

Original Paper

# A Machine Learning Approach to Differentiate Cold and Hot Syndrome in Viral Pneumonia Integrating Traditional Chinese Medicine and Modern Medicine: Machine Learning Model Development and Validation

Xiaojie Jin<sup>1,2\*</sup>, Prof Dr; Yanru Wang<sup>1\*</sup>, Dr med; Jiarui Wang<sup>3\*</sup>, MM; Qian Gao<sup>2</sup>, MSc; Yuhan Huang<sup>1</sup>, MM; Lingyu Shao<sup>4</sup>, MM; Jiali Zhao<sup>1</sup>, MM; Jintian Li<sup>1</sup>, Prof Dr; Ling Li<sup>1\*</sup>, Prof Dr; Zhiming Zhang<sup>5\*</sup>, Prof Dr Med; Shuyan Li<sup>4\*</sup>, Prof Dr; Yongqi Liu<sup>1\*</sup>, Prof Dr, Prof Dr Med

<sup>1</sup>Key Laboratory of Dunhuang Medicine, Ministry of Education, Gansu University of Chinese Medicine, Lanzhou, China

<sup>2</sup>College of Pharmacy, Gansu University of Chinese Medicine, Lanzhou, China

<sup>3</sup>The Second Affiliated Hospital of Xuzhou Medical University, Xuzhou, China

<sup>4</sup>School of Medical Information and Engineering, Xuzhou Medical University, Xuzhou, China

<sup>5</sup>Gansu Provincial Hospital of Traditional Chinese Medicine, Lanzhou, China

\*these authors contributed equally

## Corresponding Author:

Yongqi Liu, Prof Dr, Prof Dr Med  
Key Laboratory of Dunhuang Medicine  
Ministry of Education, Gansu University of Chinese Medicine  
Dingxi East Road, 35th  
Lanzhou, 730000  
China  
Phone: 86 13919019578  
Email: [liuyongqi73@163.com](mailto:liuyongqi73@163.com)

## Abstract

**Background:** Syndrome differentiation in traditional Chinese medicine (TCM) is an ancient principle that guides disease diagnosis and treatment. Among these, the cold and hot syndromes play a crucial role in identifying the nature of the disease and guiding the treatment of viral pneumonia. However, differentiating between cold and hot syndromes is often considered esoteric. Machine learning offers a promising avenue for clinicians to identify these syndromes more accurately, thereby supporting more informed clinical decision-making in the treatment.

**Objective:** This study aims to construct a diagnostic model for differentiating cold and hot syndromes in viral pneumonia by integrating TCM and modern medical features using machine learning methods.

**Methods:** The application of 8 machine learning algorithms (gradient boosting machine [GBM], logistic regression, random forest, extreme gradient boosting [XGB], light gradient boosting machine [LGB], ridge regression, least absolute shrinkage and selection operator, and support vector machine) generated and validated (both internally and externally) a model for differentiating cold and hot syndromes in viral pneumonia, based on clinical data from 1484 patient samples collected at 2 medical centers between 2021 and 2022.

**Results:** The GBM model, which combines TCM and modern medicine features, outperformed models using only TCM features or only modern medicine features in distinguishing cold and hot syndromes in patients with viral pneumonia. The optimal discrimination model comprised 13 optimal features (temperature, red cell distribution width-SD, creatinine, total bilirubin, globulin, C-reactive protein, unconjugated bilirubin, white blood cell, neutrophil percentage, aspartate transaminase/alanine transaminase, total cholesterol, thrombocytocrit, and age) and the GBM algorithm, achieving an area under the curve (AUC) of 0.7788. Under internal and external testing, the AUCs were 0.7645 and 0.8428, respectively. Moreover, significant differences were observed between the cold and hot syndrome groups in temperature ( $P=.02$ ), red cell distribution width-SD ( $P<.001$ ), neutrophil percentage ( $P=.01$ ), total cholesterol ( $P=.003$ ), thrombocytocrit ( $P<.001$ ), and age ( $P<.001$ ).

**Conclusions:** This pioneering study integrates the theory of TCM cold and hot syndromes with modern laboratory-based tests through machine learning. The developed model offers a novel approach for differentiating cold and hot syndromes in viral

pneumonia, enabling practitioners to identify the syndrome quickly and efficiently, thereby supporting more informed clinical decision-making. Additionally, this research provides new insights into the modernization and scientific interpretation of TCM syndrome differentiation.

*JMIR Med Inform* 2025;13:e64725; doi: [10.2196/64725](https://doi.org/10.2196/64725)

**Keywords:** syndrome differentiation; traditional Chinese medicine; viral pneumonia; dialectical treatment; machine learning; Chinese medicine; pneumonia; lungs; diagnostic model; retrospective study; artificial intelligence; China; model training; performance evaluation; laboratory tests; modern medicine

## Introduction

Traditional Chinese medicine (TCM), a personalized and holistic approach, treats diseases using natural medical products tailored to a patient's TCM syndrome patterns [1], that is, dialectical treatment of TCM. Specifically, TCM syndrome, a concept unique to TCM, is an abstraction of a variety of signs and symptoms. Dialectical treatment is the fundamental principle by which physicians diagnose, understand, and treat diseases. It involves comprehensively collecting patients' clinical signs and symptoms through "inspection, auscultation, olfaction, inquiry, and palpation" to determine the TCM syndrome pattern. Subsequently, corresponding therapies, such as acupuncture, cupping, qigong, massage, and diet, are selected [2] to correct maladjustments and restore the body's self-regulatory ability [3,4]. Guided by this unique principle, TCM has demonstrated remarkable clinical efficacy in treating various acute and chronic diseases, thus attracting increasing attention [5]. Especially in the fight against viral pneumonia, the importance of TCM has been widely recognized [6,7]. Consequently, promptly and accurately understanding the TCM syndrome pattern of a disease can reshape global health and welfare.

In recent years, with the successive outbreaks of SARS, influenza A (H1N1), influenza A (H7N9), COVID-19, and other viruses, viral pneumonia has become a sustained and widespread disease burden in many countries and globally. As with other diseases, the accurate identification of syndrome patterns is a prerequisite for the TCM-based prevention and treatment of viral pneumonia. Numerous studies have reported that the cold and hot syndromes are the pivotal TCM syndrome patterns in cases of viral pneumonia, such as SARS and COVID-19 [8-10]. Cold and hot syndromes, originating from the *Huangdi Neijing*, are 2 key components of the Eight Principle Syndromes—the general principles of dialectical treatment, which include cold and hot, sthenia and asthenia, exterior and interior, and yin and yang—and are used to identify the nature of a disease, reflecting disruptions in body homeostasis [11-13]. Cold syndrome-related expressions include coldness, cold pain, tastelessness, and clear, abundant urine, whereas hot syndrome is mainly characterized by heat, diaphoresis, a flushed face, burning pain, and deep-colored urine. The cold and hot syndrome framework has been widely applied in the diagnosis and treatment of viral pneumonia. Treating cold and hot syndromes within TCM has been shown to significantly alleviate clinical symptoms in patients with viral pneumonia [8,14]. For example, symptomatic treatments with

Chinese medicines such as Lianhua Qingwen capsules and Xuanfeibaidu granules can markedly reduce the incidence of severe or critical events and improve clinical recovery [6,7], indicating that the prediction and differentiation of cold and hot syndromes play a vital role in the treatment of viral pneumonia.

Nevertheless, although clinical data and studies have indicated that TCM treatments for the cold and hot syndromes of viral pneumonia are effective, a major challenge remains in distinguishing between these syndromes—similar to the challenges in identifying other TCM syndrome patterns. Specifically, (1) the differentiation of cold and hot syndrome is recondite and lacks a solid scientific foundation in classical TCM theory, and (2) the accuracy of cold and hot syndrome diagnosis in TCM relies heavily on the physicians' skills and years of clinical experience. Subjective factors play a decisive role in TCM diagnosis, which inevitably leads to a lack of standardized criteria and an objective diagnostic basis for cold and hot syndromes. This unique diagnostic model not only supports the open and divergent development of TCM, but also hinders the understanding of cold and hot syndromes in viral pneumonia from a modern biomedical perspective. It further restricts the domestic and international dissemination of TCM syndrome differentiation, making it one of the key bottlenecks in the development of TCM. Therefore, in the process of integrating TCM with modern medicine, it is worth considering how to combine TCM syndrome theory with modern medicine to enhance the scientific rigor and objectivity of cold and hot syndrome differentiation in the treatment of viral pneumonia.

Recently, with the rapid development of machine learning [15-18], artificial intelligence has brought unprecedented challenges and opportunities to the diagnosis of TCM syndrome. Interdisciplinary models combining TCM and artificial intelligence have been proposed to model TCM knowledge, diagnosis, and treatment in clinical practice. Among these, numerous models have been developed to simulate the syndrome differentiation process of human TCM doctors for automatic syndrome diagnosis. For example, Huang et al [19] successfully identified the liver-gallbladder dampness-heat syndrome in patients with breast cancer using a dataset based on the prescriptions of TCM clinical practitioners. Zhuo et al [20] effectively differentiated Qi deficiency blood stasis syndrome and phlegm stasis in channels syndrome in patients with stroke through artificial intelligence-assisted retinal feature analysis. These examples demonstrate that machine learning enables a modern interpretation of the scientific connotation of TCM syndrome differentiation, supporting improved clinical decision-making

in disease treatment. However, currently, there is still a lack of machine learning–based differentiation of cold and hot syndromes in viral pneumonia.

Among the machine learning algorithms in artificial intelligence, gradient boosting machine (GBM), logistic regression (LR), random forest (RF), extreme gradient boosting (XGB), and light gradient boosting machine (LGB), among others, have demonstrated promising performance [21–24]. Consequently, this study aimed to compare 8 machine learning methods, namely, GBM, LR, RF, XGB, LGB, ridge regression (RIDGE), least absolute shrinkage and selection operator (LASSO), and support vector machine (SVM), by integrating laboratory test indicators from modern medicine with the TCM cold and hot syndrome, to explore the connection between them and construct a diagnostic model for differentiating the cold and hot syndromes in viral pneumonia. This approach seeks to provide a new diagnostic method for the clinical differentiation of cold and hot syndromes in viral pneumonia and to offer new perspectives for the scientization, standardization, and objectivity of TCM syndrome differentiation theory in guiding treatment.

## Methods

### Source of Materials

In this study, a retrospective review of case sheets was conducted on samples from 1401 patients diagnosed with viral pneumonia between January 18 and November 6, 2021, at the Second People's Hospital of Lanzhou City in China. Samples were excluded based on the following criteria: (1) the diagnosis was not made using the Eight Principle Syndromes of TCM, and (2) the patient was diagnosed with both cold and hot syndrome patterns simultaneously. A total of 602 samples met the inclusion criteria. Samples with more than 20% (ie, 19/93 features) missing values were further excluded, resulting in 382 samples ultimately included for analysis. These comprised 97 negative samples (cold syndrome) and 285 positive samples (hot syndrome). Additionally, a stratified sampling strategy was applied to split the original dataset into training and internal test cohorts at an 8:2 ratio (details are provided in the “Performance Evaluation and Model Validation” section).

The external test cohort consisted of 83 patients with viral pneumonia diagnosed with either cold or hot syndrome at Lanzhou Heavy Ion Hospital between July and August 2022. This cohort included 36 patients with the hot syndrome and 47 patients with the cold syndrome.

### TCM and Modern Medicine Features Collection and Cold and Hot Syndrome Diagnosis

As shown in [Multimedia Appendix 1](#), a total of 93 features related to TCM and modern medicine were collected in

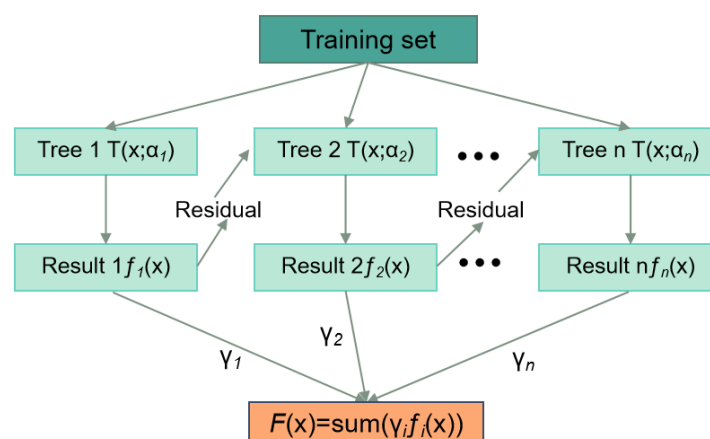
this study, including 4 general information items, 19 TCM symptoms, 2 blood gas values, 10 viral pneumonia indicators, 27 biochemical indicators, 30 blood routine indicators, and 1 coagulation indicator. Among these, patients' TCM symptoms were further quantified using a TCM symptom scoring scale for viral pneumonia ([Multimedia Appendix 2](#)).

All patients with viral pneumonia were independently diagnosed with either the TCM cold or hot syndrome by 2 TCM chief physicians. If both physicians provided the same diagnosis, the patient was enrolled. In cases of disagreement, a third investigator participated in the discussion to reach a consensus. The diagnostic criteria for syndrome differentiation were based on the textbooks *TCM Diagnostics* [25] and *Guiding Principles of Clinical Research on New Drugs of Traditional Chinese Medicine* [26]. These criteria include the 4 classical diagnostic methods used to determine TCM cold and hot syndrome patterns: inspection, auscultation, inquiry, and palpation, to accurately record the TCM characteristics.

### Model Training

In this study, 8 machine learning algorithms (GBM, LR, RF, XGB, LGB, RIDGE, LASSO, and SVM) were used to train a dataset containing all features, and GBM was selected as the final classification model. As shown in [Figure 1](#), GBM operates on the principle of ensemble learning, constructing the final prediction model by iteratively training a series of weak learners, typically decision trees. The core mechanism of GBM lies in its ability to correct the prediction errors of previous decision trees. Each newly generated decision tree is trained to fit the residuals—the differences between the actual values and the predicted values—of the previous model, thereby continuously reducing the overall loss function. Specifically, GBM begins by constructing an initial decision tree based on the training data, and then calculates the prediction error of this tree. A second tree is built to model the residuals, with its predictions aimed at correcting the deficiencies of the first tree. This process continues iteratively until a preset number of iterations is reached or a stopping condition is met. [Figure 1](#) illustrates the structure and decision-making process of multiple decision trees in the GBM model. Each decision tree branches based on input features, and the final prediction is obtained through a weighted sum of the predictions from all individual trees.

The number of top-ranking indexes was gradually increased, and the hyperparameters `n_estimators` and `max_depth` were tuned. The final model consisted of 13 top-ranking features, with `n_estimators` and `max_depth` set to 797 and 8, respectively. The GBM was implemented using the scikit-learn library (version 0.24.1) in Python (version 3.8.8; Python Foundation).

**Figure 1.** Gradient boosting machine decision tree composition and prediction flowchart.

### Optimal Feature Ascertainment

“Feature importance,” “coef,” and “SHAP values” are 3 commonly used methods to assess the relationship between features and machine learning models, and they have received significant attention in recent research [27]. These methods calculate the relevance of each feature to the model’s predictions, which are then ranked in descending order to produce a feature importance ranking.

In tree-based models (eg, GBM, RF, XGB, LGB), feature importance is typically evaluated from 2 perspectives: split count and split gain. Split count refers to how often a particular feature is selected as a splitting point during tree construction; a higher count indicates greater importance. By contrast, split gain measures the improvement in the model’s objective function (eg, squared error loss) resulting from splitting on a given feature. A larger gain implies a greater contribution to model performance and hence higher importance. Additionally, SHAP (Shapley Additive Explanations) values offer a unified measure of feature importance based on game theory by calculating the marginal contribution of each feature to the model’s prediction. SHAP values provide both global and local interpretability of the model. In linear models (eg, RIDGE, LASSO, SVM, and LR), “coef” represents the coefficients assigned to each feature. These coefficients indicate the strength and direction of the linear relationship between each feature and the target variable. SHAP values can complement these coefficients by revealing nonlinear effects and feature interactions that may not be captured by standard linear models.

The insights gained from these methods, especially when combining traditional feature importance metrics with SHAP analysis, can aid researchers in uncovering hidden data patterns and insights, thereby supporting scientific investigations and informing practical applications.

Accordingly, these methods are used to rank the feature importance for cold and hot syndromes of viral pneumonia and to ascertain the optimal features.

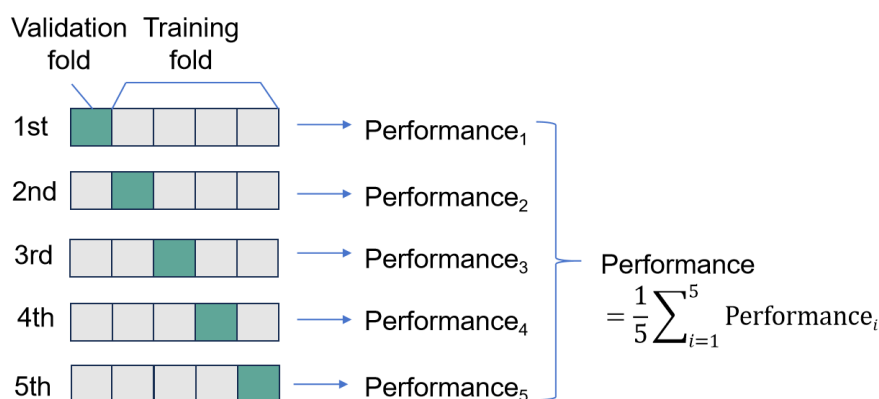
### Performance Evaluation and Model Validation

During the data preprocessing stage, a stratified sampling strategy was adopted to divide the original dataset, ensuring an 8:2 ratio between the training and validation sets. This approach preserved the proportional distribution of each class (cold and hot syndromes) as in the original dataset, thereby avoiding model evaluation bias due to imbalanced category distributions.

Given the relatively small sample size, 5-fold cross-validation was applied to ensure model accuracy and reliability. As illustrated in Figure 2, the dataset was divided into 5 nonoverlapping subsets of equal size. In each round of validation, 1 subset was selected as the validation set (the blue portion in the figure), while the remaining 4 subsets constituted the training set (gray portions). This process was repeated for 5 rounds. In the first round, the first subset was the validation set, and the remaining 4 were used as the training set, resulting in Performance<sub>1</sub>. In the second round, the second subset was used for validation, and so on, resulting in Performance<sub>2</sub> through Performance<sub>5</sub>. The final model performance was calculated as the average of these 5 evaluation results, with the formula

$$\text{Performance} = \frac{1}{5} \sum_{i=1}^5 \text{Performance}_i.$$

This strategy ensured that each sample was used for both training and validation, reduced the variance caused by random sample division, and improved the robustness and reliability of the model evaluation [28].

**Figure 2.** The 5-fold cross-validation process diagram.

In addition, a comprehensive evaluation of the model was conducted using commonly used metrics, including sensitivity, specificity (SPE), accuracy (ACC), Matthews correlation coefficient (MCC), and area under the curve (AUC). These metrics are calculated based on the values of true positives (TPs), true negatives (TNs), false positives (FPs), and false negatives (FNs), as defined below:

$$\text{Sens} = \frac{TP + FN}{TP} \quad (1)$$

$$\text{Spec} = \frac{TN}{TN + FP} \quad (2)$$

$$\text{ACC} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

$$\text{MCC} = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (4)$$

The receiver operating characteristic curve, also known as the sensitivity curve, is a graphical representation of the performance of a classification model across various evaluation thresholds. It plots the true-positive rate against the false-positive rate at different threshold settings. The true-positive rate reflects the proportion of actual positive cases correctly identified by the model, while the false-positive rate represents the proportion of negative cases incorrectly classified as positive. The AUC provides a single summary metric of the classifier's overall performance, with values ranging from 0 to 1. A higher AUC value indicates better classification performance. Therefore, AUC is widely used as a standard metric to evaluate the performance of classification models. The formula for calculating AUC is as follows:

$$\text{AUC} = \int_0^1 \text{TPR}(\text{FDR}) d\text{FDR} \quad (5)$$

## Ethical Considerations

The study was approved by the Medical Ethics Review Committee of the Second People's Hospital of Lanzhou City and Lanzhou Heavy Ion Hospital (approval number 2021-026-01). Informed consent was obtained on an opt-out basis, and all data were anonymized. No financial or material incentives were provided to participants.

## Statistical Analysis

All data were analyzed using SPSS version 27.0 (IBM Corp). Categorical variables are presented as counts and percentages, whereas continuous variables are reported as means and SDs or IQRs, as appropriate. Normality was assessed using the Kolmogorov-Smirnov test. Student 2-tailed *t* test was used to compare parametric continuous variables, while the Mann-Whitney *U* test was used for nonparametric variables, the  $\chi^2$  test was applied for categorical variables, and the Fisher exact test was used for 2x2 contingency tables. Binomial distribution tests were performed for each feature within the cold syndrome and hot syndrome groups. No correction for multiple comparisons was applied. A 2-sided  $P < .05$  was considered statistically significant.

## Results

### Data Extraction

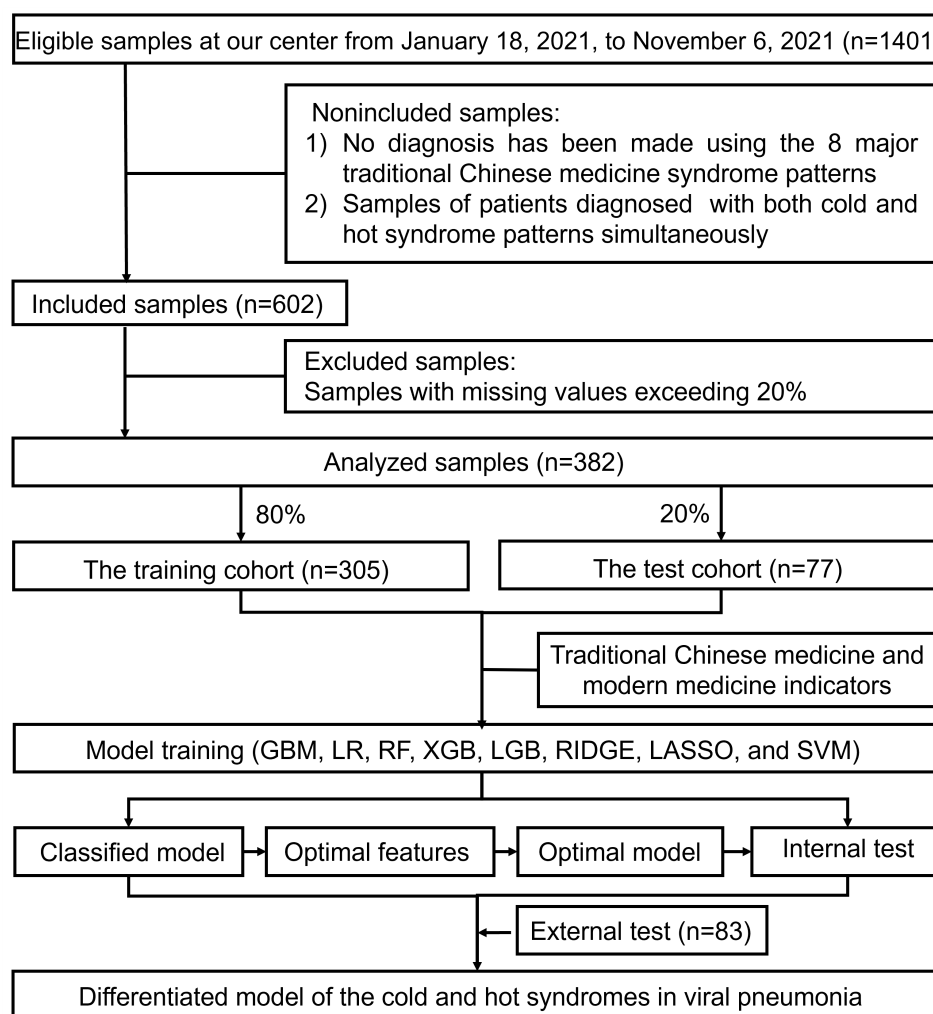
We included only individual samples from patients with viral pneumonia and excluded those that were not diagnosed using the Eight Principle Syndromes (cold and hot, sthenia and asthenia, exterior and interior, and yin and yang), as well as those simultaneously diagnosed with cold and hot syndrome. A total of 602 eligible samples were retained. Samples with more than 20% (ie, 19/93 features) missing values were further excluded, and the remaining missing values were imputed using the median of each variable. Ultimately, 382 complete samples were included in the final analysis and used to construct the mathematical model. The dataset was split into training and testing sets in an 8:2 ratio. Subsequently, by integrating indicators from TCM and modern medicine



and applying 8 machine learning algorithms, we identified the optimal model for differentiating TCM cold-hot syndromes in viral pneumonia through 5-fold cross-validation. The model's

performance was further assessed using an external validation cohort. The data processing workflow and model construction process are illustrated in Figure 3.

**Figure 3.** Research flow for data processing and predictive model construction. After preliminary screening of the collected data, 382 samples of patients with viral pneumonia with TCM cold and hot syndromes were obtained. The dataset was randomly divided into training and test sets in an 8:2 ratio. Traditional Chinese medicine and modern medical indicators were extracted through data processing and statistical analysis. The performance of the 8 machine learning models (gradient boosting machine [GBM], logistic regression [LR], random forest [RF], extreme gradient boosting [XGB], light gradient boosting machine [LGB], ridge regression, least absolute shrinkage and selection operator [LASSO], and support vector machine [SVM]) was evaluated. Based on a comprehensive evaluation of all models, the optimal classification algorithm and feature set were identified. The final model's performance was further assessed using internal and external test cohorts.



### Model Training Based on TCM and Modern Medicine Features

Modern medical indicators based on laboratory testing provide the possibility for intelligent discrimination of TCM syndromes. Therefore, in this study, modern medical indicators were integrated with TCM features for further exploration. A total of 93 features were used to construct the model. The evaluation metrics for models using all 93 features are presented in Table 1 and Figure 4A. We found that GBM and LGB achieved the highest AUC scores of 0.8329 and 0.7693, respectively, indicating their superior overall discriminative performance. GBM also obtained the highest ACC score of 0.8000, closely followed by RIDGE with a score of 0.7475, suggesting strong overall classification accuracy. LASSO, RIDGE, and

GBM achieved the highest sensitivity scores of 0.8430, 0.8388, and 0.8283, respectively, indicating good performance in correctly identifying positive samples and reducing false negatives. SVM, LGB, RF, and GBM achieved the highest specificity scores, suggesting these models had the lowest false-positive rates. Both GBM and LGB also achieved relatively high MCC scores of 0.5043 and 0.4068, respectively, demonstrating their effectiveness in handling imbalanced datasets.

Internal validation showed that GBM, RF, and XGB achieved the highest AUC scores of 0.7085, 0.7477, and 0.7146, respectively. Additionally, we calculated the confusion matrix for each model in the internal validation cohort, as shown in Figure 4B. The sum of correctly predicted cold and hot samples reflects the overall predictive accuracy

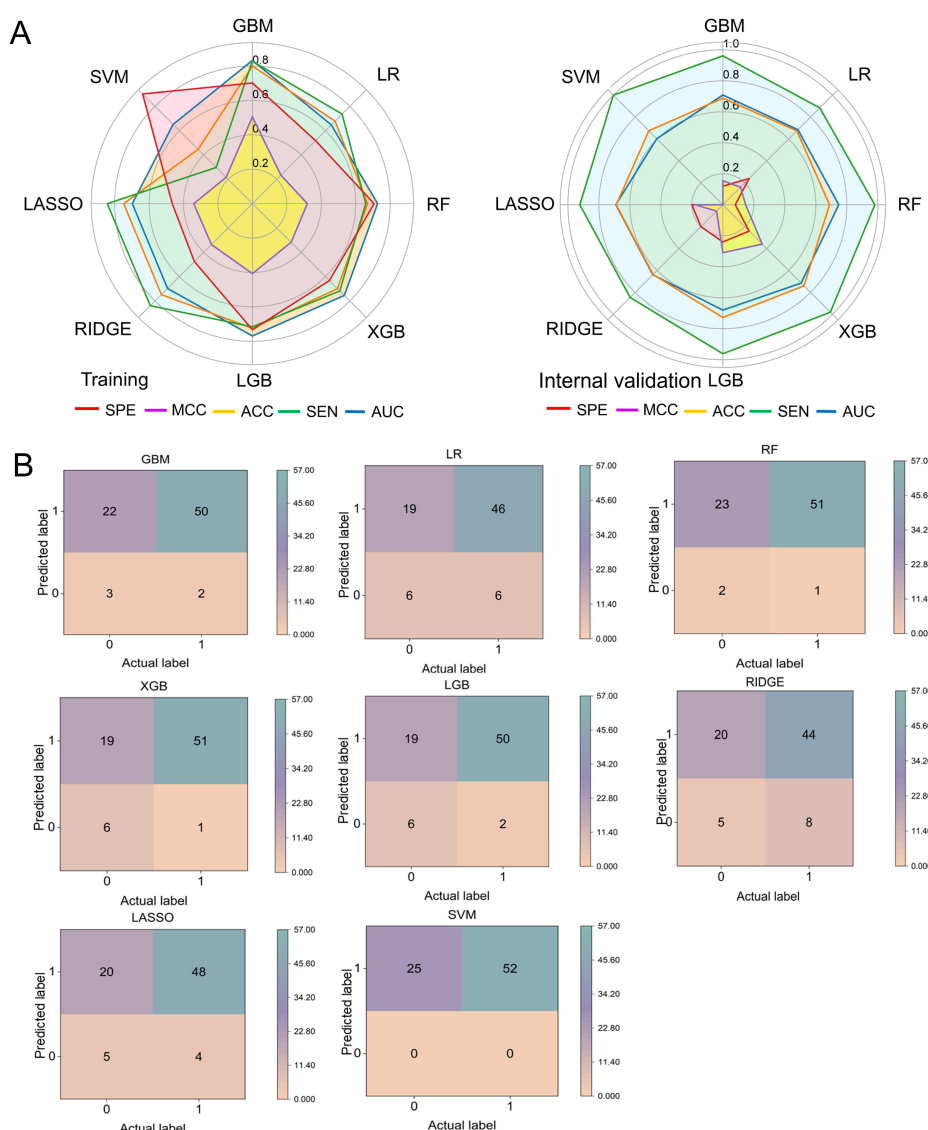
of each model. GBM, LR, RF, XGB, LGB, RIDGE, LASSO, and SVM correctly predicted 53, 52, 53, 57, 56, 49, 53, and 52 cold and hot samples, respectively. These findings further demonstrate the effectiveness of these models in distinguishing cold and heat patterns in viral pneumonia.

Overall, the results suggest that models incorporating both TCM and modern medicine features exhibit strong predictive performance for differentiating between cold and hot syndromes in viral pneumonia. Among them, the GBM algorithm demonstrated a particularly prominent discriminative capability. Therefore, GBM was selected as the optimal classified model based on the full set of 93 features for subsequent analysis.

**Table 1.** Comparison of models based on 5-fold cross-validation results using 93 features from traditional Chinese medicine and modern medicine.

Data cohort and evaluation	Gradient boosting machine	Logistic regression	Random forest	Extreme gradient boosting	Light gradient boosting machine	Ridge regression	Least absolute shrinkage and selection operator	Support vector machine
Training								
Area under the curve	0.8329	0.6477	0.7272	0.7546	0.7693	0.6970	0.6974	0.6514
Accuracy	0.8000	0.6787	0.6689	0.6984	0.7246	0.7475	0.7475	0.4459
Sensitivity	0.8283	0.7365	0.6582	0.7172	0.7158	0.8388	0.8430	0.2975
Specificity	0.7016	0.5187	0.7066	0.6326	0.7327	0.4764	0.4656	0.9018
Matthews correlation coefficient	0.5043	0.2341	0.3190	0.3177	0.4068	0.3364	0.3407	0.2161
Internal validation								
Area under the curve	0.7085	0.6854	0.7477	0.7146	0.6800	0.6392	0.6908	0.6031
Accuracy	0.6883	0.6753	0.6883	0.7403	0.7273	0.6364	0.6883	0.6753
Sensitivity	0.9615	0.8846	0.9808	0.9808	0.9615	0.8462	0.9231	1.0000
Specificity	0.1200	0.2400	0.0800	0.2400	0.2400	0.2000	0.2000	0.0000
Matthews correlation coefficient	0.1549	0.1609	0.1471	0.3596	0.3093	0.0577	0.1794	0.0000

**Figure 4.** Performance evaluation of the cold and hot syndrome identification models based on integrated traditional Chinese medicine (TCM) and modern medical features. (A) Radar chart comparing the performance of 8 machine learning models in the training and internal validation cohorts. (B) Confusion matrix analysis of the 8 models in the internal validation cohort. ACC: accuracy; AUC: area under the curve; GBM: gradient boosting machine; LASSO: least absolute shrinkage and selection operator; LGB: light gradient boosting machine; LR: logistic regression; MCC: Matthews correlation coefficient; RF: random forest; RIDGE: ridge regression; SEN: sensitivity; SPE: specificity; SVM: support vector machine; XGB: extreme gradient boosting.



## Optimal Feature Ascertainment

Based on the optimum classified model (GBM) and the ranking of “feature importance,” we identified 23 key features with importance values more than twice the median. These were temperature (T), red cell distribution width-SD (RDW-SD), creatinine (CREA), total bilirubin (TBIL), globulin (GLO), C-reactive protein (CRP), unconjugated bilirubin (IBIL), white blood cell (WBC), neutrophil percentage (NEU%), aspartate transaminase/alanine transaminase (AST/ALT), total cholesterol (TCHO), thrombocytocrit (PCT), oxygen saturation under load (SaO<sub>2</sub>), age, platelets, mean platelet volume, lactate dehydrogenase, mean corpuscular hemoglobin concentration, urea, erythrocyte/red blood cell, Na, ORF1ab gene of novel coronavirus by nose test (nCoVORF1ab), and albumin.

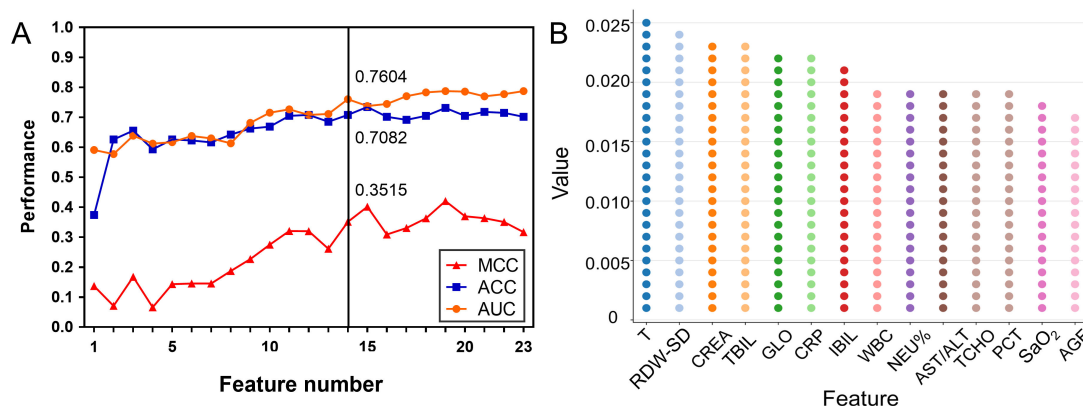
Subsequently, to identify the best-performing model with the fewest features, we examined how the number of top-ranked features influenced model performance (AUC, ACC, and MCC) using the GBM algorithm. As shown in Figure 5A and B, as the number of features increases, model performance gradually improves. Notably, at the 14-feature mark, the model achieved strong performance with an AUC of 0.7604, an ACC of 0.7082, and an MCC of 0.3515. To validate the robustness of our feature selection approach, we also used the SHAP feature importance ranking to select the top 14 features and construct a model. This model achieved an AUC of 0.7406, which was slightly lower than that obtained using the feature importance method (0.7604). Therefore, the feature importance method was ultimately retained as the feature selection approach for this study.



Finally, considering both the practicality of indicator testing and the feasibility of the clinical application, this study selected the top 14 features, excluding SaO<sub>2</sub> (under load), as the optimal set for differentiating between cold

and hot syndromes in viral pneumonia. These features are temperature (T), RDW-SD, CREA, TBIL, GLO, CRP, IBIL, WBC, NEU%, AST/ALT, TCHO, PCT, and age (AGE), as illustrated in (Multimedia Appendix 3).

**Figure 5.** Identification of key features based on feature importance analysis and model cross-validation. (A) Effect of the number of selected key features (n=23) on model performance using the gradient boosting machine (GBM) algorithm. (B) Feature importance scores of the top 14 key features identified by the GBM model. ACC: accuracy; AUC: area under the curve; AST/ALT: aspartate transaminase/alanine transaminase; CREA: creatinine; CRP: C-reactive protein; GLO: globulin; IBIL: unconjugated bilirubin; MCC: Matthews correlation coefficient; NEU%: neutrophil percentage; PCT: thrombocytocrit; RDW-SD: red cell distribution width-SD; SaO<sub>2</sub>: oxygen saturation under load; T: temperature; TBIL: total bilirubin; TCHO: total cholesterol; WBC: white blood cell.



## Performance Evaluation and Optimal Model Establishment

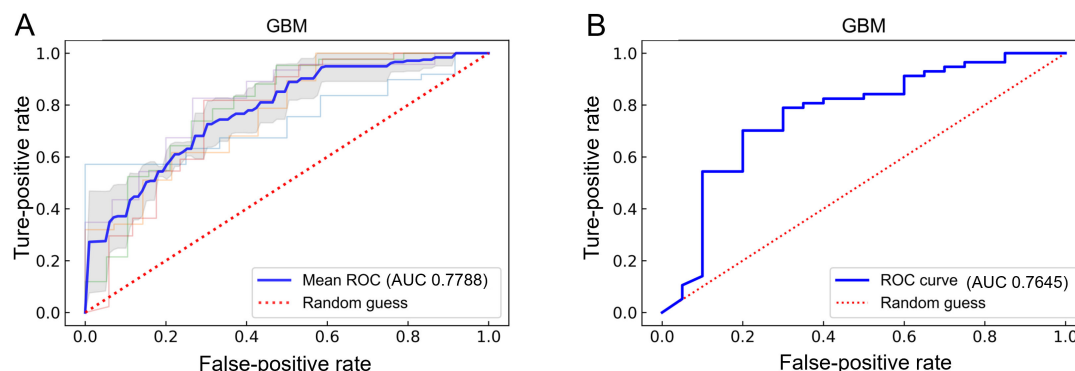
To achieve optimal GBM model performance, we reconstructed the GBM model based on 13 optimizing features. Ultimately, the model was adjusted to threshold=0.99999, n\_estimators=797, max\_depth=8, min\_samples\_split=2, and min\_samples\_leaf=1. Simultaneously, the performance of the model with 13 features was further evaluated using 5-fold cross-validation, and the results suggested that the model performance was outstanding, as the average AUC value reached 0.7788 (Figure 6A). Model verification results showed that the AUC was 0.7645 by the internal test cohort (Figure 6B). Furthermore, the additional external test results showed that the model performance for AUC, ACC, sensitivity, specificity, and MCC is 0.8428, 0.6627, 0.7222, 0.6170, and 0.3369, respectively. These results validated the excellent performance of the model.

Moreover, we found that the temperature in the hot syndrome group was significantly higher than in the cold

syndrome group ( $P=.02$ ). By contrast, RDW-SD ( $P<.001$ ), NEU% ( $P=.01$ ), TCHO ( $P=.003$ ), PCT ( $P<.001$ ), and AGE ( $P<.001$ ) were significantly lower in the hot syndrome group compared with the cold syndrome group, as shown in Table 2 and Figure 7. These results align with previous studies, which have reported that temperature, AGE, CREA, and RDW-SD are closely associated with the severity of viral pneumonia [29-33], supporting the validity of our findings.

These results suggest that machine learning algorithms perform well in differentiating between TCM cold and hot syndromes in viral pneumonia. The optimal model was constructed using 13 key features (T, RDW-SD, CREA, TBIL, GLO, CRP, IBIL, WBC, NEU%, AST/ALT, TCHO, PCT, and AGE), combined with the GBM algorithm. Among these, temperature (T), RDW-SD, NEU%, TCHO, PCT, and AGE appear to be the most relevant features for predicting cold and hot syndrome.

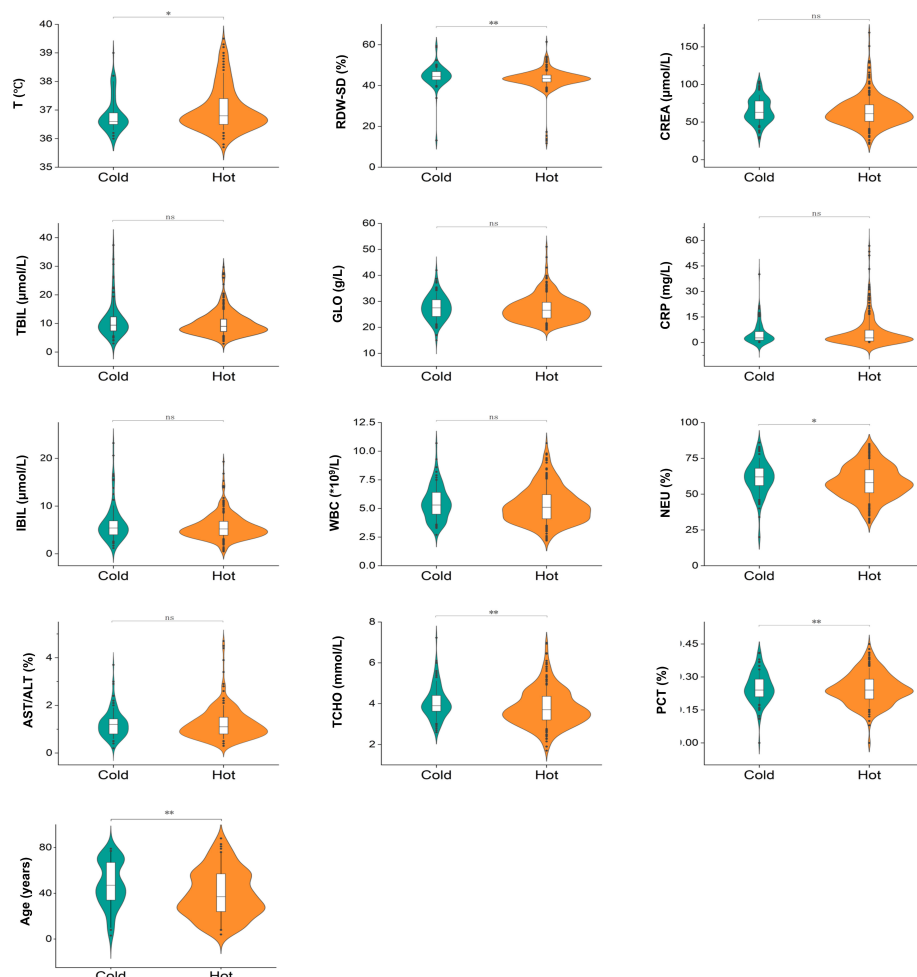
**Figure 6.** ROC curves and corresponding AUCs for cross-validation on the training and test sets. (A) ROC curve and classification performance of the final model set on the training set using 5-fold cross-validation. The x-axis represents the false-positive rate (FPR), and the y-axis represents the true-positive rate (TPR). (B) ROC curve and classification performance of the final model on the test set. The x-axis represents the FPR, and the y-axis represents the TPR. AUC: area under the curve; GBM: gradient boosting machine; ROC: receiver operating characteristic.



**Table 2.** The 13 optimal features for distinguishing between cold and hot syndromes in analysis.

Clinical features	Cold (n=97), median (IQR)	Hot (n=285), median (IQR)	P value
Temperature (°C)	36.600 (36.500-36.950)	36.800 (36.500-36.95)	.02
Red cell distribution width (%)	44.500 (42.800-46.850)	43.400 (41.80-45.100)	<.001
Creatinine (μmol/L)	63.000 (54.000-78.000)	61.000 (51.000-73.500)	.13
Total bilirubin (μmol/L)	9.400 (7.350-12.450)	9.000 (7.200-11.550)	.31
Globulin (g/L)	27.500 (24.150-30.750)	26.600 (23.600-29.600)	.18
C-reactive protein (mg/L)	2.670 (1.355-6.530)	2.670 (0.865-7.235)	.41
Unconjugated bilirubin (μmol/L)	5.400 (4.000-6.900)	5.200 (3.900-6.800)	.28
White blood cell (×10 <sup>9</sup> /L)	5.300 (4.450-6.400)	5.100 (4.100-6.250)	.22
Neutrophil percentage (%)	62.000 (56.000-68.500)	58.000 (50.500-67.000)	.01
Aspartate transaminase/alanine transaminase (%)	1.200 (0.800-1.445)	1.100 (0.800-1.500)	.68
Total cholesterol (mmol/L)	3.900 (3.630-4.400)	3.700 (3.190-4.365)	.003
Thrombocytocrit (%)	0.240 (0.210-0.290)	0.240 (0.200-0.290)	<.001
Age (years)	47.000 (34.000-67.000)	37.000 (24.000-57.000)	<.001

**Figure 7.** Violin plots showing the relationships between cold and hot syndromes and the expression levels of optimal clinical features. The horizontal axis represents the different syndromes (cold and hot), while the vertical axis indicates the expression levels of the corresponding clinical features. The plots illustrate the distribution and density of feature values, highlighting differences between the 2 syndrome groups. AST/ALT: aspartate transaminase/alanine transaminase; CREA: creatinine; CRP: C-reactive protein; GLO: globulin; IBIL: unconjugated bilirubin; NEU%: neutrophil percentage; ns: not significant; PCT: thrombocytocrit; RDW-SD: red cell distribution width-SD; T: temperature; TBIL: total bilirubin; TCHO: total cholesterol; WBC: white blood cell; \* $P < .05$ , \*\* $P < .01$  versus the cold.



## Interactive Web Server

To enhance user experience, an interactive web server named ACHVP was developed in this study, allowing users to test, explore, and experience the proposed method. The platform is designed to be highly convenient and user-friendly—users simply input relevant numerical data into designated textboxes and click the “Submit” button. The system then performs comprehensive data processing and analysis, presenting the predicted syndrome classification results on the output interface. The home page of the web server, developed for identifying cold and hot syndrome patterns in patients with viral pneumonia, is illustrated in Figure 8. The server is deployed in a Linux environment. The

backend is developed in Python using the FastAPI framework to ensure efficient data processing and application programming interface management, while the front end is built with JavaScript and the Vue.js framework for a responsive and interactive user interface. Users can access the web server via any standard web browser, such as Google Chrome, Mozilla Firefox, or Safari, using the link specified in [34]. This web-based tool demonstrates significant potential as a practical diagnostic aid, facilitating early detection of cold and hot syndromes in patients with viral pneumonia. It can support TCM practitioners in syndrome differentiation and guide more precise Chinese herbal medicine prescriptions in clinical practice.

**Figure 8.** Web server of the ACHVP method.

CPPDD

HOME SERVICES RESOURCES

ACHVP

AI-assisted discrimination of the Cold or Hot Syndrome of Traditional Chinese Medicine for viral pneumonia based on clinical features

HOME DOCUMENT CONTACT US

Please input single sample data in the text form:

**Blood indicators :**

\* White Blood Cell Count (WBC) 3.6 10<sup>9</sup>/L

\* Percentage of Neutrophils (NEU) 74 %

\* Red cell distribution width standard deviation(RDW-SD) 42.1 %

\* Platelet Hematocrit(PCT) 0.19 %

\* C-reactive protein(CRP) 2.87 mg/L

**Biochemical indicators :**

\* glutamic oxaloacetic transaminase/glutamic pyruvic transaminase (AST/ALT) 0.8 μmol/L

\* Total Bilirubin(TBIL) 15.3 μmol/L

\* Indirect bilirubin (IBIL) 8.2 μmol/L

\* Globulin(GLO) 24.6 g/L

\* Creatinine(CREA) 69 μmol/L

\* Total Cholesterol (TCHO) 3.7 mmol/L

**Baseline :**

\* Age 43

\* Temperature(T) 38.2 °C

Or upload multiple samples (Please strictly adherence to the csv format as [sample file](#)):

Drop file here or click to upload

Submit Reset

## Discussion

### Principal Findings

Dialectical treatment is a fundamental principle in TCM for diagnosing and treating diseases, with cold and hot syndromes serving as 2 key components for identifying the nature of illness [11,35-38]. In recent years, the study of TCM cold and hot syndrome has become a focal point in disease research. Numerous studies have shown that cold and hot syndromes also exhibit major TCM diagnostic patterns in various viral pulmonary diseases, including COVID-19 and SARS [8,14]. Based on these syndrome classifications, TCM has demonstrated a prominent role in alleviating symptoms, shortening treatment duration, and reducing the progression to severe pneumonia in patients with viral pneumonia [39-43]. For example, formulas such as Hanshiyi for cold syndrome [44] and Jinhua Qinggan Granules for hot syndrome have shown favorable clinical efficacy [45]. Therefore, the rapid, accurate, and specific identification of cold or hot syndrome is essential for the effective application of TCM in the treatment of viral pulmonary diseases. However, the abstract and complex nature of TCM syndrome differentiation poses challenges in accurately distinguishing between cold and hot syndromes in clinical practice. Machine learning offers a promising solution to this challenge by supporting clinicians in syndrome differentiation and facilitating more precise treatment decisions. Consequently, there is an urgent need to integrate TCM theories of cold and hot syndromes with modern laboratory diagnostics and machine learning approaches to advance the modernized differentiation and treatment of viral pneumonia.

In this study, 8 machine learning algorithms were applied to develop and evaluate a model for differentiating TCM cold and hot syndromes in viral pneumonia, using clinical data from 1484 patients across 2 medical centers. Among the models incorporating both TCM and modern medical features, the GBM model demonstrated superior learning capability, achieving an AUC of 0.8329. This indicates that GBM was particularly effective in distinguishing cold and hot syndromes based on a given dataset. The “feature importance” method of the GBM model was subsequently used to rank all features, from which the top 13 features were selected based on the highest AUC value. To validate the model’s effectiveness and generalizability, both internal and external test sets were used, each demonstrating strong discriminatory performance. As a result, the optimal model was constructed using the GBM algorithm and 13 selected features (T, RDW-SD, CREA, TBIL, GLO, CRP, IBIL, WBC, NEU%, AST/ALT, TCHO, PCT, and AGE), achieving an AUC of 0.7788. These findings suggest that effectively combining modern medical indicators, primarily based on laboratory tests, with TCM features primarily based on clinical observation may offer a viable pathway toward the modernization of TCM syndrome differentiation. This approach provides a foundation for the objectification and standardization of TCM diagnostic practices. Furthermore, the proposed model holds promise as a practical tool to assist clinicians in accurately identifying cold and hot syndromes in viral pneumonia, thereby informing appropriate and individualized treatment strategies.

### Comparison With Prior Work

Machine learning, as an advanced artificial intelligence technology, has demonstrated broad potential in the field of “artificial intelligence+TCM diagnoses.” For example,

Zhang et al [46] proposed a comprehensive model using convolutional neural networks to classify 187 types of TCM diseases. Wen et al [47] also applied deep neural networks to identify tongue image data and determine patients' constitutions, effectively addressing the imbalance problem among constitution categories. However, limited research has focused on identifying TCM cold and hot syndromes in viral pneumonia by integrating both TCM theory and modern medicine through machine learning approaches. Although some scholars have attempted to explore the cold and hot attributes of diseases to guide clinical treatment, most studies have been limited to network analysis, clustering, or laboratory testing, without incorporating artificial intelligence methods. For instance, Li et al's [11] network analysis revealed that the leptin biomarker indicated low levels of energy metabolism in patients with cold syndrome, while the human monocyte chemoattractant protein-1 (CCL2/MCP1) biomarker suggested intensified immune regulation in patients with hot syndrome, based on a study involving patients with chronic superficial gastritis and chronic atrophic gastritis. Wang et al [13] differentiated heat and cold patterns in rheumatoid arthritis through cluster and factor analyses to guide clinical medication. Wu et al [48] investigated TCM cold and hot constitutions using pulse wave parameters such as augmentation index and heart rate variability. Although some individual studies have applied artificial intelligence to differentiate cold and hot properties, most efforts have been limited to identifying the properties of Chinese herbal medicines. For example, Lin et al [49] developed a classification strategy for the cold and hot properties of Chinese herbal medicines using artificial intelligence and biological experiments. Wei et al [50] applied machine learning

techniques to intelligently identify the cold-hot nature of Chinese herbal medicines based on the similarity of their volatile oil ingredients. Therefore, this study innovatively constructs a model for differentiating cold and hot syndromes in viral pneumonia using machine learning approaches that integrate TCM and modern medicine, aiming to assist in diagnosis and medication guidance for viral pneumonia in clinical practice.

It is worth noting that in this study, we collected a large number of clinical features, including patients' symptoms, signs, laboratory examination results, and other relevant information, which were essential for training and testing the machine learning models. As shown in Table 3, based on the model training results using only 19 TCM features, we found that the AUC values of GBM, LR, RF, XGB, LGB, RIDGE, LASSO, and SVM were 0.6942, 0.6317, 0.6783, 0.6787, 0.6055, 0.6151, 0.6145, and 0.5874, respectively. Meanwhile, the study also attempted to build models using only 70 laboratory results. The AUC values for GBM, LR, RF, XGB, LGB, RIDGE, LASSO, and SVM were 0.6799, 0.6282, 0.6797, 0.6910, 0.7345, 0.6637, 0.6157, and 0.6618, respectively. Among them, the GBM model constructed using TCM features achieved an AUC of 0.6942, outperforming the GBM model constructed using laboratory features, which had an AUC of 0.6799. Internal validation and the confusion matrix results also demonstrated performance consistent with the training results, as shown in Multimedia Appendix 4. These findings suggest that the models possess auxiliary diagnostic value for differentiating TCM cold and hot syndromes in viral pneumonia.

**Table 3.** Comparison of model results by 5-fold cross-validation based on different features of traditional Chinese medicine and modern medicine.

Features and evaluation	Gradient boosting machine	Logistic regression	Random forest	Extreme gradient boosting	Light gradient boosting machine	Ridge regression	Least absolute shrinkage and selection operator	Support vector machine
Traditional Chinese medicine								
Area under the curve	0.6942	0.6317	0.6783	0.6787	0.6055	0.6151	0.6145	0.5874
Accuracy	0.6851	0.5312	0.5509	0.6031	0.5155	0.5180	0.5049	0.7508
Sensitivity	0.7365	0.4737	0.5077	0.5908	0.4389	0.4258	0.4117	0.9928
Specificity	0.5804	0.7316	0.7457	0.7006	0.7233	0.7854	0.7723	0.2601
Matthews correlation coefficient	0.2737	0.1689	0.2145	0.2400	0.1332	0.1821	0.1563	0.2385
Modern medicine								
Area under the curve	0.6799	0.6282	0.6797	0.6910	0.7345	0.6637	0.6157	0.6618
Accuracy	0.7508	0.7508	0.7541	0.7672	0.7803	0.7639	0.7180	0.7639
Sensitivity	0.9441	0.9185	0.9742	0.9400	0.9314	0.8968	0.8539	1.0000
Specificity	0.1257	0.2105	0.0410	0.2067	0.2914	0.3324	0.2762	0.0000
Matthews correlation coefficient	0.1269	0.1585	0.0335	0.1927	0.2871	0.2672	0.1752	0.0000
Traditional Chinese medicine and modern medicine								



Features and evaluation	Gradient boosting machine	Logistic regression	Random forest	Extreme gradient boosting	Light gradient boosting machine	Ridge regression	Least absolute shrinkage and selection operator	Support vector machine
Area under the curve	0.8329	0.6477	0.7272	0.7546	0.7693	0.6970	0.6974	0.6514
Accuracy	0.8000	0.6787	0.6689	0.6984	0.7246	0.7475	0.7475	0.4459
Sensitivity	0.8283	0.7365	0.6582	0.7172	0.7158	0.8388	0.8430	0.2975
Specificity	0.7016	0.5187	0.7066	0.6326	0.7327	0.4764	0.4656	0.9018
Matthews correlation coefficient	0.5043	0.2341	0.3190	0.3177	0.4068	0.3364	0.3407	0.2161

Based on “feature importance” or “coef,” the 8 models were further used to evaluate the significance of TCM features in distinguishing cold and hot syndromes of viral pneumonia. We found that, according to the significance values, there were 7 common features with importance values equal to or greater than the median in the GBM, RF, XGB, and LGB models, namely, fatigue, sore throat, fever, expectoration, cough, body pain, and diminished sense of smell. Interestingly, sore throat and fever also appeared among the features with importance values equal to or greater than the median in LR, RIDGE, LASSO, and SVM models. Based on this, we believe that fatigue, sore throat, fever, expectoration, cough, body pain, and diminished sense of smell may be the primary distinguishing factors for differentiating between cold and hot syndromes in the TCM diagnosis of viral pneumonia, which aligns with the key symptomatology commonly used by clinical TCM practitioners [51,52]. Therefore, more attention should be given to these features, especially sore throat and fever, in the discrimination of cold and hot syndromes in viral pneumonia, as shown in [Multimedia Appendix 5](#). Unfortunately, the AUC values of these models did not exceed 0.7. We boldly speculate that this might be due to the quantification of TCM features not aligning with modern medical standards. Specifically, TCM features lack precise numerical values and are typically categorized as none, mild, moderate, or severe. Moreover, the identification of TCM features primarily relies on the subjective judgment of TCM practitioners and years of clinical experience, leading to a lack of standardization and objectivity. These factors may introduce noise and reduce the effectiveness of model training. Therefore, it is essential to develop objective indicators for TCM syndrome differentiation, further supporting the original intent of our study. In subsequent experiments, we integrated modern medicine features with TCM indicators and were pleasantly surprised to find that, compared with models trained solely on TCM features, those based on integrated TCM and modern medical features demonstrated superior performance in distinguishing cold and hot syndromes in viral pneumonia across 8 machine learning algorithms. This suggests that modern medical features can enhance the predictive value of TCM features.

In addition, the best model we developed indicated that the identification of cold and hot syndromes in viral pneumonia is closely related to 13 features: temperature, RDW-SD,

CREA, TBIL, GLO, CRP, IBIL, WBC, NEU%, AST/ALT, TCHO, PCT, and AGE. Numerous studies have reported that temperature (T), CRP, and WBC hold statistical significance in distinguishing TCM cold and hot syndromes in rheumatoid arthritis [53]. Furthermore, features such as AGE, NEU%, RDW-SD, and CRP are strongly associated with the severity of viral pneumonia and the development of chronic lung sequelae. For example, a study by Liu et al [33] suggested that AGE and NEU% could serve as predictive factors for the severity of viral pneumonia. Shen et al [31] also reported that RDW-SD is a key indicator for assessing the severity of viral pneumonia. Additionally, CRP and AGE play crucial roles in the diagnosis of viral pneumonia [54,55], which is consistent with the findings of this study.

Limitations

The accuracy of traditional TCM diagnosis of cold and hot syndromes largely depends on the skills and years of clinical experience of the practitioner. Subjective factors play a decisive role in this process. Compared with traditional TCM methods, the model developed in this study offers a more objective approach to differentiating cold and hot syndromes in viral pneumonia. It provides valuable decision support and medication guidance in clinical practice, especially for practitioners with limited TCM experience. This shift toward an objective, data-driven approach helps reduce reliance on subjective judgment, improves diagnostic accuracy, and ensures greater consistency in clinical decision-making. Additionally, the model can serve as a practical tool for TCM practitioners, especially in complex cases where traditional diagnostic methods may fall short.

Of course, this study has some limitations. First, it was a retrospective study, and all patient history data were obtained from the internal electronic medical record systems of the Second People’s Hospital of Lanzhou and Lanzhou Heavy Ion Hospital. This may have introduced inaccuracies in the collection and recording of medical history, making it difficult to avoid information bias. Moreover, due to the limited clinical sample size, the conclusions, although statistically sound, require further validation and refinement in real-world applications to enhance the model’s accuracy and stability.

In recent years, emerging large language models (LLMs) have demonstrated remarkable capabilities in natural language processing, knowledge integration, and complex reasoning [56]. Given the nature of TCM theory, which often involves complex textual descriptions and implicit knowledge, LLMs could potentially offer new perspectives for model development in this study. These models may better understand and integrate the rich, diverse information contained in TCM literature, clinical records, and expert knowledge, thereby enriching the feature set and enhancing the interpretability of the models. For instance, LLMs could analyze the semantic relationships among TCM symptoms, syndromes, and treatment methods, helping to extract more meaningful features that may be difficult to identify using traditional machine learning approaches. Moreover, LLMs could assist in standardizing and structuring the subjective and qualitative information in TCM, thereby reducing the impact of inconsistent data representation. However, applying LLMs also poses challenges, such as ensuring the reliability of the extracted knowledge, addressing potential biases inherent in pretrained models, and efficiently integrating LLM-derived features with existing laboratory test data. In future work, we plan to collect more data to establish a local database and explore the integration of LLMs with traditional machine learning algorithms. By fully leveraging the advantages of both approaches, we aim to further enhance the performance of the TCM cold and hot syndrome differentiation model and advance the modernization and standardization of TCM diagnosis.

## Conclusion

This study aimed to establish a differentiation model for TCM cold and hot syndromes in patients with viral pneumonia using machine learning by exploring correlations between cold/hot syndromes and features from both TCM and modern medicine (such as general information, TCM characteristics, blood gas values, viral pneumonia indicators, biochemical indicators, routine blood test indicators, and coagulation indicators). We found that machine learning algorithms based on integrated TCM and modern medicine features outperformed models that used only TCM or only modern medicine features. Specifically, a model composed of 13 optimal features (T, RDW-SD, CREA, TBIL, GLO, CRP, IBIL, WBC, NEU%, AST/ALT, TCHO, PCT, and AGE) and the GBM algorithm effectively differentiated TCM syndromes. This study is the first to integrate the TCM cold and hot syndrome theory with laboratory-based modern medical tests using machine learning. By establishing the relationship between TCM syndrome theory (cold and hot syndromes) and indicators from both TCM and modern medicine, we aim to provide a new auxiliary diagnostic method for the clinical differentiation of cold and hot syndromes. This can support practitioners in making comprehensive diagnoses and identifying effective Chinese herbal medicine treatments, while also offering new insights for the modernization and scientific interpretation of TCM syndrome differentiation theory to guide treatment.

## Acknowledgments

This study was supported by the Major Science and Technology Project of Gansu Province (grant 22ZD1FA001), the Provincial University Industry Support Project in Gansu (grant 2020C-36), the Young Elite Scientists Sponsorship Program by the Gansu Association for Science and Technology (grant GXH20210611-02), and the 2023 Longyuan Youth Innovation and Entrepreneurship Talent Project (Gan Group Tong Zi [2023] No. 20).

## Data Availability

The datasets generated in this study are not publicly available due to hospital institutional review board regulations and patient privacy policies, but deidentified datasets are available from the corresponding author (YL) upon reasonable request. These deidentified datasets would include the model training set and the test set to facilitate independent model replications.

## Authors' Contributions

XJ was responsible for formal analysis, investigation, project administration, writing the original draft, and reviewing and editing the manuscript. YW and JW contributed to the investigation, methodology, resources, supervision, validation, writing the original draft, and reviewing and editing. YH and QG were involved in formal analysis, investigation, methodology, writing the original draft, and reviewing and editing. LS and JZ participated in the investigation and methodology. JL, LL, and ZZ contributed to the investigation, methodology, resources, and supervision. SL was responsible for conceptualization, methodology, project administration, and reviewing and editing the manuscript. YL contributed to conceptualization, formal analysis, methodology, supervision, and reviewing and editing. All authors have read and approved the final manuscript.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

All clinical features information of patients with viral pneumonia.

[\[DOCX File \(Microsoft Word File\), 18 KB-Multimedia Appendix 1\]](#)

## Multimedia Appendix 2

The traditional Chinese medicine symptom scoring scale for patients with viral pneumonia.

[\[DOCX File \(Microsoft Word File\), 19 KB-Multimedia Appendix 2\]](#)

### Multimedia Appendix 3

The optimal features for the TCM cold and hot syndrome discrimination of viral pneumonia. AGE: age; AST/ALT: aspartate transaminase/alanine transaminase; CREA: creatinine; CRP: C-reactive protein; GLO: globulin; IBIL: unconjugated bilirubin; NEU%: neutrophil percentage; PCT: thrombocytocrit; RDW-SD: red cell distribution width-SD; T: temperature; TBIL: total bilirubin; TCHO: total cholesterol; WBC: white blood cell.

[\[PNG File \(Portable Network Graphics File\), 588 KB-Multimedia Appendix 3\]](#)

### Multimedia Appendix 4

The performance evaluation of the cold and hot syndrome identification model based on traditional Chinese medicine (TCM) features. (A) The performance evaluation's radar chart of the 8 models based on TCM features. (B) The confusion matrix analysis of the 8 models based on TCM features.

[\[DOCX File \(Microsoft Word File\), 351 KB-Multimedia Appendix 4\]](#)

### Multimedia Appendix 5

Heat diagram of traditional Chinese medicine features on 8 screened models (the values of gradient boosting machine, random forest, extreme gradient boosting, and light gradient boosting were obtained using feature importance, and those of ridge regression, least absolute shrinkage and selection operator, support vector machine, and logistic regression were obtained using coef).

[\[DOCX File \(Microsoft Word File\), 1324 KB-Multimedia Appendix 5\]](#)

## References

1. Gan X, Shu Z, Wang X, et al. Network medicine framework reveals generic herb-symptom effectiveness of traditional Chinese medicine. *Sci Adv*. Oct 27, 2023;9(43):eadh0215. [doi: [10.1126/sciadv.adh0215](https://doi.org/10.1126/sciadv.adh0215)] [Medline: [37889962](https://pubmed.ncbi.nlm.nih.gov/37889962/)]
2. Dong J. The relationship between traditional Chinese medicine and modern medicine. *Evid Based Complement Alternat Med*. 2013;2013:153148. [doi: [10.1155/2013/153148](https://doi.org/10.1155/2013/153148)] [Medline: [23983772](https://pubmed.ncbi.nlm.nih.gov/23983772/)]
3. Xu W, Zhang Z, Lu Y, Li M, Li J, Tao W. Traditional Chinese medicine Tongxie Yaofang treating irritable bowel syndrome with diarrhea and type 2 diabetes mellitus in rats with liver-depression and spleen-deficiency: a preliminary study. *Front Nutr*. 2022;9:968930. [doi: [10.3389/fnut.2022.968930](https://doi.org/10.3389/fnut.2022.968930)] [Medline: [36438735](https://pubmed.ncbi.nlm.nih.gov/36438735/)]
4. Song YN, Chen J, Cai FF, et al. A metabolic mechanism analysis of Fuzheng-Huayu formula for improving liver cirrhosis with traditional Chinese medicine syndromes. *Acta Pharmacol Sin*. Jun 2018;39(6):942-951. [doi: [10.1038/aps.2017.101](https://doi.org/10.1038/aps.2017.101)] [Medline: [29072258](https://pubmed.ncbi.nlm.nih.gov/29072258/)]
5. Liu X, Li J, Wang Y, et al. To study the protective effect of Huangqi Baihe granules on radiation brain injury based on network pharmacology and experiment. *J Ethnopharmacol*. Oct 2023;315:116610. [doi: [10.1016/j.jep.2023.116610](https://doi.org/10.1016/j.jep.2023.116610)]
6. Huang K, Zhang P, Zhang Z, et al. Traditional Chinese medicine (TCM) in the treatment of COVID-19 and other viral infections: efficacies and mechanisms. *Pharmacol Ther*. Sep 2021;225:107843. [doi: [10.1016/j.pharmthera.2021.107843](https://doi.org/10.1016/j.pharmthera.2021.107843)] [Medline: [33811957](https://pubmed.ncbi.nlm.nih.gov/33811957/)]
7. Ni L, Chen L, Huang X, et al. Combating COVID-19 with integrated traditional Chinese and Western medicine in China. *Acta Pharm Sin B*. Jul 2020;10(7):1149-1162. [doi: [10.1016/j.apsb.2020.06.009](https://doi.org/10.1016/j.apsb.2020.06.009)] [Medline: [32834946](https://pubmed.ncbi.nlm.nih.gov/32834946/)]
8. Li X, Liu W, He W, Xie Y, Li J. Meta-analysis of single-group rate of the distribution of traditional Chinese medicine syndromes in 2 139 patients with coronavirus disease 2019. *Zhonghua Wei Zhong Bing Ji Jiu Yi Xue*. Jun 2020;32(6):664-670. [doi: [10.3760/cma.j.cn121430-202000509-00372](https://doi.org/10.3760/cma.j.cn121430-202000509-00372)] [Medline: [32684209](https://pubmed.ncbi.nlm.nih.gov/32684209/)]
9. Wang SC, Sun YQ, Bian GB, Xu L, Ma R, Wang MQ. Clinical study on Qingfei oral liquid treating 507 cases of phlegm-heat obstructing lung syndrome in children virus pneumonia. *World Chin Med*. 2016;11(9):1649-1653, 1658. [doi: [10.3969/j.issn.1673-7202.2016.09.001](https://doi.org/10.3969/j.issn.1673-7202.2016.09.001)]
10. Deng TT. A discussion on diagnosis and treatment of SARS by traditional Chinese medicine. *New Chin Med*. 2003(6):3-5. [doi: [10.13457/j.cnki.jncm.2003.06.001](https://doi.org/10.13457/j.cnki.jncm.2003.06.001)]
11. Li R, Ma T, Gu J, Liang X, Li S. Imbalanced network biomarkers for traditional Chinese medicine syndrome in gastritis patients. *Sci Rep*. 2013;3:1543. [doi: [10.1038/srep01543](https://doi.org/10.1038/srep01543)] [Medline: [23529020](https://pubmed.ncbi.nlm.nih.gov/23529020/)]
12. Zhou N, Yang Y, Li K, et al. Integrating strategies of chemistry, biochemistry and metabolomics for characterization of the medication principle of "treating cold/heat syndrome with hot/cold herbs". *J Ethnopharmacol*. Jul 15, 2019;239:111899. [doi: [10.1016/j.jep.2019.111899](https://doi.org/10.1016/j.jep.2019.111899)] [Medline: [31039429](https://pubmed.ncbi.nlm.nih.gov/31039429/)]
13. Wang W, Guan J, Li Z, Wang X. Rheumatoid arthritis characteristics and classification of heat and cold patterns—an observational study. *Heliyon*. Feb 2023;9(2):e13439. [doi: [10.1016/j.heliyon.2023.e13439](https://doi.org/10.1016/j.heliyon.2023.e13439)] [Medline: [36873147](https://pubmed.ncbi.nlm.nih.gov/36873147/)]

14. Feng Z, Xie Y, Chun L, Li J. Study on traditional Chinese medicine common syndrome characteristic of coronavirus disease 2019 based on latent structure combined with system clustering analysis. *Zhonghua Wei Zhong Bing Ji Jiu Yi Xue*. May 2020;32(5):537-543. [doi: [10.3760/cma.j.cn121430-20200506-00620](https://doi.org/10.3760/cma.j.cn121430-20200506-00620)] [Medline: [32576343](#)]
15. Wu J, Yang Y, Cheng L, et al. GC discrimination: identification of gastric cancer based on a milliliter of blood. *Brief Bioinformatics*. Jan 18, 2021;22(1):536-544. [doi: [10.1093/bib/bbaa006](https://doi.org/10.1093/bib/bbaa006)]
16. Jain P, Yedukondalu J, Chhabra H, Chauhan U, Sharma LD. EEG-based detection of cognitive load using VMD and LightGBM classifier. *Int J Mach Learn Cyber*. Sep 2024;15(9):4193-4210. [doi: [10.1007/s13042-024-02142-2](https://doi.org/10.1007/s13042-024-02142-2)]
17. Panda P, Bisoy SK, Panigrahi A, et al. BIMSSA: enhancing cancer prediction with salp swarm optimization and ensemble machine learning approaches. *Front Genet*. Jan 6, 2025;15:1491602. [doi: [10.3389/fgene.2024.1491602](https://doi.org/10.3389/fgene.2024.1491602)] [Medline: [39834551](#)]
18. Houkan A, Sahoo AK, Gochhayat SP, et al. Enhancing security in industrial IoT networks: machine learning solutions for feature selection and reduction. *IEEE Access*. 2024;12:160864-160883. [doi: [10.1109/ACCESS.2024.3481459](https://doi.org/10.1109/ACCESS.2024.3481459)]
19. Huang WT, Hung HH, Kao YW, et al. Application of neural network and cluster analyses to differentiate TCM patterns in patients with breast cancer. *Front Pharmacol*. 2020;11:670. [doi: [10.3389/fphar.2020.00670](https://doi.org/10.3389/fphar.2020.00670)] [Medline: [32457636](#)]
20. Zhuo YY, Wu JM, Kuang L, et al. The discriminative efficacy of retinal characteristics on two traditional Chinese syndromes in association with ischemic stroke. *Evid Based Complement Alternat Med*. 2020;2020:6051831. [doi: [10.1155/2020/6051831](https://doi.org/10.1155/2020/6051831)] [Medline: [32308711](#)]
21. Kwon O, Na W, Kang H, et al. Electronic medical record-based machine learning approach to predict the risk of 30-day adverse cardiac events after invasive coronary treatment: machine learning model development and validation. *JMIR Med Inform*. May 11, 2022;10(5):e26801. URL: <https://medinform.jmir.org/2022/5/e26801/> [Accessed 2025-06-24] [doi: [10.2196/26801](https://doi.org/10.2196/26801)] [Medline: [35544292](#)]
22. Kim S, Park J, Son Y, et al. Development and validation of a machine learning algorithm for predicting diabetes retinopathy in patients with type 2 diabetes: algorithm development study. *JMIR Med Inform*. Feb 7, 2025;13:e58107. URL: <https://medinform.jmir.org/2025/1/e58107> [Accessed 2025-06-24] [doi: [10.2196/58107](https://doi.org/10.2196/58107)] [Medline: [39924304](#)]
23. Xu Q, Cai X, Yu R, et al. Machine learning-based risk factor analysis and prediction model construction for the occurrence of chronic heart failure: health ecologic study. *JMIR Med Inform*. Jan 31, 2025;13:e64972. URL: <https://medinform.jmir.org/2025/1/e64972> [Accessed 2025-06-24] [doi: [10.2196/64972](https://doi.org/10.2196/64972)] [Medline: [39889299](#)]
24. Chen X, Chen H, Nan S, Kong X, Duan H, Zhu H. Dealing with missing, imbalanced, and sparse features during the development of a prediction model for sudden death using emergency medicine data: machine learning approach. *JMIR Med Inform*. Jan 20, 2023;11:e38590. URL: <https://medinform.jmir.org/2023/1/e38590> [Accessed 2025-06-24] [doi: [10.2196/38590](https://doi.org/10.2196/38590)] [Medline: [36662548](#)]
25. He JC. TCM Diagnostics [Book in Chinese]. People's Medical Publishing House; 2017. ISBN: 978-7-117-23280-7
26. Zheng XY. Guiding Principle of Clinical Research on New Drugs of Traditional Chinese Medicine (Trial). China Medical Science and Technology Press; 2002. ISBN: 750672586X
27. Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Statist*. 2001;29(5):44. [doi: [10.1214/aos/1013203451](https://doi.org/10.1214/aos/1013203451)]
28. Lu J. A frequency-adaptive multi-objective suspension control strategy. *J Dyn Syst Meas Control*. Sep 1, 2004;126(3):700-707. [doi: [10.1115/1.1789979](https://doi.org/10.1115/1.1789979)]
29. Goplen NP, Wu Y, Son YM, et al. Tissue-resident CD8<sup>+</sup> T cells drive age-associated chronic lung sequelae after viral pneumonia. *Sci Immunol*. Nov 6, 2020;5(53):eabc4557. [doi: [10.1126/sciimmunol.abc4557](https://doi.org/10.1126/sciimmunol.abc4557)] [Medline: [33158975](#)]
30. Nie X, Cai Y, Liu J, et al. Mortality prediction in cerebral hemorrhage patients using machine learning algorithms in intensive care units. *Front Neurol*. 2020;11:610531. [doi: [10.3389/fneur.2020.610531](https://doi.org/10.3389/fneur.2020.610531)] [Medline: [33551969](#)]
31. Shen C, Di Y, Wang W, et al. Value of red blood cell distribution width in evaluating the severity of illness of novel coronavirus Delta variant. *Zhonghua Wei Zhong Bing Ji Jiu Yi Xue*. May 2022;34(5):475-480. [doi: [10.3760/cma.j.cn121430-20220214-00127](https://doi.org/10.3760/cma.j.cn121430-20220214-00127)] [Medline: [35728847](#)]
32. Castro-Castro MJ, García-Tejada L, Arbiol-Roca A, et al. Dynamic profiles and predictive values of some biochemical and haematological quantities in COVID-19 inpatients. *Biochem Med (Zagreb)*. Feb 15, 2022;32(1):010706. [doi: [10.11613/BM.2022.010706](https://doi.org/10.11613/BM.2022.010706)] [Medline: [35210926](#)]
33. Liu Y, Yang Y, Zhang C, et al. Clinical and biochemical indexes from 2019-nCoV infected patients linked to viral loads and lung injury. *Sci China Life Sci*. Mar 2020;63(3):364-374. [doi: [10.1007/s11427-020-1643-8](https://doi.org/10.1007/s11427-020-1643-8)] [Medline: [32048163](#)]
34. ACHVP. CPPDD. URL: <http://cppdd.cn:9999/#/service/ACHVP> [Accessed 2025-06-12]
35. Wang Y, Zhou S, Wang M, et al. UHPLC/Q-TOFMS-based metabolomics for the characterization of cold and hot properties of Chinese materia medica. *J Ethnopharmacol*. Feb 2016;179:234-242. [doi: [10.1016/j.jep.2015.12.061](https://doi.org/10.1016/j.jep.2015.12.061)]



36. Wang Z, Li L, Song M, Yan J, Shi J, Yao Y. Evaluating the traditional Chinese medicine (TCM) officially recommended in China for COVID-19 using ontology-based side-effect prediction framework (OSPF) and deep learning. *J Ethnopharmacol*. May 2021;272:113957. [doi: [10.1016/j.jep.2021.113957](https://doi.org/10.1016/j.jep.2021.113957)]
37. Xia F, Liu C, Wan JB. Characterization of the cold and hot natures of raw and processed Rehmanniae Radix by integrated metabolomics and network pharmacology. *Phytomedicine*. Aug 2020;74:153071. [doi: [10.1016/j.phymed.2019.153071](https://doi.org/10.1016/j.phymed.2019.153071)] [Medline: [31537418](https://pubmed.ncbi.nlm.nih.gov/31537418/)]
38. Ma T, Tan C, Zhang H, Wang M, Ding W, Li S. Bridging the gap between traditional Chinese medicine and systems biology: the connection of cold syndrome and NEI network. *Mol Biosyst*. Apr 2010;6(4):613-619. [doi: [10.1039/b914024g](https://doi.org/10.1039/b914024g)] [Medline: [20237638](https://pubmed.ncbi.nlm.nih.gov/20237638/)]
39. Wang Y, Jin X, Fan Q, et al. Deciphering the active compounds and mechanisms of HSBDF for treating ALI via integrating chemical bioinformatics analysis. *Front Pharmacol*. 2022;13:879268. [doi: [10.3389/fphar.2022.879268](https://doi.org/10.3389/fphar.2022.879268)] [Medline: [35721141](https://pubmed.ncbi.nlm.nih.gov/35721141/)]
40. Wei J, Sun J, Zeng J, et al. Precise investigation of the efficacy of multicomponent drugs against pneumonia infected with influenza virus. *Front Pharmacol*. 2021;12:604009. [doi: [10.3389/fphar.2021.604009](https://doi.org/10.3389/fphar.2021.604009)] [Medline: [34867309](https://pubmed.ncbi.nlm.nih.gov/34867309/)]
41. Wang ZY, Fu SZ, Xu L, et al. Impact of Shenfu injection on a composite of organ dysfunction development in critically ill patients with coronavirus disease 2019 (COVID-19): a structured summary of a study protocol for a randomized controlled trial. *Trials*. Aug 24, 2020;21(1):738. [doi: [10.1186/s13063-020-04677-5](https://doi.org/10.1186/s13063-020-04677-5)] [Medline: [32831151](https://pubmed.ncbi.nlm.nih.gov/32831151/)]
42. Wu M, Zhao Y, Dong X, et al. Artificial intelligence-based preoperative prediction system for diagnosis and prognosis in epithelial ovarian cancer: a multicenter study. *Front Oncol*. 2022;12:975703. [doi: [10.3389/fonc.2022.975703](https://doi.org/10.3389/fonc.2022.975703)] [Medline: [36212430](https://pubmed.ncbi.nlm.nih.gov/36212430/)]
43. Shi N, Liu B, Liang N, et al. Association between early treatment with Qingfei Paidu decoction and favorable clinical outcomes in patients with COVID-19: a retrospective multicenter cohort study. *Pharmacol Res*. Nov 2020;161:105290. [doi: [10.1016/j.phrs.2020.105290](https://doi.org/10.1016/j.phrs.2020.105290)] [Medline: [33181320](https://pubmed.ncbi.nlm.nih.gov/33181320/)]
44. Zheng Y, Jin D, Lin J, et al. Understanding COVID-19 in Wuhan from the perspective of cold-dampness: clinical evidences and mechanisms. *Front Med (Lausanne)*. 2021;8:617659. [doi: [10.3389/fmed.2021.617659](https://doi.org/10.3389/fmed.2021.617659)] [Medline: [33693014](https://pubmed.ncbi.nlm.nih.gov/33693014/)]
45. Zhang D, Zhang B, Lv JT, Sa RN, Zhang XM, Lin ZJ. The clinical benefits of Chinese patent medicines against COVID-19 based on current evidence. *Pharmacol Res*. Jul 2020;157:104882. [doi: [10.1016/j.phrs.2020.104882](https://doi.org/10.1016/j.phrs.2020.104882)] [Medline: [32380051](https://pubmed.ncbi.nlm.nih.gov/32380051/)]
46. Zhang H, Ni W, Li J, Zhang J. Artificial intelligence-based traditional Chinese medicine assistive diagnostic system: validation study. *JMIR Med Inform*. Jun 15, 2020;8(6):e17608. URL: <https://medinform.jmir.org/2020/6/e17608/> [Accessed 2025-06-24] [doi: [10.2196/17608](https://doi.org/10.2196/17608)]
47. Wen G, Ma J, Hu Y, Li H, Jiang L. Grouping attributes zero-shot learning for tongue constitution recognition. *Artif Intell Med*. Sep 2020;109:101951. [doi: [10.1016/j.artmed.2020.101951](https://doi.org/10.1016/j.artmed.2020.101951)] [Medline: [34756217](https://pubmed.ncbi.nlm.nih.gov/34756217/)]
48. Wu HK, Ko YS, Lin YS, Wu HT, Tsai TH, Chang HH. The correlation between pulse diagnosis and constitution identification in traditional Chinese medicine. *Complement Ther Med*. Feb 2017;30:107-112. [doi: [10.1016/j.ctim.2016.12.005](https://doi.org/10.1016/j.ctim.2016.12.005)] [Medline: [28137520](https://pubmed.ncbi.nlm.nih.gov/28137520/)]
49. Lin L, Zhao P, Chen Z, et al. Identification strategy of cold and hot properties of Chinese herbal medicines based on artificial intelligence and biological experiments. *Chin Med J*. 2025;138(6):745-747. [doi: [10.1097/CM9.0000000000003509](https://doi.org/10.1097/CM9.0000000000003509)]
50. Wei G, Fu X, He X, et al. Cold-hot nature identification based on GC similarity analysis of Chinese herbal medicine ingredients. *RSC Adv*. Jul 27, 2021;11(42):26008-26015. [doi: [10.1039/d1ra04189d](https://doi.org/10.1039/d1ra04189d)] [Medline: [35479454](https://pubmed.ncbi.nlm.nih.gov/35479454/)]
51. Gu SF, Li Y, Zhang ZX, Jin L, Li MQ. Clinical observation on the combination of traditional Chinese medicine Master Li Jiren's experience formula and olistavir in the treatment of viral pneumonia. *GmjCM*. 2023;38(8):1550-1553. [doi: [10.3969/j.issn.1003-8914.2023.08.041](https://doi.org/10.3969/j.issn.1003-8914.2023.08.041)]
52. Sun TF, Wu JW, Gao Q. The effects of Qingfei Paidu decoction on the time of disappearance of major symptoms, serum D-dimer levels, and erythrocyte sedimentation rate in patients with viral pneumonia. *J Shenyang Pharm Univ*. 2021;38(S2):70-71.
53. Wang ZZ, Fang YF, Wang Y, et al. Logistic regression analysis of damp-heat and cold-damp impeding syndrome of rheumatoid arthritis: a perspective in Chinese medicine. *Chin J Integr Med*. Aug 2012;18(8):575-581. [doi: [10.1007/s11655-012-1172-1](https://doi.org/10.1007/s11655-012-1172-1)] [Medline: [22855033](https://pubmed.ncbi.nlm.nih.gov/22855033/)]
54. Qu JX, Gu L, Pu ZH, et al. Viral etiology of community-acquired pneumonia among adolescents and adults with mild or moderate severity and its relation to age and severity. *BMC Infect Dis*. Feb 22, 2015;15:89. [doi: [10.1186/s12879-015-0808-0](https://doi.org/10.1186/s12879-015-0808-0)] [Medline: [25812108](https://pubmed.ncbi.nlm.nih.gov/25812108/)]



55. Bekdas M, Goksugur SB, Sarac EG, Erkocoglu M, Demircioglu F. Neutrophil/lymphocyte and C-reactive protein/mean platelet volume ratios in differentiating between viral and bacterial pneumonias and diagnosing early complications in children. *Saudi Med J*. May 2014;35(5):442-447. [Medline: [24825803](#)]
56. Zhuang Y, Yu L, Jiang N, Ge Y. TCM-KLLaMA: Intelligent generation model for Traditional Chinese Medicine Prescriptions based on knowledge graph and large language model. *Comput Biol Med*. May 2025;189:109887. [doi: [10.1016/j.compbimed.2025.109887](#)] [Medline: [40056842](#)]

## Abbreviations

**ACC:** accuracy  
**ALT:** alanine transaminase  
**AST:** aspartate transaminase  
**AUC:** area under the curve  
**CREA:** creatinine  
**CRP:** C-reactive protein  
**FN:** false negative  
**FP:** false positive  
**GBM:** gradient boosting machine  
**GLO:** globulin  
**IBIL:** unconjugated bilirubin  
**LASSO:** least absolute shrinkage and selection operator  
**LGB:** light gradient boosting machine  
**LLM:** large language model  
**LR:** logistic regression  
**MCC:** Matthews correlation coefficient  
**nCovORF1ab:** Na, ORF1ab gene of novel coronavirus by nose test  
**NEU%:** neutrophil percentage  
**PCT:** thrombocytocrit  
**RDW:** red cell distribution width  
**RF:** random forest  
**RIDGE:** ridge regression  
**SaO<sub>2</sub>:** oxygen saturation under load  
**SEN:** sensitivity  
**SHAP:** Shapley additive explanations  
**SPE:** specificity  
**SVM:** support vector machine  
**T:** temperature  
**TBIL:** total bilirubin  
**TCHO:** total cholesterol  
**TCM:** traditional Chinese medicine  
**TN:** true negative  
**TP:** true positive  
**WBC:** white blood cell  
**XGB:** extreme gradient boosting  
**XGB:** extreme gradient boosting

*Edited by Arriel Benis; peer-reviewed by Feng-Nan Hwang, Guang Hu, Prince Jain, Yinghong Li; submitted 25.07.2024; final revised version received 03.05.2025; accepted 09.05.2025; published 16.07.2025*

### Please cite as:

Jin X, Wang Y, Wang J, Gao Q, Huang Y, Shao L, Zhao J, Li J, Li L, Zhang Z, Li S, Liu Y  
*A Machine Learning Approach to Differentiate Cold and Hot Syndrome in Viral Pneumonia Integrating Traditional Chinese Medicine and Modern Medicine: Machine Learning Model Development and Validation*  
*JMIR Med Inform* 2025;13:e64725  
URL: <https://medinform.jmir.org/2025/1/e64725>  
doi: [10.2196/64725](#)

© Xiaojie Jin, Yanru Wang, Jiarui Wang, Qian Gao, Yuhan Huang, Lingyu Shao, Jiali Zhao, Jintian Li, Ling Li, Zhiming Zhang, Shuyan Li, Yongqi Liu. Originally published in JMIR Medical Informatics (<https://medinform.jmir.org>), 16.07.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on <https://medinform.jmir.org/>, as well as this copyright and license information must be included.