SIN-CHN SCIENTIFIC PRESS PTE.LTD.

Molecular & Cellular Biomechanics 2025, 22(4), 1696.
https://doi.org/10.62617/mcb1696

Article

Traditional Chinese medicine disease classification based on sparrow search algorithm

Xiaofo Li[1], Ningjie Ma[1], Hongxing Kan[1,*], Yongsheng Han[2]

[1] Anhui University of Chinese Medicine, Hefei 230000, China
[2] Neurology Research Institute, Anhui University of Chinese Medicine, Hefei 230008, China
*** Corresponding author:** Hongxing Kan, 984377701@qq.com

CITATION

Li X, Ma N, Kan H, Han Y. Traditional Chinese medicine disease classification based on sparrow search algorithm. Molecular & Cellular Biomechanics. 2025; 22(4): 1696.
https://doi.org/10.62617/mcb1696

ARTICLE INFO

Received: 27 February 2025
Accepted: 12 March 2025
Available online: 18 March 2025

Abstract: This study aims to explore the application of various basic machine learning algorithms in the task of TCM (Traditional Chinese Medicine) disease classification, and select the best performing model for optimization through comparative analysis. After experimental verification, the random forest model has excellent performance in various evaluation indexes, and its accuracy and recall are 65.1%, 65.1%, respectively, showing its comprehensive performance advantage in the classification of TCM disease types. In order to further improve the performance of the model, the sparrow search algorithm was introduced to optimize the random forest model. The performance of the optimized model on the test set is significantly improved, with an accuracy of 74.4%, a recall rate of 70.2%, an accuracy rate of 76.3%, and an $F1$ score of 73.1%. Compared with the random forest model before optimization, the accuracy of the optimized model increased by 9.3%, the recall rate increased by 0.51, the accuracy rate increased by 9.1%, and the $F1$ score increased by 8%. These results show that the sparrow search algorithm has a significant effect in optimizing the random forest model, and can effectively improve the performance of the model in the task of TCM disease classification. This study not only verified the applicability of random forest model in TCM disease classification, but also improved the classification effect of the model through the introduction of sparrow search algorithm.

Keywords: machine learning; TCM disease classification; random forest; sparrow search algorithm

1. Introduction

The diagnosis of TCM disease type is the core component of the TCM theoretical system. Its theoretical basis is derived from the Huangdi Neijing and other classic works, which emphasize the combined method of “looking, smelling, asking, and cutting” and combine the symptoms, signs and environmental factors of patients to conduct syndrome differentiation and treatment [1]. The core of TCM disease type diagnosis lies in “syndrome differentiation”, that is, patients are classified into specific syndrome types (such as qi deficiency, blood stasis, phlegm dampness, etc.) according to their overall performance, and personalized treatment plans are formulated accordingly [2]. However, the complexity and subjectivity of TCM diagnosis make its standardization and objectification a major challenge. In recent years, with the development of modern medical technology, research on TCM disease diagnosis has been gradually combined with emerging technologies such as artificial intelligence and big data, aiming to improve the accuracy and repeatability of diagnosis [3].

As a powerful data analysis tool, machine learning can extract rules from a large number of complex data and be used for prediction and classification tasks [4]. In the diagnosis of TCM disease types, machine learning algorithms can automatically identify and classify syndrome types by analyzing patients' multidimensional data such as symptoms, tongue, and pulse. For example, through deep learning technology, features can be extracted from tongue images to identify features such as color and thickness of tongue coating so as to assist in judging the syndrome type. In addition, machine learning can also be combined with electronic medical record data to mine the potential association between symptoms and syndrome types to provide data-driven support for TCM diagnosis [5].

In TCM disease diagnosis, commonly used machine learning algorithms include support vector machine (SVM), random forest (RF), neural network (NN), etc. [6]. These algorithms can be used to construct syndrome classification models, learn the features of different syndrome types through training data, and make predictions on new data. In addition, machine learning can also be used in the decision aid system of TCM diagnosis to help doctors quickly screen possible syndrome types and provide treatment recommendations. This paper attempts to use a variety of basic machine learning algorithms to classify TCM disease types and selects the algorithm with the best performance and uses the sparrow search algorithm for optimization [7].

2. Data set introduction

The data set in this paper is a private data set, which records the basic information of the patient, such as gender, age, age of onset, course of disease, education level, and family history, as well as various clinical characteristics of the patient. There are four categories of TCM disease types, namely syndrome of phlegm and blood stasis, syndrome of spleen deficiency, syndrome of dampness-heat, and syndrome of dampness-heat and syndrome of liver-stagnation and spleen deficiency. Select some data for display, as shown in **Table 1**.

Table 1. Presentation of some data sets.

Sex	Age	Age of onset	Course of disease	First diagnosis	Educational level	Preliminary judgment
Male	34	28	6	No	Undergraduate course	Phlegm stasis interjunction syndrome
Female	25	6	19	No	Junior high school	Phlegm stasis interjunction syndrome
Male	26	21	5	No	Junior high school	Dampness-heat syndrome of spleen deficiency
Female	31	12	9	No	Junior high school	Dampness-heat intrinsic syndrome
Male	14	12	2	No	Junior high school	Dampness-heat syndrome of spleen deficiency
Male	38	29	11	No	Undergraduate course	Dampness-heat intrinsic syndrome
Male	44	27	17	No	Undergraduate course	Dampness-heat intrinsic syndrome

For the dataset part, the study was conducted on 400 patients with Wilson's disease (WD) who were hospitalized between January 2023 and September 2023 at the Divine Research Institute of Anhui University of Traditional Chinese Medicine. Among them, 245 cases were male and 155 cases were female. Ages 10–58 years old, age of onset 6 months–54 years old, shortest duration of the disease was 3 days and the longest was 35 years, clinical typing included hepatic, cerebral, and hepatic-cerebral types, of which the hepatic type accounted for 52.75% of the cases with 211 patients, the hepatic-cerebral type accounted for 35.75% of the cases with 143 patients, and the cerebral type accounted for 11.5% of the cases with 46 patients. There were 309 cases without a family history and 91 cases with a family history.

The 400 patients were identified and typed into 38 types, including 78 cases of Damp-Heat Internalization, 55 cases of Spleen Deficiency Damp-Heat, 31 cases of Liver Depression and Spleen Deficiency, 31 cases of Spleen Deficiency Dampness, 30 cases of Phlegm and Stasis Conjugation, and the rest of the 29 types of cases were fewer in number, totaling 109 cases.

3. Method

3.1. Decision tree

A decision tree is a kind of supervised learning algorithm based on a tree structure, which realizes a classification or regression task by recursively partitioning data. Its core idea is to construct decision rules by segmenting feature space layer by layer. The algorithm starts from the root node, selects the optimal feature as the basis for splitting, divides the data set into mutually exclusive subsets, and generates branch nodes. The process is repeated until the stopping conditions are met (such as the purity of node samples is up to standard, the depth limit, or the number of samples is too small), and the final leaf node represents the prediction result. Key steps include feature selection, node splitting, and pruning optimization: Feature selection evaluated feature differentiation ability by calculating information gain (ID3 algorithm), gain rate (C4.5 algorithm), or Gini coefficient (CART algorithm) and selected features that maximized the purity of child nodes for splitting. Information gain measures features' ability to eliminate uncertainty based on information entropy reduction. The gain rate penalizes the deviation of multi-valued features by introducing split information, and the Gini coefficient is efficiently calculated by calculating the random misclassification probability. After the node splits, the algorithm recursively performs the same operation on the child nodes to form a tree structure. To avoid overfitting, the model complexity should be optimized by pre-pruning (limiting the tree depth/minimum number of nodes in advance) or post-pruning (eliminating redundant branches from bottom up after generating a complete tree) [8].

3.2. Random forest

Random forest is a decision tree combination algorithm based on Bagging (Bootstrap Aggregating) integrated learning framework. Its core principle is to improve the generalization ability and stability of the model by building multiple decision trees and voting or averaging them. Specifically, the algorithm first generates multiple differentiated training subsets from the original data set through Bootstrap sampling (random sampling with put back) to ensure the diversity of training data for each tree [9]. In the process of single tree construction, feature randomness is further introduced, that is, when each node splits, only the optimal split feature (usually the square root or logarithm of the total number of features) is selected from the randomly selected feature subset, so as to reduce the correlation between trees and enhance the anti-overfitting ability of the overall model. After the independent training of all decision trees, the classification task summarized the prediction results of each tree by the majority voting method, and the regression task used the mean value method to output the final predicted value. This "double randomness" (data perturbation and feature perturbation) combined with the "group decision" mechanism enables the random forest to effectively balance bias and variance, alleviate the problem that a single decision tree is easy to overfit and sensitive to noise, while retaining the intuitiveness and efficiency of the decision tree. [10]. The model structure of the random forest algorithm is shown in **Figure 1**.

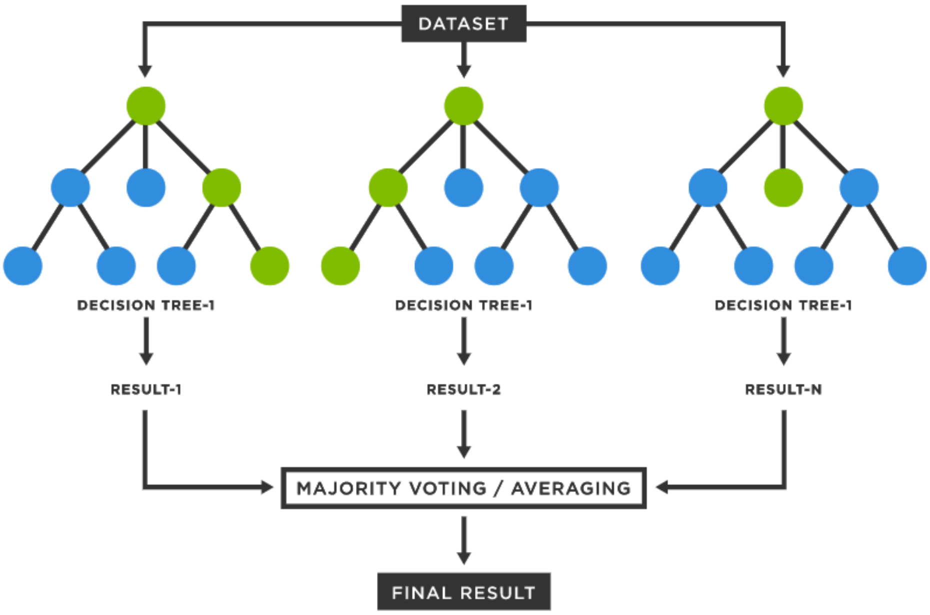

Figure 1. The model structure of the random forest algorithm.

3.3. CatBoost

CatBoost is an efficient machine learning algorithm based on gradient Boosting decision tree (GBDT), specially designed to handle category-type features. Its core innovation lies in effectively solving gradient bias and overfitting problems by "Ordered Boosting" and "symmetric tree structure". The algorithm uses ordered target coding to calculate statistics (such as the mean value of target variables) through a dynamic time series or random sequence of samples when converting the class features to a numerical type, avoiding overfitting caused by data leakage in traditional target coding [11]. In gradient lifting iteration, ordered lifting ensures that the training of each new tree only uses the preorder sample information by calculating the residuals one by one in sample order and updating the model, eliminating the prediction deviation caused by calculating the gradient and updating the model at the same time, and improving the generalization ability. In addition, CatBoost uses the Oblivious Trees structure, each layer of nodes uses the same splitting rules, simplifying model complexity and speeding up prediction, while combining features to automatically generate higher-order interactive features, enhancing expression [12]. The algorithm has built-in efficient regularization (such as L2 regularization and feature sampling) and GPU accelerated optimization, which can automatically process missing values without complicated pre-processing.

3.4. Support vector machine

Support vector machine (SVM) is a supervised learning algorithm based on statistical learning theory. Its core idea is to achieve classification or regression with high generalization ability by finding the maximum spaced hyperplane (the decision boundary that can maximize the spacing of different classes of data in a classification task). For linearly separable data, SVM constructs hard-spaced hyperplanes directly. When the data is linear and not separable, the input space is mapped to the high-dimensional feature space through the Kernel Trick, so that the data is linearly separable in the high dimension (commonly used kernel functions include linear, polynomial, Gaussian kernel, etc.), and the soft interval mechanism is introduced (allowing some samples to violate the interval constraint, The regularization parameter C balances the classification error and model complexity to deal with noise or overlapping distribution; the optimization goal of the algorithm is transformed into solving convex quadratic programming problems, and the model is finally determined by only a few "support vectors" (boundary samples close to the hyperplane), which greatly reduces the computational complexity. SVM has excellent performance in dealing with small samples, high-dimensional data, and nonlinear problems and has strong theoretical protection and anti-overfitting ability, but the computational efficiency decreases significantly with the increase of data scale. Extended forms such as support vector regression (SVR) perform the regression task with the ϵ-insensitive loss function [12].

3.5. XGBoost

XGBoost (Extreme Gradient Boosting) is an efficient machine learning algorithm based on a gradient lifting framework, which achieves high-precision prediction by integrating multiple weak decision trees and optimizing the objective function [13]. The core principle lies in regularization promotion and second-order derivative optimization: on the basis of traditional gradient promotion (using a step degree to fit the residual), the algorithm introduces regularization terms (L1/L2 regularization and tree structure complexity penalties, such as the number of leaf nodes and weight square), constructs an objective function containing loss functions and regularization terms, and effectively controls the model complexity to prevent overfitting. At the same time, the second-order Taylor expansion is used to approximate the loss function, and the first-order degree and second-order Hessian matrix information are used to update the model parameters more accurately, accelerate the convergence, and improve the optimization stability. In the process of tree generation, XGBoost finds the optimal split point by greedy algorithm, selects the feature and segmentation threshold based on the structure fraction gain (considering the loss reduction and regularization penalty after splitting), and supports the parallel feature preordering and site segmentation approximation algorithm to accelerate the calculation. In addition, the algorithm has a built-in sparse perception strategy to automatically process missing values and processes large-scale data by weighted quantile sketch optimization. The generalization ability is further enhanced by combining column sampling and row sampling [14].

3.6. Sparrow search algorithm

The Sparrow Search algorithm (SSA) is a swarm-intelligent optimization algorithm based on sparrow's foraging behavior, which simulates sparrow's searching, tracking, and escaping behavior during foraging. The algorithm gradually optimizes the solution in the search space through information sharing and cooperation among sparrow individuals. In the search process, sparrows are divided into discoverers and followers [15]. The algorithm is simple and efficient and is suitable for solving complex optimization problems [16]. The schematic diagram of the sparrow search algorithm is shown in **Figure 2**.

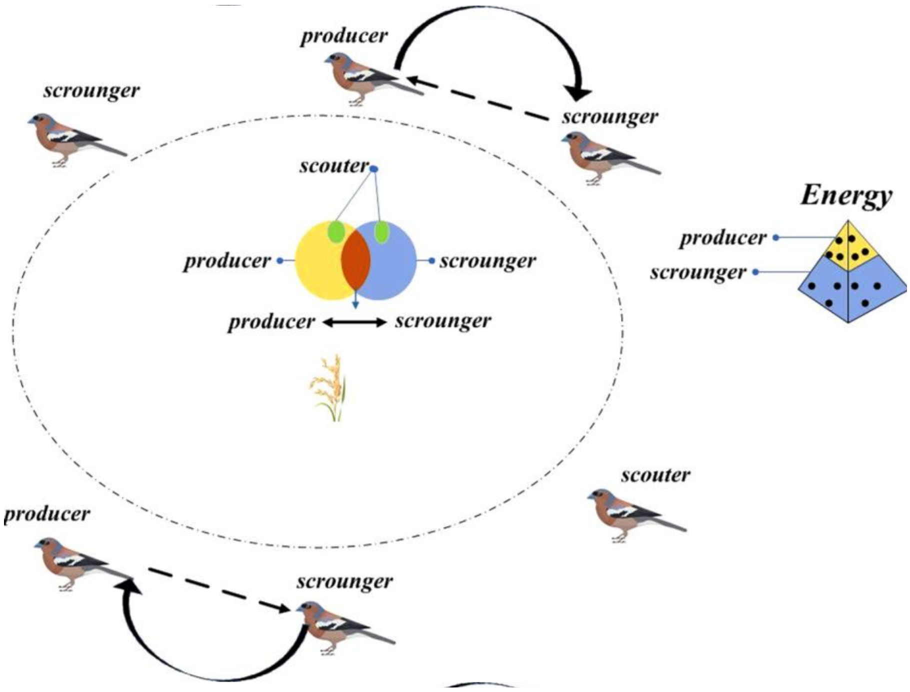

Figure 2. The schematic diagram of the sparrow search algorithm.

3.7. Optimization of random forest based on sparrow search algorithm

The principle of the sparrow search algorithm (SSA) optimization of random forest is to dynamically adjust the hyperparameters (such as the number of trees, the maximum depth, and the minimum number of node splitting samples) or the feature subset of the random forest by simulating the foraging and warning behavior of the sparrow population so as to improve the model performance. SSA regards each sparrow as a set of candidate parameter combinations and uses the discoverer-follower mechanism to explore in the solution space: the finder (high-quality solution) leads the search direction, the follower develops local optima around it, and the alert (random individual) avoids falling into local optima. The fitness function evaluates parameter quality based on random forest cross-validation accuracy or AUC (Area Under the Curve) and gradually approaches the optimal solution by iteratively updating the sparrow position (parameter adjustment). By balancing global and local search, the optimization process can automatically find random forest configurations with strong generalization ability and moderate complexity, overcome the inefficiency of traditional grid search, and inhibit overfitting, significantly improving the prediction accuracy and stability of classification/regression tasks [17].

4. Result

4.1. Machine learning result

Firstly, the decision tree, random forest, CatBoost, SVM, and XGBoost algorithms were used to classify TCM disease types, respectively. The software used in this experiment was MATLAB R2022b with 32 GB of memory. In the division of data sets, 80% of the data were randomly selected as the training set and the remaining 20% as the test set. The classification results of each model test set are output, as shown in **Table 2**.

Table 2. The predictions of various machine learning algorithms.

Model	Accuracy	Recall	Precision	***F*1**
Decision tree	0.442	0.442	0.521	0.463
Random forest	0.651	0.651	0.672	0.651
CatBoost	0.558	0.558	0.561	0.5
SVM	0.535	0.535	0.545	0.534
XGBoost	0.581	0.581	0.711	0.517

It can be seen from the above experimental results that there are significant differences in the performance of different machine learning models in the four classification tasks of TCM disease types. The random forest model has the best performance in accuracy, recall, and F1 score, which are 0.651, 0.651, and 0.672, respectively, indicating that it has a good comprehensive performance in the classification task. The XGBoost model performed particularly well in accuracy, reaching 0.711, but its F1 score was relatively low (0.517), suggesting a trade-off between recall and accuracy. CatBoost and SVM models performed relatively moderately, with accuracy and recall rates ranging from 0.535 to 0.558, and F1 scores were also relatively close, at 0.5 and 0.534, respectively. The decision tree model has the worst performance; all indexes are lower than other models, especially the *F*1 score, which is only 0.463, indicating that its comprehensive performance in the classification task is weak [18].

Overall, the stochastic forest model is the most robust in this classification task, possibly because its ensemble learning method can effectively reduce overfitting and improve generalization ability. XGBoost, while excellent in accuracy, does not perform as well as Random Forest in recall and F1 scores, and may need to be further tuned to balance accuracy and recall. The performance of CatBoost and SVM models is close, but there is still room for improvement. The performance of decision tree models is poor and may need to be improved by adjusting parameters or combining other methods [19].

4.2. The result of our model

We choose the random forest algorithm model with a good prediction effect to optimize and use the sparrow search algorithm to improve the random forest model. When optimizing random forest (RF), SSA is mainly used to adjust the hyperparameters of random forest (number of trees, maximum depth, minimum sample split number, etc.) to improve the performance of the model [20]. The principle is to search the optimal parameter combination by updating the position of individual sparrows and to approach the optimal solution step by step by using the sparrow's local search and global exploration ability. Specifically, SSA evaluates the predictive performance of the random forest (e.g., accuracy, $F1$ score, etc.) by iteratively updating the sparrow's position (i.e., parameter combination) and retains the optimal parameter combination to finally find the hyperparameter configuration that optimizes the performance of the random forest [21].

First, output the confusion matrix predicted by the model test set, as shown in **Figure 3**. According to the confusion matrix of the test set, the prediction accuracy of the random forest model optimized by the Sparrow search algorithm in TCM-type diagnosis is 0.744, the recall rate is 0.702, the accuracy rate is 0.763, and f1 is 0.731.

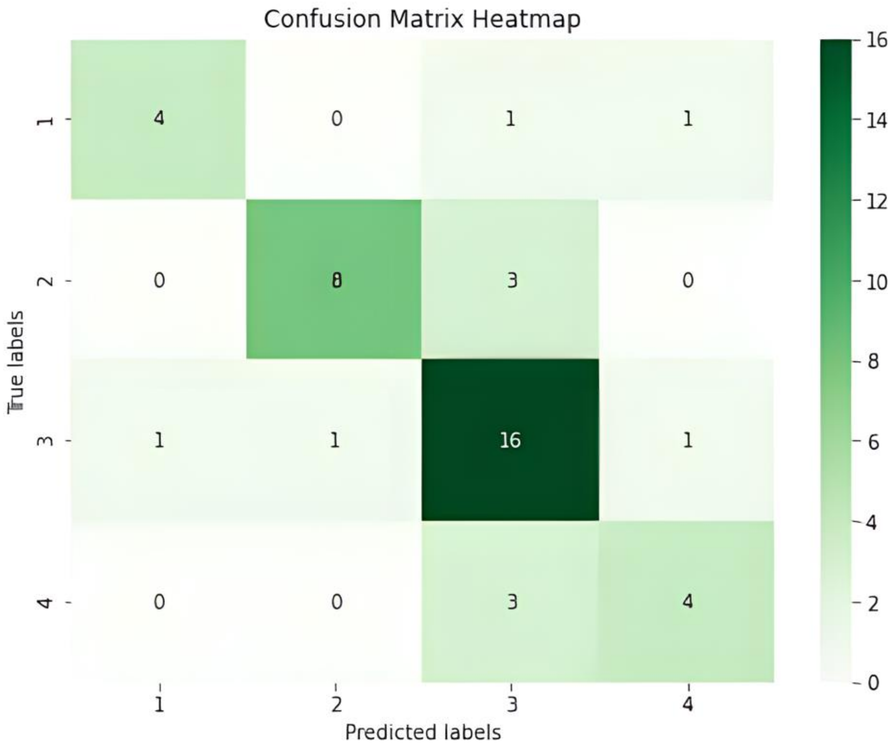

Figure 3. The confusion matrix.

The prediction effect of the randomized forest optimized by the Sparrow search algorithm was compared with that of the random forest model, as shown in **Table 3**. The comparison diagram of the prediction effect of the random forest optimized by the random forest and Sparrow search algorithm is output, as shown in **Figure 4**.

Table 3. Evaluation indicators of our model.

Model	**Accuracy**	**Recall**	**Precision**	***F*1**
Decision tree	0.442	0.442	0.521	0.463
Random forest	0.651	0.651	0.672	0.651
CatBoost	0.558	0.558	0.561	0.5
SVM	0.535	0.535	0.545	0.534
XGBoost	0.581	0.581	0.711	0.517
Our model	0.744	0.702	0.763	0.731

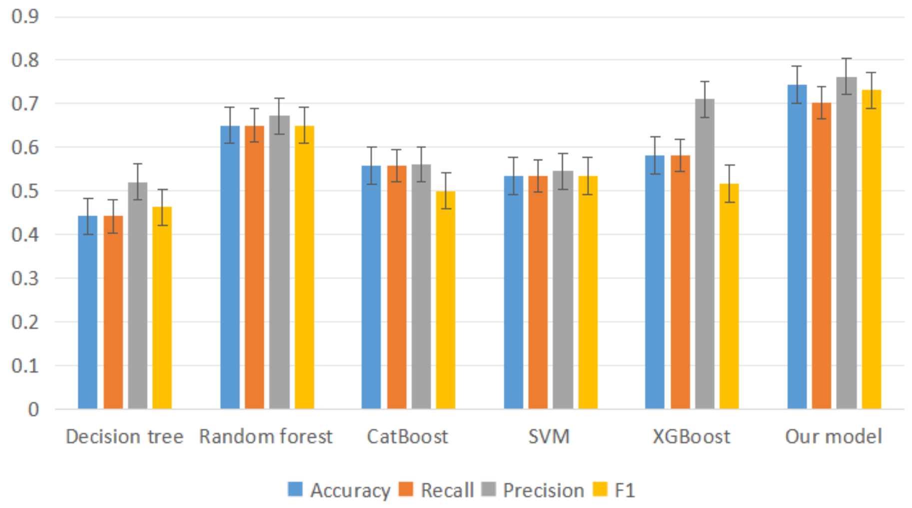

Figure 4. The performance of each model on different indicators.

Our model outperformed Random Forest by 9.3% in prediction accuracy, 5.1% in recall, 91% in accuracy, and 80% in f1. The model proposed in this paper can classify TCM disease more effectively.

5. Conclusion

This study systematically evaluated the application effectiveness of various classical machine learning algorithms in TCM disease classification tasks, innovatively introduced intelligent optimization algorithms to optimize the parameters of the optimal model, and finally built a TCM syndrome differentiation classification model with clinical application potential. In the experimental design stage, this study first compares the basic algorithms such as support vector machine, decision tree, K-nearest neighbor, naive Bayes, and random forest. The experimental results show that the random forest model shows the best comprehensive classification performance in the initial test, and its accuracy rate, recall rate, accuracy rate, and *F*1 score reach 0.651, 0.651, 0.672, and 0.651, respectively, which is about 15% higher than other algorithms on average. This finding verified that the ensemble learning method has a stronger feature-capturing ability and generalization performance when dealing with complex TCM syndromes.

In order to further improve the performance of the model, the sparrow search algorithm was creatively introduced to optimize the random forest hyperparameters. By simulating the foraging behavior and anti-predation strategy of the sparrow population, the intelligent optimization algorithm effectively balances the relationship between global exploration and local exploitation and successfully breaks through the local optimal limit of traditional grid search. The optimized SSA-RF model showed significantly improved classification performance on the test set: accuracy increased to 0.744 (+9.3%), recall rate increased to 0.702 (+5.1%), accuracy increased to 0.763 (+9.1%), and F1 score increased to 0.731 (+8.0%). In particular, the confusion matrix analysis showed that the model improved the accuracy of distinguishing basic TCM syndromes, such as cold heat and deficiency, most obviously, which is of great value for the objectification of TCM eight-line syndrome differentiation.

The results of this study have important practical significance for promoting the modernization of traditional Chinese medicine. The constructed intelligent syndrome differentiation system can not only be used as a clinical auxiliary diagnostic tool to provide reliable reference for TCM physicians, but more importantly, through big data mining technology, it can discover the evolution law of the syndrome hidden in the traditional syndrome differentiation system, providing a new research path for TCM theoretical innovation.

6. Discuss

By comparing a variety of classical machine learning algorithms, this study verified the significant advantages of random forest in the classification of TCM syndromes. Its integrated learning mechanism can effectively capture the complex non-linear feature interaction in TCM syndrome differentiation, and its index performance is improved by 15% on average compared with the traditional single model. It reflects the characteristics of high dimension and multi-association of TCM syndrome data, which is more suitable to be represented by the group decision model of the decision tree. The hyperparameter optimization strategy based on the Sparrow search algorithm successfully breaks through the local optimal limit of traditional grid search and realizes the dynamic balance between global exploration and local development through the bio-intelligent bionic mechanism, which improves the key indicators of the model by 8%–9%, confirming the value of the intelligent optimization algorithm in improving the clinical usability of the TCM syndrome differentiation model. This technical path of "integrated learning + intelligent optimization" not only effectively alleviates the overfitting risk caused by the small sample size and high noise of TCM data but also provides an interpretable basis for the quantitative extraction of key syndrome differentiation elements through feature importance analysis and provides a methodological framework with both algorithmic innovation and clinical adaptation for objectified research on TCM syndrome differentiation. In the future, multi-modal data fusion and dynamic syndrome differentiation modeling can be further explored to promote the actual landing of the TCM intelligent diagnosis and treatment system.

Author contributions: Conceptualization, XL; methodology, XL; validation, XX and NM; formal analysis, XL; investigation, XL and NM; resources, XL; data curation, XL and NM; writing—original draft preparation, XL; writing—review and editing, XL, NM and HK; visualization, XL; supervision, HK and YH; project administration,

HK and YH; funding acquisition, HK and YH. All authors have read and agreed to the published version of the manuscript.

Ethical approval: Not applicable.

Conflict of interest: The authors declare no conflict of interest.

References

1. Chen Z, Zhang D, Liu C, et al. Traditional Chinese medicine diagnostic prediction model for holistic syndrome differentiation based on deep learning. Integrative Medicine Research. 2024; 13(1): 101019. doi: 10.1016/j.imr.2023.101019
2. Hu H, Cheng C, Ye Q, et al. Enhancing traditional Chinese medicine diagnostics: Integrating ontological knowledge for multi-label symptom entity classification. Mathematical Biosciences and Engineering. 2023; 21(1): 369-391. doi: 10.3934/mbe.2024017
3. Gong A, Guo L, Yu Y, et al. Spectrum-image dual-modality fusion empowered accurate and efficient classification system for traditional Chinese medicine. Information Fusion. 2024; 101: 101981. doi: 10.1016/j.inffus.2023.101981
4. Tian D, Chen W, Xu D, et al. A review of traditional Chinese medicine diagnosis using machine learning: Inspection, auscultation-olfaction, inquiry, and palpation. Computers in Biology and Medicine. 2024; 170: 108074. doi: 10.1016/j.compbiomed.2024.108074
5. Jiang Q, Huang W, Liang J, et al. A novel intelligent model for visualized inference of medical diagnosis: A case of TCM. Artificial Intelligence in Medicine. 2024; 149: 102799. doi: 10.1016/j.artmed.2024.102799
6. Ni T, Liu Y, Huang M, et al. Association Between Anemia Status and the Risk of Different Types of Heart Failure: A RCSCD-TCM Study in China. Angiology. 2023; 75(2): 190-196. doi: 10.1177/00033197231161908
7. Tao QF, Huang YB, Yuan L, et al. Acupuncture versus tricyclic antidepressants in the prophylactic treatment of tension-type headaches: an indirect treatment comparison meta-analysis. The Journal of Headache and Pain. 2024; 25(1). doi: 10.1186/s10194-024-01776-5
8. Chen H, Wang X, Zhang J, et al. Exploration of TCM syndrome types of the material basis and risk prediction of Wilson disease liver fibrosis based on 1H NMR metabolomics. Journal of Pharmaceutical and Biomedical Analysis. 2024; 245: 116167. doi: 10.1016/j.jpba.2024.116167
9. Singh A, Dhau J, Kumar R, et al. Tailored carbon materials (TCM) for enhancing photocatalytic degradation of polyaromatic hydrocarbons. Progress in Materials Science. 2024; 144: 101289. doi: 10.1016/j.pmatsci.2024.101289
10. Wang Y, Wang DD, Pucka AQ, et al. Differential clinical characteristics across traditional Chinese medicine (TCM) Syndromes in patients with sickle cell disease. Frontiers in Pain Research. 2024; 4. doi: 10.3389/fpain.2023.1233293
11. Asghar MZ, Lajis A, Alam MM, et al. A Deep Neural Network Model for the Detection and Classification of Emotions from Textual Content. Ahmad M, ed. Complexity. 2022; 2022(1). doi: 10.1155/2022/8221121
12. Cahyani DE, Patasik I. Performance comparison of tf-idf and word2vec models for emotion text classification. Bulletin of Electrical Engineering and Informatics. 2021; 10(5): 2780-2788. doi: 10.11591/eei.v10i5.3157
13. Deng J, Ren F. A Survey of Textual Emotion Recognition and Its Challenges. IEEE Transactions on Affective Computing. 2023; 14(1): 49-67. doi: 10.1109/taffc.2021.3053275
14. Gupta V, Jain N, Katariya P, et al. An Emotion Care Model using Multimodal Textual Analysis on COVID-19. Chaos, Solitons & Fractals. 2021; 144: 110708. doi: 10.1016/j.chaos.2021.110708
15. Liu Y, Fu G. Emotion recognition by deeply learned multi-channel textual and EEG features. Future Generation Computer Systems. 2021; 119: 1-6. doi: 10.1016/j.future.2021.01.010
16. Zhou X, Dong X, Li C, et al. TCM-FTP: Fine-Tuning Large Language Models for Herbal Prescription Prediction. In: Proceedings of the 2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); 2024.
17. Wang Y, Wang Z, Tang B, et al. Discussion on the cultivation mode of professional postgraduate students of traditional Chinese medicine under the orientation of inheritance ability. Jiangsu Health Career Management. 2025; 36(1): 130-132+139.
18. Xu D, Zhou H, Quan W, et al. A new method proposed for realizing human gait pattern recognition: Inspirations for the application of sports and clinical gait analysis. Gait & Posture. 2024; 107: 293-305. doi: 10.1016/j.gaitpost.2023.10.019

19. Zhang M, Otsuki K, Li W. Molecular networking as a natural products discovery strategy. Acta Materia Medica. 2023; 2(2). doi: 10.15212/amm-2023-0007
20. Liu WC, Li MP, Huang HY, et al. Research trends of machine learning in traditional medicine: a big-data based ten-year bibliometric analysis. Traditional Medicine Research. 2023; 8(7): 37. doi: 10.53388/tmr20221113001
21. Li JT, Gu A, Tang NN, et al. Exploring anti-tumor potential of food and medicine homology substances: An in-silico evaluation of Citri Grandis Exocarpium against gallbladder cancer. Food & Medicine Homology; 2025.