ISSN: 2472-3703 DOI: 10.6911/WSRJ.202505_11(5).0019

Mycoplasma Pneumonia Diagnosis Algorithm Based on Multimodal Data Fusion

Mengxiang Xia[1], Xiaoyu Pi[4,*], Yufan Tong[1], Shunning Wang[2], Mengya Wang[3] and Qinyu Zhang[1]

[1]College of Science, North China University of Science and Technology, Tangshan 063210, China

[2]School of Nursing and Rehabilitation, North China University of Science and Technology, Tangshan 063210, China

[3]School of Economics and Management, North China University of Science and Technology, Tangshan 063210, China

[4]School of Mining Engineering, North China University of Science and Technology, Tangshan 063210, China

[*] Corresponding author

Abstract

The incidence of Mycoplasma Pneumonia (MP) has been on the rise. Traditional diagnostic methods have limitations, and existing machine - learning - based diagnostic studies also have deficiencies. In this study, a diagnostic system for Mycoplasma Pneumonia based on a multimodal visual hypergraph neural network (Trans - HGNN) was proposed. Firstly, basic processing such as denoising and standardization, as well as optimization using the VAE - GAN algorithm, were carried out on lung CT images and biochemical indicator data. Then, a hypergraph was constructed to fuse multimodal data, the VHNNs architecture was designed, and the Transformer algorithm was applied. Experiments show that the Trans - HGNN model performs best in terms of accuracy, recall rate, and F1 - score, reaching 88.69%, 0.8569, and 0.9229 respectively. The training process has good convergence and strong generalization ability. Although there may be a slight overfitting problem, it still provides reliable technical support for the diagnosis of Mycoplasma Pneumonia and has broad application prospects in the field of medical diagnosis.

Keywords

Mycoplasma Pneumonia; multimodal data fusion; hypergraph neural network; Transformer algorithm; diagnostic model.

1. INTRODUCTION

Mycoplasma pneumonia (Mycoplasma Pneumonia, MP), as an important type of community-acquired pneumonia (Community-Acquired Pneumonia, CAP), has shown a significant upward trend in incidence rates in recent years, posing a serious threat to public health. Epidemiological studies indicate that the proportion of mycoplasma pneumonia among CAP cases fluctuates between 10% and 40%, with particularly high incidence rates in children and adolescents. The clinical symptoms of this disease mainly include coughing and fever, which lack distinct specificity and can easily be confused with other types of pneumonia, making accurate clinical diagnosis a significant challenge. If the diagnostic process is delayed or erroneous, the condition

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202505_11(5).0019

may progress, leading to severe complications such as respiratory failure and myocarditis, posing a serious threat to patients' lives and health.

Traditional mycoplasma pneumonia diagnosis methods have many limitations. In laboratory microbiology testing, although MP culture is the "gold standard" for diagnosis, it requires stringent conditions and a long growth cycle, typically taking 2 to 3 weeks, which fails to meet the clinical need for rapid diagnosis. Serological tests such as cold agglutination and mycoplasma antibody detection are relatively simple to perform but lack sensitivity and specificity, with a positive detection rate of only about 60%, making misdiagnosis and missed diagnosis common. Imaging examinations like chest X-rays and CT scans can provide direct information on pulmonary lesions, but their accuracy in diagnosing early or mild lesions is limited, and relying solely on imaging findings makes it difficult to identify the pathogen.

With the rapid development of artificial intelligence technology, machine learning is increasingly applied in medical diagnosis. In pneumonia diagnosis, numerous research findings continue to emerge. Chang Minli et al. used support vector machines, random forests, and neural networks to conduct discriminant analysis on tuberculosis and pneumonia patients. The results showed that the model based on random forests performed better in distinguishing between tuberculosis and pneumonia patients, with an accuracy rate of 91%. Mao Rongzhi et al. trained eight machine learning methods on clinical manifestations and biochemical data of mycoplasma pneumoniae infection, screening out key diagnostic indicators such as Three Concave Sign.1 and Nucleic Acid PCR, providing important references for clinical diagnosis. Xu Xiang et al. constructed nine ML models to predict the risk of obstructive bronchiolitis in children with refractory mycoplasma pneumoniae pneumonia, finding that the XGBoost model had higher predictive performance, offering significant guidance for clinical treatment. Bao Xiuli et al. trained seven machine models to establish an early diagnostic prediction model for HIV/AIDS combined with pulmonary aspergillosis, where the CatBoost model performed excellently, providing new insights for disease diagnosis.

However, existing machine learning-based pneumonia diagnosis studies still have some shortcomings. Most researches only analyze data from a single modality, such as relying solely on imaging data or clinical indicators, failing to adequately integrate multisource information and thus unable to fully reflect the patient's condition. Additionally, when dealing with complex disease diagnoses, the exploration of higher-order interactions between data is not deep enough, leading to the need for further improvement in the accuracy and reliability of diagnostic models.

In response to the aforementioned issues, this study proposes a mycoplasma pneumonia diagnosis system based on multimodal visual hypergraph neural networks. The aim is to integrate multimodal data such as lung CT images and biochemical indicators, leveraging the powerful high-order relationship modeling capabilities of hypergraph neural networks to deeply mine complex interaction information between data. This approach seeks to construct more precise and efficient diagnostic models, providing stronger support for clinical mycoplasma pneumonia diagnosis. It holds promise for overcoming the limitations of traditional diagnostic methods and existing machine learning diagnostic models, thereby enhancing diagnostic accuracy and improving patient outcomes.

2. DATA PREPROCESSING

2.1. Basic image processing

During the data preprocessing stage, basic processing is applied to lung CT images to improve image quality and prepare for subsequent analysis. Noise interference in the images is removed using denoising algorithms to reduce its impact on diagnostic feature extraction. The images are also standardized, with contrast and brightness adjusted to meet uniform processing

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202505_11(5).0019

standards. Additionally, image enhancement techniques are employed to highlight key feature regions in lung CT images, such as lesion areas or abnormal structures, thereby better capturing imaging characteristics related to mycoplasma pneumonia. This provides high-quality input data for subsequent feature extraction and model training. These fundamental processing steps are crucial for ensuring that the subsequent diagnostic models can accurately identify and analyze lung CT images, laying the foundation for the efficient operation of the entire diagnostic system.

2.2. VAE-GAN algorithm

Based on data regularization, VAE-GAN can generate additional data samples to effectively address issues of data scarcity or imbalance. Furthermore, by analyzing the reconstruction error of the model, VAE-GAN can identify and correct outliers or errors in the data, further enhancing the quality of the dataset and the accuracy of analysis. This data preprocessing method based on VAE-GAN not only optimizes the processing of lung CT images and biochemical indicator data but also establishes a more accurate and reliable data foundation for pneumonia diagnosis and related research based on these data, significantly improving the overall performance of the diagnostic system.

To further optimize data quality and enhance the diversity and completeness of datasets, this project employs Variational Autoencoder Generative Adversarial Networks (VAE-GAN) for preprocessing lung CT images and biochemical indicator data. VAE-GAN combines the advantages of Variational Autoencoders (VAE) and Generative Adversarial Networks (GAN), using deep learning algorithms to achieve precise feature extraction and representation learning of the data. This method can reduce data dimensions while retaining the underlying characteristics and internal structure of the data.

In the VAE part, its core is to approximate the real posterior distribution $p_\theta(z|x)$ by introducing an approximate posterior distribution $q_\phi(z|x)$, and to realize the gradient back propagation by using the reparameterization technique. The objective function of VAE is the evidence lower bound (ELBO), which is expressed as:

$$\mathcal{L}_{\mathrm{VAE}} = \mathrm{E}_{q_\phi(z|x)}[\log p_\theta(x|z)] - \mathrm{D}_{\mathrm{KL}}(q_\phi(z|x)||p(z))$$

Among them, the reconstruction loss $\mathrm{E}_{q_\phi(x|z)}[\log p_\theta(x|z)]$ minimizes the difference between the original data x and the reconstructed data, ensuring that key information is retained in low-dimensional space; the KL divergence $\mathrm{D}_{\mathrm{KL}}(q_\phi(z|x)||p(z))$ is used to constrain the distribution of the latent variable z to be close to the prior distribution $p(z)$, typically set as a standard normal distribution $p(z)$, to ensure that the distribution of the latent variable has good properties, facilitating subsequent processing.

In the GAN part, the generator G receives noise or hidden variables obtained from VAE to generate data samples$\hat{x} = G(z)$; the discriminator D is responsible for distinguishing real data x and generated data $\hat{x}$. The adversarial loss function of GAN is:

$$\mathcal{L}_{\mathrm{GAN}} = \mathrm{E}_{x\sim p_{\mathrm{data}}(x)}[\log D(x)] + \mathrm{E}_{z\sim p(z)}[\log(1 - D(G(z)))]$$

Here, the goal of the generator G is to minimize$\mathcal{L}_{\mathrm{GAN}}$ and make the generated data closer to the real data; the goal of the discriminator D is to maximize $\mathcal{L}_{\mathrm{GAN}}$ and improve the ability to distinguish between true and false data. Through the adversarial training between the generator and the discriminator, the quality of the generated data by the generator is continuously optimized.

VAE-GAN combines the two, and the joint loss function can be expressed as:

$$\mathcal{L} = \mathcal{L}_{\mathrm{VAE}} + \lambda\mathcal{L}_{\mathrm{GAN}}$$

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202505_11(5).0019

Among them, the weight parameters λ for balancing VAE and GAN losses can be adjusted to balance the reconstruction capability of VAE and the generation capability of GAN, adapting to different data characteristics and task requirements. On the basis of data regularization, VAE-GAN can generate additional data samples, effectively addressing issues of data scarcity or imbalance. Furthermore, by analyzing the reconstruction error of the model (corresponding to the reconstruction loss term in the VAE objective function), VAE-GAN can identify and correct outliers or errors in the data, further enhancing the quality of the dataset and the accuracy of analysis. This data preprocessing method based on VAE-GAN not only optimizes the processing of lung CT images and biochemical indicator data but also builds a more accurate and reliable data foundation for pneumonia diagnosis and related research based on these data, significantly improving the overall performance of the diagnostic system.

3. DATA FUSION

3.1. Hypergraph construction

In the diagnosis system for mycoplasma pneumonia, data fusion is a critical step to achieve precise diagnosis. This project integrates data from two modalities: lung CT images and biochemical indicators, establishing a multimodal data fusion framework. Lung CT images can intuitively present the radiological characteristics of the disease in the lungs, while biochemical indicators reflect the physiological and biochemical status within the patient's body. By scientifically and reasonably analyzing and fusing these two data sources, a more comprehensive view of the patient's health can be achieved, thereby enhancing the accuracy and effectiveness of the diagnosis.

Specifically, hypergraph theory plays a crucial role. A hypergraph is a generalized graph structure that differs from traditional graphs in that its hyperedges can connect multiple nodes. This characteristic allows it to effectively capture complex interactions between data, which is significant for understanding the biological mechanisms and clinical manifestations of mycoplasma pneumonia. When constructing a hypergraph, each region in lung CT images and biochemical indicators are treated as nodes in the hypergraph. For example, lung CT images can be divided into multiple regions with specific meanings based on anatomical structures and lesion locations, with each region corresponding to a node; biochemical indicators, such as white blood cell count, C-reactive protein, and mycoplasma antibody titers, also exist as independent nodes.

For the construction of hyperedges, multiple factors should be considered comprehensively. From the perspective of lung CT images, if two regions are anatomically adjacent or share similar lesion characteristics, such as both showing ground-glass opacities and consolidation shadows typical of mycoplasma pneumonia, then a hyperedge can be used to connect the nodes corresponding to these two regions. This suggests that there may be some intrinsic connection between these two regions during the progression of the disease, and the existence of the hyperedge serves as a modeling representation of this connection. From the biochemical marker perspective, if certain biochemical markers have biological associations, for example, white blood cell count and C-reactive protein often change simultaneously during inflammatory responses, reflecting the degree of immune response in the body, then hyperedges can be used to connect the nodes corresponding to these related markers.

In addition, hyperedges are constructed between lung CT image node and biochemical indicator node based on their medical connections. For example, when a lesion appears in a specific area of the lungs, certain biochemical indicators may change accordingly. An increase in the white blood cell count and C-reactive protein levels can accompany the expansion of an inflammatory area in the lungs. At this point, hyperedges can be constructed to connect the

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202505_11(5).0019

node representing the lung region with the relevant biochemical indicator node, thereby describing the relationship between pulmonary lesions and systemic biochemical responses.

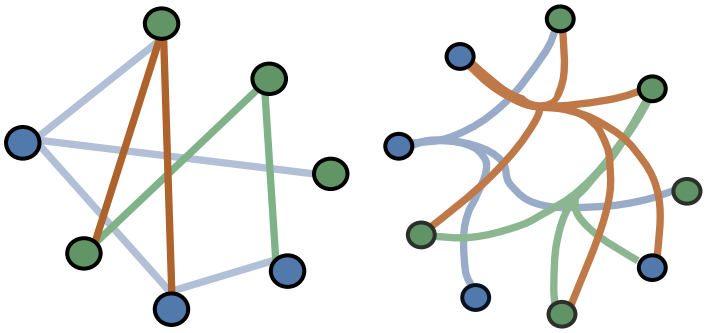

Figure 1. Schematic diagram of supergraph

By constructing such hypergraphs, it is possible to comprehensively and meticulously describe the multivariate and higher-order relationships between lung CT images and biochemical indicator data. This hypergraph structure not only reveals the internal associations within a single data modality but, more importantly, clearly presents the complex interactions between different modalities of data. It provides rich and valuable information for subsequent hypergraph-based analysis and diagnostic model training, laying a solid foundation for improving the accuracy of Mycoplasma pneumoniae diagnosis.

3.2. VHNNs Architecture design

This project also employs multimodal visual hypergraph neural network technology, integrating hypergraph theory with deep neural networks to further optimize data fusion strategies. The network can comprehensively analyze lung CT images and biochemical indicators of patients, extracting composite features to provide comprehensive and precise decision support for the diagnosis of mycoplasma pneumonia. Through this multimodal fusion approach, the system can more accurately identify disease characteristics, especially in early diagnosis and assessment of lesion severity, significantly enhancing the robustness and accuracy of diagnoses. The multimodal visual hypergraph neural network model constructed through data fusion not only effectively handles complex relationships involving multiple nodes but also captures deeper patterns and connections. This method not only aids in scientifically rigorous analysis and understanding of disease characteristics but also guides clinical doctors to make more precise diagnostic decisions, which is crucial for improving the accuracy and efficiency of pneumonia diagnosis.

3.3. Transform algorithm

Transformer algorithm is a deep learning architecture based on self-attention mechanisms, which has achieved significant results in natural language processing and computer vision in recent years. At its core lies the self-attention mechanism, which calculates the relevance between elements at different positions within an input sequence, enabling the model to dynamically focus on important parts of the sequence. The calculation formula is:

$$\text{Attention}(\text{Q}, \text{K}, \text{V}) = \text{softmax}\left(\frac{\text{QK}^{\text{T}}}{\sqrt{\text{d}_\text{k}}}\right)\text{V}$$

Among them, Q (Query), K (Key), and V (Value) are the three input matrices, typically obtained from linear transformations $d_k\sqrt{\text{d}_\text{k}}$of the input data. The dimensions of the K matrix are used to scale the dot product results, preventing excessively large values that could cause

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202505_11(5).0019

the softmax function gradient to vanish. When processing lung CT images and biochemical indicator data, specific linear transformation operations are applied to Q, K, and V based on the characteristics of the data, to accurately reflect the relationships between different positions in the input sequence, helping the model focus on key information.

In order to further improve the performance, Transformer introduces the multi-head attention mechanism. This mechanism divides the input data into multiple "heads", each of which calculates self-attention independently, and then concatenates the output. The formula is as follows:

$$\mathrm{MultiHead}(\mathrm{Q},\mathrm{K},\mathrm{V}) = \mathrm{Concat}(h\mathrm{ead}_1,\ldots,h\mathrm{ead}_h)\mathrm{W}^{\mathrm{O}}$$

In this context, h represents the number of heads

$\mathrm{head_i} = \mathrm{Attention}(\mathrm{QW_i^Q},\mathrm{KW_i^K},\mathrm{VW_i^V})\mathrm{W_i^Q}$、$\mathrm{W_i^K}$、$\mathrm{W_i^V}$ is the linear transformation matrix corresponding to the i-th head, used to map input data into different subspaces to capture information from various angles. It is used $\mathrm{W^O}$ for concatenating multi-head outputs, integrating information to enable the model to have a more comprehensive understanding of the data. When processing lung CT images, different heads can focus on different regions or features, such as texture and edges, enhancing the ability to understand the image.

The Transformer algorithm adopts an encoder-decoder architecture. The encoder is responsible for encoding input data into context vectors, while the decoder generates output results based on these vectors. Both the encoder and decoder utilize multi-head attention mechanisms and feedforward neural networks to process the data. Feedforward neural networks typically consist of two linear layers and one activation function, with the formula being:

$$\mathrm{FFN}(\mathrm{x}) = \max(0, \mathrm{xW_1} + \mathrm{b_1})\,\mathrm{W_2} + \mathrm{b_2}$$

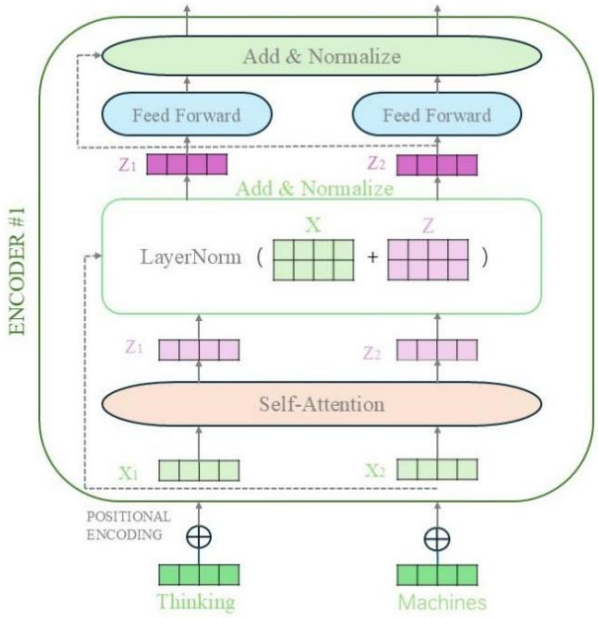

Figure 2. Transformer algorithm structure diagram

In the system for diagnosing mycoplasma pneumonia, x represents input features, $\mathrm{W_1}, \mathrm{W_2}$ is the weight matrix that determines how input features are transformed in the network, $\mathrm{b_1}, \mathrm{b_2}$ is the bias vector used to enhance the model's fitting ability, and $\max(0,\cdot)$ is ReLU activation function introduces non-linear transformations to improve the model's expressive power. In this system, CT images are segmented into multiple small blocks, treated as serialized "words."

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202505_11(5).0019

Transformer uses these mechanisms to capture spatial relationships between different regions of the image and combines numerical characteristics from biochemical indicators to enhance understanding of disease status. This multimodal data fusion approach allows Transformer to analyze data from multiple perspectives, comprehensively capturing interactions between different modalities, which is crucial for precise diagnosis of mycoplasma pneumonia. It significantly enhances the diagnostic system's capability to handle complex multimodal data, as well as its accuracy and robustness.

4. PREDICTIONS

As shown in Table 1 below, according to the provided comparative experimental data, we can compare and analyze the performance of three different neural network models. These models include traditional convolutional neural network (CNN), HGNN, and Trans-HGNN.

Table 1. compares the model results

	cnn convolutional neural network	HGNN	Trans-HGNN
Accuracy	75.00%	84%	88.69%
Recall	0.74	0.81	0.8569
F1	0.71	0.82	0.9229

According to the provided table data, the Trans-HGNN model performs best in accuracy, recall rate, and F1 value, with an accuracy of 88.69%, a recall rate of 0.8569, and an F1 value of 0.9229. This indicates that Trans-HGNN has a significant advantage in overall classification performance, especially in identifying positive samples. In contrast, the HGNN model has an accuracy of 84%, a recall rate of 0.81, and an F1 value of 0.82, which is slightly lower than Trans-HGNN but still significantly better than the CNN model, demonstrating good classification results. The CNN model, however, has an accuracy of 75%, a recall rate of 0.74, and an F1 value of 0.71, performing the worst overall, particularly with relatively low recall rates and F1 values, indicating its insufficient recognition capability in this task. Overall, Trans-HGNN is the optimal choice, suitable for tasks requiring high model performance, especially when high recall rates and F1 values are needed.

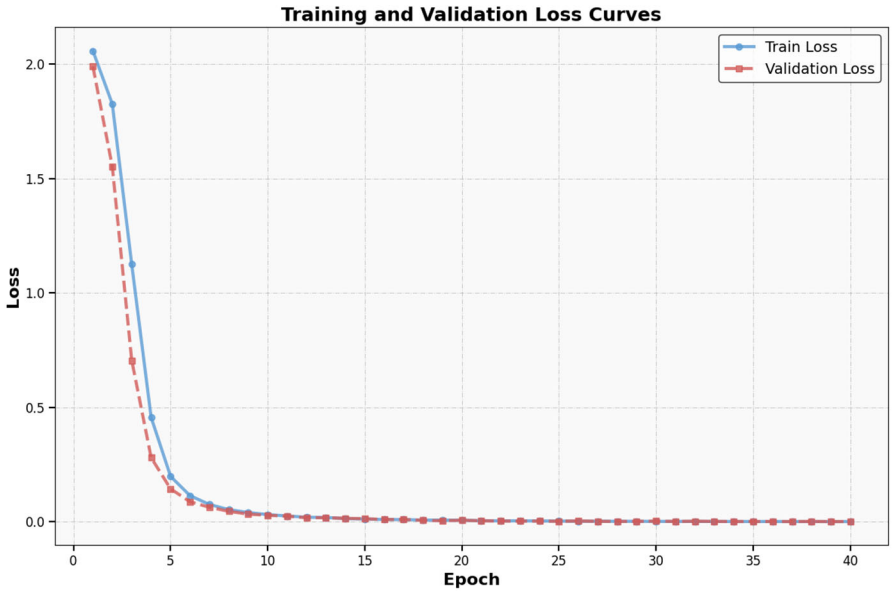

Figure 3. Trans-HGNN training loss curve

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202505_11(5).0019

According to the training and validation loss curves in the figure, it can be seen that the model exhibits good convergence during training. In the early stages, the training loss (blue curve) drops very quickly, especially in the first few epoch, with the loss value decreasing from nearly 2 to close to 0, indicating strong learning ability of the model on the training data. The validation loss (red curve) also decreases, but its rate is more gradual compared to the training loss, suggesting that the model's generalization ability on the validation set gradually improves. Overall, after about 30 epoch, both the training loss and validation loss stabilize, showing little further downward trend, which means the model has largely converged, and further training may not significantly enhance performance. Despite this, the rapid decrease in training loss and the slower decline in validation loss suggest a slight risk of overfitting, so potential overfitting risks should be monitored in subsequent training. In summary, this graph reflects the gradual optimization process of Trans-HGNN during training and demonstrates good generalization and convergence.

5. CONCLUSION

In this study, we propose a mycoplasma pneumonia diagnosis system based on multimodal visual hypergraph neural network (Trans-HGNN). Experimental results show that the Trans-HGNN model performs excellently in three key metrics: accuracy, recall rate, and F1 value, demonstrating its strong capability in pneumonia diagnosis. During training, Trans-HGNN exhibits good convergence and strong generalization ability. Although the initial training loss decreases rapidly, the validation loss declines more slowly, indicating that the model has strong adaptability to the training data and gradually improves its generalization performance on the validation set. As training progresses, the loss curve stabilizes, suggesting that the model has largely converged and reached an optimal state. Nevertheless, potential overfitting issues should still be monitored, especially when the training loss decreases quickly. Overall, Trans-HGNN not only effectively integrates data from different modalities but also accurately captures complex relationships between data, providing reliable technical support for early diagnosis and lesion assessment of mycoplasma pneumonia, showcasing the broad application prospects of this method in medical diagnostics.

ACKNOWLEDGMENTS

This study was supported by the Innovation and Entrepreneurship Training Program for College Students at North China University of Science and Technology, project number: 202410081035.

REFERENCES

[1] FAN San-cai. WEN Fei's representative suggestion: Accelerate the digital and intelligent transformation of traditional manufacturing industry [N]. China Metallurgical News, 2025-03-15(2). (in Chinese)

[2] XU Xiang, CAO Ling, ZHAO Aihong. Building a Risk Prediction Model for Bronchiolitis Obliterans in Children with Refractory Mycoplasma Pneumoniae Pneumonia Based on Explainable Machine Learning Algorithms[J]. Practical Journal of Cardiac Cerebral Pneumal and Vascular Disease, 2025, 33(02): 20 - 25.

[3] BAO Xiuli, MENG Sirun, WEI Wudi, [3] BAO Xiuli, MENG Sirun, WEI Wudi, et al. Early Diagnosis Machine Learning Prediction Model for Pneumocystis Pneumonia in HIV/AIDS Patients in

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202505_11(5).0019

Guangxi[J/OL]. China Tropical Medicine, 2025, (1 - 11)[2025 - 03 - 17]. http://link.cnki.net/urlid/46.1064.r.20250314.1918.006.

[4] CHANG Minli, YOU Shuping, CHEN Xiaodie, [4] CHANG Minli, YOU Shuping, CHEN Xiaodie, et al. A Discriminant Analysis Study of Pulmonary Tuberculosis and Pneumonia Patients Based on Machine Learning[J]. Acta Universitatis Medicinalis Anhui, 2025, 60(03): 705 - 715.

[5] MAO Rongzhi, LU Xiaohua, PEI Jialu, [5] MAO Rongzhi, LU Xiaohua, PEI Jialu, et al. Key Feature Analysis of Clinical Manifestations and Biochemical Indicators in the Diagnosis of Mycoplasma Pneumonia Based on Machine Learning[J]. Modern Computer, 2024, (14): 9 - 17.