ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 1 - 2026

Design of decision model for quality evaluation of traditional Chinese Medicine training based on improved Random Forest

Ziyue Guan[1], *Yanli Liao*[2], *Chuang Liu*[1] *and Xiaojiang Peng*[1,*]

[1] School of Chinese Medicine, Zhaoqing Medical College, Zhaoqing, 526000, Guangdong, China

[2] School of General Practice Medicine, Zhaoqing Medical College, Zhaoqing, 526000, Guangdong, China

SUMMARY: *The quality of traditional Chinese medicine (TCM) training directly affects the formation of syndrome differentiation thinking, the mastery of operational skills and the improvement of subsequent clinical competence of students. However, the existing evaluation methods mostly rely on teachers 'experience scoring and outcome assessment, which has problems such as insufficient utilization of indicators, insufficient integration of process information and unstable quality stratification. Therefore, this paper constructs the quality evaluation decision model of traditional Chinese medicine training based on improved Random Forest, and integrates multi-dimensional indicators such as theoretical test scores, manipulation standard degree, operation proficiency, case analysis integrity, classroom participation and teaching feedback into the unified feature space. Through feature screening, category weight adjustment and parameter optimization, the recognition ability of the model for complex teaching data was improved. The improved Random Forest was further compared with Decision Tree, standard Random Forest and XGBoost. The results show that the Accuracy, Precision, Recall and F1 values of the improved Random Forest reach 0.923, 0.911, 0.896 and 0.903, respectively, and the verification error is reduced to 0.112 at the 80th iteration. The overall performance is better than that of the control model. The empirical analysis found that the accuracy of syndrome differentiation judgment, the standard of manipulation, the proficiency of operation and the integrity of case analysis were the core factors affecting the quality stratification of TCM training. The research shows that the improved Random Forest can effectively reveal the key variables in the formation of traditional Chinese medicine training quality, and provide a new technical path for the optimization of training teaching, hierarchical guidance and digital teaching management.*

Key words: *TCM training quality evaluation; Random forest; Machine learning; Instructional Decision Model*

1 Introduction

Traditional Chinese medicine (TCM) practical training is the key link that can best reflect the characteristics of "understanding and practice" in the TCM talent training system. Its quality is not only related to the students 'mastery of syndrome differentiation thinking, four diagnosis and combination of parameters and clinical operation standards, but also directly affects the follow-up post competency and clinical service ability. In recent years, with the gradual shift of

*guan4425390@163.com
https://doi.org/10.65102/is2026256 Published: 30 April, 2026

traditional Chinese medicine education from experience-teaching training to capacity-oriented training, the organizational mode, assessment content and quality monitoring methods of practical teaching have all changed [6-10]. In this context, how to carry out a more accurate, stable and interpretable quality assessment of the process and results of traditional Chinese medicine training has become an unavoidable problem in teaching management and quality control of talent training.

From the existing research, the discussion around the quality of TCM education mostly focuses on the course satisfaction, clinical ability performance, the effect of standardized patient training and the comparison of teaching models. These studies provided an important reference for understanding the influencing factors of TCM teaching quality, and also showed that the quality of training was not determined by a single score, but by multi-dimensional factors such as theoretical basis, operation norms, case analysis, clinical communication, learning engagement and teacher guidance. However, the existing evaluation methods are still mainly based on questionnaire statistics, experience scoring or single index comparison. The evaluation results are susceptible to subjective judgment, sample fluctuation and index correlation interference, and it is often difficult to form a stable decision support model in the face of multi-source heterogeneous teaching data. Especially in the training scene, student performance has the characteristics of stage, nonlinear and interactive, and the ability of traditional linear analysis method to identify complex relationships is relatively limited.

With the development of educational data mining and machine learning technology, algorithm-based quality assessment model provides a new technical path for teaching decision-making. Random forest has been used in the prediction of students' learning performance and the evaluation of education quality because of its advantages of strong adaptability to high-dimensional features, good anti-overfitting ability, and the ability to deal with nonlinear relationships [18-20]. However, there are still some limitations in the practical application of standard random forest, such as feature redundancy will weaken the splitting efficiency, unbalanced class distribution will affect the recognition ability of the model for low-frequency samples, and if the parameter setting only depends on experience, it is also easy to cause fluctuations in prediction accuracy and generalization ability. For the traditional Chinese medicine training data with complex index structure and obvious cross evaluation dimensions, it is often difficult to fully release the model performance if the traditional random forest is directly applied.

Based on this, this paper constructs an improved Random Forest decision-making model around the actual needs of traditional Chinese medicine training quality evaluation. Based on the design of the quality evaluation index system of traditional Chinese medicine (TCM) training, this paper introduced the strategies of feature screening, parameter optimization and sample weight adjustment, integrated and analyzed the students 'training assessment data, process performance data and teaching feedback data, and formed an evaluation framework for quality hierarchical identification and teaching decision support. This paper hopes to improve the objectivity and operability of the quality evaluation of traditional Chinese medicine training through computer modeling method, and provide method support for the improvement of training teaching, personalized intervention and digital quality management.

2 Literature Review

In recent years, the research on teaching quality evaluation has gradually moved away from the traditional path of relying solely on final grades or teachers 'experience judgment, and began to evolve to the direction of multi-index integration, process data mining and algorithm-aided decision-making. Especially in the scenarios of medical education and professional ability training, the evaluation objects often include multiple dimensions such as knowledge mastery, operation specification, clinical communication, adaptability and professional quality at the same time, and the relationship between the indicators is not simple linear, and it is difficult for a single statistical method to fully describe the real level [11-16]. In this context, how to use computer modeling methods for structural analysis of complex teaching data has become an important extension of educational evaluation research.

In terms of traditional Chinese medicine education, the existing research is mostly carried out from the perspectives of training objectives, curriculum effects and clinical ability. Zheng et al. constructed an evaluation tool for TCM students from the perspective of competency, indicating that the cultivation of TCM talents needs to establish a more operational measurement framework between knowledge, skills and comprehensive literacy [1]. Qu et al. investigated the TCM learning situation of international students and revealed the influence of differences in learning background on the judgment of teaching effect [2]. Zhang et al. evaluated online courses of traditional Chinese medicine during the epidemic, reflecting that the judgment of teaching quality was no longer limited to offline classes, but gradually extended to multiple scenarios [3]. Zeng et al. and Huang et al. investigated the improvement effect of students' clinical ability based on standardized patient and teacher participatory training respectively, indicating that the quality of TCM training is closely related to the way of situation construction, feedback mechanism and training intensity [4, 5]. Kim introduced PBL method into undergraduate teaching of traditional Chinese medicine, and further explained that the change of teaching mode would have a significant impact on the formation of students' ability [6]. Chen et al., Zhou et al., from the perspective of transnational comparison and the integration of traditional teacher-apprentice and modern education, suggested that the quality evaluation of TCM education should not only focus on the end result, but also take into account the training path and teaching organization mode [7, 8]. Xu et al. applied cognitive diagnosis method to the evaluation of national TCM examination, which provided strong methodological inspiration for ability level identification [9]. These studies laid a foundation for understanding the quality of traditional Chinese medicine (TCM) education, but most of them were more inclined to descriptive analysis, questionnaire measurement or single teaching reform effect test, and the joint modeling of multi-source data in practical training was still insufficient.

In the study of clinical teaching quality evaluation, Snell et al., Fluit et al., Ansari et al., Vaižgėlienė et al., and Shorey et al., discussed from the aspects of teacher evaluation, questionnaire tools, teaching quality measurement and post competency activities [11-15]. Related studies have promoted the transformation of evaluation content from "whether to teach" to "whether to form competency", and also promoted evaluation tools to pay more attention to construct validity and application context [12-16]. However, most of these studies are based on scale statistics, expert scoring or institutional framework analysis, and the depth of mining the possible nonlinear associations, variable interaction effects and hidden hierarchical characteristics between evaluation objects is still limited. In other words, the existing research has been sufficient in "evaluation dimension design", but there is still room for improvement in "complex data calculation".

With the development of educational data mining, machine learning has been introduced into student performance prediction and education quality assessment scenarios. Boruta feature selection method proposed by Kursa et al provides a stable tool for high-dimensional variable identification, which can effectively reduce the interference of redundant features on model judgment [17]. Huynh-Cam et al. used decision trees and random forests to predict the learning performance of college freshmen, which proved that the tree model had good adaptability and interpretability in educational data processing [18]. Chen et al. further optimized the random forest tree structure by genetic algorithm for the prediction of students' academic performance, showing that parameter optimization can improve the accuracy and stability of the model [19]. Lu et al. introduced random forest and Logistic regression into the quality evaluation of college students' innovation and entrepreneurship education, showing that this kind of algorithm is not only suitable for performance prediction, but also suitable for multi-index comprehensive evaluation [20]. Compared with single regression or artificial weighting methods, random forest can deal with nonlinear relationships, accommodate heterogeneous variables, and output feature importance results. Therefore, it is more suitable for scenes with complex evaluation dimensions and obvious indicators intersection such as traditional Chinese medicine training.

As shown in Table 1, the existing research has made progress in the evaluation of TCM education, clinical teaching measurement and the application of educational machine learning, but there are still two prominent gaps. First, there are few modeling studies on the specific object of TCM "training quality", and the existing results focus on the curriculum, examination or teaching mode, rather than the comprehensive evaluation of the whole process of training. Second, although the existing teaching evaluation research emphasizes multi-dimensional indicators, it rarely introduces computer strategies such as feature selection, class balance adjustment and parameter optimization, which leads to the lack of discrimination ability of the model on complex samples.

Table 1: Main contents and limitations of related studies

Reference	Research Topic	Main Method / Tool	Key Findings	Limitations
Zheng et al. [1]	Competency Assessment of TCM Students	Competency Evaluation Tool Development	Provided a structured framework for measuring TCM education quality	Focused on tool development; lacks intelligent predictive modeling
Qu et al. [2]	Evaluation of TCM Learning for International Students	Survey Analysis	Reflected the impact of different learning backgrounds on teaching effectiveness	Results mainly based on questionnaire statistics; limited computational depth
Zhang et al. [3]	Online TCM Course Evaluation	Course Evaluation Analysis	Demonstrated that changes in teaching scenarios affect quality judgment	More focused on course level; did not deeply address practical training evaluation
Zeng et al. [4]	Standardized Patients in TCM Education	Prospective Randomized Study	Helped improve the authenticity of clinical competency measurement	Focused on teaching intervention effects; not a comprehensive assessment model
Huang et al. [5]	Teacher-led Standardized Patient Training	Randomized Controlled Study	Positively influenced clinical skill improvement	Evaluation indicators relatively focused; hard to cover full training scope
Xu et al. [9]	Diagnostic Assessment of TCM Exam Skills	Cognitive Diagnostic Methods	Improved accuracy in ability level identification	Exam-focused scenarios; not fully adapted to practical training data
Fluit et al. [12]	Clinical Teacher Quality Measurement	Systematic Review & Scale Analysis	Emphasized content validity and structural soundness of evaluation tools	Highly reliant on scales and subjective scoring
Kursa et al. [17]	Feature Selection	Boruta Algorithm	Useful for identifying key variables and reducing redundant interference	Not a complete predictive model by itself
Huynh-Cam et al. [18]	Student Learning Performance Prediction	Decision Tree, Random Forest	Demonstrated Random Forest suitability for educational data prediction	Focused on general higher education; insufficient adaptation for professional practical training
Chen et al. [19]	Student Grade Prediction Optimization	Genetic Algorithm-optimized Random Forest	Parameter optimization improves model performance	Mainly aimed at academic grades; did not involve TCM practical training
Lu et al. [20]	Comprehensive Evaluation of Educational Quality	Random Forest, Logistic Regression	Tree models can be applied for multi-indicator education quality evaluation	Focused on innovation and entrepreneurship education; scenario transfer needs verification

Based on the above research context, it can be seen that the research on the quality evaluation of TCM education has formed a certain foundation, and random forest and its optimization method have also shown good application potential in educational data analysis, but the two have not yet formed a close combination on the quality evaluation of TCM training. Therefore, this paper integrates the evaluation index system of traditional Chinese medicine (TCM) training with the improved Random Forest algorithm. While preserving the interpretability and robustness of the tree model, the evaluation accuracy is improved through feature screening, parameter optimization and sample adjustment, and then a quality evaluation decision model more suitable for TCM training scenarios is constructed.

3 The quality evaluation decision model of traditional Chinese medicine training based on improved random forest

3.1 Improving random forest and its optimization strategy

Random forest is an ensemble learning method based on decision trees. Its core idea is to construct multiple training subsets by sampling with replacement on the original sample set, and obtain the final output through the collaborative discrimination of multiple classification trees. Compared with the single decision tree, the random forest has better robustness in dealing with high-dimensional features, nonlinear relationships and interaction effects between variables, especially in the scene of traditional Chinese Medicine training quality evaluation where indicators have multiple sources, variable structures are complex, and category boundaries are not clear [18-20]. The process of traditional Chinese medicine (TCM) training includes not only skill indicators such as pulse diagnosis, tongue diagnosis, massage manipulation and acupuncture operation, but also process indicators such as case analysis, syndrome differentiation thinking, doctor-patient communication and classroom participation. There are often coupling relationships between different variables, and it is difficult to completely identify its internal structure by traditional linear models. Based on this, this paper uses random forest as the basic classifier, and improves it by combining feature screening, class weight adjustment and parameter optimization strategy, so as to improve the recognition accuracy and interpretation ability of the model for the quality level of TCM training. In the standard random forest, let the training sample set be

$$D = \{(x_i, y_i)\}_{i=1}^{N} \tag{1}$$

Here, xi represents the training feature vector of the ith student, and yi represents the corresponding quality level label. The model repeatedly draws samples from D by Bootstrap, generates T sub-training sets, and builds classification trees respectively. For each split node, the algorithm does not traverse all the features, but searches for the optimal partition in the randomly selected candidate feature subset, which can reduce the correlation between trees. Node purity is usually measured by the Gini index, which is expressed as

$$\mathrm{Gini}(D) = 1 - \sum_{c=1}^{C} p_c^2 \tag{2}$$

Here, p_c represents the proportion of samples of class c in the sample set D, and C is the number of classes. A smaller Gini value means that the samples inside the node are more pure and the splitting effect is better. For the quality evaluation of traditional Chinese medicine training, this division mechanism can better distinguish different quality levels such as "excellent", "good", "qualified" and "to be improved".

However, if the standard random forest is directly used, there are still three problems. First, there are a large number of evaluation indicators in TCM training, and there may be strong redundancy between different indicators. For example, class attendance, homework completion and stage test scores have a common representation role to a certain extent, and if they are all included in the modeling, it is easy to weaken the tree splitting efficiency. Secondly, the class distribution of training samples is often unbalanced, the number of students in the "to be improved" level is usually less than that of the "qualified" and "good" levels, and the model is easy to bias to the majority class judgment. Third, if the parameters such as the number of trees, the maximum depth and the number of candidate features are set only empirically, the model performance is often unstable. To address these issues, we perform three optimizations on top of the basic random forest.

One is to introduce the idea of Boruta for feature screening. By constructing shadow features and comparing the importance scores of real variables and random perturbation variables, this method retains the core indicators that have a stable contribution to the classification results, thereby reducing the interference of redundant information. Let the importance of the JTH feature be I_j, then the core feature screening can be expressed as

$$S = \{x_j \mid I_j > I_{shadow}^{max}\} \tag{3}$$

Here, I_{shadow}^{max} is the maximum importance value in the shadow feature. The variables entered into the model after screening were more concentrated in the key dimensions of operation standardization, syndrome differentiation accuracy, case analysis integrity and training feedback quality, and the model discrimination boundary was also clearer.

The second is to add a class weight adjustment mechanism to reduce the bias caused by sample imbalance. Let the number of samples of class c be n_c, the total number of samples be N, and the number of classes be C, then its weight can be defined as

$$w_c = \frac{N}{C \cdot n_c} \tag{4}$$

After the introduction of weights in the tree node splitting and final voting stage, the model's attention to low-frequency categories is improved, which is more conducive to identifying students with weak training quality, thereby enhancing the practical value of the model in teaching early warning scenarios.

The third is the combination of grid search and out-of-bag error for parameter optimization. The out-of-bag error of the random forest can be written as

$$OOB\_error = \frac{1}{N}\sum_{i=1}^{N} I\,(\hat{y}_i \neq y_i) \tag{5}$$

Here, $\hat{y}_i$ is the prediction result of out-of-bag samples, and $I(\cdot)$ is the indicative function. In this paper, the number of trees, the maximum depth, the minimum leaf node sample number and the maximum number of features are jointly searched to minimize the out-of-bag error and take into account the macro-average F1 value, so that the model achieves a better balance between accuracy, generalization and computational efficiency.

The improved model process is shown in Figure. 1. The model first cleaned, standardized and coded the original TCM training data, and then completed the core feature screening. Then it trained multiple classification trees by combining category weight and parameter optimization, and finally output the quality level results through weighted voting. Compared with the traditional random forest, this optimization framework does not change the basic structure of the ensemble tree model, but enhances the adaptation ability of the model to the complex data of traditional Chinese medicine training in three levels of feature expression, category recognition and parameter adaptation, which lays the algorithm foundation for the subsequent quality evaluation system construction and empirical analysis.

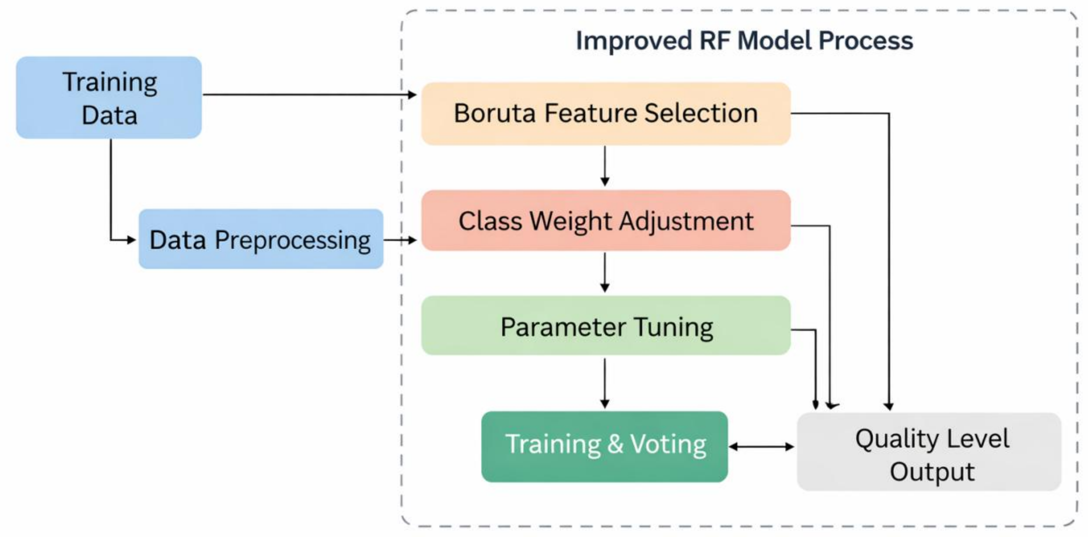

Figure 1: Flow chart of TCM training quality evaluation decision model based on improved random forest

3.2 Construction of TCM training quality evaluation system

The key of TCM training quality evaluation is not to simply summarize a number of scores, but to transform the information scattered in the teaching process, such as knowledge mastery, operation execution, syndrome differentiation analysis and feedback correction, into a computable, comparable and traceable index structure. TCM training has obvious complex characteristics. On the one hand, it involves operational skills such as acupuncture, massage, tongue diagnosis and pulse diagnosis, and on the other hand, it includes cognitive content such as pathogenesis judgment, syndrome differentiation and treatment, case expression and docer-patient communication. If we only rely on teacher experience to make overall judgment, it is easy to be affected by subjective bias and grading scale differences. Based on this, this paper constructs a quality evaluation system for the improved Random Forest model based on the analysis of the teaching process of traditional Chinese medicine (TCM) training, so as to transform the quality of TCM training from experience judgment into a structured decision problem. From the system perspective, the TCM training quality evaluation system can be expressed as

$$Q = (I, R, W) \tag{6}$$

where Q represents the quality evaluation system of TCM training, I represents the set of indicators, R represents the association relationship between indicators, and W represents the strength of each indicator in model discrimination. This expression shows that the evaluation system is not a simple juxtaposition of several indicators, but an organic structure composed of multi-level variables. Different indicators not only have mutual support relationship, but also may have overlap and redundancy. Therefore, the hierarchical sorting and variable constraint must be completed in the system construction stage.

Combined with the teaching organization mode and data availability of traditional Chinese medicine training, this paper divided the evaluation indicators into four first-level dimensions, that is, students 'basic and learning status, training operation performance, syndrome differentiation analysis and clinical expression, teaching support and process feedback. The set of first-class dimensions can be written as

$$I = \{I_1, I_2, I_3, I_4\} \tag{7}$$

Among them, I_1 represents students' basic and learning status, I_2 represents practical operation performance, I_3 represents syndrome differentiation analysis and clinical expression, and I_4 represents teaching support and process feedback. This division method, on the one hand, can cover the main links of "learning, practice, judgment, and correction" in traditional Chinese medicine training, on the other hand, it is also convenient for subsequent mapping of data from different sources into a unified model input space.

Below the first-level indicators, this paper further sets up a number of second-level indicators. Students' basic and learning status mainly reflected the learning starting point and training input, including theoretical test scores, attendance, autonomous practice frequency and classroom participation. The training operation performance mainly measures the standardization and stability of students in the operation level, including the standard degree of manipulation, the operation proficiency, the accuracy of acupoint positioning and the standardization of instrument use. Syndrome differentiation analysis and clinical expression mainly correspond to the quality of TCM thinking, including the integrity of four diagnosis information extraction, the accuracy of syndrome differentiation judgment, the rationality of treatment ideas and the clarity of case expression. Teaching support and process feedback are used to describe the external teaching conditions, including the timeliness of teacher guidance, the effectiveness of feedback, the completeness of training resources and the response degree of stage error correction. If the KTH student's observation on the JTH index is assumed to be x_{kj}, its original eigenvector can be written as

$$X_k = (x_{k1}, x_{k2}, \dots, x_{km}) \tag{8}$$

where, m is the total number of evaluation indicators. The vector can contain both numerical variables and coded rank variables, so as to meet the processing requirements of random forest for mixed educational data.

Because the dimensions of the different metrics are not consistent, such as attendance as a percentage, performance rating as a score, and frequency of self-directed practice as a number of times, they need to be standardized before entering the model. In this paper, the range normalization method is used to map each index to a unified numerical interval, which is expressed as

$$x'_{kj} = \frac{x_{kj} - \min(x_j)}{\max(x_j) - \min(x_j)} \tag{9}$$

where $\min(x_j)$ and $\max(x_j)$ represent the minimum and maximum value of the JTH index in all samples, respectively. After processing, different indicators remain consistent in numerical scale, which is conducive to node splitting and importance ranking of the model.

Considering that some indicators belong to negative indicators, such as the number of operation errors, the number of missing items in cases and the number of lateness, larger values of such variables do not mean higher quality, so direction consistency processing is also needed.

If the JTH index is negative, its transformed form can be expressed as

$$z_{kj} = 1 - x'_{kj} \tag{10}$$

After this transformation, all the indicators meet the unified logic of "the larger the value is, the better the quality is", and avoid the misjudgment of the model due to the conflict of the variable direction in the learning process.

In terms of quality level labeling, combined with the practice assessment practice of traditional Chinese medicine, the quality of student training is divided into four categories: excellent, good, qualified and to be improved. Let the set of quality labels be

$$Y = \{y_1, y_2, y_3, y_4\} \tag{11}$$

Among them, y_1, y_2, y_3, y_4 correspond to the above four levels respectively. The label generation can be completed according to the comprehensive rating interval, and can also refer to the fusion result of the teacher's final rating and stage performance, so as to ensure that the supervised learning samples have high teaching interpretability.

In order to reduce the impact of index redundancy on model training efficiency and discrimination stability, this paper introduces a correlation constraint mechanism in the evaluation system construction stage. If the correlation coefficient between two indicators x_a and x_b is too high, the indicator with more explanatory power and clearer teaching implications will be retained. Its correlation coefficient can be expressed as

$$r_{ab} = \frac{\sum_{k=1}^{N}(x_{ka} - \bar{x}_a)(x_{kb} - \bar{x}_b)}{\sqrt{\sum_{k=1}^{N}(x_{ka} - \bar{x}_a)^2}\sqrt{\sum_{k=1}^{N}(x_{kb} - \bar{x}_b)^2}} \tag{12}$$

Here, N is the number of samples, and $\bar{x}_a$ and $\bar{x}_b$ are the mean values of the corresponding indices, respectively. Through this processing, the integrity of the evaluation system can be maintained while reducing repeated information input and improving the efficiency of subsequent model training.

According to the above ideas, the TCM training quality evaluation system constructed in this paper is shown in Figure 2. The system is composed of a three-layer structure of "first-level dimensions-second-level indexes-quality labels", which can not only reflect the core competence requirements of traditional Chinese medicine training, but also connect with the data organization method of computer model. Different from traditional evaluation tables, the indicators in this system are not regarded as isolated scoring items, but as a set of features that can be entered into the machine learning model, so as to provide a unified data basis for the subsequent improvement of the training, prediction and decision output of Random Forest.

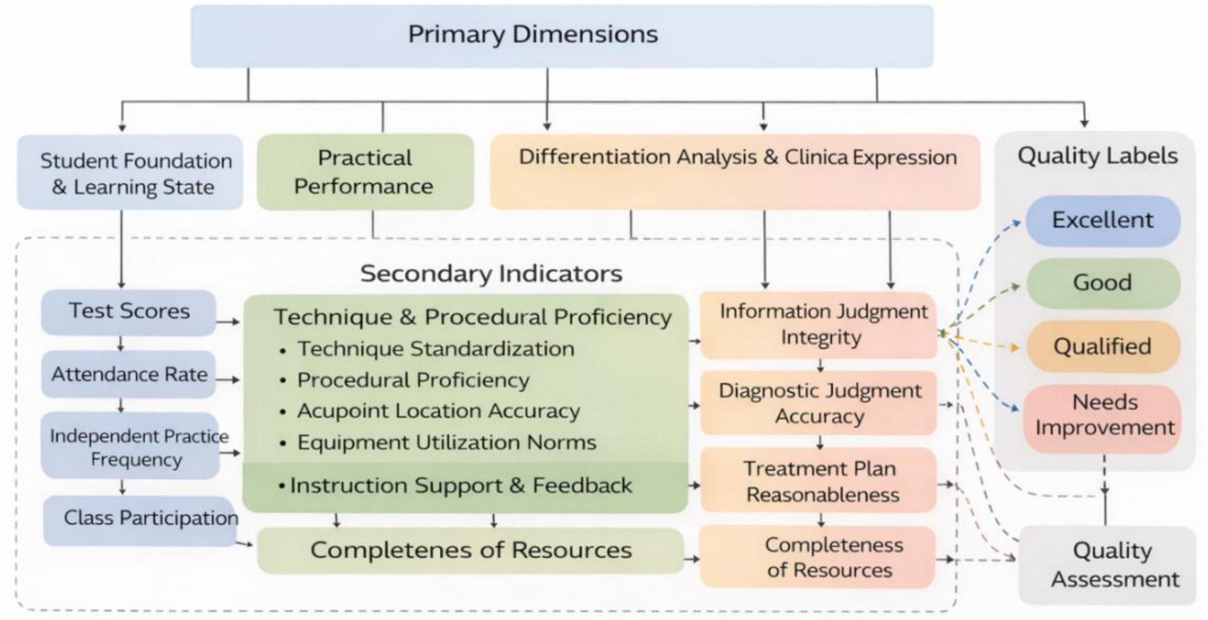

Figure 2: Architecture of TCM training quality evaluation based on improved random forest

After completing the construction of the evaluation index system and improving the random forest optimization, the key to the quality evaluation of traditional Chinese medicine training turned to how to run the model. For the teaching management scenario, the classification algorithm alone is not enough to support quality decision-making. It is also necessary to establish a complete implementation chain from data collection, feature processing, model training to result output, so that the evaluation results can be used by teachers, teaching researchers and management. Based on this consideration, the decision-making process of the quality evaluation model of traditional Chinese medicine training is designed as four interconnected modules of "data access, feature conversion, model discrimination and result feedback", and a computable, callable and iterative implementation framework is constructed. Let the original sample set be

$$D = \{X, Y\} \tag{13}$$

Among them, X represents the feature matrix composed of students 'training results, operation records, classroom behavior, teacher evaluation and stage feedback, and Y represents the set of quality level labels. When the model is running, the system first extracts data from the training teaching platform, the score management module and the classroom record port, completes missing value filling, outlier identification and category variable coding, and then inputs the processed results into the feature screening module. The valid features retained after the screening are fed into the improved random forest classifier, which is used to generate the student quality rating and the corresponding probability distribution.

In order to enhance the decision significance of the model output, this paper does not only retain a single class result, but also outputs the posterior probability of each quality level. Let the predicted probability of a student sample x on class c be $P(y = c|x)$, then the final decision result of the model can be expressed as

$$\hat{y} = \arg\max_{c \in C} P\,(y = c|x) \tag{14}$$

Here, C is the set of quality levels. This expression means that the model does not simply give a static label of "excellent" or "qualified", but determines the optimal decision by comparing the probabilities of each class, so as to provide a more detailed basis for teaching intervention. When the probability of "good" and "qualified" is close, the system can identify this kind of students as boundary samples, prompting teachers to focus on their subsequent performance.

At the implementation level, the model framework needs to balance computational efficiency and readability for teaching. The data layer is mainly responsible for the collection and standardization of the original training data. The feature layer is responsible for index coding, normalization and importance screening. The model layer completed the random forest training and prediction after parameter optimization. The application layer is used to display the evaluation levels, key influencing factors and early warning suggestions. Table 2 presents the main components of the proposed model implementation framework. It can be seen from Table 2 that the framework does not regard the algorithm as a closed black box, but integrates data processing, model calculation and teaching feedback into the same link, so as to ensure that the evaluation results can be truly transformed into the basis for teaching management.

Table 2: The module composition of the model decision-making process and the implementation framework

Module Layer	Main Content	Function / Description
Data Layer	Score data, operational records, attendance information, teacher evaluations, feedback logs	Aggregate multi-source training data to form the raw sample set
Processing Layer	Missing value imputation, outlier detection, variable encoding, normalization	Improve data consistency and construct computable features
Feature Layer	Boruta selection, correlation testing, feature subset generation	Retain key indicators and reduce redundant inputs
Model Layer	Improved Random Forest training, parameter optimization, probability output	Perform quality level identification and prediction
Application Layer	Result visualization, risk alerting, teaching recommendation generation	Support teacher diagnostics and instructional decision-making

After the model is trained, the prediction results need to be remapped back to the teaching context to form executable intervention information. Let the importance of the JTH feature be Ij, then its relative contribution can be defined as

$$R_j = \frac{I_j}{\sum_{k=1}^{m} I_k} \tag{15}$$

where, m is the total number of features. By calculating the relative contribution degree of each index, the system can identify the key factors that affect the quality of TCM training, such as insufficient operation standardization, weak syndrome differentiation analysis or low feedback participation. In this way, the output of the model can no longer stay at the result discrimination, but can further serve the process improvement.

4 Algorithm performance comparison and model empirical analysis

4.1 Source and preprocessing of experimental data

In order to verify the applicability of the improved Random Forest in the quality evaluation task of traditional Chinese medicine (TCM) training, this paper constructs an experimental data set around the teaching process of TCM major in a medical college. The data mainly came from three parts. The first part was the stage assessment records of students 'practical training courses, including structured result data such as theoretical test scores, operation assessment scores, and case analysis scores. The second is the process data generated by the training teaching platform and the classroom management system, including attendance, independent exercise frequency, classroom interaction records, homework submission and stage feedback times. The third is the evaluation data formed by teachers in the training link, including the standard score of manipulation, the integrity evaluation of dialectical thinking, communication expression performance and error correction response records. The above data can completely reflect the students 'knowledge base, operational ability, thinking performance and learning engagement in TCM training, and also provide stable sample support for subsequent model training.

In terms of sample organization, this paper obtained a total of 286 students 'training records. After eliminating the samples with serious missing key fields, incomplete evaluation chains and label conflicts, 268 effective samples were finally retained. Each sample corresponds to the comprehensive performance of a student in a complete training cycle, and contains a total of 16 evaluation features and 1 quality label variable. The quality label is determined according to the comprehensive score of the training and the final evaluation results of the teachers, which is divided into four levels of "excellent", "good", "qualified" and "to be improved". Such a way of label construction not only preserves the actual context of teaching evaluation, but also enhances the goal clarity in the supervised learning task. Table 3 shows the main sources and field composition of the experimental data.

Table 3: Experimental data sources and field composition

Data Source	Key Fields	Data Type	Usage / Purpose
Training Assessment Records	Theoretical Test Scores, Operational Assessment Scores, Case Analysis Scores	Numerical	Reflect students’ stage-wise learning outcomes
Learning Platform Logs	Attendance Rate, Frequency of Independent Practice, Assignment Submission, Interaction Count	Numerical / Count	Describe learning engagement and process participation
Teacher Evaluation Records	Technique Standardization, Diagnostic Accuracy, Expression Clarity, Error Correction Responsiveness	Ordinal / Numerical	Represent practical training skills and performance quality
Comprehensive Evaluation Results	Excellent, Good, Pass, Needs Improvement	Categorical	Serve as supervised labels for the model

From the original data characteristics, the data from different sources have obvious differences in dimension, format and stability. Theoretical test scores and operational assessment scores belong to continuous numerical variables, classroom interaction and autonomous practice frequency belong to counting variables, and teacher evaluation has both grade scores and textual records. If these heterogeneous data are directly input into the model, it will not only affect the consistency of feature expression, but also weaken the classifier's ability to recognize key patterns. Therefore, in this paper, the raw data are systematically preprocessed before modeling.

The preprocessing work mainly includes four aspects. First, for a small number of missing values, the combination of median imputation and mode imputation is used to complete the imputation, in which the median is used to fill the continuous variable, and the mode is used to fill the rank variable, so as to reduce the interference of extreme values. Secondly, for abnormal samples, boxplot rules are combined with teaching record verification to identify, so as to avoid model deviation caused by input errors or abnormal fluctuations. Thirdly, for the graded evaluation variables, the ordinal coding method is used to convert them into computable values, so that the "to-improve-passing-good-excellent" sequence can be maintained clearly. Fourth, for different dimensional variables, a unified normalization process is carried out, so that the eigenvalues are mapped to similar intervals, and the influence of scale differences on the model splitting process is reduced. At the same time, considering that the number of "to be improved" categories in the quality evaluation samples of traditional Chinese medicine training is relatively small, and there is a certain imbalance in the class distribution, this paper synchronously counts the proportion of samples of each level in the preprocessing stage, and introduces the class weight mechanism in the subsequent model training to improve the recognition ability of the model for low-frequency categories. After cleaning, coding and standardization, the data set is divided into training set and test set according to the ratio of 8：2, and five-fold cross validation is used in the training phase to control the fluctuation of experimental results caused by the sample division method.

4.2 Comparative analysis of algorithm performance

In order to test the applicability of the improved Random Forest in the quality evaluation of traditional Chinese medicine training, four models of Decision Tree, standard Random Forest, XGBoost and improved Random Forest are selected for comparative experiments. Each model was run under the same data set and the same partition of training set and test set. The ratio of training set to test set was set to 8：2, and five-fold cross validation was used to reduce the impact of chance error. Considering that the quality evaluation of TCM training requires not only a high overall recognition Accuracy, but also a better recognition of the "to be improved" class samples, this paper selects Accuracy, Precision, Recall and F1 score as the main evaluation indicators.

From the classification results, the improved Random Forest is superior to other control models in four indicators. It can be seen from Figure 3 that the Accuracy, Precision, Recall and F1 value of Decision Tree are 0.821, 0.804, 0.781 and 0.792, respectively, showing a relatively weak overall performance. The Accuracy of the standard Random Forest is improved to 0.884, and the F1 value reaches 0.862, indicating that the stability of the model is improved after multi-tree integration. The Accuracy, Precision, Recall and F1 value of XGBoost are 0.901, 0.889, 0.874 and 0.881, respectively, and the overall performance is further improved. In contrast, the Accuracy of the improved Random Forest reaches 0.923, the Precision is 0.911, the Recall is 0.896, and the F1 value reaches 0.903, which are the highest in all four indicators. Especially in the Recall index, the improved Random Forest is 0.043 higher than the standard Random Forest, indicating that the model has stronger recognition ability for the weak quality samples of traditional Chinese medicine training.

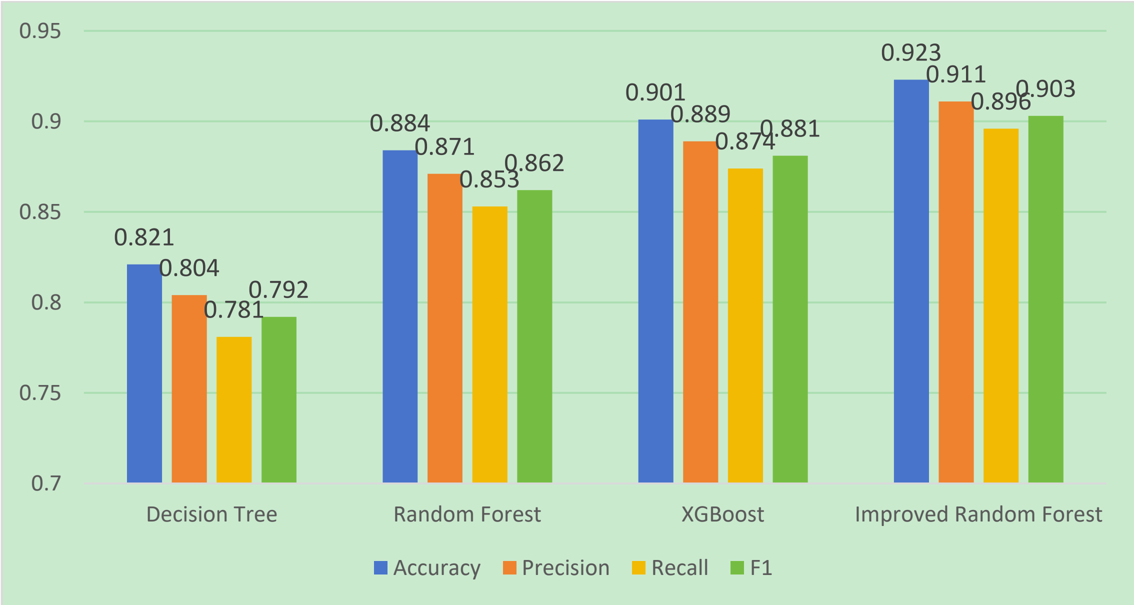

Figure 3: Comparison of classification performance of different algorithms

From the error convergence process, the stability difference of different models in the training phase is also obvious. It can be seen from Figure. 4 that in the early stage of the training process, the validation error of each model is high. As the model complexity increases and the training progresses, the error gradually decreases. Taking the corresponding stage in the figure as an example, at the 10th observation node, the verification errors of Decision Tree, standard Random Forest, XGBoost and improved Random Forest are 0.321, 0.284, 0.266 and 0.241, respectively. At the 40th observation node, the corresponding errors decreased to 0.231, 0.187, 0.161 and 0.134, respectively. At the 80th observation node, the Decision Tree error is 0.208, the standard Random Forest error is 0.156, the XGBoost error is 0.135, and the improved Random Forest error is reduced to 0.112. It can be seen that the improved Random Forest not only has a faster error decline speed, but also enters the stable interval with fewer iterations, showing better training efficiency and generalization ability.

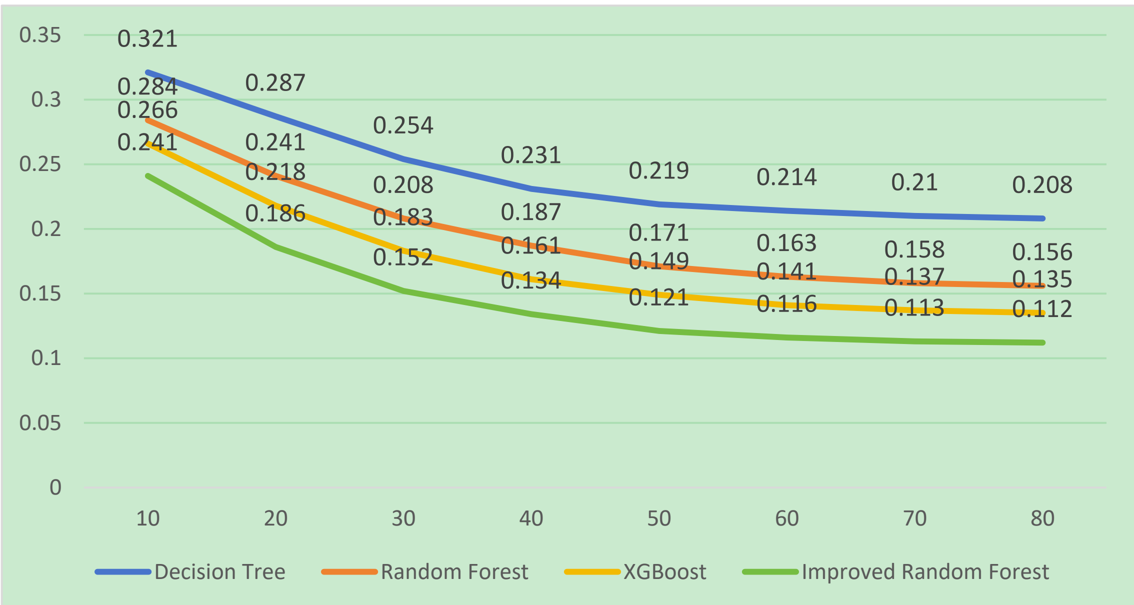

Figure 4: Validation error convergence curves for different algorithms

On the one hand, the reason for the above differences is that the proposed model uses the feature screening method to compress the redundant variables before training, so that the key features are more concentrated. On the other hand, the class weight adjustment alleviates the imbalance of sample distribution, thereby enhancing the sensitivity of the model to low-frequency categories. In addition, parameter optimization further improves the matching relationship between tree depth, the number of leaf node samples and the number of feature extraction, and reduces invalid splitting and local fluctuations. For the quality evaluation of traditional Chinese medicine training, this improvement is not only reflected in the rise of classification values, but also means that the model can identify the student groups with insufficient operation specifications, weak syndrome differentiation analysis and low feedback participation earlier, so as to provide a more reliable basis for subsequent teaching intervention. Based on the results of Figure 3 and Figure 4, it can be considered that the improved Random Forest shows higher accuracy, faster error convergence speed and better balance of category recognition in the quality evaluation task of TCM training, which can better meet the actual needs of hierarchical evaluation of TCM training quality and teaching assistant decision-making.

4.3 Empirical analysis of quality evaluation model

After completing the performance comparison of the algorithm, this paper further applies the improved Random Forest to the actual samples of the quality evaluation of traditional Chinese medicine training, in order to test its interpretability and decision-making value in teaching diagnosis. The focus of empirical analysis is no longer the comparison between the pros and cons of the models, but to identify which evaluation indicators have a greater impact on the quality stratification of TCM training, and reveal the key points of action in current training teaching. To enhance the readability of the results, the core indicators entering the final model are ranked and analyzed based on the feature importance output, as shown in Figure 5.

The results showed that the accuracy of syndrome differentiation judgment, the standard degree of manipulation, the proficiency degree of operation and the completeness of case analysis had the highest importance, with the scores of 0.148, 0.136, 0.129 and 0.118, respectively, all exceeding 0.110. This shows that in the quality evaluation of TCM training, what really separates the differences between students' levels is not only the general attendance or homework completion, but also the dialectical thinking and operation execution level directly related to the core competence of TCM. In particular, the accuracy of syndrome differentiation judgment, whose importance score reached 0.148, ranked first among all indicators, indicating that whether students could transform the information of four diagnoses into more accurate syndrome judgment was the primary factor affecting the classification of training quality. The importance of manipulation standard and operation proficiency were 0.136 and 0.129, respectively, indicating that traditional Chinese medicine training did not stop at the knowledge memory level, and the standardization and stability of practical actions also determined the model's judgment of students' quality level.

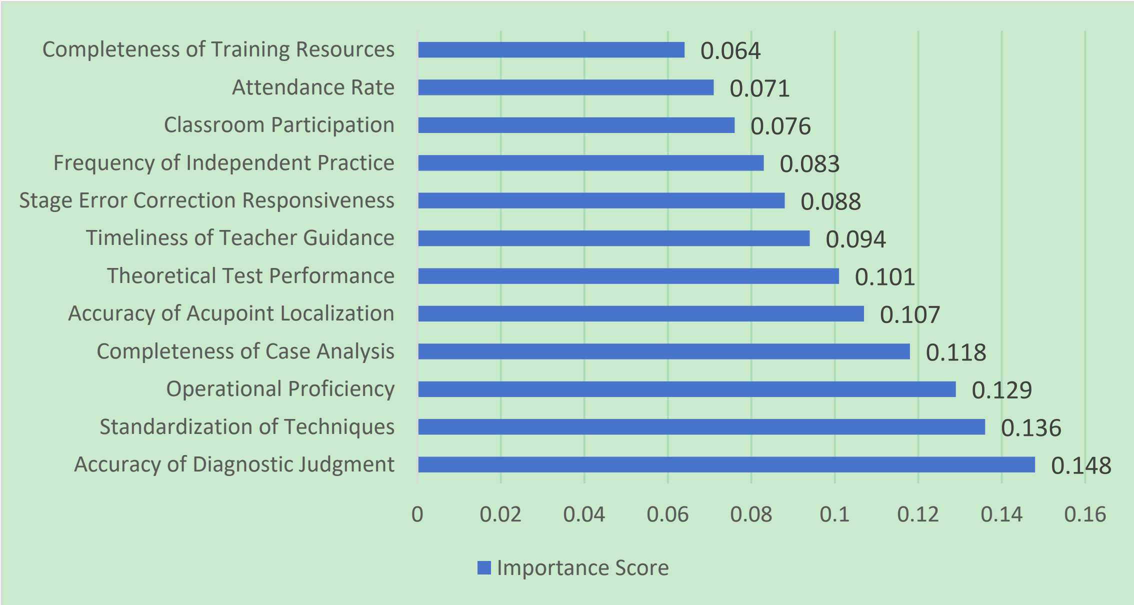

Figure5: Importance ranking of core indicators of TCM training quality evaluation

In contrast, although attendance, frequency of self-regulated practice and classroom participation also had a certain explanatory effect, their importance scores were 0.071, 0.083 and 0.076, respectively, which were lower than the core competence index as a whole. This does not mean that process indicators are not important, but that in the existing training samples, such variables are more reflected as basic supporting factors, which can affect students' performance, but are not enough to determine their final quality level alone. The score of teacher guidance timeliness was 0.094, and the response degree of stage error correction was 0.088, indicating that the teaching support variables still had a role that could not be ignored in the formation process of traditional Chinese medicine training quality, especially for the improvement of low and middle level students.

Combined with the model output results, it can be believed that the current TCM training quality improvement should focus on two aspects. First, strengthen the training of syndrome differentiation analysis and case deduction, and improve the coherence of students from symptom observation to syndrome induction. The second is to strengthen the technique demonstration, operation feedback and repeated practice design, and reduce the cumulative impact of skill errors in training evaluation. The empirical results of the improved Random Forest not only give the quality stratification results, but also reveal the critical path of quality formation through the feature importance, which makes the model turn from a simple classification tool to a decision support tool with teaching interpretation power. For the practical teaching of traditional Chinese medicine, this result can provide more direct data basis for subsequent course adjustment, hierarchical counseling and process intervention.

5 Discussion

Focusing on the quality evaluation task of traditional Chinese medicine training, this paper makes a comparative analysis of Decision Tree, standard Random Forest, XGBoost and improved Random Forest, and further carries out empirical test combined with real teaching samples. The experimental results show that the improved Random Forest is superior to other control models in terms of classification accuracy, low-frequency category recognition ability and training stability, indicating that the adaptation ability of the model to multi-dimensional heterogeneous data of traditional Chinese medicine training has been significantly improved after feature screening, category weight adjustment and parameter optimization. This shows that after feature screening, category weight adjustment and parameter optimization, the adaptation ability of the model to multi-dimensional data of traditional Chinese medicine training has been significantly improved. This result is consistent with the basic conclusions of the existing educational data mining research. Compared with the single tree model, ensemble learning usually has higher robustness when dealing with complex teaching variables, nonlinear relationships and cross features. On the basis of random forest, the feature compression and class balance mechanism are further introduced, which is helpful to improve the recognition effect of low-frequency categories. The results of this paper show that the improved Random Forest improves the Recall index by 0.043 compared with the standard Random Forest, indicating that it has stronger recognition ability for the "to be improved" class samples. This is particularly important for TCM practical teaching, because the purpose of quality assessment is not only to distinguish excellent levels, but also to find students with operation deviation, weak differentiation and insufficient training investment as early as possible, so as to buy time for teaching intervention.

From the results of empirical analysis, the importance scores of the accuracy of syndrome differentiation judgment, the standard degree of manipulation, the proficiency degree of operation and the completeness of case analysis were 0.148, 0.136, 0.129 and 0.118, respectively, ranking in the forefront of all indicators. This shows that the core differences in the quality of TCM training are still mainly concentrated in two aspects of "syndrome differentiation" and "operation". The former determines whether students can organize the information of the four diagnoses into reasonable syndrome judgments, and the latter determines whether students have the ability to transform theoretical cognition into standardized actions. In contrast, although the process indicators such as attendance, classroom participation and frequency of self-directed practice also play a role, they are closer to the supporting conditions, and their effects are usually indirectly reflected through the core competence variables. It can be seen that the quality improvement of traditional Chinese medicine training should not stop at the general strengthening of teaching management, but should focus more pertinence on syndrome differentiation analysis training, case deduction and skill action feedback mechanism. At the same time, although the model in this paper has achieved good recognition results in the current sample, its results are still constrained by the data scale, sample source and label generation method. The evaluation of traditional Chinese medicine training itself has strong context. Different colleges, different course modules and different teachers 'scoring styles will have an impact on the sample distribution. In particular, although the teacher evaluation data has been entered into the model through coding and standardization, it still inevitably retains some subjectivity. This means that the improved Random Forest can improve the consistency and computational efficiency of evaluation, but it cannot completely replace high-quality teaching observation and professional judgment. Follow-up studies need to introduce cross-year, cross-course and cross-institution data in a larger sample range to further verify the generalization ability of the model.

6 Conclusions

With the gradual digitalization and refinement of traditional Chinese medicine education evaluation, how to realize objective identification and effective decision-making of traditional Chinese medicine training quality has become an important issue in teaching reform. Focusing on this practical demand, this paper constructs a quality evaluation decision model of traditional Chinese medicine training based on improved Random Forest, and carries out research from four aspects: evaluation index system design, model optimization strategy, algorithm performance comparison and empirical analysis. The results show that after introducing feature screening, class weight adjustment and parameter optimization into random forest, the adaptation ability of the model to multi-dimensional data of traditional Chinese medicine training is significantly enhanced, and the quality level division and key influencing factors identification can be better completed. In the comparison of algorithm performance, the Accuracy, Precision, Recall and F1 value of the improved Random Forest reach 0.923, 0.911, 0.896 and 0.903, respectively. It is better than Decision Tree, standard Random Forest and XGBoost. In the error convergence analysis, the verification error is reduced to 0.112 at the 80th iteration, which shows good stability and generalization ability. This shows that the optimized model not only improves the classification accuracy, but also improves the recognition effect of low-frequency quality level samples. For TCM training scenarios, this is particularly important, because the significance of quality evaluation is not only to distinguish between good and bad, but also to find the weak links in training in time. The empirical analysis further showed that the accuracy of syndrome differentiation judgment, the standard degree of manipulation, the proficiency degree of operation and the completeness of case analysis were the core indicators affecting the quality stratification of TCM training, and the importance scores were 0.148, 0.136, 0.129 and 0.118, respectively. It can be seen that the key to improving the quality of TCM training still lies in the training of dialectical thinking and the strengthening of operational ability, and the process input indicators play more supporting and regulating roles. This study shows that the computer modeling method can be used to integrate the information originally scattered in teaching assessment, classroom records and teacher evaluation into a decision-making basis with explanatory power, so as to provide data support for hierarchical guidance, process intervention and teaching quality improvement.

References

[1] Zheng B, Sun G, Wang H. Evaluation tool for traditional Chinese medicine students in China: a competency perspective[J]. Sage Open, 2019, 9(3): 2158244019861494.

[2] Qu F, Zhang Q, Dai M, et al. An evaluation survey of traditional Chinese medicine learning among international students majoring in conventional medicine: a study from a university in China[J]. BMC complementary medicine and therapies, 2021, 21(1): 16.

[3] Zhang Q, He Y J, Zhu Y H, et al. The evaluation of online course of Traditional Chinese Medicine for MBBS international students during the COVID-19 epidemic period[J]. Integrative Medicine Research, 2020, 9(3): 100449.

[4] Zeng J, Liang S, Zhang X, et al. Assessment of clinical competency among TCM medical students using standardized patients of traditional Chinese medicine: A 5-year prospective randomized study[J]. Integrative Medicine Research, 2022, 11(2): 100804.

[5] Huang M, Yang H, Guo J, et al. Faculty standardized patients versus traditional teaching method to improve clinical competence among traditional Chinese medicine students: a prospective randomized controlled trial[J]. BMC Medical Education, 2024, 24(1): 793.

[6] Kim Y J. The PBL teaching method in neurology education in the traditional Chinese medicine undergraduate students: An observational study[J]. Medicine, 2023, 102(39): e35143.

[7] Chen J, Loyeung B, Zaslawski C, et al. Comparison of traditional Chinese medicine education between mainland China and Australia—a case study[J]. Journal of Integrative Medicine, 2016, 14(4): 291-296.

[8] Zhou X, Yang Q, Bi L, et al. Integrating traditional apprenticeship and modern educational approaches in traditional Chinese medicine education[J]. Medical teacher, 2024, 46(6): 792-807.

[9] Xu L, Jiang Z, Han Y. Evaluating a National Traditional Chinese Medicine Examination via Cognitive Diagnostic Approaches[J]. Sustainability, 2024, 16(13): 5400.https://doi.org/10.3390/su16135400

[10] Yeh Y C, Chung U L. An investigation into competence in TCM of BSN graduates from technological universities in Taiwan[J]. Journal of Nursing Research, 2007, 15(4): 310-318.

[11] Snell L, Tallett S, Haist S, et al. A review of the evaluation of clinical teaching: new perspectives and challenges[J]. Medical education, 2000, 34(10): 862-870.

[12] Fluit C R M G, Bolhuis S, Grol R, et al. Assessing the quality of clinical teachers: a systematic review of content and quality of questionnaires for assessing clinical teachers[J]. Journal of general internal medicine, 2010, 25(12): 1337-1345.

[13] Ansari A A, Arekat M R, Salem A H. Validating the modified system for evaluation of teaching qualities: a teaching quality assessment instrument[J]. Advances in medical education and practice, 2018: 881-886.

[14] Vaižgėlienė E, Padaiga Ž, Rastenytė D, et al. Evaluation of clinical teaching quality in competency-based residency training in Lithuania[J]. Medicina, 2017, 53(5): 339-347.

[15] Shorey S, Lau T C, Lau S T, et al. Entrustable professional activities in health care education: a scoping review[J]. Medical Education, 2019, 53(8): 766-777.

[16] Schuwirth L W T, van der Vleuten C P M. A history of assessment in medical education[J]. Advances in Health Sciences Education, 2020, 25(5): 1045-1056.

[17] Kursa M B, Rudnicki W R. Feature selection with the Boruta package[J]. Journal of statistical software, 2010, 36: 1-13.

[18] Huynh-Cam T T, Chen L S, Le H. Using decision trees and random forest algorithms to predict and determine factors contributing to first-year university students’ learning performance[J]. Algorithms, 2021, 14(11): 318.

[19] Chen M, Liu Z. Predicting performance of students by optimizing tree components of random forest using genetic algorithm[J]. Heliyon, 2024, 10(12).

[20] Lu Q, Chai Y, Ren L, et al. Research on quality evaluation of innovation and entrepreneurship education for college students based on random forest algorithm and logistic regression model[J]. Peerj computer science, 2023, 9: e1329.