Implementation of XGBoost Algorithm in Zang Organs Syndrome Diagnosis and Therapy Recommendations

Yosefina Finsensia Riti[1*], Andre Hartanto[2], Onny Priskila[3], Benaya Azareel Oentoro[4]

[1,2,4]Informatics Department, Faculty of Engineering, Universitas Katolik Darma Cendika, Surabaya, East Java, Indonesia
[3]Acupuncture and Herbal Medicine Department, Faculty of Engineering, Universitas Katolik Darma Cendika, Surabaya, East Java, Indonesia
E-mail: [1*]yosefina.riti@ukdc.ac.id, [2]andre.hartanto@ukdc.ac.id, [3]onny.priskila@ukdc.ac.id, [4]benaya.oentoro@student.ukdc.ac.id

(Received: 2 Dec 2025, revised: 15 Dec 2025, accepted: 17 Dec 2025)

Abstract

In Traditional Chinese Medicine (TCM), syndrome is primarily categorized into Zang and Fu organ syndromes. Zang organs include the heart, liver, spleen, lungs, and kidneys, and diagnosis is derived from patient complaints, symptoms, and examination findings. However, Zang organ syndrome determination is highly dependent on clinical experience, resulting in diagnostic subjectivity and variability among practitioners. In addition, symptom overlap across different organs further complicates the diagnostic process and challenges standardization. To address these issues, this study proposes an artificial intelligence (AI)-based intelligent system to assist Zang organ syndrome diagnosis and provide acupuncture therapy recommendations. The system applies machine learning to improve diagnostic accuracy and consistency, utilizing the Extreme Gradient Boosting (XGBoost) algorithm due to its effectiveness in classifying complex and nonlinear patterns commonly found in TCM diagnostics. The dataset was constructed based on standardized TCM references and expert-defined symptom–syndrome relationships, consisting of 39 symptom indicators, 72 Zang organ syndrome classes, and 15,532 data instances. To address class imbalance, the Synthetic Minority Oversampling Technique (SMOTE) was applied during training. Experimental results show that the proposed model achieved an accuracy of 98.6% and a macro-average F1-score of 83% for both syndrome diagnosis and therapy recommendation tasks. The results demonstrate that the proposed AI approach effectively supports Zang organ syndrome identification and acupuncture point recommendation, contributing to improved diagnostic efficiency and decision support in TCM practice.

Keywords: Syndrome, Zang Organ, Machine Learning, XGBoost

I. INTRODUCTION

Traditional medicine is a type of treatment widely used by people for healthcare, wellness, prevention, and treatment of various diseases. One treatment that is currently developing in Indonesia is TCM, better known as Traditional Chinese Medicine. TCM is a healthcare system originating from China and plays a vital role in the prevention and treatment of diseases in traditional medicine worldwide [1]. TCM is a form of complementary and alternative medicine that has been proven effective in preventing and treating various diseases [2]. In its development, TCM and conventional medicine can be combined to produce complementary treatment effects [3]. TCM treatment techniques include acupuncture, herbs, *tuina* (massage therapy), acupressure, *qigong* (breathing exercises), and diet therapy [4], where each treatment technique is adjusted based on the results of the diagnosis of various TCM factors.

In TCM, body organs are divided into two concepts, which are *Zang* and *Fu*. The *Zang* organs are the heart, liver, spleen, lungs, and kidneys, while the *Fu* organs are the small intestine, large intestine, bladder, gallbladder, and stomach. The *Zang* and *Fu* organs not only describe physiological functions, but also emotional and energy/*Qi* functions which influence each other. Hindrance in one of these organs will cause a series of symptoms in the human body [5]. In TCM treatment, diagnostic treatment is based on four examination methods: *wang* (observation), *wen* (hearing and smell), *wun* (anamnesis), and *qie* (palpation). The results of the diagnosis are called syndromes and are based on the *Zang* and *Fu* organs. The difficulty faced by practitioners in TCM diagnosis is

DOI: 10.34148/teknika.v15i1.1430

CC BY SA

determining the syndrome. Determining the syndrome requires clinical experience, so diagnostic results can vary between practitioners. TCM diagnosis requires a thorough analysis of the patient's overall condition, and the validity of a TCM diagnosis can vary greatly depending on the practitioner's experience [6].

TCM syndrome diagnosis methods play a crucial role in the selection of treatment methods and the expected outcomes. When dealing with the syndrome diagnosis of various diseases, TCM practitioners typically rely on individual subjective experiences, which poses challenges for standardization of TCM syndromes [7]. Some syndromes sometimes overlap across different organs, posing unique challenges for practitioners [7]. Therefore, an intelligent system capable of processing data on complaints, symptoms, and examination findings is needed to assist TCM practitioners in determining syndrome patterns more quickly. This system is expected to be a tool that can combine TCM treatment with modern technology, thereby improving the quality of care and therapeutic outcomes for patients.

Studies related to the implementation of intelligent systems in the diagnosis of TCM syndromes have been conducted in previous studies, including research for the classification of several TCM syndromes by comparing five deep learning models, namely BERT, TextCNN, LSTM RNN, LSTM ATTENTION, and TCM-BERT-CNN, where the TCM-BERT-CNN model is the best model with a precision of 92.6% [7]. The classification of liver cancer syndromes uses a machine learning algorithm, namely a hybrid of ELM and PSO with an accuracy of 87.5% [8]. Another study combined Bayesian Optimization and XGBoost (Extreme Gradient Boosting), abbreviated as BO-XGBoost for the prediction of metabolic syndrome (Mets) with good results among other models [9].

However, most existing studies focus on single diseases or limited syndrome categories and are primarily restricted to diagnostic prediction without providing integrated therapy recommendations. In addition, detailed evaluation of diagnostic errors and the impact of class imbalance are rarely discussed. These limitations highlight the need for a comprehensive and standardized AI-based approach that simultaneously supports Zang organ syndrome diagnosis and therapy recommendation. Based on these considerations, this study differs from previous works by addressing comprehensive Zang organ syndromes and providing therapy recommendations in the form of appropriate acupuncture points.

The aim of this study is to develop an intelligent system (AI) to identify *Zang* syndrome based on patient complaints input into the system and provide treatment recommendations in the form of appropriate acupuncture points. The application of AI, especially machine learning, is carried out to improve the accuracy of diagnosis and therapy recommendation for the treatment, as a form of acupuncture points. The machine learning algorithm proposed in this study is XGBoost because of its advantages in having high accuracy in classifying complex patterns found in traditional medicine diagnostics, good processing speed, as well as good accuracy [9], [10]. In addition, several cases of disease classification using the XGBoost algorithm produced good classification accuracy compared to other algorithms, such as breast cancer classification [11], [12], prostate cancer classification [13], and thyroid [14]. The results of this study can be used by therapists or acupuncture practitioners and can also be used by students who are studying in the Acupuncture and Herbal Medicine study program, with the aim of being a reference for diagnosis and speed of diagnosis.

II. RESEARCH METHODOLOGY

The aim of this study is to determine *Zang* organ syndromes and acupuncture therapy recommendations using the XGBoost algorithm. The research methodology consists of several stages, which are data collection, data preprocessing, model creation, and model evaluation.

A. Dataset Sources and Collection

The dataset used in this study was constructed from standardized Traditional Chinese Medicine (TCM) reference materials rather than direct clinical patient records. Specifically, the symptom database was derived from The Foundations of Chinese Medicine: A Comprehensive Text, a World Health Organization (WHO)-recognized reference that systematically describes symptom patterns associated with organ syndromes, particularly Zang organ syndromes. The dataset was compiled and curated by an academic expert, namely a lecturer from the Acupuncture and Herbal Medicine Study Program at Universitas Katolik Darma Cendika (UKDC), who has expertise in TCM diagnostics and acupuncture practice. This process involved summarizing, organizing, and structuring the symptom–syndrome–therapy relationships documented in the reference text into a format suitable for machine learning analysis. Although the dataset was literature-based and expert-curated rather than derived from real patient records, it provides a standardized representation of TCM diagnostic knowledge that is suitable for initial AI model development and evaluation

The symptom indicators in the dataset are based on the four fundamental TCM examination methods: inquiry or anamnesis, observation, palpation, and listening and smelling. The reference text provides comprehensive descriptions of syndrome-specific symptoms, including core (focal) symptoms that are essential for diagnosis and supporting symptoms that may vary between cases. For example, lung Qi deficiency syndrome is characterized by symptoms such as pale complexion, weak voice, reluctance to speak, shortness of breath, mild cough, spontaneous sweating, pale tongue, tongue fissures, and an empty pulse. In clinical practice, the presence of core symptoms—such as shortness of breath, mild cough, and an empty pulse—is sufficient to support the diagnosis, whereas the absence of these key indicators precludes syndrome confirmation.

This standardized data serves as the basis for an artificial intelligence (AI) system to assist therapists in establishing a syndrome diagnosis. Standardization minimizes variation in

interpretation between therapists, resulting in more consistent and objective diagnostic results, which can be used as a basis for TCM syndrome identification applications.

The data collected were then classified based on the syndrome category based on the four examination methods. The total data classification was 39 data indicators, which are awareness, expression, facial color, body condition, body shape, head condition, skin, hair, eye complaints, ears, mouth, tongue muscle color, tongue muscle thickness, specific conditions of the tongue, tongue membrane color, tongue membrane thickness, tongue membrane condition, defecation, urination, hot/cold conditions, sweat, eating/drinking habits, sleep, pulse, special organ conditions (lungs, large intestine, stomach, spleen, heart, small intestine, bladder, kidney, pericardium, *san jiao*, gallbladder, liver), special conditions for women and children, and blood pressure.

Based on these indicators, a syndrome database consisting of 72 Zang organ syndromes was constructed, comprising 16 lung syndromes, 12 spleen syndromes, 15 heart syndromes, 11 kidney syndromes, and 18 liver syndromes. From this syndrome formulation, a total of 15,532 data instances were generated. Corresponding therapeutic principles and treatment recommendations were then developed for each syndrome. As a result, the dataset includes 72 therapy principles and 72 acupuncture point recommendations aligned with the identified Zang organ syndromes.

B. Pre-processing

After the data is collected and stored in Excel format, the next step is data preprocessing so that the data is ready to be processed, analyzed and used for model training. Preprocessing includes removing irrelevant/unnecessary columns (for example, column numbers), handling missing values, changing data formats such as converting all text data to lowercase, correcting typos, merging similar data classes, and simplifying excessively long text data. Then, the data is label encoded to convert categorical data into numbers.

In model training, the training data from the dataset will be oversampled using the SMOTE (Synthetic Minority Oversampling Technique) method to handle imbalance of target syndrome classes and therapy recommendations in the dataset, as well as preventing overfitting in models using imbalanced datasets. After oversampling the training data with SMOTE, the data is normalized using standard scaling.

C. XGBoost Algorithm Modeling

XGBoost is an ensemble learning method based on the gradient boosting framework, proposed by Tianqi Chen et al. in 2016. This algorithm has demonstrated good performance and wide application in machine learning, such as classification, regression, and ranking [15]. The main idea of XGBoost is to train the model using residuals, where the error is reduced by using the output of the most recent training tree as input for the next iteration, performed over several consecutive iterations [16]. XGBoost combines several weak classifiers to create a more robust model, capable of handling complex data [13]. In the case of *Zang* organ syndrome classification in this study, the XGBoost model was able to identify patterns between symptoms, syndromes and therapy recommendations effectively.

The XGBoost model is enhanced based on gradient boosting trees so that its objective function contains the loss function and model complexity. By minimizing the objective function, the optimal decision tree at a given time is generated iteratively using the Greedy algorithm. When generating the first K trees, the model learns the difference between the actual value and the previous $K-1$, and the model will learn the residual between the previous value and the previous predicted value, so that the model predicts gradually closer to the actual value [11].

When the XGBoost model has K tree with samples x_i, the predicted result $\hat{y}_i$ is as shown in Equation 1.

$$\hat{y}_i = F_K(x_i) = F_{K-1}(x_i) + f_K(x_i) = \sum_{k=1}^{K} f_k(x_i) \tag{1}$$

Where $f_K(x_i)$ is the prediction from the k-th tree, and $F_{K-1}(x_i)$ is the sum of the prediction results from the previous tree [11]. The prediction results $\hat{y}_i$ are used in the objective function of XGBoost, which are as shown in Equation 2.

$$\begin{aligned} \Gamma(\theta) &= \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \\ \Omega(f_k) &= \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} \omega_j^2 \end{aligned} \tag{2}$$

Where function l is the calculated training error (such as squared error, log loss, etc.) and $\Omega(f_k)$ is the regularization for the k-th tree, which includes the tree complexity T and the weights ω_j for the leaf nodes. Next, by performing a second-series Taylor expansion on the loss function, the approximate optimization objective can be obtained as shown in Equation 3.

$$\begin{aligned} \Gamma^{(t)} &\approx \sum_{i=1}^{n} \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t(x_i)^2 \right] + \Omega(f_t) \\ g_i &= \frac{\partial l\left(y_i, \hat{y}_i^{(t-1)}\right)}{\partial \hat{y}_i^{(t-1)}} \quad h_i = \frac{\partial^2 l\left(y_i, \hat{y}_i^{(t-1)}\right)}{\partial \hat{y}_i^{(t-1)}} \end{aligned} \tag{3}$$

Where g_i is the first derivative and h_i is the second derivative. By performing optimization, a new decision tree can be trained efficiently [15].

The XGBoost model used in this study aims to classify Zang organ syndromes and the appropriate therapy recommendations in the form of acupuncture points as shown in Figure 1.

DOI: 10.34148/teknika.v15i1.1430

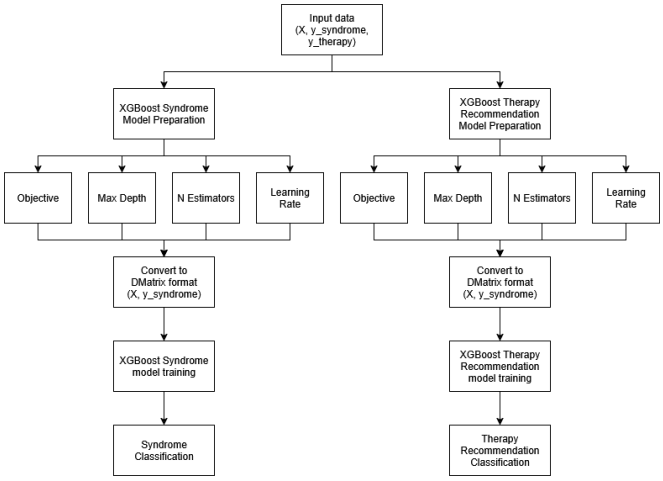


Figure 1. XGBoost Algorithm Modeling in This Study

D. Model Evaluation

The performance of the trained model will be evaluated using the confusion matrix, accuracy score, and F1 score metrics. These metrics are used to measure how effectively the model can classify syndromes based on their symptoms.

A confusion matrix is a table used to describe the performance of a classification model. The table represents the model's predictions compared to the actual results, separating the results into four categories: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). True positives are conditions correctly identified as positive, true negatives are conditions correctly identified as negative, false positives are cases incorrectly identified as positive, and false negatives are cases incorrectly identified as negative [17].

In the multiclass classification in this study, the values of TP, TN, FP, and FN are calculated per class in the target column, for example in the syndrome of "lung *qi* deficiency", TP means the predicted values and actual values are the "lung *qi* deficiency" syndrome, FP means the predicted value is "lung *qi* deficiency", but the actual value is not "lung *qi* deficiency", FN means the predicted value is not "lung *qi* deficiency", but the actual value is "lung *qi* deficiency", and TN means the predicted and actual values are not "lung *qi* deficiency".

Accuracy is the value of the correct predictions (TP and TN) divided by the sum of all samples. In healthcare, accurately identifying good or bad conditions is very important, but in imbalanced classes, when one class is significantly more common than the other, accuracy is not the most appropriate metric due to the impact of the dominant class [17]. The formula for accuracy in multiclass classification is as shown in Equation 4.

$$Accuracy = \frac{\sum_{c=1}^{C} TP_c}{N} \tag{4}$$

Where C is the number of classes in the target attribute, and N is the number of predictions made.

The F1 score is the harmonic mean of accuracy and recall and provides a balanced assessment in prediction such as disease. The F1 score takes into account both false positive and false negative predictions, which is particularly useful when working with uneven class distributions [17]. The F1 score formula is as shown in Equation 5.

$$F_1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}$$
$$Precision = \frac{TP}{TP + FP} \tag{5}$$
$$Recall = \frac{TP}{TP + FN}$$

III. RESULTS AND DISCUSSION

A. Pre-processing

Pre-processing was performed on the dataset used, including removing unused columns such as numbers and filling columns with null or NaN values with values for normal symptom properties. Each categorical data in the dataset was converted to lowercase, with text trimming, typo correction, and merging or changing similar data in attributes, such as the facial color attributes "red" and "red face" were combined into "red", while "redness in the face (malar flush)" was simplified to "malar flush". Once the categorical data in the dataset was in order, the categorical data in each column was converted into numeric values using the label encoder from the Python scikit-learn library.

After pre-processing, the dataset is divided for the model preparation process using the *train_test_split* function from scikit-learn, where the dataset is divided into training data (train) and test data (test), with training data measuring 80% of the entire dataset, and test data measuring 20%. To ensure that all syndromes and therapy recommendations as target attributes are trained in a balanced manner, the SMOTE (Synthetic Minority Oversampling Technique) oversampling method is used to increase the amount of data for minority classes in the target to prevent overfitting using datasets with unbalanced classes. After the training data is balanced for each class, the standard scaler from scikit-learn is used on the *x_train* and *x_test* variables to ensure that each attribute to determine the target attribute is normalized.

B. XGBoost Algorithm Modeling

XGBoost model training used the hyperparameters listed in Table 1. To ensure reproducibility of the model training with the same results, the *random_state* hyperparameter was used with *random_state* = 42. The model used the objective for multiclass classification with *multi:softmax*, with *num_class* = 67 for the syndrome classification model and *num_class* = 66 for the therapy recommendation classification model. The evaluation metrics used for this model were mlogloss (multiclass log loss) and merror (multiclass error), with the model accuracy value obtained from $1 - merror$ for each round of model training. Model training used data that had been oversampled with SMOTE.

CC BY SA

Table 1. XGBoost Model Hyperparameters

Hyperparameter	Value
max_depth	3
n_estimators	200
learning_rate	0.2

C. Model Evaluation

a) Classification Accuracy and Loss Results

The XGBoost model trained to learn symptom patterns for each syndrome and corresponding therapy recommendations yielded an accuracy for 0.986 for both syndromes and therapy recommendations, a log loss of 0.050 for both syndromes and therapy recommendations, and an F1 score of 0.83 for both syndromes and therapy recommendations.

The accuracy graph in Figure 2 and log loss graph in Figure 3 show that the model learned the data quickly in the first 20-50 rounds, then plateaued, with model performance not improving after 50-75 rounds of training. Because the test accuracy was close to the training accuracy and the test loss was consistently higher than the training loss, the model did not overfit at the specified hyperparameters as show in Table 1.

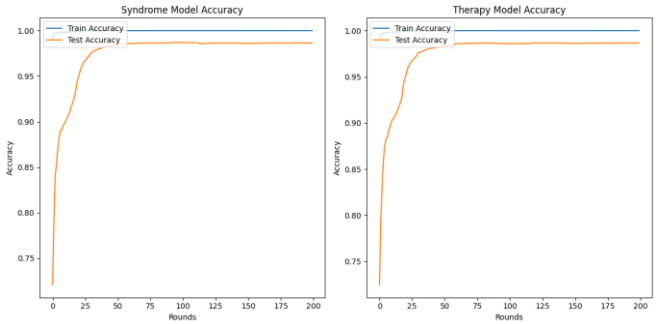


Figure 2. XGBoost Model Accuracy for Syndrome and Therapy Recommendations During Training

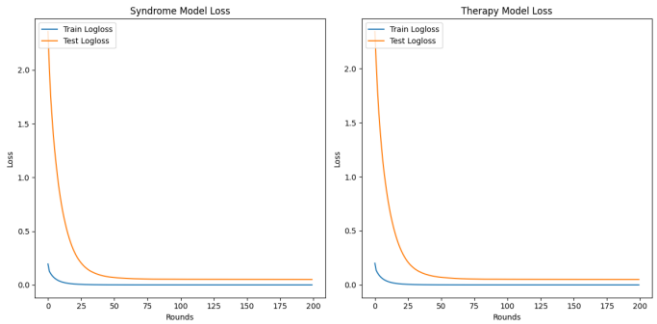


Figure 3. XGBoost Model Log Loss for Syndrome and Therapy Recommendations During Training

b) Classification Testing Confusion Matrix

Figure 4 shows the confusion matrix for the classification test using 10 syndrome types, with each syndrome having 10 data points for classification, except for one syndrome, lung yin deficiency, which has four data points for classification. The confusion matrix in Figure 4 shows that the XGBoost model successfully identified the syndrome for the 10-data-point class, without any prediction errors. However, for syndromes with limited data in the dataset, such as lung yin deficiency, the XGBoost model sometimes classified the given symptoms as normal or no syndrome. Figure 5 shows that this misclassification also extends to the classification of therapy recommendations, where two data points that should have been for lung yin deficiency were classified as normal.

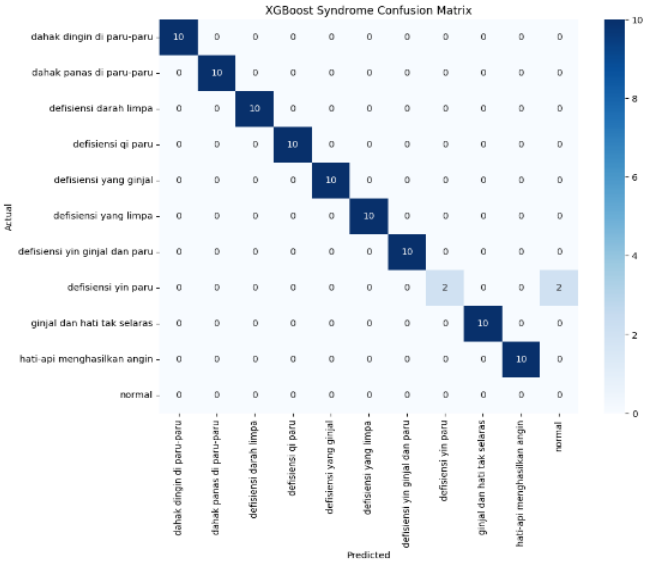


Figure 4. Syndrome Classification Confusion Matrix Using 10 Syndromes

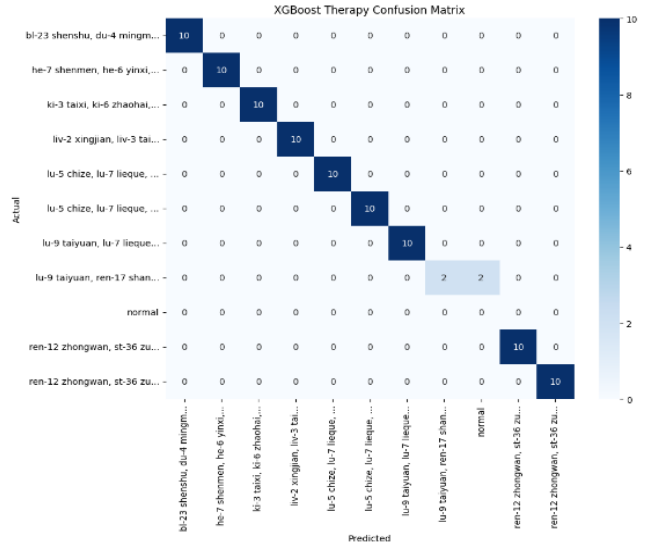


Figure 5. Therapy Recommendation Confusion Matrix From The 10 Syndromes

c) Classification Report

Based on the classification report results from the XGBoost model, both for syndrome classification and therapy recommendation classification, it can be explained that the accuracy obtained is 0.99 for both syndrome and therapy recommendation. This 0.99 accuracy is similar to the results of the training model accuracy graph in Figure 2. With syndrome

DOI: 10.34148/teknika.v15i1.1430

and therapy recommendation having a macro average F1 score of 0.83 and recall of 0.89, and precision of 0.81 for syndrome and 0.80 for therapy recommendation, this indicates that not all target classes have the same good performance. Syndrome and therapy recommendation also get a weighted average of 0.99 for precision, recall, and F1 score, indicating that for the majority class, predictions can be done well, and errors in classification only occur in the minority class. Because the recall of syndrome and therapy recommendation is higher than precision, the model has low false negatives, but some classes have false positives.

In the detailed classification report in Figure 6, it was found that syndrome classes with low numbers such as "liver fire insulting the lungs" and "*qi* and blood deficiency" failed to classify the following class, having a support value greater than 0, indicating that the following classes require additional training data. The following classes are only in the training data, so there is no data on actual precision and recall. Furthermore, there are syndrome classes that have low precision but high recall, such as the syndrome classes "liver-kidney disharmony" and "kidney *jing*/essence deficiency" had a precision of 0.67 and a recall of 1.00, and the class "heat and dampness in the liver" had a precision of 0.33 and a recall of 1.00, indicating that the following classes have many false positives in the following classes. Additionally, there are syndrome classes that have equally poor precision and recall performance, such as "liver and heart blood deficiency" having a precision of 0.25 and a recall of 0.50, indicating the model is unable to classify this class correctly for now, with low support values such as 2 for the "liver and heart blood deficiency" class.

Therapy recommendations also follow a pattern similar to the syndrome, with classes having low precision and high recall, equally low precision and recall, and equally low precision and recall of 0.00 with one class having a support value of 0 indicating it is not present in the test data, and the other having a support value greater than 0, indicating the need for additional training data.

kekurangan jing / esensi ginjal	0.67	1.00	0.80	4
kekurangan qi paru-paru dan jantung	1.00	1.00	1.00	4
limpa tidak dapat mengendalikan darah	1.00	0.50	0.67	2
liver-yang meningkat berasal dari liver & kekurangan yin ginjal	1.00	1.00	1.00	3
menghaluskan hati, menundukkan qi hati.	0.50	1.00	0.67	1
normal	1.00	0.99	0.99	2678
panas dan lembab di hati	0.33	1.00	0.50	1
panas paru - paru	0.00	0.00	0.00	1
panas-lembap menyerang limpa	0.88	1.00	0.93	7
paru - paru kering	1.00	1.00	1.00	3
pembuluh darah jantung tersumbat	1.00	1.00	1.00	4
penyumbatan limpa oleh kelembaban dengan stagnasi qi hati	1.00	1.00	1.00	3
qi ginjal lemah	1.00	1.00	1.00	1
qi hati berhenti dan berubah menjadi panas	0.33	1.00	0.50	1
qi hati yang memberontak menyerang lambung	1.00	1.00	1.00	2
qi limpa turun	1.00	1.00	1.00	3
stagnasi darah hati	0.00	0.00	0.00	1
stagnasi dingin di saluran hati	1.00	1.00	1.00	2
stagnasi qi hati	1.00	1.00	1.00	2
stagnasi qi jantung	1.00	1.00	1.00	1
stasis darah jantung	0.33	1.00	0.50	1
yang hati naik	1.00	1.00	1.00	5
yang jantung collaps	1.00	1.00	1.00	2
accuracy			0.99	2913
macro avg	0.81	0.89	0.83	2913
weighted avg	0.99	0.99	0.99	2913

Figure 6. Classification Report for The Syndrome Xgboost Model

d) Testing User Inputs

To show the model performance in testing user input, Figure 7 shows the correct classification results, and Figure 8 shows the incorrect classification results.

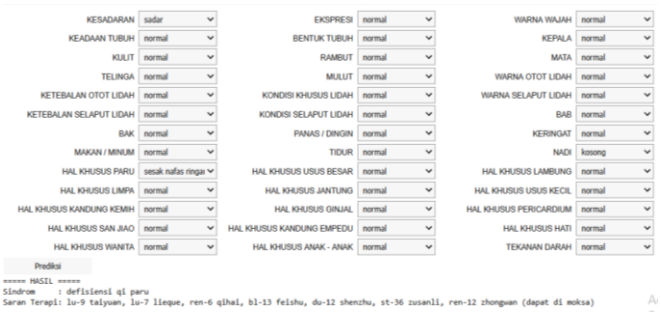


Figure 7. Correct Syndrome and Therapy Recommendation Classification (Lung Qi Deficiency)

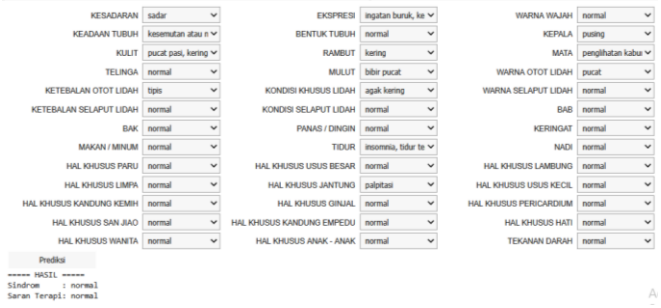


Figure 8. Incorrect Syndrome and Therapy Recommendation Classification (Liver and Heart Blood Deficiency)

The performance of the proposed XGBoost-based model demonstrates competitive results when compared with previous studies on TCM syndrome classification. Earlier research primarily employed deep learning approaches such as BERT-based models and CNN–LSTM hybrids, which reported precision values exceeding 90% for specific or limited syndrome classification tasks [7] . In contrast, the proposed approach focuses on comprehensive Zang organ syndrome classification rather than a single disease or a limited syndrome group.

Although the proposed model achieved high overall accuracy, variations in classification performance across different syndrome classes were observed. Syndromes with overlapping symptom characteristics—particularly among liver, spleen, and kidney-related disorders—pose greater classification challenges due to shared digestive, emotional, and Qi-related manifestations. Such overlap increases the likelihood of misclassification, as reflected in the confusion matrix and classification report results.

Dataset imbalance also significantly influences model performance. Syndromes with larger sample sizes demonstrate higher precision and recall, while minority syndromes tend to exhibit reduced F1-scores and higher misclassification rates. Despite the application of SMOTE to mitigate imbalance during training, classes with very limited support remain difficult to classify accurately. Misclassification in these minority classes may lead to inappropriate therapy recommendations, highlighting the importance of cautious

CC BY SA

interpretation when applying the system in practical TCM settings. These findings suggest that future improvements should focus on expanding minority class data and incorporating expert validation to enhance robustness.

IV. CONCLUSION

In this study, an artificial intelligence (AI)-based system was developed to identify Zang organ syndromes and recommend acupuncture point treatments based on patient complaints using the XGBoost machine learning algorithm. The proposed model achieved an accuracy of 98.6% and an F1 score of 83% for both syndrome classification and therapy recommendation.

The system supports the standardization of TCM syndrome diagnosis, which traditionally relies on practitioner experience and subjective interpretation. By producing consistent diagnostic results and corresponding therapy recommendations, the system can function as a decision-support tool for acupuncture practitioners and students while remaining complementary to clinical expertise.

Despite its promising performance, this study has several limitations. The dataset was derived from literature-based and expert-curated sources and validated by a single domain expert, which may limit inter-rater reliability. In addition, class imbalance affected model performance in minority syndrome classes.

Future research will focus on incorporating real patient data, involving multiple experts for dataset validation, expanding minority class representation, and applying explainable AI techniques to enhance interpretability and clinical reliability.

REFERENCES

[1] Z. Wang et al., "TCMEval-SDT: a benchmark dataset for syndrome differentiation thought of traditional Chinese medicine," Sci Data, vol. 12, no. 1, p. 437, Mar. 2025, doi: 10.1038/s41597-025-04772-9.

[2] Z. Huang et al., "A Traditional Chinese Medicine Syndrome Classification Model Based on Cross-Feature Generation by Convolution Neural Network: Model Development and Validation," JMIR Med Inform, vol. 10, no. 4, p. e29290, Apr. 2022, doi: 10.2196/29290.

[3] V. C. H. Chung, C. H. L. Wong, C. C. W. Zhong, Y. Y. Tjioe, T. H. Leung, and S. M. Griffiths, "Traditional and complementary medicine for promoting healthy ageing in WHO Western Pacific Region: Policy implications from utilization patterns and current evidence," Integrative Medicine Research, vol. 10, no. 1, p. 100469, Mar. 2021, doi: 10.1016/j.imr.2020.100469.

[4] A. Xu, "Traditional Chinese Medicine in Modern Healthcare: A Bridge between Ancient Wisdom and Contemporary Practice," IJEH, vol. 12, no. 1, pp. 69–72, Jan. 2024, doi: 10.54097/bwd69y85.

[5] L. C. Matos, J. P. Machado, F. J. Monteiro, and H. J. Greten, "Understanding Traditional Chinese Medicine Therapeutics: An Overview of the Basics and Clinical Applications," Healthcare, vol. 9, no. 3, p. 257, Mar. 2021, doi: 10.3390/healthcare9030257.

[6] K. A. O'Brien et al., "An Investigation into the Reliability of Chinese Medicine Diagnosis According to Eight Guiding Principles and Zang-Fu Theory in Australians with Hypercholesterolemia," The Journal of Alternative and Complementary Medicine, vol. 15, no. 3, pp. 259–266, Mar. 2009, doi: 10.1089/acm.2008.0204.

[7] Z. Chen et al., "Traditional Chinese medicine diagnostic prediction model for holistic syndrome differentiation based on deep learning," Integrative Medicine Research, vol. 13, no. 1, p. 101019, Mar. 2024, doi: 10.1016/j.imr.2023.101019.

[8] L. Ding, X. Zhang, D. Wu, and M. Liu, "Application of an extreme learning machine network with particle swarm optimization in syndrome classification of primary liver cancer," Journal of Integrative Medicine, vol. 19, no. 5, pp. 395–407, Sept. 2021, doi: 10.1016/j.joim.2021.08.001.

[9] J. Zheng et al., "Metabolic syndrome prediction model using Bayesian optimization and XGBoost based on traditional Chinese medicine features," Heliyon, vol. 9, no. 12, p. e22727, Dec. 2023, doi: 10.1016/j.heliyon.2023.e22727.

[10] H. Gong, H. Zhang, L. Zhou, Y. Liu, and L. Zhang, "An Interpretable Artificial Intelligence Model of Chinese Medicine Treatment Based on XGBoost Algorithm," in 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Seoul, Korea (South): IEEE, Dec. 2020, pp. 1550–1554. doi: 10.1109/bibm49941.2020.9313424.

[11] D. Yifeng, L. Jinsong, and M. Wentao, "Study on Breast Cancer Classification Prediction based on XGBoost," in 2024 IEEE 6th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), Chongqing, China: IEEE, May 2024, pp. 411–416. doi: 10.1109/IMCEC59810.2024.10575645.

[12] X. Y. Liew, N. Hameed, and J. Clos, "An investigation of XGBoost-based algorithm for breast cancer classification," Machine Learning with Applications, vol. 6, p. 100154, Dec. 2021, doi: 10.1016/j.mlwa.2021.100154.

[13] S. Degadwala, D. Vyas, A. Jadeja, and D. D. Pandya, "Enhancing Prostate Cancer Diagnosis: Leveraging XGBoost for Accurate Classification," in 2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS), Trichy, India: IEEE, Aug. 2023, pp. 1776–1781. doi: 10.1109/ICAISS58487.2023.10250511.

[14] S. Sankar, A. Potti, G. N. Chandrika, and S. Ramasubbareddy, "Thyroid Disease Prediction Using XGBoost Algorithms," JMM, Feb. 2022, doi: 10.13052/jmm1550-4646.18322.

[15] Y. Dou and W. Meng, "Predicting Cancer Classification Based on the Improved XGBoost Algorithms," in 2024

6th International Conference on Frontier Technologies of Information and Computer (ICFTIC), Qingdao, China: IEEE, Dec. 2024, pp. 1013–1016. doi: 10.1109/ICFTIC64248.2024.10913271.

[16] Rajitha. M and S. S. Megala, "A Novel Approach for Heart Disease Classification using Cascaded XGBoost Algorithm," in 2024 3rd International Conference on Automation, Computing and Renewable Systems (ICACRS), Pudukkottai, India: IEEE, Dec. 2024, pp. 905–912. doi: 10.1109/ICACRS62842.2024.10841637.

[17] E. K. Attipoe, A. S. Yussiff, and M. G. Asante-Mensah, "A Comparative Analysis of Machine Learning Classification Algorithms for Cardiovascular Disease Prediction," in 2024 IEEE 9th International Conference on Adaptive Science and Technology (ICAST), Accra, Ghana: IEEE, Oct. 2024, pp. 1–5. doi: 10.1109/ICAST61769.2024.10856496.