Application of Model Combination Methods in Medical Diagnostics

Yuanan Lu *

Ulink college of shanghai, 310000, China

*Corresponding author's e-mail: 2252779690@qq.com

Abstract. With the explosive growth of medical data and the rapid development of AI technology, medical diagnosis is transforming from traditional experience-driven to data-driven, and a single machine learning model is often difficult to meet the clinical requirements in medical diagnosis due to problems such as incomplete feature capture and insufficient generalization capability. Model combination methods have become a hot research topic in the field of medical diagnosis by integrating the advantages of multiple models, effectively balancing the bias and variance, and significantly improving the accuracy and stability of medical diagnosis. This paper systematically expounds the theoretical foundations of model combination methods, comprehensively detailing the principles and application scenarios of common approaches including simple averaging, weighted averaging, bagging, boosting, and stacking, thoroughly investigates the core advantages of model combination methods in enhancing medical diagnostic performance, analyzes current challenges such as high computational costs and poor interpretability, and prospects future integration with deep learning and multimodal fusion technologies. The model combination method provides scientific and reliable technical support for medical diagnosis and has important clinical application value and academic research significance.

Keywords: Model combination methods, Medical diagnosis, Accuracy

1 Introduction

In the field of disease diagnosis, accurate and efficient diagnostic results are crucial for formulating treatment plans and making prognostic assessments for patients. With the rapid development of artificial intelligence and machine learning technologies, various diagnostic models have emerged. However, single models in disease diagnosis often exhibit certain limitations, such as insufficient capability to capture complex pathological features and suboptimal generalization performance. Model combination methods, by integrating the advantages of multiple independent models, can effectively improve the accuracy and reliability of disease diagnosis, making them a current research hotspot.

Model combination methods refer to strategies that integrate prediction results from multiple independent models to reduce variance, bias, or overfitting risks of

K. Othman et al. (eds.), *Proceedings of the 2026 5th International Conference on Mathematical Statistics and Economic Analysis (MSEA 2026)*, Advances in Economics, Business and Management Research 402,
https://doi.org/10.2991/978-94-6239-774-3_2

individual models, thereby enhancing overall predictive performance. The theoretical foundation of these methods stems from ensemble learning concepts in statistics and machine learning. The core philosophy of ensemble learning follows the "wisdom of crowds" principle, where combining multiple moderately-performing models yields a superior composite model.

In disease diagnosis, the characteristics of diseases are often relatively complex. Consider the challenge of early lung cancer detection: sub-centimeter nodules in CT scans are easily obscured by blood vessels or lung tissue, leading to a 20–30% missed diagnosis rate with single radiologists. Different models are sensitive to different features: for example, some models are good at capturing subtle changes in image data, some models perform well in processing clinical text data, and some models are good at processing indicator data. Model combination allows you to combine the strengths of different models to fully utilize the various characteristics of a disease in order to improve diagnostic accuracy. In addition, the model combination approach can capitalize on the diversity of data. In disease diagnosis, data sources can be very diverse, including different hospitals and different testing equipment, and there may be differences in the data that may be difficult for a single model to accommodate. The model combination method can be used to train and integrate data from different sources to increase the power of the model to adapt to the diversity of data and enhance the stability of the diagnosis.

This paper focuses on the application of model combination methods in medical diagnosis. Firstly, it describes the theoretical foundation of model combination methods; secondly, it categorizes and introduces the technical principles and adaptive scenarios of model combination methods. It demonstrates the practical value of model combination methods by combining with the diagnosis cases of multiple types of diseases, and then it introduces the challenges faced by the model combination methods, and looks forward to the direction of technological development, to provide references for the promotion of the clinical transformation of the model combination methods.

2 Domestic and international research status

Model combination methods demonstrate a series of advantages in disease diagnosis, leading to their widespread application in this field, with the benefits primarily manifested in the following aspects. First, they improve diagnostic accuracy by integrating the strengths of multiple models to comprehensively utilize various disease features and information, thereby enhancing diagnostic precision. Different models exhibit sensitivity to different features, and their combination can compensate for the limitations of individual models in feature capture, providing a more comprehensive reflection of disease characteristics. Second, they enhance model generalization performance. Individual models are often susceptible to training data influences and may perform poorly on new datasets (i.e., demonstrate weak generalization capability), while model combination methods mitigate this dependency on training data and improve adaptability to new data by integrating predictions from multiple models. Third, they accommodate complex disease features. Many diseases present highly complex characteristics involving multifaceted information that single models struggle to fully characterize across different data types. Model combination approaches effectively combine different models' capabilities in processing diverse features to adapt to complex disease manifestations. Fourth, they increase diagnostic stability. In disease diagnosis, variations in data sources and quality may cause significant fluctuations in single-model performance, resulting in unstable diagnostic outcomes. Model combination methods synthesize results from multiple models to balance the impacts of data variability and improve diagnostic consistency. Consequently, model combination methods have attracted extensive attention from researchers worldwide in the field of disease diagnosis.

2.1 Current status of foreign research

Zhang et al. [1] proposed an improved Bayesian combination model using medical record data downloaded from the Ancient and Modern Medical Records Cloud Platform developed by the Institute of Medical Information at the China Academy of Chinese Medical Sciences (CACMC) and practice guideline data from the Traditional Chinese Medicine Clinical Decision Support System as corpora, demonstrating significantly enhanced performance and notable clinical utility. Nafiseh et al [2] found that the latest AI techniques performed more accurately in diagnosing and predicting numerous diseases compared to experts without human error by using a combination of machine learning and deep learning AI techniques for medical image processing. Zhou et al [3] used the RETFound model to learn generalizable representations from unlabeled retinal images and provide a basis for label-efficient model adaptation in a variety of applications. It was found that RETFound provides a generalized solution that improves model performance and reduces the annotation workload of experts, enabling a wide range of clinical AI applications for retinal imaging implementations. Chiranjibi et al [4] integrated 13 pre-trained deep learning models for detecting monkeypox virus by fine-tuning them and analyzing the metrics to determine the optimal model. Experiments were then conducted on publicly available datasets with average precision, recall, F1 score and accuracy of 85.44%, 85.47%, 85.40% and 87.13%, respectively, and this approach was found to have better performance and be suitable for large-scale screening. Ahsan et al [5] identified diseases by using machine learning. A literature study of publications using data from databases summarizes the latest trends and approaches in Machine Learning Based Disease Diagnosis (MLBDD) considering the following factors: algorithms, disease types, data types, applications, and evaluation metrics. Xie et al [6] introduced some basic deep learning frameworks and summarized the deep learning prediction methods corresponding to different diseases, pointed out a series of current problems in disease prediction, and applied deep learning models to disease prediction and drug response to improve the accuracy of medical disease prediction. Kavitha et al [7] proposed a new machine learning approach to predict heart disease. In the implementation, 3 machine learning algorithms were used: random forest, decision tree, and hybrid model (a mixture of random forest and decision tree). The experimental results showed that the accuracy of the heart disease prediction model was 88.7% by using the model with hybrid model. Saif et al [8] proposed a new model DNLR-NET designed for automated and accurate diagnosis of MPox disease and achieved 97.55% accuracy, which is superior to the basic DenseNet201 model that achieved only 95.91%, finding its effectiveness and potential for clinical implementation in the diagnosis of MPox.

2.2 Status of domestic research

Yan Yanling et al [9] formed the final prediction value by weighting and combining the predicted values obtained from the gray model and autoregressive movement, and found that the combined model was more accurate than the single prediction model, and was able to predict the incidence trend of occupational diseases better. Lin Yuting et al [10] compared the advantages and disadvantages of the combined prediction model with the single prediction model by analyzing the simulation data and Arrhythmia dataset. The simulation study and empirical analysis showed that overall the combined prediction model has higher classification prediction accuracy.

Overall, the model combination approach has achieved certain results in the field of medical diagnosis, providing more reliable support for medical decision-making, whether it is to deal with the diagnosis of complex diseases or to help medical image analysis. However, in order to promote its wider and more effective application in clinical diagnosis, subsequent research needs to solve the problems of high computational cost and difficult data fusion, and continuously improve the performance, interpretability and reliability of the model.

3 Common model combination methods

Single-model approaches in medical diagnostics face inherent limitations that stem from a fundamental mismatch between their fixed architectures and the complex, variable nature of clinical data and decision-making demands, severely hindering their translation into reliable clinical tools. Firstly, their data adaptability is markedly insufficient. Secondly, single models suffer from unresolvable performance bottlenecks, failing to balance accuracy, robustness, and error control. Thirdly, single models are poorly suited for clinical deployment due to interpretability, cost, and uncertainty gaps. As a result, this paper mainly introduces six common model combination methods.

3.1 Simple average method

The simple average method is the most basic and intuitive model combination method, which directly averages the prediction results of multiple independent models to obtain the final prediction results. The principle of this method is that each model has a certain prediction energy, and the accuracy of the prediction can be improved by averaging the random errors. The specific formula is as follows:

$$\hat{y}=\frac{1}{M}\sum_{i=1}^{M}\hat{y}_i \tag{1}$$

where M is the number of models, $\hat{y}_i$ is the predicted value of the i-th model, and $\hat{y}$ is the predicted value after combination.

3.2 Weighted average method

The weighted averaging method is an extension of the simple averaging approach, which assigns different weights to individual models based on their performance and then combines their outputs through weighted summation to obtain the final prediction. Models with superior performance receive higher weights. The weight determination typically requires a validation set, where each model's prediction error is calculated and weights are allocated accordingly, with models demonstrating smaller prediction errors being assigned larger weights. The specific formula is as follows:

$$\hat{y}=\sum_{i=1}^{M} w_i \cdot \hat{y}_i,\ \sum_{i=1}^{M} w_i=1\ , w_i\geq 0 \tag{2}$$

w_i is the weight of the i-th model, $\hat{y}_i$ is the predicted value of the i-th model, and $\hat{y}$ is the combined predicted value.

3.3 Stacking method

Stacking is a hierarchical model combination method consisting of two layers: base models and a meta-model. The first layer comprises multiple diverse base models that independently process input data and generate predictions. The second layer contains the meta-model, which takes the base models' predictions as input for retraining and final prediction. The key to stacking lies in base model selection and meta-model training - base models should exhibit diversity to capture data features from different perspectives, while the meta-model must effectively integrate base model predictions and uncover their underlying relationships. The specific formula is as follows

$$z(x)=[h_1(x),h_2(x),\ldots,h_M(x)]\ , \hat{y}=g(z(x)) \tag{3}$$

where $h_i(x)$ is the prediction of sample x by the i-th base model, $z(x)$ is the prediction of M base models as a new feature, and the meta-model g learns the mapping relation from x to the true label.

3.4 Blending method

The blending method is similar to the stacking method, which is also a hierarchical model combination method. However, it has a relatively simple structure, and the Blending method divides the dataset into two parts: the training set and the validation set. First, several base models are trained using the training set, and then these base models are utilized to make predictions on the validation set to get the prediction results. Finally, using these predictions as new features combined with the true labels of the validation set, a meta-model is trained for final prediction. The advantages of the Blending method are that it is simple to operate, has low computational cost, and does not require complex division and cross-validation of the dataset. However, since it only uses the validation set to train the meta-model, this may lead to the underutilization of information, which will affect the model's performance to a certain extent. The specific process is as follows:

The first step is to train the basic model with the training set D_{train} : $h_k \leftarrow Train(D_{train})$.

The second step generates meta-features using the base model for validation set prediction:

$$z_i^{(k)}=h_k(x_i) \tag{4}$$

The third step is to train the meta-model:

$$g \leftarrow \mathrm{Train}(Z, y_{\mathrm{val}}) \tag{5}$$

The fourth step is to get the final result:

$$\hat{y}(x)=g(h_1(x),h_2(x),\ldots,h_K(x)) \tag{6}$$

3.5 Bagging method

The Bagging method generates several different training sets using the bootstrap sampling technique. Then, the same type of base model is trained on each training set, and finally, the predictions of these models are integrated by voting or averaging. Since each training set is obtained by random sampling, there are some differences between them, and the models trained on these training sets will be different. Integrating the results of multiple models can reduce the sensitivity of a single model to the training data and improve the generalization performance of the model. The specific process is as follows:

The first step generates the bootstrap sample set from the original training set with put-back sampling:

$$D_m=\{(x_1^{(m)},y_1^{(m)}),\ldots,(x_N^{(m)},y_N^{(m)})\}, \quad m=1,2,\ldots,M.$$

The second step is to train an independent base model based on each training sample set:

$$h_m \leftarrow \mathrm{Train}(D_m), \quad m=1,2,\ldots,M$$

The third step is to synthesize the predictions from each model to get the final result.

3.6 Boosting method

The boosting method is a string type of combinatorial method with the core idea of combining multiple weak learners into a single strong learner that can improve the overall predictive power. It involves iteratively training multiple base models, with each new model focusing on correcting the errors of the predecessor model. During the training process, higher weights are given to the samples that are incorrectly predicted by the predecessor models, enabling the new model to focus more on these samples, thus improving the overall performance of the model. The specific process is as follows:

Initialize sample weights:

$$\mathrm{w}_{1,\mathrm{i}}=\frac{1}{\mathrm{N}} \tag{7}$$

The error rate is obtained based on the weighting model:

$$\epsilon_t=\sum_{i=1}^{N} w_{t,i} \cdot I(\mathrm{h}_t(x_i)\neq y_i) \tag{8}$$

Calculate model weights based on error rates:

$$\alpha_t=\frac{1}{2}\ln\left(\frac{1-\epsilon_t}{\epsilon_t}\right) \tag{9}$$

Update the sample weights:

$$w_{t+1,i}=\frac{w_{t,i}\cdot \exp\left(-\alpha_t \cdot y_i \cdot \mathrm{h}_t(x_i)\right)}{Z_t} \tag{10}$$

Final prediction after performing multiple rounds of weight updates:

$$\hat{y}=\mathrm{sign}\left(\sum_{t=1}^{T} \alpha_t \cdot \mathrm{h}_t(x)\right) \tag{11}$$

4 Practical application of model combination methods in medical diagnosis

This paper focuses on three major applications of model combination methods in disease diagnosis as shown in Table 1.

Table 1. The model combination methods for disease diagnosis

Diseases	How to diagnose with model combination methods
cancer diagnosis	model combination, weighted averaging method
chronic diseases	random forest, neural network
infectious disease diagnosis	The blending method, the logistic regression model

In cancer diagnosis, where early screening and precise classification are paramount, these methods integrate multimodal data (imaging, pathology, genomics, etc.) to surpass the limitations of single data sources. For colorectal cancer subtype diagnosis, model combination approaches simultaneously process gene expression data from TCGA to train autoencoders for mutation feature extraction and analyze colonoscopy images for polyp morphology characteristics. The weighted averaging method dynamically combines these two models' outputs, with weights adjusted based on validation set AUC performance.

In the diagnosis of chronic diseases, which have a long course and complex etiology, the model combination approach improves the predictive diagnostic efficacy by integrating long-term monitoring data with multidimensional features. In the diagnosis of type 2 diabetes, fasting blood glucose, glycosylated hemoglobin, waist circumference, BMI, and other characteristic inputs were used. Random forest and neural network were used to construct the basic model. The results were integrated by the stacking method, which improved the diagnostic effect compared with the single model.

In infectious disease diagnosis, where both timeliness and accuracy are critical, model combination methods play a pivotal role in fusing multi-source testing data. In novel coronavirus diagnosis, the blending method achieves efficient screening, and the training set is used to train two basic models, in which the convolutional neural network analyzes the features of milled glass shadows and solid shadows in the chest CT images, the logistic regression model processes the nucleic acid detection values and the clinical symptom data, and the detection results on the validation set are input into the meta-model, which outputs the final diagnosis results. Utilizing a model combination approach can greatly reduce the rate of missed diagnosis due to false-negative nucleic acid tests and reduce the time-consuming nature compared to traditional methods.

To ground the theoretical analysis in real-world clinical practice, integrating case studies of model combination applications across diverse medical scenarios is critical to demonstrating their tangible value and adaptability. In medical imaging diagnostics—the most mature application area—lung nodule detection via chest CT exemplifies how stacking-based ensembles address single-model limitations: a team at a tertiary cancer center fused outputs from a 3D CNN, a Vision Transformer, and a U-Net to tackle the challenge of missed sub-centimeter nodules and benign/malignant misclassification. This ensemble achieved an AUROC of 0.96 on a multi-center dataset of 12,000 scans, reducing false negatives by 32% compared to the best single model (ResNet-50) and matching the performance of a panel of three radiologists. In dermatology, a Bagging-based ensemble of 50 lightweight CNNs trained on heterogeneous skin image datasets improved skin cancer classification accuracy to 92%, up from 83% with a single CNN, by mitigating overfitting to clinic-specific image artifacts.

In laboratory data-driven diagnosis, a study on diabetic nephropathy prediction leveraged a weighted boosting ensemble to integrate sequential glucose trends, static lab markers, and clinical demographics (via SVM). Trained on 5 years of electronic health record data from 8,000 patients, the ensemble predicted nephropathy onset 24 months in advance with 89% accuracy—outperforming single-model approaches

(78% for LSTM alone) and enabling proactive intervention (e.g., early ACE inhibitor therapy) in high-risk patients. For multi-modal integrated diagnostics in complex diseases, researchers at a neurology institute combined MRI-derived brain atrophy features (CNN), cerebrospinal fluid (CSF) biomarker levels (feedforward network), and cognitive test scores (gradient boosting) into a cross-modal attention ensemble for early Alzheimer's disease (AD) detection. This model identified preclinical AD in 87% of patients with mild cognitive impairment (MCI), compared to 79% accuracy for clinician-only assessments and 84% for single-modal AI, by synthesizing complementary biological and functional evidence.

5 Challenges Facing Model Combination Methods

While model combination methods demonstrate significant advantages in disease diagnosis, they also face several challenges. The primary issue is their high computational cost. These methods require training multiple base models and meta-models, substantially increasing computational demands compared to single-model approaches. It is particularly evident when processing large-scale medical data such as massive imaging datasets and genomic data, which impose stringent hardware requirements, including substantial computing power and storage capacity. These computational demands have, to some extent, limited the application of model combination methods in resource-constrained healthcare institutions. Another critical challenge lies in the poor interpretability of model combination methods. Particularly for complex approaches like stacking and boosting methods, their intricate internal architectures often make it extremely difficult to interpret and explain the diagnostic decision-making process. In medical applications, the interpretability of diagnostic results is paramount, as clinicians must understand the model's decision-making rationale to effectively integrate it with their clinical expertise. Poor model interpretability may significantly erode physicians' trust in the system and hinder its practical adoption in clinical settings. Additionally, data integration poses significant challenges. Medical diagnostics involve diverse data sources, including imaging data, textual data, and genomic data, which vary substantially in format and characteristics, making their integration particularly difficult. While model combination methods can synthesize outputs from different modalities, the effective preprocessing and fusion of raw heterogeneous data to enhance model performance remains a critical unresolved issue. Finally, the selection of models and parameter tuning present considerable complexity in model combination methods, as they involve choosing and optimizing multiple models - a highly intricate process. Different base models may be suitable for distinct tasks, requiring substantial expertise and extensive experimentation to determine both the optimal combination of base models and the appropriate meta-model type with its parameters. Improper model selection or suboptimal parameter configuration may result in combined models underperforming individual models or even developing overfitting issues.

6 Conclusion and Outlook

By integrating the advantages of multiple models, the model combination approach demonstrates significant advantages in the accuracy, generalization, and stability of disease diagnosis, and has become a key technology for medical AI to move from the laboratory to the clinic. It has a wide range of applications in early cancer screening, chronic disease risk prediction, rapid diagnosis of infectious diseases, and other scenarios, providing a technical path to solve the problems of uneven medical resources and differences in diagnostic levels. However, challenges such as high computational cost, insufficient interpretability, and data heterogeneity still need to be broken through, and lightweight, interpretability, and standardization will be a feasible direction for future development. With the technology iteration, the model combination method is expected to realize three breakthroughs. First, it is upgraded from an assisted diagnosis to a decision-making partner, deeply integrating into clinical work through a trusted AI framework. The second is to evolve from a general model to a personalized system, dynamically optimizing the reasoning strategy based on patient characteristics. The third is to expand from single-disease diagnosis to multi-disease co-occurrence assessment, integrating multi-organ correlation features to realize comprehensive health management.

For researchers and practitioners navigating the limitations of single-model approaches in medical diagnostics, targeted, actionable strategies are critical to advancing reliable AI-enabled clinical tools. For researchers, prioritize data-centric innovation: address small-sample and multi-center heterogeneity by adopting federated ensemble frameworks that train base models locally across institutions without sharing sensitive data, while developing modular architectures to seamlessly integrate imaging, clinical, and genomic data. For practitioners, implement phased validation and deployment workflows: first validate base models individually against clinical gold standards before testing fusion logic, and prioritize edge-compatible solutions for primary care settings. Additionally, build iterative feedback loops—integrate radiologist/pathologist corrections to refine model weights and document fusion rationale with XAI tools (e.g., SHAP values linking predictions to specific features like "spiculated margins" or "AFP levels") to enhance clinical trust. By aligning technical innovation with clinical workflows and regulatory requirements, stakeholders can systematically overcome single-model limitations and unlock the full potential of model combination as a trustworthy adjunct to patient care.

Under the synergistic promotion of policy and industry, the model combination method will provide core support for the construction of the precision medical system. In the future, it is necessary to strengthen interdisciplinary cooperation, join hands with computer scientists, clinicians, and epidemiologists to optimize the technical solutions, ensure safety and reliability while improving performance, and ultimately achieve the goal of “technology-enabled healthcare, combined to enhance the effectiveness of the”, to provide patients with higher-quality diagnostic services.

In AI-driven medical diagnostics—especially with complex model combination approaches—ethical considerations and transparency are not merely auxiliary concerns but foundational to building clinical trust and ensuring accountable care. A primary ethical risk lies in privacy vulnerability: while federated ensemble learning mitigates raw data exposure, the aggregation of model parameters across institutions still carries risks of indirect privacy leakage, demanding transparent data governance frameworks that explicitly disclose how data is anonymized, shared, and protected.

References

1. Zhang, Zhulv, et al. Research on diagnosis prediction of traditional Chinese medicine diseases based on improved Bayesian combination model [J]. Evidence-Based Complementary and Alternative Medicine, 2021.1 (2021): 5513748.
2. Ghaffar Nia, Nafiseh, Erkan Kaplanoglu, and Ahad Nasab. Evaluation of artificial intelligence techniques in disease diagnosis and prediction[J]. Discover Artificial Intelligence ,3.1 (2023): 5.
3. Zhou, Yukun, et al. A foundation model for generalizable disease detection from retinal images[J]. Nature ,622.7981 (2023): 156-163.
4. Sitaula, Chiranjibi, and Tej Bahadur Shahi. Monkeypox virus detection using pre-trained deep learning-based approaches[J]. Journal of Medical Systems ,46.11 (2022): 78
5. Ahsan, Md Manjurul, Shahana Akter Luna, and Zahed Siddique. Machine-learning-based disease diagnosis: A comprehensive review[J]. Healthcare. Vol. 10. No. 3. MDPI, 2022.
6. Xie, Shuxuan, Zengchen Yu, and Zhihan Lv. Multi-disease prediction based on deep learning: a survey[J]. Computer Modeling in Engineering & Sciences ,128.2 (2021): 489-522.
7. Kavitha, M., et al. Heart disease prediction using hybrid machine learning model[C]. 2021 6th international conference on inventive computation technologies (ICICT). IEEE, 2021.
8. Khan, Saif Ur Rehman, et al. Deep hybrid model for Mpox disease diagnosis from skin lesion images[J]. International Journal of Imaging Systems and Technology ,34.2 (2024): e23044.
9. Yan Yanling, He Xiaoxin. Application of combined model in prediction of occupational diseases [J].Industrial health and occupational diseases, 2013,39 (04) : 209-212.DOI : 10.13692 / j.cnki.gywsyzyb.2013.04.014.
10. Lin Yuting, Zhuang Hongli, Li Liting, etc. Research on disease diagnosis based on logistic regression combination prediction [J].China Health Statistics, 2018,35 (01) : 146-151.