

Research article

Predicting TCM patterns in PCOS patients: An exploration of feature selection methods and multi-label machine learning models

Jiekee Lim^a, Jieyun Li^a, Xiao Feng^b, Lu Feng^a, Xinang Xiao^a, Mi Zhou^a, Hong Yang^c, Zhaoxia Xu^{a,d,*}

^a School of Traditional Chinese Medicine, Shanghai University of Traditional Chinese Medicine, Shanghai, PR China

^b The First Affiliated Hospital, Guangzhou University of Traditional Chinese Medicine, Guangzhou, PR China

^c Shanghai Municipal Hospital of Traditional Chinese Medicine, Shanghai University of Traditional Chinese Medicine, Shanghai, PR China

^d Shanghai Key Laboratory of Health Identification and Assessment, Shanghai, PR China

ARTICLE INFO

Keywords:

Diagnosis
Feature selection
Machine learning
Pattern differentiation
Polycystic ovary syndrome
Prediction model
Traditional Chinese medicine

ABSTRACT

Background: Traditional Chinese Medicine (TCM) offers individualized treatment for Polycystic Ovary Syndrome (PCOS) through pattern differentiation, but the subjectivity of TCM diagnoses can lead to inconsistent outcomes. Integrating machine learning (ML) offers an objective basis to support TCM diagnoses. This study aims to evaluate various feature selection techniques and multi-label ML algorithms to develop an effective predictive model for classifying TCM patterns in PCOS patients, thereby enhancing diagnostic standardization and treatment personalization.

Methods: The study utilized a dataset comprising 432 patients with PCOS, exhibiting one or more of five TCM patterns. Feature selection began with Variance Thresholding (VT), followed by a comparison of five advanced techniques: Statistical Analysis Test, Recursive Feature Elimination with Cross-Validation (RFECV), Least Absolute Shrinkage and Selection Operator Regression, BorutaShap, and ReliefF. To ascertain the most effective model for predicting PCOS TCM patterns, four ML algorithms—Support Vector Machine, Logistic Regression, Extreme Gradient Boosting (XGBoost), and Artificial Neural Networks—were evaluated against the identified feature set.

Results: VT reduced the feature count from 224 to 174. RFECV emerged as the most effective feature selection method, identifying 67 key features. XGBoost emerged as the top-performing model, demonstrating superior testing accuracy (0.7870), F1 score (0.9519), and Hamming loss (0.0481) with RFECV-optimized features.

Conclusions: The RFECV-XGBoost model proved effective for classifying TCM patterns in PCOS. It emphasizes the necessity of precise feature selection and the significant capabilities of ML in advancing TCM pattern diagnostics, marking a significant step toward enhancing precise and personalized healthcare in biomedical studies.

* Corresponding author. School of Traditional Chinese Medicine, Shanghai University of Traditional Chinese Medicine, Shanghai, 201203, PR China.

E-mail addresses: jiekeelim@outlook.com (J. Lim), zhaoxia7001@shutcm.edu.cn (Z. Xu).

<https://doi.org/10.1016/j.heliyon.2024.e35283>

Received 22 July 2024; Accepted 25 July 2024

Available online 26 July 2024

2405-8440/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Polycystic Ovary Syndrome (PCOS) is a prevalent endocrine disorder affecting adolescents and women of reproductive age, with a global prevalence estimated to be between 8 and 13 % [1]. The diagnosis and treatment of PCOS have long been challenging and remain central focal points in clinical research due to its varied hormonal imbalances and metabolic issues, which manifest in diverse and often overlapping symptoms with other conditions [2].

The practice of Traditional Chinese Medicine (TCM) offers unique advantages through its holistic approach. Central to TCM diagnosis is the concept of pattern differentiation (also known as syndrome differentiation), a principle that recognizes the unique constellation of symptoms and underlying imbalances in each individual [3,4]. This enables TCM practitioners to tailor treatment plans to the specific needs of each patient, thereby addressing not just the symptoms but the root cause of the condition [5]. TCM utilizes personalized treatments like herbal formulas, acupuncture, and dietary and lifestyle adjustments to improve health outcomes and maintain long-term well-being [6]. Thus, TCM offers a valuable and complementary perspective to contemporary PCOS healthcare by prioritizing individualized and proactive treatment strategies [7].

However, the inherent subjectivity of traditional pattern differentiation presents considerable challenges in consistently diagnosing and treating patients. Despite the keen observational skills of TCM practitioners, diagnostic conclusions can differ, leading to potential variations in treatment effectiveness and clinical outcomes [8]. Addressing this subjectivity is essential to ensure a higher standard of care for PCOS patients within the field of TCM.

Machine learning (ML) offers a powerful solution to this issue, providing a quantitative and empirical approach to TCM pattern identification [9]. Through ML algorithms, the complex task of TCM pattern differentiation can become more standardized and replicable, thereby enhancing consistency in patient care [10–12].

Recent studies have applied ML to identify PCOS using hybrid algorithms and commonly known diseases like obesity, diabetes, high blood pressure, and heart disease [13–15], Fractional Factorial Design has been employed to identify crucial parameters for PCOS diagnosis [16]. Although ML studies using TCM features are still limited, they show promising results. For instance, Wang et al. used the Least Absolute Shrinkage and Selection Operator (LASSO) regression for feature selection and a Support Vector Machine (SVM) model for PCOS identification based on tongue and pulse parameters [17]. Our team’s previous studies have shown that the utility of Voting and Long Short-Term Memory models was competitive in predicting PCOS based on specific pulse parameters [18], while SVM

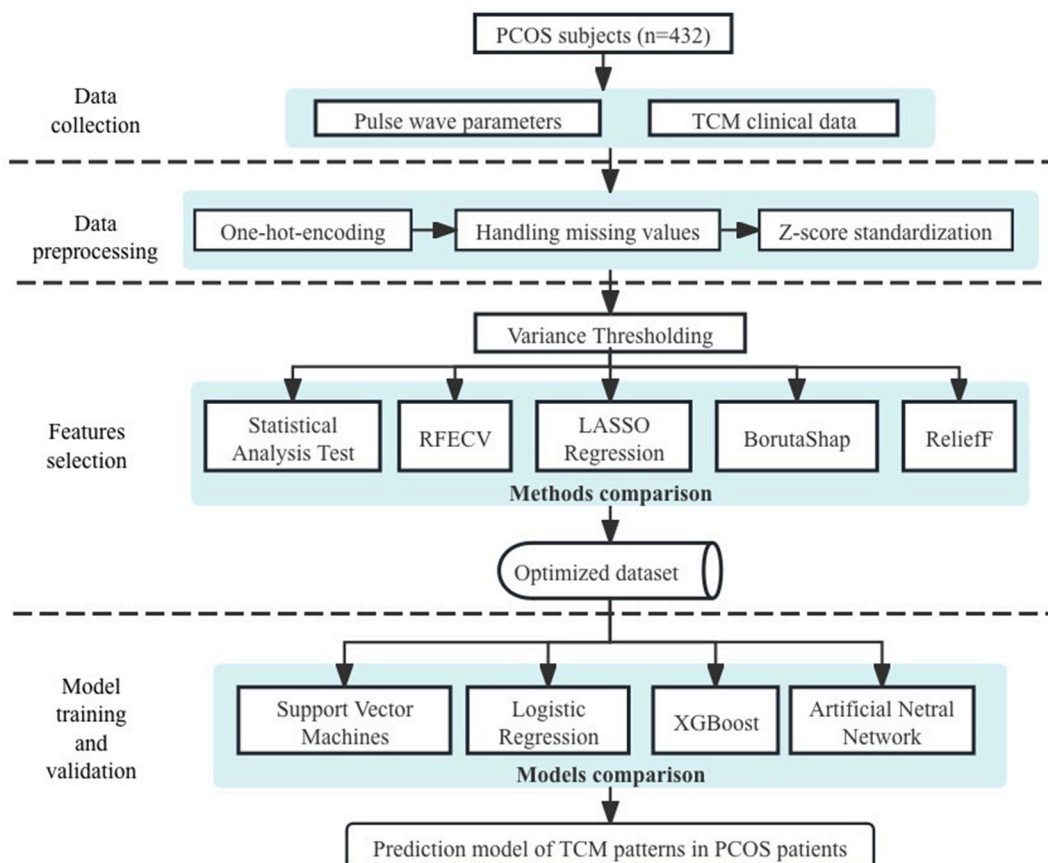


Fig. 1. Workflow of the study for predicting TCM patterns in PCOS patients.

aided with Recursive Feature Elimination with Cross-Validation (RFECV) obtained high performance based on pulse parameters and TCM clinical indices [19]. These studies underscore the contribution of ML to the field of TCM for PCOS, however, there is still a lack of comprehensive investigations into the optimal ML models for predicting TCM patterns in PCOS patients. Our research seeks to bridge this gap by focusing on the identification of five specific TCM patterns that PCOS may manifest as, which could be present individually or in combination, to enhance the personalization of PCOS treatment. These patterns will be discussed in a subsequent section.

It is also important to note that the objectification of TCM diagnostic methods provides a broad range of complex and diverse data [20]. The data encompasses a wide spectrum of patient information, from physical symptoms, and tongue and pulse readings to environmental factors and emotional states, reflecting the interconnectedness of various health determinants in TCM philosophy [12]. Thus, effective feature selection is crucial in simplifying this complex data and improving ML models' predictive capabilities [21]. There are studies demonstrating how feature selection methods can effectively capture the diagnostic essence of the TCM patterns in various diseases [22–24], but comparisons between different feature selection methods specific to TCM pattern classification are lacking.

This study evaluated various feature selection techniques and multi-label ML models to accurately classify TCM patterns specific to PCOS patients. The feature selection process commenced with an initial reduction through variance thresholding. Subsequently, five advanced feature selection techniques were applied in parallel to pinpoint the most predictive feature set, and then we assessed the performance of four ML algorithms. The overall workflow of the proposed study is illustrated in Fig. 1.

Our objective is to identify the most effective feature selection and multi-label ML model combinations for TCM pattern prediction in PCOS patients. By determining optimal methodologies for predictive accuracy, the study endeavors to support the advancement of personalized medicine for PCOS patients and to improve the standardization of TCM diagnostics, thereby integrating traditional medical knowledge with modern computational strategies.

2. Methods

2.1. Subjects

The study involved a cohort of 432 female subjects, all diagnosed with PCOS. These participants were recruited from the Shanghai Municipal Hospital of Traditional Chinese Medicine and Shuguang Hospital affiliated with the Shanghai University of Traditional Chinese Medicine (SHUTCM). The recruitment phase spanned from August 2018 to January 2024. All experimental protocols were reviewed and approved by the institutional review board of SHUTCM, under the approval number 2024-1-16-05. All research methods were conducted in strict adherence to relevant guidelines and regulations. Informed consent was obtained from all participants.

2.2. Inclusion, exclusion, and diagnostic criteria

The inclusion criteria for the study stipulate that participants must be female, aged between 18 and 40 years, diagnosed with PCOS, and have provided signed informed consent forms. The exclusion criteria include (1) failure to meet the above inclusion criteria; (2) use of hormonal drugs (including contraceptives, ovulation stimulators, and glucocorticoid drugs) within the last month; (3) inability to cooperate fully with the research plan due to pregnancy, history of infectious diseases, mental illness, or other conditions; (4) patients with incomplete TCM clinical data or unstable pulse wave diagrams, which could compromise the reliability of the data.

Diagnosis of PCOS is based on the Rotterdam European Society of Human Reproduction and Embryology (ESHRE) and the American Society for Reproductive Medicine (ASRM) consensus (revised 2003) [25], which requires the presence of at least two of the following three criteria for a PCOS diagnosis: (1) Oligo- and/or anovulation; (2) Clinical and/or biochemical signs of hyperandrogenism; (3) Ovarian polycystic changes, evidenced by ultrasound revealing ≥ 12 follicles with a diameter of 2–9 mm in one section, and/or ovarian volume ≥ 10 mL.

The TCM patterns diagnostic criteria for PCOS were established by referencing authoritative texts in the field, including “Gynecology of Traditional Chinese Medicine” [26], “Diagnostic of Traditional Chinese Medicine” [27], and “Guidelines for Diagnosis and Treatment of Common Diseases in Gynecology of Traditional Chinese Medicine” [28]. According to these guidelines, PCOS can manifest in one or more of the following five TCM patterns: Kidney Yin Deficiency, Kidney Yang Deficiency, Phlegm-Dampness, Blood Stasis, and Qi Stagnation. The specific diagnostic criteria for each pattern are based on TCM symptoms and are comprehensively detailed in Appendix A. The classification into five patterns facilitates targeted treatment strategies aimed at correcting the underlying imbalances.

2.3. Data collection

Data collection involved gathering clinical data and pulse wave parameters to provide a comprehensive overview of participants' health status from a TCM perspective.

The TCM clinical data were collected using a form designed for the four diagnostic methods of TCM related to menstrual diseases, developed earlier by our group (Copyright registration number: 2021-A-00037274). The form encompasses three main sections: recording the patient's demographics, gynecology specialized consultation (including menstrual cycle, menstrual period, menstrual flow, color of menstruation, and symptoms of menstrual discomfort), general consultation (covering responses to cold and heat, perspiration, general body symptoms, diet, bowel and urination habits, sleep, etc.), and observations of the patient's facial complexion and the diagnostic information on the tongue and pulse.

To collect pulse wave parameters from participants, we used the Z-BOX pulse meter, developed by the SHUTCM. This device captured essential cardiovascular data from both the left and right wrists, identified as "L" and "R" respectively. The parameters collected include amplitude ratios (As/Ad_L and As/Ad_R), height ratios (H2/H1_L, H4/H1_L, H5/H1_L, H2/H1_R, H4/H1_R, and H5/H1_R), time ratios (T1/T_L, T4/T_L, T1/T4_L, T5/T4_L, T1/T_R, T4/T_R, T1/T4_R, and T5/T4_R), and width ratios (W1/T_L and W1/T_R). These measurements provided a comprehensive overview of the cardiac ejection activity and the dynamics of the pulse wave along the vascular tree. The detailed collection procedure, the pulse map, the pulse wave parameters, and their implications were further detailed in [Appendix B](#), offering a basis for the cardiovascular examination of PCOS from a TCM perspective [29].

2.4. Data preprocessing

The dataset includes various data types: numerical (discrete counts like pregnancies and continuous measures like age, BMI, and pulse wave parameters), categorical (nominal groups like blood type and marital status), and binary (yes/no options for symptoms). To prepare categorical data for ML models, all categorical variables were transformed into dummy/binary variables using one-hot encoding.

Handling missing values in a dataset is essential to ensure its completeness and reliability for analysis. The steps were as followed: (1) Missing values in pulse wave parameters were filled using the K-Nearest Neighbors (KNN) imputation method, which calculates missing values based on the mean of the nearest neighbors, k (in this study, k = 5). This approach was particularly useful for preserving the overall distribution and relationships within the data [30]. (2) Median imputation was used for missing values of numerical variables in TCM clinical data to provide a robust measure for central tendency without being skewed by outliers. (3) Missing values in categorical data were filled with the mode, or the most common category, to maintain the distribution. (4) Missing values in binary data were assumed to be "no" and are thus filled with 0. This assumes a negative response where no information is provided.

In this study, the Z-Score was used to standardize numerical features [31,32]. The Z-Score is calculated using the formula:

$$Z = \frac{(X - \mu)}{s} \tag{1}$$

where X is the value of the data point, μ is the mean of the samples, and s is the standard deviation of the samples. This standardization process, implemented using the StandardScaler from Scikit-Learn, normalized the numerical features to have a mean of zero and a standard deviation of one. For each feature in the dataset, the Z-Score was calculated to standardize the values, enabling comparison across different scales and distributions. This scaling ensured all numerical variables contribute equally to the analysis, preventing features with larger scales from skewing results. This improved model performance by aiding faster convergence and enhancing predictive accuracy. Standardized features allowed easier interpretation of model coefficients and improve the feature selection and ML process.

After completing the preprocessing steps, the final dataset contained 224 features (the features were detailed in [Appendix C](#)) and 432 instances. The target variables were the five TCM patterns.

Table 1
Comparison of the different advanced feature selection methods.

Method	Description	Implementation Details
SAT	Uses statistical tests to find features related to TCM patterns.	With SciPy package, T-test or Kruskal-Wallis was applied for continuous data depending on the normality of data distribution; Chi-Square for binary data. P-value <0.05 is considered significant.
RFECV	Recursive feature elimination method integrated with cross-validation [35]. A process that iteratively removes features to find the optimal subset of features that maximizes model performance.	With Scikit-learn package, XGBoost estimator was used with a 0.1 learning rate, max depth of 1, and 200 estimators. RFECV with 5-fold cross-validation and "f1_micro" scoring metric.
LASSO	A regularization method that performs automatic feature selection during model fitting. It penalizes the absolute size of the coefficients, which effectively sets those of lesser importance to zero [36,37].	With Scikit-learn package, the LassoCV was employed to determine the optimal regularization strength with 'alphas = None' and 'cv = 5', allowing it to automatically choose a range of alpha values via 5-fold cross-validation. Additionally, 'n_alphas = 100', 'eps = 0.001', and 'selection = cyclic' were set. Features with non-zero coefficients were retained, iteratively refining a set of key features predictive of the TCM patterns.
BorutaShap	A feature selection method that combines the strengths of the Boruta algorithm with SHapley Additive exPlanations (SHAP) values [38]. It identifies relevant features by creating shadow features for comparison and applies a tree-based model and SHAP values for evaluation [39]. Features are accepted, rejected, or tentative based on statistical testing against shadow features.	With BorutaShap package, XGBoost was used as feature selector with 0.1 learning rate, max depth of 1, and 200 estimators for 100 trials. Accepted features were identified for each target and a set of unique features were retained.
Relieff	An advanced feature selection algorithm that enhances the original Relief algorithm [40,41]. It evaluates features based on how well they differentiate between instance classes, is designed for handling multi-class problems, and can effectively deal with noisy and incomplete data [42].	With Scikit-rebate package, specified 'n_neighbors = 10', importance scores of features were assigned. Thresholds (0.015, 0.02, 0.025, 0.03, 0.035) of importance scores were tested to refine the features set. Each feature set trained a logistic regression model; the best-performing model dictated the optimal feature set. The process was repeated for each target.

3. Feature selection

The process of feature selection involved two steps: an initial reduction phase followed by a detailed comparison of five distinct feature selection methods. These steps were integral to finding out the most relevant and informative features. All computations were performed using Python.

3.1. Variance thresholding

We employed Variance Thresholding (VT) as our initial feature selection technique [33,34]. It is designed to remove features with low variance, under the assumption that variables with little variation do not contribute significantly to the predictive power of the model. The VT operation was independent of the data's distribution and the correlation between features and target variables, making it a universally applicable method for initial feature reduction across different types of data, including both numerical and binary variables. Utilizing the 'VarianceThreshold' function from the Scikit-Learn library, we set a threshold of 0.01, features with variance below this threshold were removed. This simplification ensures more targeted and effective subsequent feature selection.

3.2. Advanced feature selection methods

Following the initial reduction via VT, we explored five advanced feature selection methods to refine the feature set further, including the filter-based Statistical Analysis Test (SAT), the wrapper-based RFECV, embedded-based LASSO, the ensemble-based BorutaShap, and the hybrid-based ReliefF. Table 1 summarizes the advanced feature selection methods.

4. Machine learning models and evaluation metrics

Upon determining the best feature selection method and the optimal feature set, the multi-label ML model training process began with an iterative train-test split, at a 75 %–25 % ratio, where the larger portion was used for training the model and the smaller for testing its performance. This division ensured that the model could be evaluated on unseen data, providing insights into its generalization capabilities.

To optimize model performance, GridSearchCV was employed to systematically explore a range of hyperparameters. The process incorporated 10-fold cross-validation to ensure robustness, whereby the training set is split into ten parts. The model trained on nine parts and validated on one, rotating through all parts for validation. This approach thoroughly evaluated model performance and averaged results to reduce variance.

Four supervised learning models were used. Models from the Scikit-Learn library were adeptly instantiated within a OneVsRestClassifier framework (except for ANN), a strategy crucial for multi-label scenarios. This approach created a separate classifier for each class, treating the problem as multiple binary classification tasks, thereby enabling detailed analysis and handling of each class independently.

The models used were (1) SVM, which was particularly effective for classification tasks by finding the hyperplane that best separated different classes in the feature space [43]. The model's complexity and margin were controlled by hyperparameters like C (penalty parameter), gamma, and kernel type, which were optimized using grid search and cross-validation. (2) LR, though simpler, excels in binary classification problems by estimating probabilities, there was literature showing LR was as good as other ML models in predicting the risk of major chronic diseases [44]. The hyperparameter tuning concentrated on optimizing the regularization strength 'C' and selecting the appropriate solver from 'liblinear' and 'saga' which are compatible with the 'l1' penalty. (3) XGBoost, short for eXtreme Gradient Boosting, was an implementation of gradient-boosted decision trees [45], which offered excellent performance and speed and had been widely adopted. To optimize the XGBoost model, a comprehensive grid search was conducted over a predefined range of hyperparameters, including the number of estimators, learning rate, and maximum depth. (4) ANN, particularly the deep learning model, leveraged the Keras library to construct a sequential architecture tailored for classification. The model's architecture and parameters like layers, neurons per layer, activation functions, learning rate, epochs, and batch size were optimized. The model was adept at capturing complex patterns in large datasets but required careful tuning of the architecture and parameters.

For multi-label classification tasks, where each instance may belong to multiple classes simultaneously, choosing the right evaluation metric is more complex due to the need to accurately capture model performance across several labels. In this context, the key metrics used are:

1. **Hamming Loss:** Calculates the proportion of incorrectly predicted labels to the total number of labels, offering an overall error rate. This metric is valuable for its direct measurement of performance across an entire label set, highlighting the model's predictive accuracy on a granular level. The range of Hamming Loss is between 0 and 1, where 0 indicates perfect prediction with no incorrect labels, and 1 means all predictions are incorrect.
2. **Subset Accuracy (Exact Match):** This strict metric demands a complete match between the predicted and actual label sets for an instance to be considered correct. Although this method is thorough, it offers a precise measure of model accuracy but may not fully account for partially correct predictions. For simplicity, we refer to it as "accuracy."
3. **F1 Score:** As the harmonic mean of precision and recall, the F1 score merges these metrics into a singular measure to balance their contribution, especially useful for comparing models with varying trade-offs. Precision is defined as the ratio of true positives to the sum of true and false positives, precision quantifies the model's ability to accurately predict positive labels, emphasizing the

importance of relevancy in the predictions. Recall measures the proportion of actual positives that have been correctly identified, calculated as the ratio of true positives to the sum of true positives and false negatives. Recall is crucial for scenarios where missing a positive instance carries significant consequences.

- Area Under the Receiver Operating Characteristic (ROC) Curve (AUC):** AUC evaluates a model's capacity to differentiate between classes across all thresholds. By plotting the true positive rate against the false positive rate (ROC), it assesses separability and predictive reliability. AUC value, ranging from 0 to 1, with higher values representing better classification performance.

5. Results

All experiments for this research were conducted on a MacBook Air equipped with an Apple M2 chip and 16 GB of memory, running macOS Sonoma version 14.1.2. Python version 3.9.16 was used for all computational tasks.

The PCOS dataset includes 432 subjects categorized into 5 TCM patterns, showcasing considerable variation in their distribution. Specifically, the Kidney Yin Deficiency pattern was observed in 230 subjects, the Kidney Yang Deficiency pattern in 259 subjects, the Phlegm-Dampness pattern in 286 subjects, the Blood Stasis pattern in 156 subjects, and the Qi Stagnation pattern in 154 subjects. This distribution indicated a diverse representation of TCM patterns within the dataset.

5.1. Feature selection results

The number of features reduced by each method is indicated in Table 2. The details of the features selected by each method were stated in Appendix D. The initial feature reduction technique, VT, reduced the number of features from 224 to 174, indicating that 50 features with low variance were excluded. This established a baseline for further analysis using the refined dataset of 174 features, upon which five advanced feature selection methods were simultaneously applied and compared.

The SAT was conducted to assess the significance of features in the dataset. The Kruskal-Wallis test was used for the 21 continuous features, identifying 6 as statistically significant ($P < 0.05$). For the 203 binary and categorical features, chi-squared tests found 89 significantly associated with the target variables ($P < 0.05$). Ultimately, 95 features showing statistical significance were retained for predicting the target variables.

The RFECV procedure determined that the optimal number of features for the dataset was 67. Fig. 2 shows the relationship between the number of features selected and the mean test score of RFECV. The plot demonstrates how the mean test score evolves as more

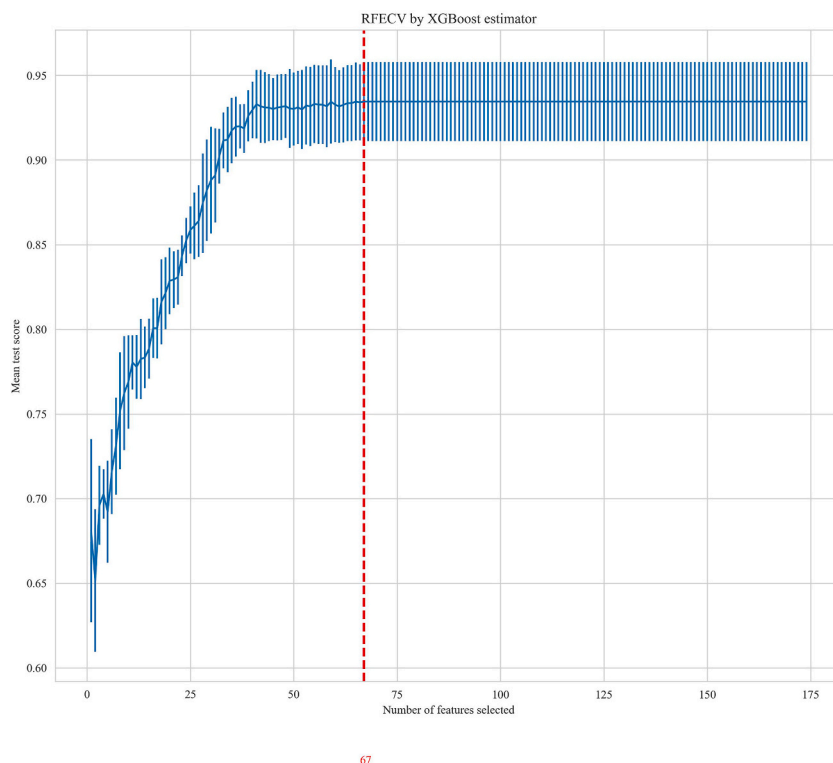


Fig. 2. Relationship between the number of selected features and mean test scores of RFECV. The dashed red line indicates the optimal number of features, determined to be 67. As the number of features increases, the mean test score stabilizes around 0.9346, with a standard deviation of 0.0234, suggesting the point of optimal feature selection. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

features are included in the model.

The LASSO feature selection method was employed across the five patterns, resulting in the identification of 106 unique features. Each pattern exhibited varying numbers of selected features: Kidney Yin Deficiency had 47 features, Kidney Yang Deficiency had 15 features, Phlegm-Dampness had 16 features, Blood Stasis had 12 features, and Qi Stagnation had 11 features, highlighting the diverse set of indicators associated with different TCM patterns.

The BorutaShap feature selection method resulted in the retention of 52 unique features for the dataset. Fig. 3 displays the accepted features (those deemed significantly predictive of the target variables) determined by BorutaShap, providing insights into the selected features for each target.

Results of the ReliefF method showed the Kidney Yin Deficiency pattern with 13 features at 90.74 % accuracy (threshold: 0.02), the Kidney Yang Deficiency with 16 features at 87.96 % (threshold: 0.02), and the Phlegm-Dampness pattern with 11 features at 81.48 % (threshold: 0.035). The Qi Stagnation pattern reaches 99.07 % accuracy with 18 features (threshold: 0.03), and the Blood Stasis pattern achieves 96.30 % with 36 features (threshold: 0.015). Overall, the combined ReliefF selection across all patterns resulted in 62 unique features being selected.

5.2. Comparison of feature selection methods

In comparing the five feature selection methods, VT as the initial reduction was the baseline of comparison. Table 2 summarizes the Hamming loss achieved by each feature selection method across SVM, LR, XGBoost, and ANN classifiers. The average Hamming loss for each method is also provided, lower Hamming loss indicates a lower rate of incorrect predictions.

The SAT method showed a marked improvement in performance over the VT baseline, notably reducing the Hamming loss in SVM

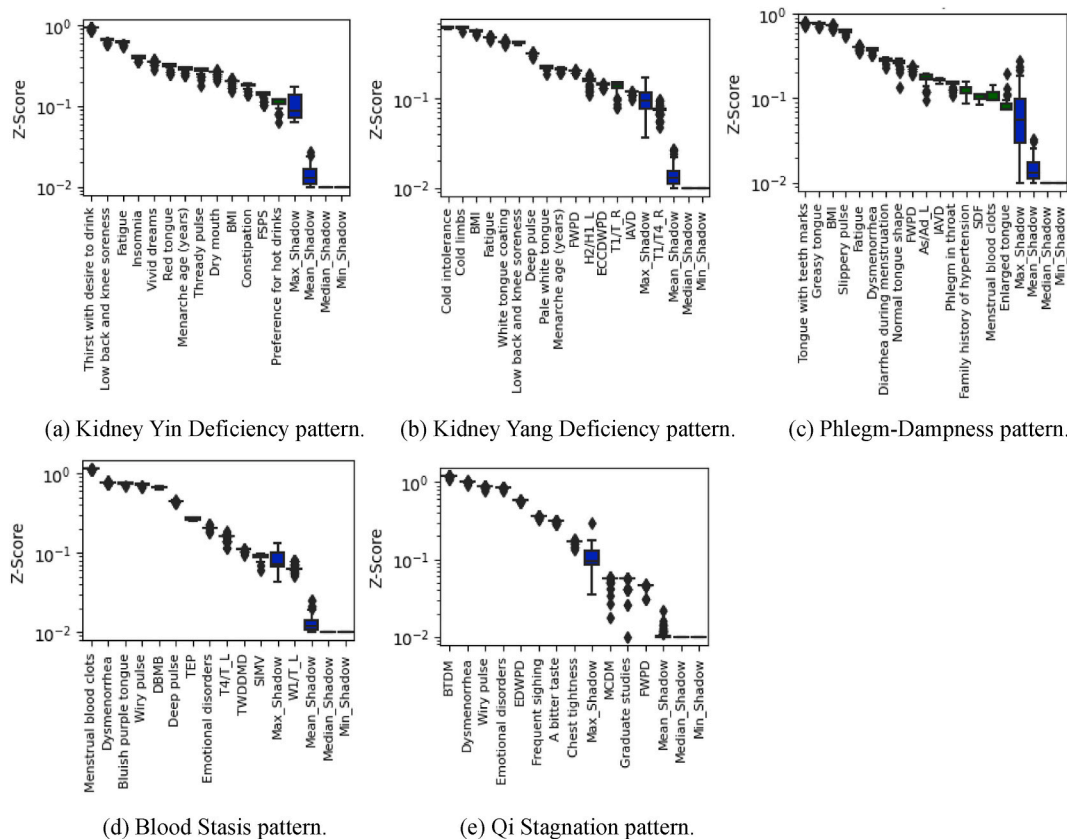


Fig. 3. Feature selection results of BorutaShap. This figure presents a comprehensive visualization of the importance of accepted features across five distinct patterns in TCM diagnosis: Kidney Yin Deficiency (13 features), Kidney Yang Deficiency (15 features), Phlegm-Dampness (16 features), Blood Stasis (12 features), and Qi Stagnation (11 features). The accepted features are represented with green boxplots.

Abbreviations: BTDM, Breast tenderness during menstruation; DBMB, Dark brown menstrual blood; ECCDWPDP, Exposure to cold or cold diet worsens perimenstrual disorders; EDWPD, Emotional distress worsens perimenstrual disorders; FSPS, Feverish sensations in palms and soles; FWPD, Fatigue worsens perimenstrual disorders; IAVD, Increased amount of vaginal discharge; MCDM, Mood changes during menstruation; SDF, Stomach distension and fullness; SIMV, Significantly increased menstrual volume; TEP, Tongue with ecchymosis/petechiae; TWDDMD, Thirst without desire to drink or minimal drinking. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 2
Comparison of Hamming loss across different feature selection methods for various classifiers.

	VT	SAT	RFECV	LASSO	BorutaShap	Relieff
No. of features	174	95	67	106	48	62
SVM	0.1722	0.1352	0.0944	0.1389	0.0852	0.1130
LR	0.0759	0.0759	0.0704	0.0759	0.0778	0.0870
XGBoost	0.0556	0.0537	0.0481	0.0630	0.0537	0.0759
ANN	0.2222	0.1667	0.1019	0.1519	0.1037	0.1426
Average	0.1315	0.1079	0.0787	0.1074	0.0801	0.1046

(from 0.1722 to 0.1352) and ANN (from 0.2222 to 0.1667). However, it does not outperform all other methods.

The RFECV emerged as the most effective method, offering the most substantial improvement in Hamming loss across LR, XGboost, and ANN when compared to VT. This method achieved the lowest average Hamming loss among the methods evaluated (0.0787), highlighting its robustness in selecting highly relevant features for a variety of models.

LASSO provided a moderate improvement in average Hamming loss relative to VT, with its performance being somewhat mixed across different classifiers. LASSO demonstrated its ability to effectively reduce the feature count while maintaining (LR) or slightly improving (SVM and ANN) model performance, indicating a balanced approach to feature reduction. Despite the general effectiveness, LASSO did not outperform the VT when applied to the XGBoost classifier.

BorutaShap emerged as another top performer, particularly in lowering the Hamming loss for SVM and ANN, and significantly outperforming VT. Remarkably, it managed to select the fewest number of features (48) while achieving the second lowest average Hamming loss among the methods compared. This efficiency in identifying the most impactful features across different classifiers positions BorutaShap as an excellent choice for feature selection, especially in scenarios requiring a stringent feature reduction without compromising model performance.

Relieff improved upon the VT baseline only in SVM and ANN, but its benefits were not consistently observed across all models. Despite this, it succeeded in reducing the average Hamming loss compared to VT, underscoring its potential utility in specific contexts or with certain types of classifiers.

Fig. 4 shows a radar chart visualizing the performance (measured by Hamming loss) of different ML classifiers across several feature selection methods. In summary, both RFECV and BorutaShap were highly effective feature selection methods. However, in our study aimed at identifying TCM patterns in PCOS patients, RFECV was chosen as the final feature selection method due to its better

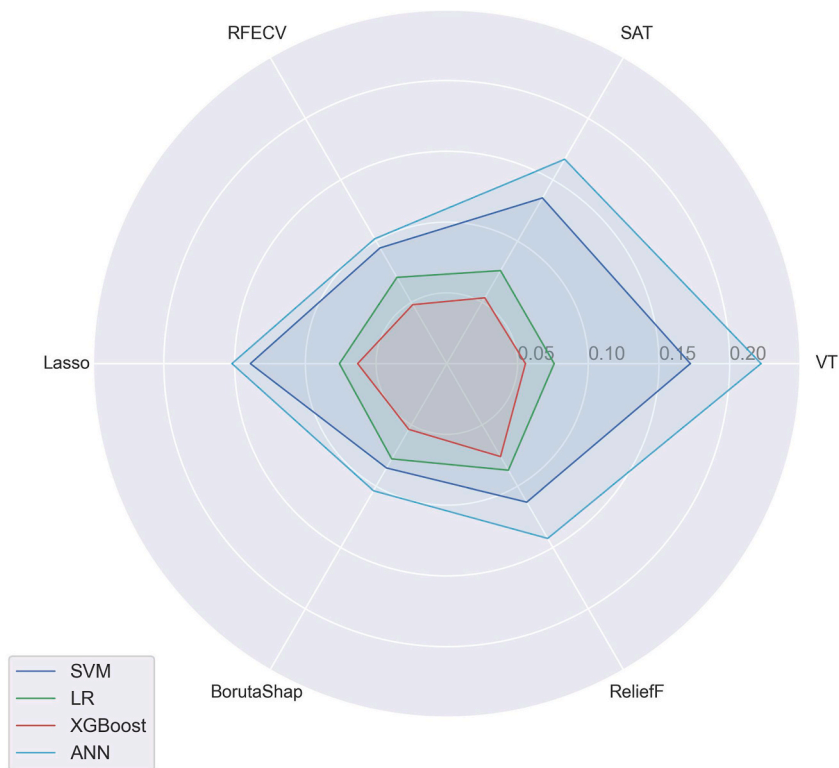


Fig. 4. Performance comparison of ML classifiers across feature selection methods. The plot visualizes the comparative analysis of how each classifier performs with each feature selection method, with the Hamming loss values plotted on the radar chart.

performance in minimizing Hamming loss and its enhanced versatility and robustness across various classifiers. Although RFECV selected more features than BorutaShap, its ability to ensure optimal and generalizable feature selection, which is crucial for the complex nature of TCM pattern identification, made it the preferred choice.

5.3. RFECV-optimized model performance

Utilizing RFECV across different ML models yielded varied impacts on their performance metrics, including accuracy, F1 scores, and Hamming loss. Table 3 presents the comprehensive comparison of performance metrics across four ML models: SVM, LR, XGBoost, and ANN. Fig. 5 visualizes the comparison of Hamming loss across various classification models. Fig. 6 illustrates the efficacy of RFECV-optimized models, contrasting training and testing outcomes through boxplots and line plots for F1 scores and accuracy. The details of the models' classifiers and their hyperparameters are stated in Appendix E.

SVM showed moderate training accuracy (0.5777 ± 0.0866) and a high F1 score (0.8922 ± 0.0264), with an improvement in testing accuracy (0.6296) and a relatively low Hamming loss (0.0944), indicating good generalization. LR outperformed SVM in both training and testing phases, achieving higher accuracy (0.7530 ± 0.0660 , 0.7130) and F1 scores (0.9453 ± 0.0151 , 0.9294) with a notably lower Hamming loss (0.0704), showcasing its efficiency and predictive performance. XGBoost emerged as the standout model, delivering the highest accuracy (0.7811 ± 0.1088 , 0.7870) and F1 scores (0.9525 ± 0.0220 , 0.9519) in both training and testing, alongside the lowest Hamming loss (0.0481), highlighting its robustness and generalizability. Conversely, ANN had the lowest performance, with training and testing accuracies of 0.5188 ± 0.1029 and 0.5833, respectively, and F1 scores of 0.8653 ± 0.0310 and 0.8980, it also showed the highest Hamming loss at 0.1019. The model exhibited challenges in learning from the training dataset.

Overall, the RFECV feature selection method proved to enhance model performance significantly, with XGBoost as the most effective model for TCM pattern classification requiring high accuracy and predictive efficiency. The AUC plot in Fig. 7 visually represents the performance of the XGBoost model trained to classify TCM patterns. The plot illustrates the model's strong predictive ability to distinguish between the patterns.

6. Discussion

Our study initiated a thorough investigation to assess the impact of various feature selection techniques on the predictive performance of multi-label ML models, specifically in identifying TCM patterns in PCOS patients by utilizing a dataset that richly represents five distinct TCM patterns and diverse data types.

Existing studies had used tongue and pulse parameters for PCOS identification [17–19] and TCM symptoms for pattern classification [24,46], such as Xie et al., who selected 42 features out of 200 for rheumatoid arthritis patients [24]. Building on this, this study used pulse parameters and TCM clinical data (mostly TCM symptoms), selecting 67 features through a comprehensive comparison of multiple feature selection techniques.

In the initial phase, VT was employed to streamline the feature space, resulting in a significant reduction of features. This step eliminated features with minimal variance, setting the stage for more refined and efficient feature selection processes to follow.

Among the advanced feature selection methods evaluated, RFECV emerged as the premier method, surpassing SAT, LASSO, BorutaShap, and ReliefF. It achieved an optimal balance between feature reduction and predictive performance enhancement, evidenced by the lowest average Hamming loss, particularly with the XGBoost model. The success of RFECV lies in its iterative refinement and validation of the feature set, retaining only the most impactful features. This process ensures the model's robustness and adaptability across various classifiers, solidifying RFECV's position as the most effective strategy for our study's objective.

BorutaShap demonstrated outstanding capability by significantly lowering the number of features while improving model performance (LR being the exception), showcasing its efficiency in feature selection. For instance, Silva et al. successfully presented the BorutaShap-Random Forest algorithm to enhance the PCOS diagnosis by identifying the most relevant clinical and laboratory variables [47]. Although SAT, LASSO, and ReliefF showed improvements over the baseline, their effectiveness varied depending on the classifier used. This variation emphasizes the importance of matching the feature selection technique with the specific characteristics of the classifier.

In ML model evaluations using RFECV refined features, XGBoost emerged as a top performer by demonstrating superior accuracy, F1 scores, and the lowest Hamming loss, highlighting its robustness and effectiveness in classifying TCM patterns. This finding aligns with existing literature that highlights XGBoost's effectiveness in handling complex classification problems using TCM features [48]. This makes it a valuable asset for healthcare practitioners looking to incorporate data-driven insights into diagnostic practices. Its achievement in obtaining the lowest Hamming loss demonstrates its precision in reducing incorrect predictions across all classes, a

Table 3
Performance metrics of ML models with RFECV feature selection.

Model	Training results (Mean \pm SD)		Testing results		
	Accuracy	F1 score	Accuracy	F1 score	Hamming loss
SVM	0.5777 ± 0.0866	0.8922 ± 0.0264	0.6296	0.9057	0.0944
LR	0.7530 ± 0.0660	0.9453 ± 0.0151	0.7130	0.9294	0.0704
XGBoost	0.7811 ± 0.1088	0.9525 ± 0.0220	0.7870	0.9519	0.0481
ANN	0.5188 ± 0.1029	0.8653 ± 0.0310	0.5833	0.8980	0.1019

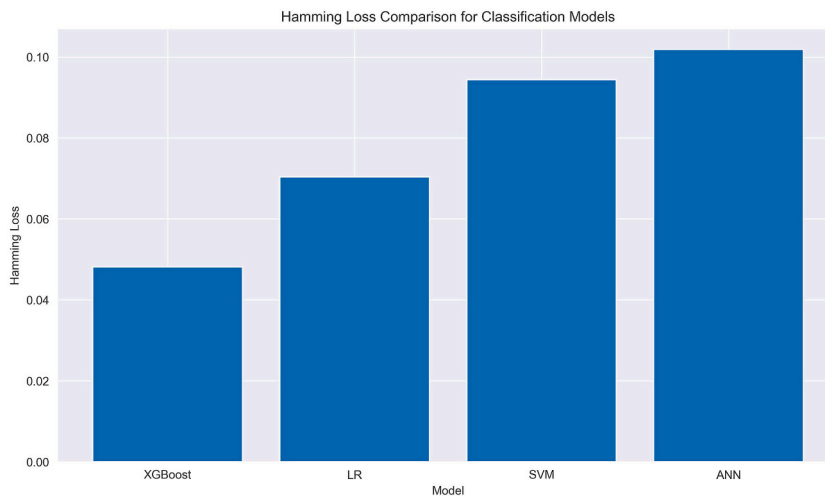


Fig. 5. Hamming loss comparison between models optimized with RFECV. Lower Hamming loss indicates a lower rate of incorrect predictions, and XGBoost emerged as the top performer.

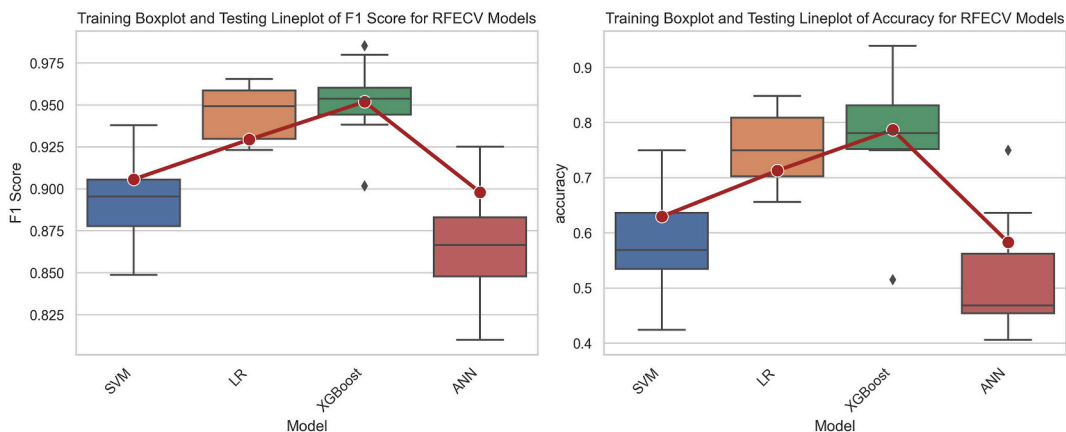


Fig. 6. Model performance (F1 score and accuracy) with RFECV optimization. The left panel focuses on F1 scores, comparing training distributions to testing results, while the right panel does the same for accuracy. This visualization highlights each model’s training variability and testing performance, aiding in the evaluation of their generalization to unseen data.

vital aspect in medical diagnostics where accuracy is paramount.

Following closely, LR demonstrated superior performance over the SVM, marking it as the second-best algorithm. This highlights LR’s robust predictive performance and efficiency [49], making it a strong candidate for ML model selection, particularly when interpretability and computational efficiency are crucial. Meanwhile, SVM displayed moderate generalization abilities and performance compared to XGBoost and LR. Conversely, the ANN exhibited the lowest performance across the board. These outcomes may reflect ANN’s sensitivity to the dataset’s complexity, suggesting a need for optimization or architectural refinement to discern TCM patterns in PCOS patients more effectively. While ANN possessed significant potential for TCM pattern recognition for dysmenorrhea [50] and lung cancer [51] diagnosis, customizing its structure to better accommodate the complex nature of TCM diagnostic criteria for PCOS could unlock its full capabilities in this domain.

In summary, the RFECV-XGBoost algorithm proved to be highly effective for identifying TCM patterns in patients with PCOS. This outcome underscores the substantial promise of applying ML techniques within TCM diagnostics. Our study represents a significant advancement in applying ML to TCM diagnostics compared to existing research. While previous research often focused on single feature selection algorithm [24,46] or conventional ML methods [52], our investigation evaluates a comprehensive array of feature selection techniques to assess their impact on the predictive performance of multi-label ML models, particularly in the context of TCM pattern identification for PCOS patients. This broader spectrum approach allowed us to identify the RFECV-XGBoost combination as notably superior in enhancing diagnostic accuracy. By conducting such detailed analyses, our study not only advances the methodological framework for TCM diagnostics but also sets a precedent for future research to explore more sophisticated ML strategies.

The RFECV-XGBoost algorithm can significantly enhance clinical practice by promoting intelligent pattern differentiation apps in

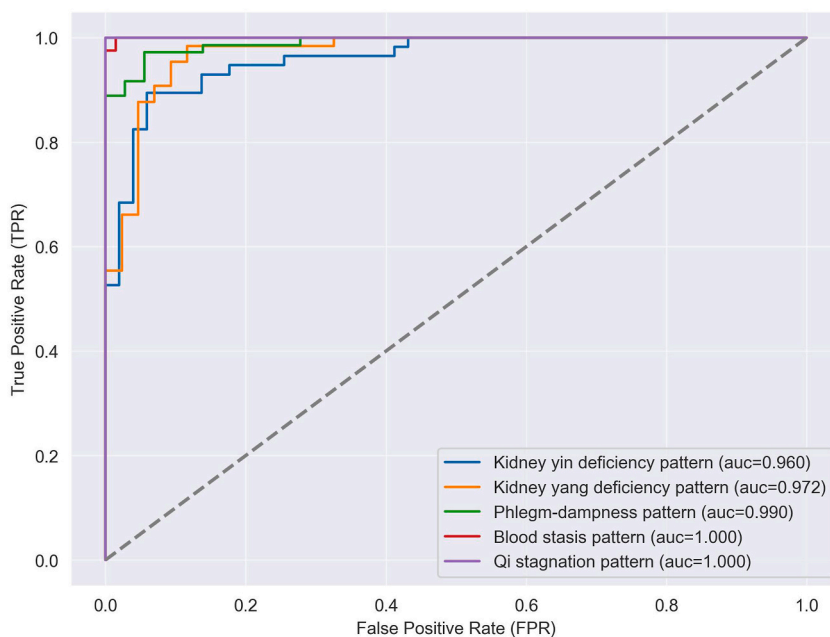


Fig. 7. ROC curves for TCM pattern classification using RFECV-optimized XGBoost.

TCM, advancing personalized medicine. It facilitates the accurate identification of TCM patterns, crucial for tailoring effective treatment plans, and contributes to the standardization of diagnostic practices, ensuring more consistent and reliable patient care. Furthermore, this approach integrates traditional knowledge with modern data analytics, making TCM diagnostics and treatments more evidence-based and data-driven. This methodological advancement marks a significant step forward in merging modern computational techniques with traditional medical practices, paving the way for more precise, reliable, and personalized TCM diagnostics.

While our study provides valuable insights into the application of ML for TCM pattern classification, it has its limitations. Firstly, the study's reliance on a dataset comprising only 432 subjects representing five distinct TCM patterns of PCOS may not encompass the full spectrum of variability seen in clinical practice, potentially limiting the generalizability of our findings. The relatively small dataset size could limit the model's learning capacity and may not adequately represent the broader patient population. Future studies should aim to include a larger and more diverse sample to validate and extend our findings. Additionally, the performance of ANN highlighted the challenges of model complexity and the need for further optimization, which may not have been fully addressed within the scope of this study. Furthermore, the comparative analysis of feature selection methods, though extensive, was conducted within a specific context, and results may vary with different datasets or medical conditions. Acknowledging these limitations is crucial for interpreting our findings and guiding future research efforts.

7. Conclusions

This study demonstrated that feature selection is crucial for simplifying TCM data and boosting predictive performance in diagnosing TCM patterns in PCOS. Among the various methods evaluated, the RFECV feature selection method excelled in identifying key predictive features, and the XGBoost model showed superior performance. The combined RFECV-XGBoost formed the most effective algorithm for predicting TCM patterns in PCOS patients, achieving the highest accuracy, F1 scores, and the lowest Hamming loss, underscoring its robustness and precision. Our findings highlight the potential of integrating advanced data analytics into TCM, paving the way for more effective and personalized healthcare solutions. This study can also promote the research and development of objectifying and standardizing TCM diagnostics, providing a more reliable and precise approach to diagnosing and treating PCOS. Future research should continue to refine these computational methods and explore their applicability across a broader spectrum of medical conditions, further enhancing diagnostic precision and personalized treatment strategies in TCM.

Funding

This work was supported by the Shanghai Key Laboratory of Health Identification and Assessment [No.21DZ2271000]; the Shanghai Municipal Science and Technology Commission [No.21Y11923400]; and the Shanghai Municipal Health Commission [No.202240007].

Data availability

The Ethics Committee of Shanghai University of Traditional Chinese Medicine limited the data used to support the results of this study to protect the privacy of patients. For researchers who meet the criteria for obtaining confidential data, the data of this study are available from the corresponding author upon request.

CRedit authorship contribution statement

Jiekee Lim: Formal analysis, Methodology, Software, Writing – original draft. **Jieyun Li:** Visualization, Writing – review & editing. **Xiao Feng:** Investigation. **Lu Feng:** Investigation. **Xinang Xiao:** Validation. **Mi Zhou:** Data curation. **Hong Yang:** Funding acquisition, Resources. **Zhaoxia Xu:** Conceptualization, Funding acquisition, Project administration, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in the paper.

Acknowledgments

Not applicable.

Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.heliyon.2024.e35283>.

References

- [1] Polycystic ovary syndrome, (n.d.). <https://www.who.int/news-room/fact-sheets/detail/polycystic-ovary-syndrome> (accessed March 4, 2024).
- [2] F.R. Tehrani, M. Amiri, Polycystic ovary syndrome in adolescents: challenges in diagnosis and treatment, *Int. J. Endocrinol. Metabol.* 17 (2019) 91554, <https://doi.org/10.5812/IJEM.91554>.
- [3] A.S. Ferreira, A.J. Lopes, Chinese medicine pattern differentiation and its implications for clinical practice, *Chin. J. Integr. Med.* 17 (2011) 818–823, <https://doi.org/10.1007/S11655-011-0892-Y/METRICS>.
- [4] A.P. Lu, K.J. Chen, Integrative medicine in clinical practice: from pattern differentiation in traditional Chinese medicine to disease treatment, *Chin. J. Integr. Med.* 15 (2009) 152, <https://doi.org/10.1007/S11655-009-0152-6/METRICS>.
- [5] J. gen Li, H. Xu, Incorporating pattern identification of Chinese medicine into precision medicine: an integrative model for individualized medicine, *Chin. J. Integr. Med.* 21 (2015) 807–810, <https://doi.org/10.1007/S11655-015-2058-9/METRICS>.
- [6] M. Jiang, C. Lu, C. Zhang, J. Yang, Y. Tan, A. Lu, K. Chan, Syndrome differentiation in modern research of traditional Chinese medicine, *J. Ethnopharmacol.* 140 (2012) 634–642, <https://doi.org/10.1016/J.JEP.2012.01.033>.
- [7] L.W. Fu, Z. Gao, N. Zhang, N. Yang, H.Y. Long, L.Y. Kong, X.Y. Li, Traditional Chinese medicine formulae: a complementary method for the treatment of polycystic ovary syndrome, *J. Ethnopharmacol.* 323 (2024) 117698, <https://doi.org/10.1016/J.JEP.2023.117698>.
- [8] A. Sá Ferreira, Misdiagnosis and undiagnosis due to pattern similarity in Chinese medicine: a stochastic simulation study using pattern differentiation algorithm, *Chin. Med.* 6 (2011) 1–13, <https://doi.org/10.1186/1749-8546-6-1/TABLES/3>.
- [9] D. Tian, W. Chen, D. Xu, L. Xu, G. Xu, Y. Guo, Y. Yao, A review of traditional Chinese medicine diagnosis using machine learning: inspection, auscultation-olfaction, inquiry, and palpation, *Comput. Biol. Med.* 170 (2024) 108074, <https://doi.org/10.1016/J.COMPBIOMED.2024.108074>.
- [10] Z. Chen, D. Zhang, C. Liu, H. Wang, X. Jin, F. Yang, J. Zhang, Traditional Chinese medicine diagnostic prediction model for holistic syndrome differentiation based on deep learning, *Integr Med Res* 13 (2024) 101019, <https://doi.org/10.1016/J.IMR.2023.101019>.
- [11] H. Pang, S. Wei, Y. Zhao, L. He, J. Wang, B. Liu, Y. Zhao, Effective attention-based network for syndrome differentiation of AIDS, *BMC Med Inform Decis Mak* 20 (2020) 1–10, <https://doi.org/10.1186/S12911-020-01249-0/FIGURES/4>.
- [12] W. Te Huang, H.H. Hung, Y.W. Kao, S.C. Ou, Y.C. Lin, W.Z. Cheng, Z.R. Yen, J. Li, M. Chen, B.C. Shia, S.T. Huang, Application of neural network and cluster analyses to differentiate TCM patterns in patients with Breast cancer, *Front. Pharmacol.* 11 (2020) 529736, <https://doi.org/10.3389/FPHAR.2020.00670/BIBTEX>.
- [13] S. Aggarwal, K. Pandey, Early identification of PCOS with commonly known diseases: obesity, diabetes, high blood pressure and heart disease using machine learning techniques, *Expert Syst. Appl.* 217 (2023) 119532, <https://doi.org/10.1016/J.ESWA.2023.119532>.
- [14] S. Aggarwal, K. Pandey, PCOS diagnosis with commonly known diseases using hybrid machine learning algorithms, in: *Proceedings of International Conference on Contemporary Computing and Informatics, IC3I*, vol. 2023, 2023, pp. 1658–1662, <https://doi.org/10.1109/IC3I59117.2023.10397717>.
- [15] S. Aggarwal, K. Pandey, An analysis of PCOS disease prediction model using machine learning classification algorithms, *Recent Pat. Eng.* 15 (2020) 6–16, <https://doi.org/10.2174/1872212115999201224130204>.
- [16] S. Aggarwal, K. Pandey, Determining the representative features of polycystic ovary syndrome via Design of Experiments, *Multimed. Tool. Appl.* 81 (2022) 29207–29227, <https://doi.org/10.1007/S11042-022-12913-0/METRICS>.
- [17] W. Wang, W. Zeng, S. He, Y. Shi, X. Chen, L. Tu, B. Yang, J. Xu, X. Yin, A new model for predicting the occurrence of polycystic ovary syndrome: based on data of tongue and pulse, *Digit Health* 9 (2023), <https://doi.org/10.1177/20552076231160323>.
- [18] J. Lim, J. Li, X. Feng, L. Feng, Y. Xia, X. Xiao, Y. Wang, Z. Xu, Machine learning classification of polycystic ovary syndrome based on radial pulse wave analysis, *BMC Complement Med Ther* 23 (2023) 1–15, <https://doi.org/10.1186/S12906-023-04249-5/FIGURES/12>.
- [19] J. Lim, J. Li, X. Feng, L. Feng, X. Xiao, Y. Xia, Y. Wang, L. Qian, H. Yang, Z. Xu, Machine learning-based evaluation of application value of traditional Chinese medicine clinical index and pulse wave parameters in the diagnosis of polycystic ovary syndrome, *Eur J Integr Med* 64 (2023) 102311, <https://doi.org/10.1016/J.EUJIM.2023.102311>.
- [20] S. Dong, Z. Lei, Y. Fei, Data-driven based four examinations in TCM: a survey, *Digital Chinese Medicine* 5 (2022) 377–385, <https://doi.org/10.1016/J.DCMED.2022.12.004>.

- [21] J. Li, K. Cheng, S. Wang, F. Morstatter, R.P. Trevino, J. Tang, H. Liu, Feature selection, *ACM Comput. Surv.* 50 (2017), <https://doi.org/10.1145/3136625>.
- [22] Y. Wang, L. Ma, P. Liu, Feature selection and syndrome prediction for liver cirrhosis in traditional Chinese medicine, *Comput Methods Programs Biomed* 95 (2009) 249–257, <https://doi.org/10.1016/j.cmpb.2009.03.004>.
- [23] J. Xu, Z. xia Xu, P. Lu, R. Guo, H. xia Yan, W. jie Xu, Y. qin Wang, C. ming Xia, Classifying syndromes in Chinese medicine using multi-label learning algorithm with relevant features for each label, *Chin. J. Integr. Med.* 22 (2016) 867–871, <https://doi.org/10.1007/S11655-016-2264-0/METRICS>.
- [24] J. Xie, Y. Li, N. Wang, L. Xin, Y. Fang, J. Liu, Feature selection and syndrome classification for rheumatoid arthritis patients with Traditional Chinese Medicine treatment, *Eur J Integr Med* 34 (2020) 101059, <https://doi.org/10.1016/j.eujim.2020.101059>.
- [25] B.C.J.M. Fauser, Fauser Tarlatzis, Aziz Chang, Dewailly Legro, Balen Franks, Dahlgren Bouchard, Diamanti Devoto, Filicori Dunaif, Ibanez Homburg, Magoffin Laven, Norman Nestler, Pugeat Pasquali, Tan Strauss, Wild Taylor, Guzik Wild, Lobo Ehrmann, Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome (PCOS), *Hum. Reprod.* 19 (2004) 41–47, <https://doi.org/10.1093/HUMREP/DEH098>.
- [26] T. Yong, *Gynecology of Traditional Chinese Medicine*, China Traditional Chinese Medicine, Press, Beijing, 2016.
- [27] L. Can Dong, *Diagnostic of Traditional Chinese Medicine*, China Traditional Chinese Medicine, Press, Beijing, 2016.
- [28] China Association of Chinese Medicine, *Guidelines for Diagnosis and Treatment of Common Diseases in Gynecology of Traditional Chinese Medicine*, China Traditional Chinese Medicine Press, Beijing, 2012.
- [29] F. Zhao Fu, *Contemporary Sphygmology in Traditional Chinese Medicine*, People's Medical Publishing House, Beijing, 2003.
- [30] A. Jadhav, D. Pramod, K. Ramanathan, Comparison of performance of data imputation methods for numeric dataset, *Appl. Artif. Intell.* 33 (2019) 913–933, <https://doi.org/10.1080/08839514.2019.1637138>.
- [31] StandardScaler — scikit-learn 1.5.1 documentation, (n.d.). <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html> (accessed July 21, 2024).
- [32] A.B. Majumder, S. Gupta, S. Majumder, D. Singh, A heart disease prediction model using merged XGBoost-SVM classifier and particle swarm optimization, in: *Proceedings - 2024 5th International Conference on Mobile Computing and Sustainable Informatics, ICMCSI*, vol. 2024, 2024, pp. 241–248, <https://doi.org/10.1109/ICMCSI61536.2024.00042>.
- [33] sklearn.feature_selection.VarianceThreshold — scikit-learn 1.4.0 documentation, (n.d.). https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.VarianceThreshold.html (accessed February 6, 2024).
- [34] Y. Wang, L. Wang, B. Qin, X. Hu, W. Xiao, Z. Tong, S. Li, Y. Jing, L. Li, Y. Zhang, Preoperative prediction of sonic hedgehog and group 4 molecular subtypes of pediatric medulloblastoma based on radiomics of multiparametric MRI combined with clinical parameters, *Front. Neurosci.* 17 (2023) 1157858, <https://doi.org/10.3389/FNINS.2023.1157858/BIBTEX>.
- [35] J. Sung, S. Han, H. Park, S. Hwang, S.J. Lee, J.W. Park, I. Youn, Classification of stroke severity using clinically relevant symmetric gait features based on recursive feature elimination with cross-validation, *IEEE Access* 10 (2022) 119437–119447, <https://doi.org/10.1109/ACCESS.2022.3218118>.
- [36] R. Tibshirani, Regression shrinkage and selection via the lasso, *J R Stat Soc Series B Stat Methodol* 58 (1996) 267–288, <https://doi.org/10.1111/J.2517-6161.1996.TB02080.X>.
- [37] J. Ranstam, J.A. Cook, LASSO regression, *Br. J. Surg.* 105 (2018), <https://doi.org/10.1002/BJS.10895>, 1348–1348.
- [38] E. Keany, BorutaShap 1.0.17, (n.d.). <https://pypi.org/project/BorutaShap/> (accessed March 4, 2024).
- [39] S.M. Lundberg, G. Erion, H. Chen, A. DeGrave, J.M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, S.I. Lee, From local explanations to global understanding with explainable AI for trees, *Nat. Mach. Intell.* 2 (1 2) (2020) 56–67, <https://doi.org/10.1038/s42256-019-0138-9>, 2020.
- [40] K. Kira, L.A. Rendell, A practical approach to feature selection, in: *Proceedings of the 9th International Workshop on Machine Learning, ICML*, vol. 1992, 1992, pp. 249–256, <https://doi.org/10.1016/B978-1-55860-247-2.50037-1>.
- [41] I. Kononenko, M. Robnik-Sikonja, U. Pompe, *RelieFF for estimation and discretization of attributes in classification, regression. ILP Problems.*, IOS Press, Amsterdam, The Netherlands, 1996.
- [42] I. Kononenko, Estimating attributes: analysis and extensions of RELIEF, *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 784 LNCS (1994) 171–182, https://doi.org/10.1007/3-540-57868-4_57/COVER.
- [43] D.A. Pisner, D.M. Schnyer, Support Vector Machine, *Machine Learning: Methods and Applications to Brain Disorders*, 2020, pp. 101–121, <https://doi.org/10.1016/B978-0-12-815739-8.00006-7>.
- [44] S. Nusinovi, Y.C. Tham, M.Y. Chak Yan, D.S. Wei Ting, J. Li, C. Sabanayagam, T.Y. Wong, C.Y. Cheng, Logistic regression was as good as machine learning for predicting major chronic diseases, *J. Clin. Epidemiol.* 122 (2020) 56–69, <https://doi.org/10.1016/j.jclinepi.2020.03.002>.
- [45] T. Chen, C. Guestrin, XGBoost: a scalable tree boosting system, in: *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 13-17-August-2016*, 2016, pp. 785–794, <https://doi.org/10.1145/2939672.2939785>.
- [46] L. Dai, J. Zhang, C. Li, C. Zhou, S. Li, Multi-label feature selection with application to TCM state identification, *Concurr. Comput.* 31 (2019) e4634, <https://doi.org/10.1002/CPE.4634>.
- [47] I.S. Silva, C.N. Ferreira, L.B.X. Costa, M.O. Sôter, L.M.L. Carvalho, J. de, M.F. Sales, A.L. Candido, F.M. Reis, A.A. Veloso, K.B. Gomes, Polycystic ovary syndrome: clinical and laboratory variables related to new phenotypes using machine-learning models, *J. Endocrinol. Invest.* 45 (2022) 497–505, <https://doi.org/10.1007/S40618-021-01672-8/FIGURES/3>.
- [48] J. Zheng, Z. Zhang, J. Wang, R. Zhao, S. Liu, G. Yang, Z. Liu, Z. Deng, Metabolic syndrome prediction model using Bayesian optimization and XGBoost based on traditional Chinese medicine features, *Heliyon* 9 (2023) e22727, <https://doi.org/10.1016/j.heliyon.2023.e22727>.
- [49] E. Christodoulou, J. Ma, G.S. Collins, E.W. Steyerberg, J.Y. Verbakel, B. Van Calster, A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models, *J. Clin. Epidemiol.* 110 (2019) 12–22, <https://doi.org/10.1016/j.jclinepi.2019.02.004>.
- [50] Z. Huang, J. Miao, J. Chen, Y. Zhong, S. Yang, Y. Ma, C. Wen, A traditional Chinese medicine syndrome classification model based on cross-feature generation by convolution neural network: model development and validation, *JMIR Med Inform* 10 (4) (2022) E29290, <https://doi.org/10.2196/29290>. *Medinform.Jmir.Org/2022/4/E29290* 10 (2022) e29290.
- [51] Z. Liu, H. He, S. Yan, Y. Wang, T. Yang, G.Z. Li, End-to-End models to imitate traditional Chinese medicine syndrome differentiation in lung cancer diagnosis: model development and validation, *JMIR Med Inform* 8 (6) (2020) E17821, <https://doi.org/10.2196/17821>. *Medinform.Jmir.Org/2020/6/E17821* 8 (2020) e17821.
- [52] C. Zhao, G.Z. Li, C. Wang, J. Niu, Advances in patient classification for traditional Chinese medicine: a machine learning perspective, *Evid Based Complement Alternat Med* 2015 (2015), <https://doi.org/10.1155/2015/376716>.