Lung Cancer Detection Model Integrating Orthogonal Views and Blood Analysis Indicators for Patients with Pleural Effusion

Dongxun Jiang
Tongji University
Shanghai 201804, China

Dongdong Zhang*
Tongji University
Shanghai 201804, China

***Abstract*—Pleural effusion, as a highly prevalent clinical symptom, has garnered wide-spread attention. However, most models are designed for lung cancer detection, but the detection of tumor lesions within pleural effusion still requires improvement. This paper proposes a lung cancer detection model integrating orthogonal views and blood analysis indicators for patients with pleural effusion(OVBI-LCDM), in which feature enhancement network based on view complementarity is designed to more accurately depict hidden lesions in orthographic views. Blood indicators analysis module is introduced to analyze correlation and importance of six blood indicators and determine their weights. Feature fusion module based on analysis results is proposed to dynamically re-evaluate the contributions of orthogonal view images and blood indicators, thereby achieving mutual enhancement of these features. The experimental results show that the model achieved good classification performance, with an accuracy of 82.76%, outperforming other models.**

***Index Terms*—Pleural Effusion, Lung Cancer Detection, Blood Analysis Indicators, Orthogonal Views, Cross-Modal Fusion**

I. Introduction

Pleural effusion (PE) is characterized by abnormal fluid accumulation within the pleural cavity. It may obscure underlying thoracic diseases such as lung tumors, prompting clinicians to prioritize drainage procedures to expose tissues for clearer imaging diagnosis. However, drainage delays diagnosis and treatment, potentially missing the opportunity for optimal intervention. Therefore, rapidly and accurately identifying the cause of pleural effusion without waiting for fluid drainage—particularly malignant effusions caused by lung cancer—holds significant clinical value.

With the continuous advancement of deep learning architectures, lung cancer detection models are also undergoing further development and refinement. Lung-EffNet [14] effectively handles the complexity of lung images, such as the diverse appearance of lung tissue and the presence of multiple pathological conditions. It efficiently identifies and classifies different types of lung diseases, including pneumonia, tuberculosis, and lung cancer, through advanced feature extraction techniques. LungConVT-Net [15] utilizes convolutional layers to capture local features from chest images, while the VIT component enables the model to understand the overall context of the image and the relationships between its different parts.

* Corresponding author: Dongdong Zhang (email:ddzhang@tongji.edu.cn)

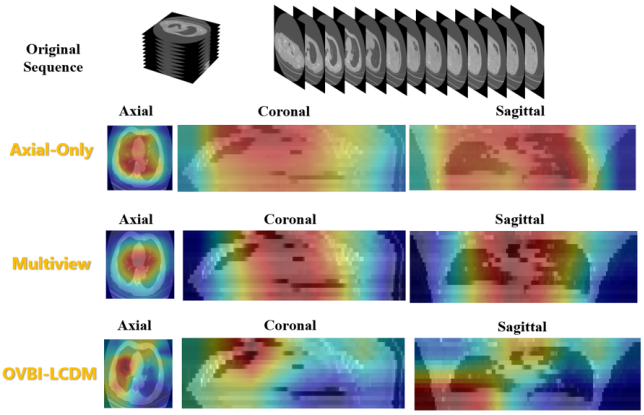

Fig. 1. Comparison of different methods for focusing on the lesion area. The Axial-Only approach utilizes solely axial information, resulting in poor consistency between coronal and sagittal heatmaps. This leads to large red regions that fail to accurately localize lesions. While the Multiview method processes three views in parallel, cross-view semantic misalignment persists, causing conflicting weights and retaining extensive red lesion areas. OVBI-LCDM achieves multimodal feature-level interaction, enabling consistent highlighting across three-plane heatmaps and more precise lesion localization (precise red lesion areas).

Although the current model demonstrates good detection performance for general tumors and nodules, they still struggle to extract features from lesions obscured by fluid accumulation, as illustrated in Figure 1. Furthermore, most existing pleural effusion identification models focus on distinguishing benign from malignant conditions, yet exhibit relatively weak recognition capabilities for specific diseases. Current research also lacks sufficient attention to blood analysis indicators critical for clinical diagnosis. [1] To address these challenges, this paper proposes a lung cancer detection model for patients with pleural effusion based on orthogonal views of CT sequences and blood analysis indicators. The main contributions of this study are as follows:

- We propose a feature enhancement network based on view complementarity to improve efficiency and observe lesions from different angles, which avoids the issue of continuous effusion regions in 3D volumes that may obscure lesions.

- We analyze the correlation and importance of six blood analysis indicators and determine the weights to pay more attention to indicators indicating lung cancer.
- We introduce a feature fusion module to dynamically reweight the contributions of orthogonal views images and blood analysis indicators.

II. Related Work

A. Lung Disease Detection Methods Based on Images

Axial lung cancer detection primarily involves analyzing CT scans in the axial plane to identify and classify lung nodules. This method leverages the detailed cross-sectional images provided by CT scans to detect abnormalities in lung tissue. For instance, one study utilized axial CT images from the LIDC-IDRI database to detect lung nodules using adaptive thresholding and watershed segmentation techniques, achieving an accuracy of 96% [16]. Another approach involved a 2D object detection neural network trained on axial slices, which demonstrated high precision in nodule detection with a low false positive rate [17]. Additionally, a study employed a single frequency holographic electromagnetic induction imaging method to detect small lung tumors in axial CT images, showing potential for early detection [18]. These methods highlight the effectiveness of axial imaging in providing detailed and accurate lung cancer detection, although they may be limited by the need for extensive manual analysis and potential overdiagnosis.

Multi-view lung cancer detection enhances diagnostic accuracy by integrating information from multiple imaging planes or modalities. This approach addresses the limitations of single-view methods by providing a more comprehensive analysis of lung nodules. For example, a multi-view aspect model using digital image processing techniques and Convolutional Neural Networks (CNNs) was proposed to improve diagnostic accuracy [19]. Another study introduced a Multi-View Soft Attention-Based CNN (MVSA-CNN) model, which classified lung nodules into benign, primary, and metastatic stages using three different views, achieving an accuracy of 97.10% [20]. These multi-view methods demonstrate significant improvements in detection accuracy and robustness. However, for patients with pleural effusion, tumors may adhere closely to the chest wall while being surrounded or obscured by large amounts of pleural fluid. This makes it difficult for traditional methods to accurately discern the precise relationships between the effusion, pleura, and lung tissue.

B. Lung Disease Detection Based on Blood Indicators

The importance of blood test indicators such as SCC, CEA, CK19, NSE, ProGRP, and AFP in the diagnosis of lung cancer lies in their ability to serve as noninvasive diagnostic tools that can help with early detection, monitoring, and evaluation of the treatment of the disease [7].

Squamous Cell Carcinoma Antigen (SCC) is commonly used for the diagnosis of squamous cell carcinoma, a subtype of non-small cell lung cancer (NSCLC). It has a high positive rate in patients with squamous cell carcinoma [7]. Carcinoembryonic antigen (CEA) is frequently elevated in various types of lung cancer, especially adenocarcinoma, another subtype of NSCLC. Cytokeratin 19 fragments (CK19) is particularly sensitive for NSCLC and is often elevated in patients with lung adenocarcinoma and squamous cell carcinoma. Neuron-Specific Enolase (NSE) is a key marker for small cell lung cancer (SCLC) and is used to assess the extent of the disease and monitor the response to treatment. It is highly specific for SCLC compared to other types of lung cancer. Pro-Gastrin-Releasing Peptide (ProGRP) is also specific for SCLC and has higher sensitivity and specificity compared to NSE. [7], [8] Alpha-Fetoprotein (AFP) can be useful in specific contexts, such as identifying certain germ cell tumors that may exist in the lungs [9]. The combined use of these indicators improves diagnostic accuracy and sensitivity, as individual indicators may have limitations in specificity and sensitivity, which is crucial for early detection [7].

III. Methods

A. Overview

In this paper, the lung cancer detection model integrating orthogonal views and blood analysis indicator for patients with pleural effusion is proposed and Figure 2 illustrates the overall framework of the model. Feature enhancement network based on view complementarity is devised, which extracts 3D feature information from orthogonal views using neural networks to achieve complementary perspectives, thereby more accurately depicting hidden lesions. In the module dedicated to the analysis of blood indicators, a correlation and importance analysis of six blood analysis indicators is conducted, with weights being assigned to emphasize those indicators that are more indicative of lung cancer. Feature fusion module based on the analysis results is proposed, which dynamically reweights the contributions of orthogonal views images and structured features. This enables data to guide features and features to validate data at the feature level.

B. Feature Enhancement Network Based on View Complementarity

Unlike traditional multi-view methods that perform "static stitching-averaging" fusion at the decision or feature layer, the proposed feature enhancement network based on view complementarity introduces a cross-view self-attention mechanism. This enables the axial, coronal, and sagittal 3D encoders to exchange global contextual information in real-time during each forward pass. This dynamic collaborative strategy transforms "view complementarity" from a fixed prior structure after training into a data-driven adaptive process. Consequently, even when information is lost due to pleural effusion occlusion or single-view artifacts, the model can perform macro-level "self-repair" of lesion features through complementary semantics from other views, thereby enhancing its overall robustness in detecting concealed lesions.

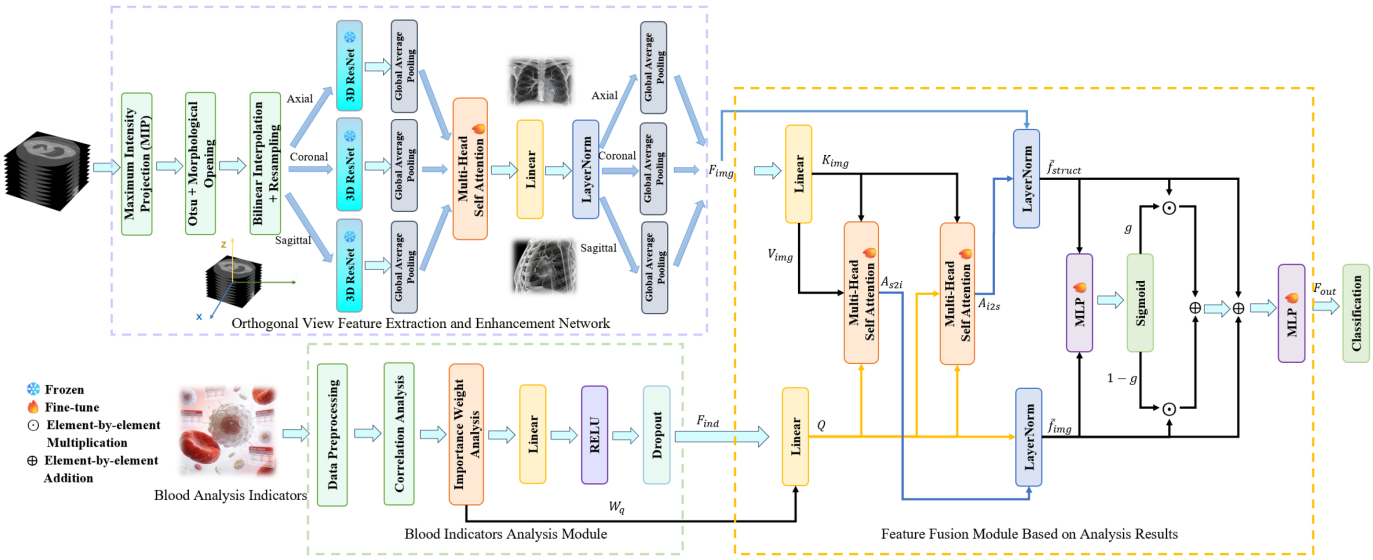

Fig. 2. Overview of our proposed lung cancer detection model integrating orthogonal views and blood analysis indicators for patients with pleural effusion(OVBI-LCDM).

1) Construction of Three-Dimensional Volume and Orthographic Views: Before the data are fed into the network, each CT scan is first loaded into memory as a raw grayscale volume $V \in \mathbb{R}^{H_0 \times W_0 \times D_0}$, where H_0 and W_0 are the in-plane resolutions and D_0 is the original number of slices. To guarantee that the network always receives a tensor of fixed spatial size, we perform a Maximum Intensity Projection (MIP) on V and obtain $M \in \mathbb{R}^{H_0 \times W_0}$.

M is binarized into S by Otsu's adaptive thresholding followed by a morphological opening; the largest connected component is then extracted and its axis-aligned bounding box $bbox = (x_1, y_1, x_2, y_2)$ is computed.

The axis-aligned bounding box is then mapped back to the original volume, yielding the lesion-localized sub-volume $\hat{V} = V[x_1 : x_2, y_1 : y_2, :]$ with dimensions(h, w, D_0), where $h = y_2 - y_1$ and $w = x_2 - x_1$. Each slice is then bilinearly interpolated in the xy-plane and resized to 128 × 128, yielding $\tilde{V} \in \mathbb{R}^{128 \times 128 \times D_0}$. Subsequently, resampling is performed along the depth direction with a target depth of T = 16, using linear indexing

$$k_i = \left\lfloor \frac{i \cdot (D_0 - 1)}{T - 1} \right\rfloor, \quad i = 0 \ldots T - 1 \tag{1}$$

To extract slices, resulting in the final normalized standard volume $X \in [0,1]^{\mathbb{R}^{128 \times 128 \times 16}}$. X is horizontally flipped in the xy-plane with probability 0.5, then rotated by a random angle $\theta \in [-10°, 10°]$ and translated $(\Delta x, \Delta y) \in [-0.05w, 0.05w] \times [-0.05h, 0.05h]$ and X' is obtained.

To create orthogonal views input, the same X' is treated as three separate views: Axial, where the native orientation is kept, yielding the tensor $A = X' \in \mathbb{R}^{128 \times 128 \times 16}$; Coronal, where the dimension order is first swapped so that the y-axis becomes the depth dimension: $C_0 = X'.\text{transpose}(0, 2, 1) \Rightarrow C_0 \in \mathbb{R}^{128 \times 16 \times 128}$. Trilinear interpolation is applied to resize the coronal plane back to 128×128, yielding $C \in \mathbb{R}^{128 \times 16 \times 128}$; Sagittal, where the x-axis is moved to the depth position: $S_0 = X'.\text{transpose}(2, 1, 0) \Rightarrow S_0 \in \mathbb{R}^{16 \times 128 \times 128}$. After the same interpolation step, $S \in \mathbb{R}^{16 \times 128 \times 128}$. Finally, each of the three view-volumes is replicated into 3-channel format and concatenated along the batch dimension to form a 5D tensor $X \in \mathbb{R}^{B \times 3 \times 3 \times 16 \times 128 \times 128}$.

2) Orthogonal View Feature Extraction and Enhancement: After performing multi-view segmentation on the CT sequences, let the input volume in one batch be denoted as $X \in \mathbb{R}^{B \times V \times C \times D \times H \times W}$, where B is the batch size, V is the number of views, C is the number of channels (the grayscale volume is stacked into three channels), and (D, H, W) denote the depth, height, and width dimensions, respectively. The tensor is then reshaped into a single-view tensor sequence $\{X_i\}_{i=1}^{V}, \quad X_i \in \mathbb{R}^{B \times C \times D \times H \times W}$ Each view X_i is fed into a 3D ResNet for feature extraction. At the network head, global average pooling (GAP) collapses the spatio-temporal dimensions, yielding the per-view feature vector f_i.

The feature vectors of all views are stacked into a matrix $F = [f_1; f_2; \cdots; f_V] \in \mathbb{R}^{B \times V \times E}$. To capture complementary information across the different views, we introduce a multi-head self-attention mechanism. The matrix F is linearly projected to obtain the query, key, and value matrices.

This operation is performed in parallel over all H heads and the outputs are concatenated and then linearly projected back to the original dimension, denoted as MHSA(F).Subsequently, a residual connection followed by layer normalization is applied:

$$Z = \text{LayerNorm}(F + \text{MHSA}(F)) \in \mathbb{R}^{B \times V \times E} \tag{2}$$

Finally, global average pooling is performed along the view

TABLE I
WEIGHT TABLE OF IMPORTANT FACTORS

Number	Inspection Items	Importance weight
1	Carcinoembryonic Antigen (CEA)	0.298
2	Cytokeratin 19 Fragment (CK19)	0.177
3	Neuron-specific Enolase (NSE)	0.171
4	Alpha-fetoprotein (AFP)	0.153
5	Pro Gastrin Releasing Peptide (ProGRP)	0.133
6	Squamous Cell Carcinoma Antigen (SCC)	0.069

dimension to obtain the ultimate image-level feature vector:

$$f_{\text{img}} = \frac{1}{V}\sum_{i=1}^{V} Z_i \in \mathbb{R}^{B\times E} \tag{3}$$

At this point, f_{img} is the image representation obtained after orthogonal views interaction and self-attention-based fusion. It can be fed, together with structured indicators features, into the subsequent cross-modal fusion and classification network.

C. Blood Indicators Analysis Module

We perform a correlation and importance analysis on a clinical dataset that contains 6 blood indicators, and obtain the importance weights for lung cancer for key factors, which are used for subsequent fusion of characteristics.

To analyze the relationships among all features in the dataset including blood indicators and the lung cancer label, we calculate the Pearson correlation coefficient for each pair of features. For any two features i and j, we first compute their mean values, denoted as X_i and X_j. Using these means, we then calculate the covariance between the two features across all samples. To determine the relative contribution of each modal factor, we utilize the Random Forest algorithm in conjunction with a correlation matrix analysis. The features are then ranked according to their importance scores to identify the most influential factors, with the sorted results presented in Table I.

The raw 6 dimensional structured vector is first fed into a fully-connected layer that linearly projects it into a higher-dimensional semantic space (64D). A ReLU activation is then applied to introduce non-linearity. Finally, Dropout regularization is employed to mitigate over-fitting. The resulting vector serves as the deep representation of the structured clinical data and will be fused with imaging features in the subsequent cross-modal fusion stage.

D. Feature Fusion Module Based on Analysis Results

Traditional methods completely isolate information from different modalities during feature extraction. The relationships between modalities can only be passively learned by the final few layers of fully connected classifiers, resulting in extremely limited interaction. In contrast, our proposed feature fusion module based on analysis results leverages the complementary nature of information across modalities to uncover deep-seated relationships.

Let the batch size be B, the image-feature dimension be E, and the structured-feature dimension be S. First, the image encoder outputs $F_{\text{img}} = \left[f_{\text{img}}^{(1)} \cdots f_{\text{img}}^{(B)}\right]^{\top} \in \mathbb{R}^{B\times E}$, where each $f_{\text{img}}^{(i)}$ is produced by the orthogonal views encoder and represents the fused representation of the three orthogonal views. The structured features $h_{\text{drop}}^{(i)}$ from Equation (3) are mapped to a lower-dimensional embedding via a fully-connected layer followed by activation and dropout, and then assembled into the matrix $f_{\text{ind}}^{(i)} = h_{\text{drop}}^{(i)} \in \mathbb{R}^{S'}, \quad S' \ll E$, $F_{\text{ind}} = \left[f_{\text{ind}}^{(1)} \cdots f_{\text{ind}}^{(B)}\right]^{\top} \in \mathbb{R}^{B\times S'}$.

In the cross-modal attention module, the structured features are first projected into the same dimensional space as the image features:

$$Q = F_{\text{ind}}W_q + b_q \in \mathbb{R}^{B\times E} \tag{4}$$

Here, W_q denotes the importance weight of blood indicators. b_q is the learnable bias term of the query vector, which together with the weight matrix (W_q) are used for linear transformation. This projection ensures that the two modalities reside in the same semantic space, facilitating subsequent similarity and attention computations. The module then performs bidirectional cross-attention. On one direction, the structured features serve as queries, while the image features act as keys and values. Attention weights from the structured modality to the image modality are computed, allowing the model to identify image regions that are highly relevant to the structured indicators. This direction is formalized as

$$A_{\text{s2i}} = \text{softmax}\left(\frac{Q(K_{\text{img}})^{\top}}{\sqrt{d_k}}\right)\cdot V_{\text{img}}, \quad d_k = \frac{E}{H} \tag{5}$$

Here, K_{img} and V_{img} are obtained from the image features F_{img} through linear transformations, and d_k is a scaling factor used to prevent excessively large dot-product values that could destabilize gradients. Conversely, the module performs the attention mechanism in the reverse direction: image features serve as queries, while structured features act as keys and values. This computes attention responses from the image modality to the structured modality, thereby identifying key indicators in the structured data that are strongly reinforced by image content. This direction is expressed as

$$A_{\text{i2s}} = \text{softmax}\left(\frac{K_{\text{img}}Q^{\top}}{\sqrt{d_k}}\right)\cdot Q, \quad d_k = \frac{E}{H} \tag{6}$$

Through attention computed in both directions, the model achieves deep information exchange and complementary enhancement between image and structured features. Subsequently, the attention outputs are separately combined with the original features via residual connections and layer normalization, yielding the enhanced representations:

$$\tilde{f}_{\text{img}} = \text{LayerNorm}(F_{\text{img}} + A_{\text{i2s}}) \tag{7}$$

This operation not only preserves the information contained in the original features, but also introduces context-rich enhancements from the other modality, effectively boosting the discriminative power of the representation. To further control the relative contribution of the two modalities in the final fusion, the module incorporates a gating mechanism built

TABLE II
PERFORMANCE COMPARISON OF DIFFERENT METHODS ON OUR DATASET

Methods		Accuracy(%)	Sensitivity(%)	AUC
Traditional Machine Learning Methods	Logistic Regression [5]	36.78	24.56	0.3193
	K-Nearest Neighbor [6]	58.62	82.46	0.4576
	Decision Tree [5], [6]	51.72	47.37	0.5804
	Logistic Regression* [5]	64.37	49.12	0.6573
	Decision Tree* [5], [6]	67.82	66.67	0.7257
Ensemble Learning Methods	Random Forest [4]	66.67	84.21	0.6673
	Random Forest* [4]	70.11	59.65	0.7947
	XGBoost* [3]	77.01	87.72	0.8392
	AdaBoost* [3]	73.56	78.95	0.8006
	LightGBM* [3]	71.26	70.18	0.8193
Deep Learning Methods	VIT [11]	70.11	78.95	0.7319
	Diffusion Model [12]	63.22	57.89	0.7269
	Lung-EffNet [14]	79.31	80.70	0.8959
	3D U-Net* [2]	68.97	80.70	0.7404
	Transformer* [10]	72.41	80.70	0.8076
	VIT* [11]	65.52	52.63	0.7140
	Diffusion Model* [12]	67.82	61.40	0.7611
	DenseNet121* [13]	74.71	73.68	0.8193
	Lung-EffNet* [14]	78.16	77.19	0.8228
	LungConVT-Net* [15]	67.82	80.70	0.7573
Ours	**OVBI-LCDM**	**82.76**	**87.72**	**0.9135**

* Indicates that the method used our special multimodal data and weights.

upon a multi-layer perceptron (MLP) followed by a Sigmoid activation. Concretely, the enhanced image and structured features are first concatenated into a joint vector

$$\tilde{f}_{\text{ind}} = \text{LayerNorm}(Q + A_{\text{s2i}}), [\tilde{f}_{\text{img}}; \tilde{f}_{\text{ind}}] \in \mathbb{R}^{B\times 2E} \tag{8}$$

The joint vector is then passed through a two-layer MLP with non-linear activation to produce a gating weight vector whose dimension matches that of the image features:

$$g^{(i)} = \sigma(\text{MLP}([\tilde{f}_{\text{img}}; \tilde{f}_{\text{ind}}])) \in (0,1)^E \tag{9}$$

where σ denotes the element-wise Sigmoid function, ensuring that each dimension of the gate lies between 0 and 1, thereby enabling soft, per-dimensional selection between image and structured features.

$$f_{\text{fused}} = g \odot \tilde{f}_{\text{img}} + (1-g) \odot \tilde{f}_{\text{ind}} + \tilde{f}_{\text{img}} + \tilde{f}_{\text{ind}} \tag{10}$$

This expression not only incorporates the gated, weighted combination of features, but also—through its residual form—retains the bidirectionally enhanced original features, further enriching the representational power of the fused embedding. Finally, the fused feature is passed through a feed-forward network composed of a linear layer, ReLU activation, and layer normalization, which performs additional transformation and normalization to produce the ultimate multi-modal fusion feature vector f_{out}.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

In this study, we use Python 3.9 and utilize an NVIDIA GeForce RTX 4060 Laptop GPU to accelerate the training of neural networks. A total of 585 sets of CT images and their multimodal indicators of various types of lung cancer are obtained from the First Affiliated Hospital of Bengbu Medical University(382 groups are diseased and 203 groups were normal). We randomly divide the dataset according to the ratio of training set: test set: validation set=14:3:3 and conduct subsequent experiments.

TABLE III
ABLATION STUDY

Orthogonal Views	Blood Indicators	Fusion Module	Accuracy(%)	Sensitivity(%)	AUC
✓	×	×	75.86	80.70	0.8690
Axial	✓	×	66.67	96.49	0.5532
Coronal	✓	×	57.47	78.95	0.4895
Sagittal	✓	×	44.83	24.56	0.5018
✓	✓	×	81.61	84.21	0.8942
✓	✓	✓	**82.76**	**87.72**	**0.9135**

A. The Performance of Our Model

After the model's training and learning process, we obtain the results as shown in the Table II.

Our model, OVBI-LCDM, achieve the best performance across all evaluation metrics. This model attain an accuracy of 82.76%, a sensitivity of 87.72%, and an AUC of 0.9135, significantly outperforming all other compared models. This indicates that the OVBI-LCDM model possesses higher accuracy and reliability when processing multimodal data, effectively identifying positive samples while maintaining a high AUC value, demonstrating excellent classification capabilities.

B. Ablation Study

Based on the model design process, we conduct ablation experiments on orthogonal views, multimodal information and fusion module on the same test set. The results are shown in Table III. Overall best setting reached 82.76% accuracy, 87.72% sensitivity, and AUC 0.9135, showing the model most effective when it integrates multi-view and multi-source

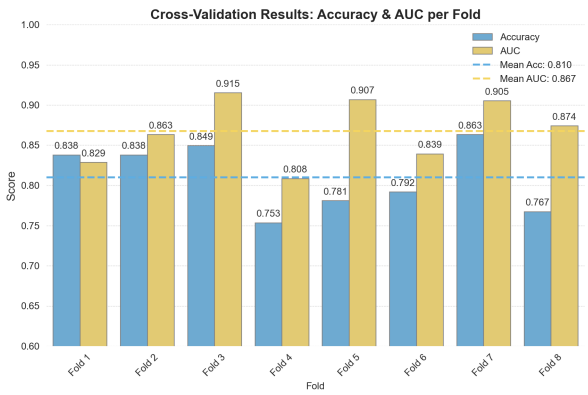

Fig. 3. Cross validation results

information via an explicit fusion stage. Without the fusion module but keeping orthogonal views and blood indicators the model still outperforms any single view, which indicates blood indicators and multi-view provide rich complementary cues; the fusion module's role is to optimally combine those cues and reduce the remaining performance gap. Single-view results are poor, demonstrating each view is partial and insufficient—orthogonal views supply complementary spatial context that improves detectability and reduces ambiguities. Using orthogonal views without blood indicators also degrades performance, confirming blood indicators supply important nonredundant information that boosts discrimination.

C. Cross Validation

To prevent the model from overfitting, we use cross-validation to verify its stability. In this study, we use stratified 8-Fold cross-validation, dividing the data into 8 subsets while maintaining the same class proportions as the overall data set.

From the Figure 3, it is evident that the average accuracy in all eight folds exceeds 81%, indicating that the model possesses strong generalizability in different subsets of data. Furthermore, the model's AUC scores are also commendable, with all fold AUC values above 0.80 and an average AUC as high as 0.867, further confirming the model's effectiveness in distinguishing between positive and negative samples.

V. Conclusion

In this paper, we propose a lung cancer detection model integrating orthogonal views and blood analysis indicator for patients with pleural effusion(OVBI-LCDM). Feature enhancement network based on view complementarity is designed to extract three-dimensional features from different orthogonal perspectives, creating a comprehensive view that enhances the detection of subtle lesions. Blood indicators analysis module is introduced and conducts a thorough examination of the correlation and significance of six key blood parameters. This analysis has allowed us to assign weights to these indicators, prioritizing those that are more strongly associated with lung cancer. Feature fusion module based on analysis results is proposed to adjust the importance of features from orthogonal views and structured data dynamically.

References

[1] L. Li, Y. Xu, Y. Wang, et al. *The Diagnostic and Prognostic Value of the Combination of Tumor M2-Pyruvate Kinase, Carcinoembryonic Antigen, and Cytokeratin 19 Fragment in Non-Small Cell Lung Cancer*, Technology in Cancer Research & Treatment, 23, 2024.

[2] H. Rikhari, E. B. Kayal, S. Ganguly, et al. *Lung Nodule Segmentation in CT Scans Using 3D U-Net Models with Inception and ResNet Architectures*, Proceedings of InC4 2024 - 2024 IEEE International Conference on Contemporary Computing and Communications, 2024.

[3] C. Arunkumar Madhuvappam, D. Vinod Kumar, S. Kanna, et al. *Enhanced Lung Cancer Detection using Ensemble Learning Algorithms: A Comparative Study of LightGBM & CatBoost*, Proceedings - 2024 5th International Conference on Image Processing and Capsule Networks, ICIPCN 2024, pp. 324–328, 2024.

[4] S. Makubhai, G. R. Pathak, P. R. Chandre, *Comparative analysis of explainable artificial intelligence models for predicting lung cancer using diverse datasets*, IAES International Journal of Artificial Intelligence, 13 (2), pp. 1978–1989, 2024.

[5] K. E. Narayana, K. Jayashree, A. ThavaVinu, et al. *Evaluation of Logistic Regression and Decision Tree Models for Lung Cancer Prediction Using Survey Data*, IEEE International Conference on Electronic Systems and Intelligent Computing, ICESIC 2024 - Proceedings, pp. 196–201, 2024.

[6] M. Yunianto, S. Suparmi, C. Cari, et al. *Comparative Performance Analysis of Lung Cancer Detection using Naïve Bayes, Support Vector Machine, K-Nearest Neighbor and Decision Tree*, International Journal of Intelligent Systems and Applications in Engineering, 11 (2), pp. 425–436, 2023.

[7] *Expert consensus on the clinical utilization of lung cancer serum biomarkers and establishment of standardized reference intervals for Chinese population*, Chinese Journal of Clinical Oncology, 48 (22), pp. 1135 - 1140, 2021.

[8] L. Wang, D. Wang, G. Zheng, et al. *Clinical evaluation and therapeutic monitoring value of serum tumor markers in lung cancer*, International Journal of Biological Markers, 31 (1), pp. e80 - e87, 2016.

[9] H. Ohkura, *Clinical usefulness of circulating tumor markers*, Gan to kagaku ryoho. Cancer & chemotherapy, 31 (7), pp. 1131 - 1134, 2004.

[10] A. Vaswani, N. Shazeer, N. Parmar, et al. *Attention is all you need*, *Advances in Neural Information Processing Systems*, 2017-December, pp. 5999–6009, 2017.

[11] A. Dosovitskiy, L. Beyer, A. Kolesnikov, et al. *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale*, *ICLR 2021 - 9th International Conference on Learning Representations*,pp. 5999–6009, 2021.

[12] J. Ho, A. Jain, P. Abbeel, *Denoising diffusion probabilistic models*, *Advances in Neural Information Processing Systems*, pp. 1–9, 2020.

[13] G. Huang, Z. Liu, L. Van Der Maaten, et al. *Densely connected convolutional networks*, *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017-January, pp. 2261–2269.

[14] R. Raza, F. Zulfiqar, M. O. Khan, et al. *Lung-EffNet: Lung cancer classification using EfficientNet from CT-scan images*, *Engineering Applications of Artificial Intelligence*, 126, pp. 1-14, 2023.

[15] A. Lasker, M. Ghosh, M. Sk, et al. *LungConVT-Net: A visual transformer network with blended features for Pneumonia detection*, *Pattern Recognition*, 171, pp. 112–150, 2026.

[16] K. T. Navya, G. Pradeep, *Lung nodule segmentation using adaptive thresholding and watershed transform*, in *Proc. 2018 3rd IEEE Int. Conf. Recent Trends Electron. Inf. Commun. Technol. (RTEICT)*, pp. 630–633, 2018.

[17] Y. R. Chillakuru, K. Kranen, V. Doppalapudi, Z. Xiong, et al. *High precision localization of pulmonary nodules on chest CT utilizing axial slice number labels*, *BMC Medical Imaging*, vol. 21, no. 1, pp. 1–11, 2021.

[18] L. Wang, A. M. Al-Jumaily, *Imaging of Lung Structure Using Holographic Electromagnetic Induction*, *IEEE Access*, vol. 5, pp. 20313–20318, 2017.

[19] P. Deepa, M. Arulselvi, M. Meenakshi Sundaram, *An Effective Method for Lung Cancer Classification Using Convolutional Neural Network*, *Int. J. Intell. Syst. Appl. Eng.*, vol. 11, no. 9s, pp. 461–467, 2023.

[20] J. F. Esha, T. Islam, M. A. Pranto, et al. *Multi-View Soft Attention-Based Model for the Classification of Lung Cancer-Associated Disabilities*, *Diagnostics*, vol. 14, no. 20, pp. 1–18, 2024.