model evaluation

In the PCOS and healthy female subgroups, the characteristic parameters of tongue and pulse data obtained by screening were used to establish a PCOS prediction model based on four machine learning classifiers: MLP, XGBoost, and SVM. The ROC mean and SD of each model were calculated by repeated sampling; the selection was ten times. The validation set of each resampling training accounted for 20% of the total sample, and the training set accounted for 80%, indicating that each group was selected as the test set in turn to ensure that the evaluation results were not accidental. Then, the ten evaluation results were averaged to reduce the errors due to unreasonable choices in the test set. The DeLong test was performed to examine the differences between the ROC curves of different machine learning classifiers, and its clinical application was explored by DCA.

The parameters of the MLP classifier model were as follows: activation selected relu, the hidden layer width was (12, 12), and the number of iterations was 100. The specific parameters of the XGBoost classifier model were as follows: the optimized objective function was binary logistic, the learning rate was 0.002, the max depth was 10, the minimum child weight was 10, and the regularized coefficient was 1. The SVM model parameters were as follows: the regularization factor was 0.1, the kernel type

Table 2. Baseline characteristics after correction in the dataset.

Baseline characteristics	Healthy group (n = 201)	PCOS group (n = 201)	Statistical magnitude	P-value
Height (m, median (IQR))	1.640 (0.080)	1.620 (0.030)	3.653	<0.001
Weight (kg, median (IQR))	61.000 (9.200)	62.400 (7.900)	−2.738	0.006
BMI (kg/m^2, median (IQR))	22.684 (2.534)	24.221 (2.360)	−6.183	<0.001
Age (years, median (IQR))	28.000 (5.000)	28.000 (5.000)	−1.820	0.068

PCOS: polycystic ovary syndrome; IQR: interquartile range; BMI: body mass index.

was sigmoid, and the convergence metric was set to 0.01. The training and test set results of the four model classifications are summarized in Tables 4 and 5.

The machine learning ROC curve graphs are shown in Figures 3 and 4. The AUCs of the three machine learning models were >0.9, indicating high predictive values. Then, the DeLong test was performed on the ROC curves of the three machine learning methods. The results showed no statistically significant differences between the ROC curves of the above machine learning models ($P >$ 0.05) (Table 6). The calibration curves for the validation sets of multiple models (Figure 5) showed that, except for the XGBoost model, the predicted probabilities of the other machine classification models were close to the true probabilities. Moreover, the SVM model was closer to the reference value of the diagonal. Subsequently, we used DCA to assess whether the model could guide the net gain of the validation cohort. The DCA curve of the validation set is illustrated in Figure 6. The SVM model had significant utility in clinical decision-making in the entire validation cohort. Taken together, SVM is considered the best model choice for this dataset.

Table 3. Statistical analysis of tongue and pulse indexes.

Data of tongue and pulse	Healthy group (n = 201)	PCOS group (n = 201)	Statistical magnitude	P-value
TC-L, median (IQR)	104.964 (4.366)	90.028 (1.994)	17.103	<0.001
TC-a, median (IQR)	16.862 (2.008)	8.700 (2.867)	17.274	<0.001
TC-b, median (IQR)	4.570 (1.328)	7.715 (8.249)	−6.103	<0.001
TB-L, median (IQR)	103.820 (1.930)	97.789 (2.028)	16.685	<0.001
TB-a, median (IQR)	23.280 (2.463)	11.653 (2.922)	17.322	<0.001
TB-b, median (IQR)	6.228 (1.480)	9.083 (7.867)	−6.781	<0.001
w_1/t, median (IQR)	0.190 (0.060)	0.200 (0.070)	−2.369	0.018
w_2/t, median (IQR)	0.130 (0.040)	0.130 (0.040)	−1.212	0.224
t_1/t, median (IQR)	0.150 (0.030)	0.160 (0.040)	−5.237	<0.001
t_4/t_5, median (IQR)	0.840 (0.080)	0.850 (0.060)	−1.688	0.091
h_1/t_1, median (IQR)	3.970 (1.830)	3.820 (2.550)	4.067	<0.001
h_3/h_1, median (IQR)	0.590 (0.120)	0.510 (0.120)	6.022	<0.001
h_4/h_1, median (IQR)	0.260 (0.090)	0.320 (0.090)	−7.405	<0.001
t_1, median (IQR)	0.120 (0.020)	0.120 (0.020)	2.072	0.035
t_4, median (IQR)	0.350 (0.030)	0.330 (0.030)	9.262	<0.001
t_5, median (IQR)	0.420 (0.020)	0.390 (0.030)	9.104	<0.001
t, median (IQR)	0.850 (0.140)	0.750 (0.160)	8.523	<0.001
h_1, median (IQR)	12.470 (6.06)	7.890 (7.02)	8.921	<0.001
h_3, median (IQR)	7.370 (3.650)	3.400 (2.060)	13.037	<0.001
h_4, median (IQR)	3.281 (1.350)	2.100 (1.390)	9.064	<0.001
h_5, median (IQR)	0.550 (0.590)	0.290 (0.310)	6.541	<0.001

PCOS: polycystic ovary syndrome; IQR: interquartile range; TB-a: tongue body a value; TB-b: tongue body b value; TB-L: tongue body L value; TC-a: tongue coating a value; TC-b: tongue coating b value; TC-L: tongue coating L value.
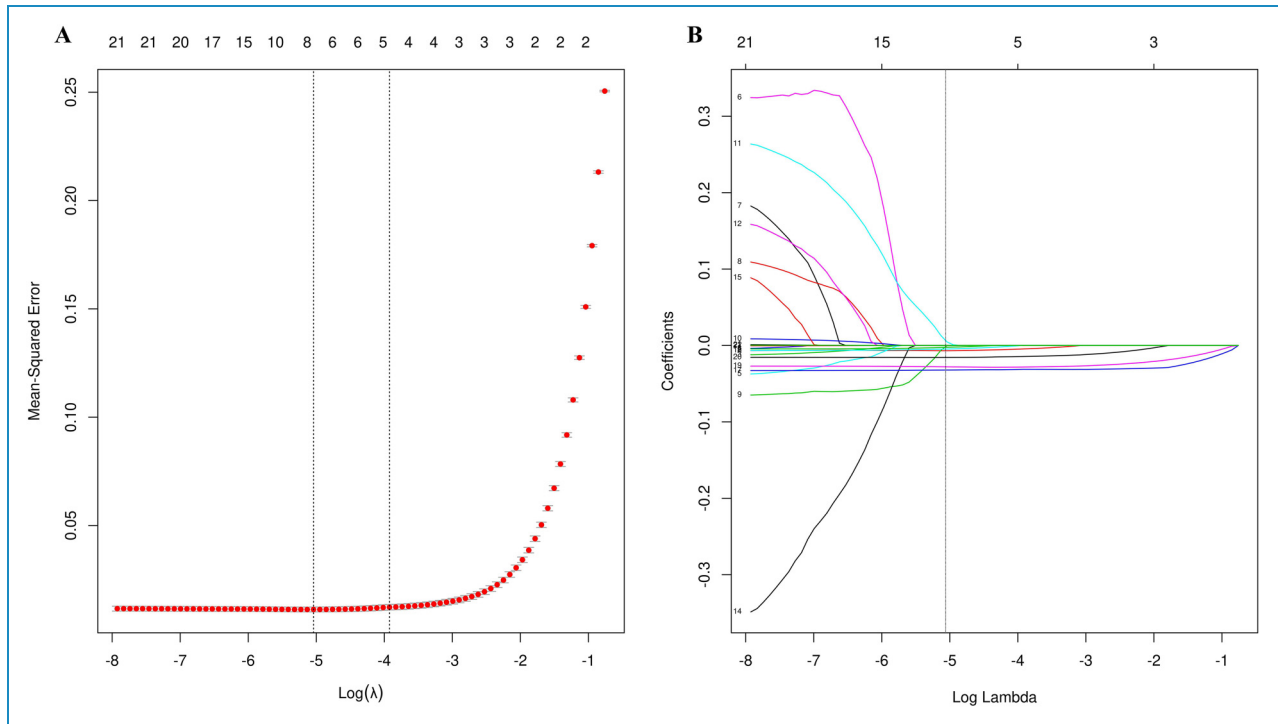
Figure 2. Tongue and pulse features selection using the LASSO binary logistic regression model. (A) The LASSO model's tuning parameter (λ) was chosen to satisfy the minimum criteria using ten-fold cross-validation. The AUC curve versus log(λ) was plotted. The two vertical dashed lines in the graph represent the logarithm λ of the minimum mean squared error (left dashed line, λ = 0.006) and the logarithm λ of the standard error of the minimum distance (right dashed line, λ = 0.02). (B) LASSO coefficient profiles of the 21 tongue and pulse features and BMI baseline characteristics. A coefficient profile plot was produced against the log(λ) sequence. A vertical line was drawn at the value selected using the minimum mean squared error, where optimal λ resulted in seven nonzero coefficients.
LASSO: least absolute shrinkage and selection operator; AUC: area under the curve; BMI: body mass index.

Discussion

PCOS is a common gynecological endocrine disease that severely affects the quality of life of patients. Although ultrasound and serum biochemical tests can be used to diagnose some types of polycystic diseases according to the Rotterdam guidelines,[20] these were operator- and woman's menstrual cycle-dependent methods. Ultrasound cannot reliably detect polycystic ovary morphology in patients with type B PCOS.[32] Furthermore, it is difficult to ensure that all patients receive daytime serology in high-dose screening. Also, the asynchrony in time and cost of ultrasound and serum biochemical assays greatly diminishes their clinical application value.

Tongue and pulse diagnoses in TCM are simple and direct methods to obtain clinical diagnosis and treatment information. Also, these are noninvasive techniques to detect the disease, determine disease severity, and evaluate treatment effects. In recent years, the rapid development of computer technology has facilitated various machine-learning methods in medical research. Using machine learning models to predict and classify the objective information of TCM diagnosis is beneficial to the development of artificial intelligence for TCM diagnosis.[33–35] Hence, we conducted a series of studies on this aspect. Tongue and pulse

diagnoses are valuable in classifying and predicting non-alcoholic fatty liver disease,[36] type 2 diabetes,[17,34] and hypertension.[33,37] However, the TCM tongue-pulse diagnostic technique for diagnosing and predicting PCOS has not yet been reported. The present study aimed to establish a machine learning model based on lingual pulse characteristic parameters as a noninvasive method to predict the onset of PCOS.

We analyzed and documented the data of 285 PCOS patients and 201 healthy women. To minimize multicollinearity between tongue and pulse data and avoid overfitting, we applied LASSO regression to screen the variables. The results showed that seven characteristic parameters of tongue and pulse (TB-L, TB-a, TB-b, TC-L, TC-a, h_3, and h_4/h_1) were closely related to PCOS occurrence. These features were added as independent factors to establish the PCOS prediction models.

CIE-Lab was selected as the color space for feature extraction from tongue imaging data. The most prominent feature of CIE-Lab is the smooth changes in color space, which defines all the colors that can be perceived by the human eye. The amount of spatial color shift in the CIE-Lab color space does not change with devices and is

**Table 4.** Summary of the training set results of multiple model classifications.

Machine learning model	AUC (SD)	Accuracy (SD)	Sensitivity (SD)	Specificity (SD)
MLP classifier	0.910 (0.179)	0.904 (0.158)	0.857 (0.251)	0.978 (0.027)
XGBoost classifier	0.998 (0.003)	0.761 (0.282)	0.996 (0.005)	0.993 (0.007)
SVM	0.979 (0.003)	0.957 (0.003)	0.979 (0.014)	0.933 (0.013)

AUC: area under the curve; SD: standard deviation; SVM: support vector machine; XGBoost: eXtreme gradient boosting; MLP: multi-layer perceptron.

Table 5. Summary of the validation set results of multiple model classifications.

Machine learning model	AUC (SD)	Accuracy (SD)	Sensitivity (SD)	Specificity (SD)
MLP classifier	0.926 (0.157)	0.900 (0.158)	0.872 (0.228)	0.979 (0.036)
XGBoost classifier	0.993 (0.010)	0.751 (0.287)	0.991 (0.009)	0.983 (0.019)
SVM	0.980 (0.016)	0.944 (0.015)	0.971 (0.019)	0.956 (0.024)

AUC: area under the curve; SD: standard deviation; SVM: support vector machine; XGBoost: eXtreme gradient boosting; MLP: multi-layer perceptron.
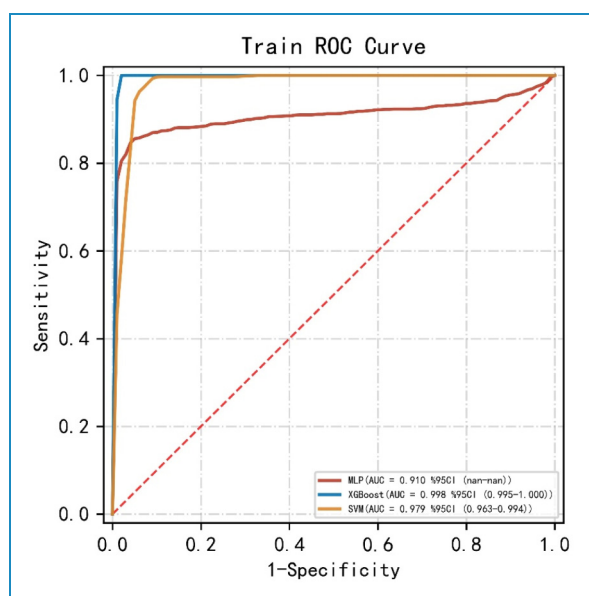
Figure 3. Receiver operating characteristic (ROC) curve for multiple model classifications of the training set.
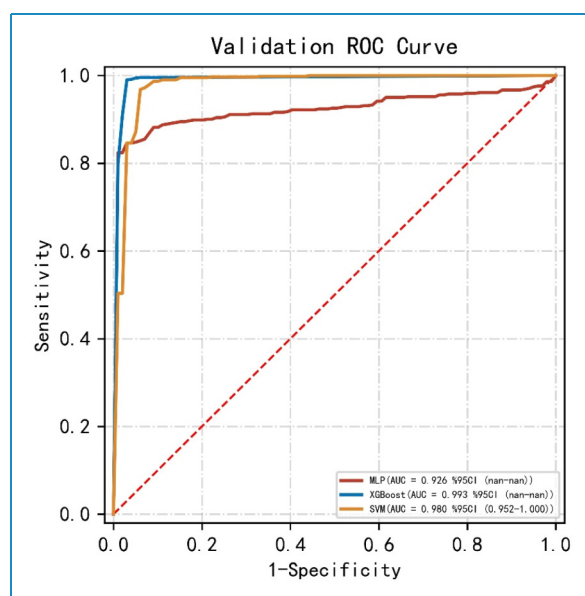
Figure 4. Receiver operating characteristic (ROC) curve for multiple model classifications of the validation set.

more valuable than the RGB color space in digital images for compensating for the lack of color space.[38]

As shown in Figure 2, this study found that TB-L, TB-a, TB-b, TC-L, and TC-a were essential parameters of tongue characteristics to distinguish PCOS patients from healthy women. The tongue colors of TB-L and TC-L were lesser in the PCOS group than in the healthy group ($P < 0.05$), suggesting that the colors of the tongue body and tongue coating of the PCOS patients were darker.

According to the TCM theory, these results indicated that the tongue of healthy women is moister and brighter than that of PCOS patients, indicating that the qi and blood circulation of PCOS patients are not as smooth as that of healthy individuals. The TB-a and TC-a values of the tongue coating in the PCOS group were lower than those in the healthy group ($P < 0.05$), indicating that the tongue body and tongue coating in the PCOS group were not as high as the red value in the healthy group. The tongue

consists of blood vessels and nerves. The tongue of a healthy person is full of blood, and hence, the color of the front of the tongue is reddish, implying that the blood flow of the tongue of patients with PCOS is not as smooth as that of healthy people, and the tongue's color reflects the lower value of red color. These findings also suggested that the tongue coating in the PCOS group is

thick, and the color exposure of the tongue is not apparent. The TB-b value of the tongue coating in the PCOS group was higher than that in the healthy group ($P < 0.05$), indicating that the tongue body of PCOS women was yellower than that of healthy women. The indicators of tongue yellowness in PCOS patients may be influenced by their thick tongue coating color. The tongue coating results from a mixture of bacteria and food residues with the tongue mucosa and the elongated stratum corneum of the hyperkeratinized epithelium.[39] Hsieh et al.[14] proposed that the yellow color of the tongue coating is due to the yellowish appearance of sulfur compounds, such as dimethyl disulfide, dimethyl trisulfide, and 2,4-dithiopentane.[40] Thus, it can thus be suggested that sulfide concentrations in the oral cavity of PCOS were higher than those in healthy women.

In addition, h_3 and h_4/h_1, the characteristic indicators of pulse condition, were also included in the analysis of the machine learning model. The h_3 value means heavy wavefront wave amplitude and reflects arterial elasticity. These findings indicated that PCOS has a lower h3 value than healthy women. The weakening of arterial elasticity is

Table 6. Delong detection results for multiple model classification.

Machine learning model		MLP classifier	XGBoost classifier
XGBoost classifier	Z-value	2.149	
	P-value	0.246	
SVM	Z-value	2.149	1.122
	P-value	0.269	0.283

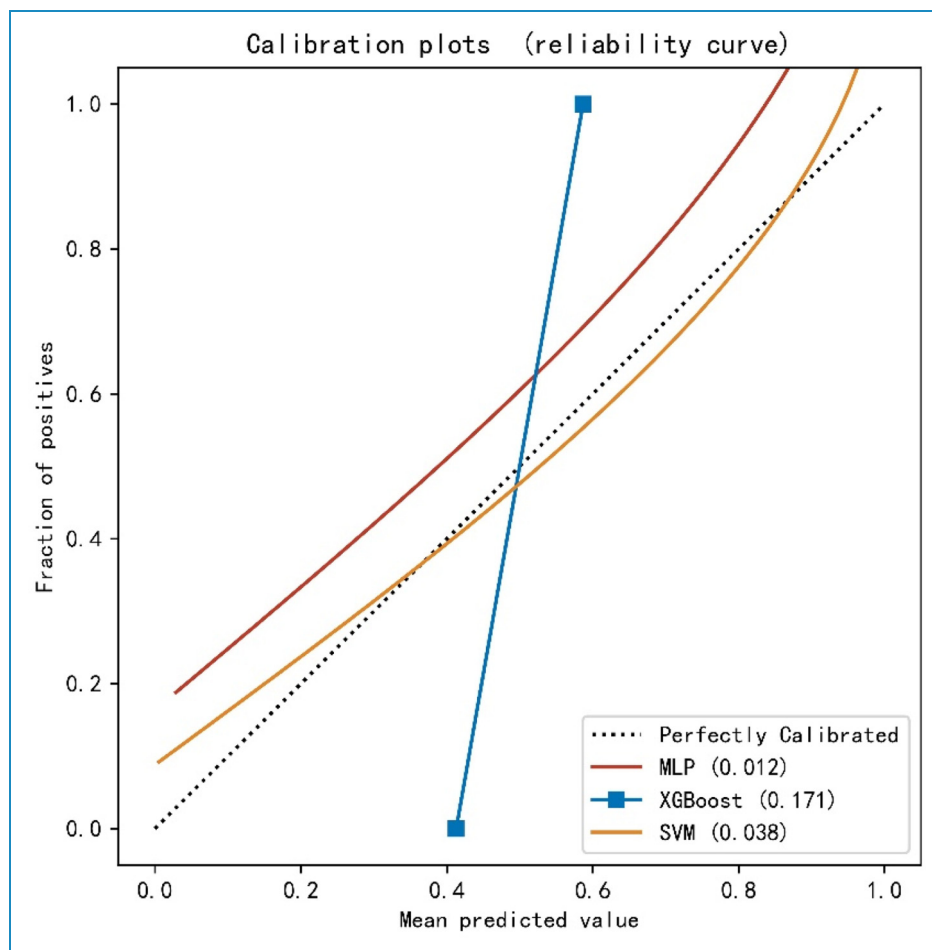SVM: support vector machine; XGBoost: eXtreme gradient boosting; MLP: multi-layer perceptron.

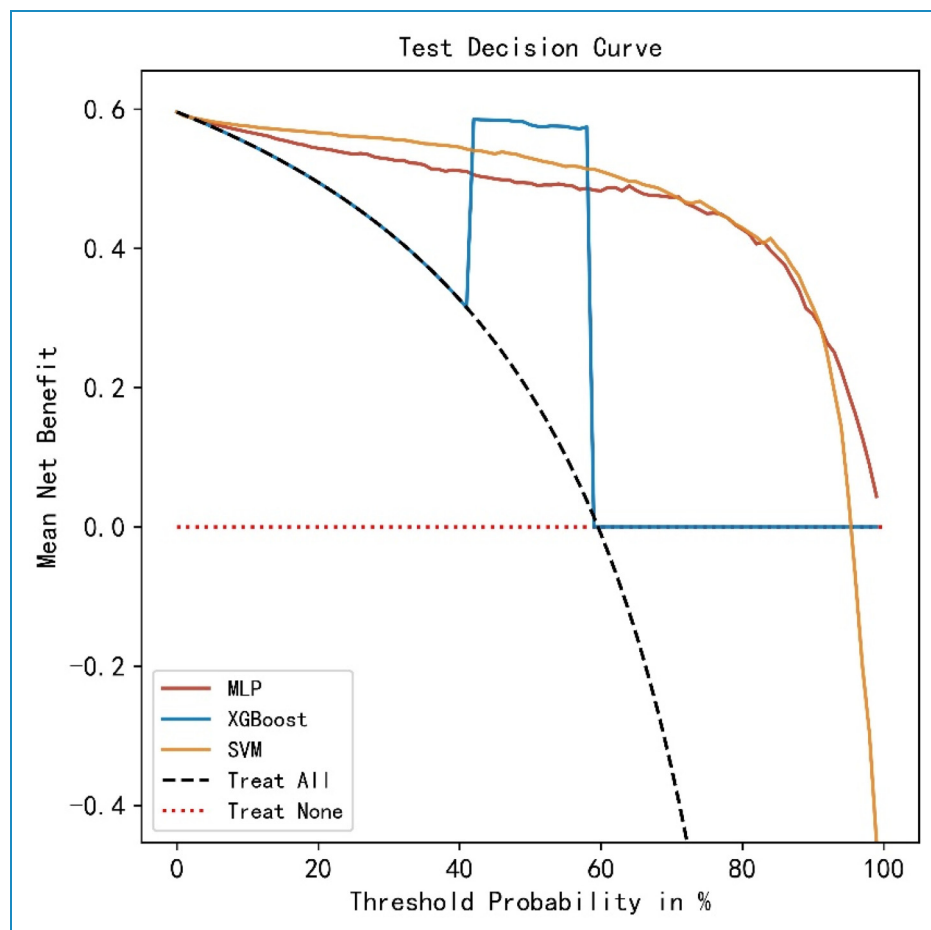**Figure 5.** Calibration curves for the validation sets of multiple models.

Figure 6. Decision curve analysis (DCA) on the validation set.

due to the gradual deposition of excessive lipids in the blood, forming plaques in the arterial intima. Newman et al.[41] showed that women with PCOS disease often exhibit an atherogenic lipid profile characterized by increased triglycerides, low high-density lipoprotein cholesterol, and standard or increased low-density lipoprotein cholesterol, which is consistent with our findings. The h_4/h_1 mainly reflects the level of peripheral resistance. The h_4/h_1 value of pulse in the PCOS group was higher than that in the healthy group, suggesting that resistance to peripheral vascular contraction increased in patients with PCOS. Interestingly, increased peripheral resistance restricts the aortic blood flow to the periphery. Hence, the blood remains in the arterial blood vessels, and the pressure on the blood vessel wall increases, resulting in significantly increased diastolic blood pressure. Our findings demonstrated that individuals with PCOS are more at risk of high blood pressure than healthy people. This finding is consistent with that of previous studies, wherein women with PCOS were more hypertensive and frequently used antihypertensive medication than the controls.[42]

Machine learning is an implementation method of artificial intelligence and a major research direction in life sciences. In this study, three algorithms, MLP classifier,

XGBoost classifier, and SVM, were used for classification modeling. Combined with AUROC, DeLong test, and DCA curve evaluation, the SVM prediction model has the best classification effect and net income. It is a novel and applicable few-shot learning method with a solid theoretical foundation. Also, it is not sensitive to outliers and can help us capture critical samples. The advantage of the method is the simple algorithm and robustness. Herein, we focused on the AUROC obtained by these predictive models and compared the DCA curves of different models. In addition to the diagnostic accuracy of the predictive model, we also considered the clinical utility of a specific model. Therefore, the obtained optimal model fulfills the clinical needs and is suitable for clinical decision-making.

Conclusion

Machine learning techniques can estimate the risk of PCOS with high accuracy. These methods provide physicians with a convenient and noninvasive method for screening and diagnosing PCOS patients. TB-L, TB-a, TB-b, TC-L, and TC-a in tongue parameters and h_3, h_4/h_1 in pulse parameters are closely related to PCOS occurrence. This study

evaluated the performance of the MLP classifier, XGBoost classifier, and SVM methods in identifying PCOS occurrence. Taken together, the SVM model shows the best performance. Therefore, we speculated that the machine learning classification model based on tongue and pulse data has glorious performance and can adequately predict PCOS risk. In the future, we will collect additional data and explore other machine learning algorithms to obtain PCOS risk prediction models with even higher prediction accuracy.

Abbreviations

AUC	area under the curve
AUROC	area under the receiver operating characteristic curve
BMI	body mass index
DCA	decision curve analysis
IQR	interquartile range
LASSO	least absolute shrinkage and selection operator
MLP	multi-layer perceptron
PCOS	polycystic ovary syndrome
PSM	propensity score matching
ROC	receiver operating characteristic
SD	standard deviation
SVM	support vector machine
TB-a	tongue body a value
TB-b	tongue body b value
TB-L	tongue body L value
TC-a	tongue coating a value
TC-b	tongue coating b value
TC-L	tongue coating L value
TCM	traditional Chinese medicine
TDAS	tongue diagnosis analysis system
XGBoost	eXtreme gradient boosting

Contributorship: Investigation: Weiying Wang, Weiwei Zeng, Xinmin Chen. Methodology: Xiuqi Yin, Jiatuo Xu. Analysis: Liping Tu, Weiying Wang, Shunli He. Resources: Weiwei Zeng, Xiuqi Yin, Xiatuo Xu. Supervision: Liping Tu, Yulin Shi. Writing-original draft: Weiying Wang, Weiwei Zeng. Writing-review and editing: Weiying Wang, Weiwei Zeng, Yulin Shi, Xinmin Chen, Bingyi Yang. All authors have read and approved the manuscript.

ORCID iD: Weiying Wang https://orcid.org/0000-0002-2422-8195

References

1. Bozdag G, Mumusoglu S, Zengin D, et al. The prevalence and phenotypic features of polycystic ovary syndrome: a systematic review and meta-analysis. *Hum Reprod* 2016; 31: 2841–2855.
2. Manco M, Castagneto-Gissey L, Arrighi E, et al. Insulin dynamics in young women with polycystic ovary syndrome and normal glucose tolerance across categories of body mass index. *PLoS One* 2014; 9: e92995.
3. Dumesic DA, Oberfield SE, Stener-Victorin E, et al. Scientific statement on the diagnostic criteria, epidemiology, pathophysiology, and molecular genetics of polycystic ovary syndrome. *Endocr Rev* 2015; 36: 487–525.
4. Escobar-Morreale HF. Polycystic ovary syndrome: definition, aetiology, diagnosis and treatment. *Nat Rev Endocrinol* 2018; 14: 270–284.
5. Dapas M, Lin FTJ, Nadkarni GN, et al. Distinct subtypes of polycystic ovary syndrome with novel genetic associations: an unsupervised, phenotypic clustering analysis. *PLoS Med* 2020; 17: e1003132.
6. Wolf WM, Wattick RA, Kinkade ON, et al. The current description and future need for multidisciplinary PCOS clinics. *J Clin Med* 2018; 7: 95.
7. Jedel E, Waern M, Gustafson D, et al. Anxiety and depression symptoms in women with polycystic ovary syndrome compared with controls matched for body mass index. *Hum Reprod* 2010; 25: 450–456.
8. Wild RA, Carmina E, Diamanti-Kandarakis E, et al. Assessment of cardiovascular risk and prevention of

cardiovascular disease in women with the polycystic ovary syndrome: a consensus statement by the Androgen Excess and Polycystic Ovary Syndrome (AE-PCOS) Society. *J Clin Endocrinol Metab* 2010; 95: 2038–2049.

9. Wekker V, van Dammen L, Koning A, et al. Long-term cardiometabolic disease risk in women with PCOS: a systematic review and meta-analysis. *Hum Reprod Update* 2020; 26: 942–960.

10. Barry JA, Azizia MM and Hardiman PJ. Risk of endometrial, ovarian and breast cancer in women with polycystic ovary syndrome: a systematic review and meta-analysis. *Hum Reprod Update* 2014; 20: 748–758.

11. Cui J, Cui H, Yang M, et al. Tongue coating microbiome as a potential biomarker for gastritis including precancerous cascade. *Protein Cell* 2019; 10: 496–509.

12. Hsu PC, Wu HK, Huang YC, et al. The tongue features associated with type 2 diabetes mellitus. *Medicine (Baltimore)* 2019; 98: e15567.

13. Yu G and Wang J. Blood stasis syndrome of coronary heart disease: a perspective of modern medicine. *Chin J Integr Med* 2014; 20: 300–306.

14. Hsieh SF, Shen LL and Su SY. Tongue color changes within a menstrual cycle in eumenorrheic women. *J Tradit Complement Med* 2015; 6: 269–274.

15. He QY, Wang J and Zhang YL. Study on the Chinese medicine syndrome characteristics and coronary artery lesion specialty of female coronary heart disease. *Zhongguo Zhong Xi Yi Jie He Za Zhi* 2009; 29: 879–882.

16. Rubio DM, Schoenbaum EE, Lee LS, et al. Defining translational research: implications for training. *Acad Med* 2010; 85: 470–475.

17. Li J, Chen Q, Hu X, et al. Establishment of noninvasive diabetes risk prediction model based on tongue features and machine learning techniques. *Int J Med Inform* 2021; 149: 104429.

18. Shi Y, Yao X, Xu J, et al. A new approach of fatigue classification based on data of tongue and pulse with machine learning. *Front Physiol* 2022; 12: 708742.

19. Zhang H, Wang K, Zhang D, et al. Computer aided tongue diagnosis system. *Conf Proc IEEE Eng Med Biol Soc* 2005; 2005: 6754–6757.

20. Rotterdam ESHRE/ASRM-Sponsored PCOS consensus workshop group. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome (PCOS). *Hum Reprod* 2004; 19: 41–47.

21. Fei ZF. *Modern pulse diagnosis of traditional Chinese medicine*. Beijing: People's Health Publishing House, 2003.

22. Hashem S, ElHefnawi M, Habashy S, et al. Machine learning prediction models for diagnosing hepatocellular carcinoma with HCV-related chronic liver disease. *Comput Methods Programs Biomed* 2020; 196: 105551.

23. Luo Z, Gardiner JC and Bradley CJ. Applying propensity score methods in medical research: pitfalls and prospects. *Med Care Res Rev* 2010; 67: 528–554.

24. Tibshirani R. Regression shrinkage and selection via the Lasso. *J. R Stat. Soc. Ser. B Methodol* 1996; 58: 267–288. 1996.tb02080. x

25. DeLong ER, DeLong DM and Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988; 44: 837–845.

26. Vickers AJ and Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 2006; 26: 565–574.

27. Song K, Li Q, Gao W, et al. Allodriver: a method for the identification and analysis of cancer driver targets. *Nucleic Acids Res* 2019; 47: W315–W321.

28. Chang S, Lee U, Hong MJ, et al. Time-series growth prediction model based on U-net and machine learning in *Arabidopsis*. *Front Plant Sci* 2021; 12: 721512.

29. Gutiérrez S, Fernández-Novales J, Diago MP, et al. On-the-go hyperspectral imaging under field conditions and machine learning for the classification of grapevine varieties. *Front Plant Sci* 2018; 9: 1102.

30. Klöppel S, Stonnington CM, Chu C, et al. Automatic classification of MR scans in Alzheimer's disease. *Brain* 2008; 131: 681–689.

31. Zhou C, Liang D, Yang X, et al. Wheat ears counting in field conditions based on multi-feature optimization and TWSVM. *Front Plant Sci* 2018; 9: 1024.

32. Azziz R. Polycystic ovary syndrome. *Obstet Gynecol* 2018; 132: 321–336.

33. Luo ZY, Cui J, Hu XJ, et al. A study of machine-learning classifiers for hypertension based on radial pulse wave. *Biomed Res Int* 2018; 2018: 2964816.

34. Zhang J, Xu J, Hu X, et al. Diagnostic method of diabetes based on support vector machine and tongue images. *Biomed Res Int* 2017; 2017: 7961494.

35. Jiang T, Hu XJ, Yao XH, et al. Tongue image quality assessment based on a deep convolutional neural network. *BMC Med Inform Decis Mak* 2021; 21: 47.

36. Jiang T, Guo XJ, Tu LP, et al. Application of computer tongue image analysis technology in the diagnosis of NAFLD. *Comput Biol Med* 2021; 135: 104622.

37. Hu XJ, Zhang L, Xu JT, et al. Pulse wave cycle features analysis of different blood pressure grades in the elderly. *Evid Based Complement Alternat Med* 2018; 2018: 1976041.

38. Mokrzycki W and Maciej T. Perceptual difference in L* a* b* color space as the base for object colour identification. *Image Process Commun. Challenges* 2009; 1: 403–412.

39. Joseph BK and Savage NW. Tongue pathology. *Clin Dermatol* 2000; 18: 613–618.

40. Guichard H and Bonnarme P. Development and validation of a plate technique for screening of microorganisms that produce volatile sulfur compounds. *Anal Biochem* 2005; 338: 299–305.

41. Newman CB, Blaha MJ, Boord JB, et al. Lipid management in patients with endocrine disorders: an endocrine society clinical practice guideline. The published correction appears in *J Clin Endocrinol Metab* 2021; 106: e2465.

42. Ollila ME, Kaikkonen K, Järvelin MR, et al. Self-reported polycystic ovary syndrome is associated with hypertension: a Northern Finland birth cohort 1966 study. *J Clin Endocrinol Metab* 2019; 104: 1221–1231.