Lung Cancer Subtype Image Classification Based on Multi-Model Fusion

HaoLiang Ji, Xiaoxia Zhang

Abstract—**Adenocarcinoma, a malignant tumor derived from glandular epithelial cells, is the most prevalent subtype of non-small cell lung cancer (NSCLC). Although it is commonly observed in smokers, it also occurs with notable frequency in non-smoking females and younger individuals. Typically, adenocarcinoma originates in the glandular cells of the small alveoli and is predominantly located in the peripheral regions of the lungs. Early diagnosis is crucial for improving treatment outcomes, and chest CT scans play a pivotal role in facilitating early detection. This paper presents an optimized multi-model fusion framework for the efficient classification of adenocarcinoma. The framework first employs weighted feature fusion to extract three-dimensional CT image features, followed by an enhanced Firefly Algorithm (FFA) to select the most discriminative features. Classification is then performed using the optimized feature set. Experimental results on the publicly available TCIA dataset demonstrate that the framework achieves an impressive accuracy (ACC) of 0.99, highlighting its exceptional performance in adenocarcinoma classification. Furthermore, the proposed approach holds significant potential for integration into computer-aided diagnostic systems, offering state-of-the-art capabilities for clinical applications.**

Index Terms—**Adenocarcinoma; Image Classification; CT Images; Convolutional Neural Networks; Deep Learning; Model Fusion**

I. Introduction

According to data from the World Health Organization (WHO), lung cancer is the leading cause of cancer-related deaths worldwide, with the highest mortality rates among both men and women. Lung adenocarcinoma, a major subtype of non-small cell lung cancer (NSCLC), accounts for 40%-50% of lung cancer cases and is particularly prevalent among individuals exposed to long-term air pollution or second-hand smoke [1]. Treatment options for advanced lung adenocarcinoma are limited, making early detection crucial for improving survival rates. Adenocarcinoma typically originates from bronchial epithelial cells in the distal lung and is characterized by glandular structures or mucin secretion. Early detection is critical for initiating treatment and preventing tumor progression [2]. CT scans are widely used to identify tumor-affected regions, not only for direct visual assessment but also for semi-quantitative analysis. In addition to providing basic imaging information, CT scans contain numerous features associated with lesions [3]. However, these features are often challenging to quantify or assess directly [4], requiring extraction and analysis for effective utilization. Although advancements in technology have improved lung cancer treatments, early detection remains a key strategy for reducing lung cancer mortality. Studies have shown that low-dose CT screening can effectively detect lung cancer at an early stage [5].

Manuscript received January 19, 2025; revised April 3, 2025.

H. L. Ji is a postgraduate student of School of Computer Science and Software Engineering, University of Science and Technology LiaoNing, Anshan, 114051, China (e-mail: 2546376312@qq.com).

X. X. Zhang is a Professor of School of Computer Science and Software Engineering, University of Science and Technology LiaoNing, Anshan, 114051, China (corresponding author, phone: 86-0412-5929812; e-mail: aszhangxx@163.com).

In recent years, advancements in artificial intelligence (AI) and deep learning have significantly enhanced chest imaging analysis, making it a rapidly evolving field [6]. These technologies can extract critical features from images that are imperceptible to the human eye, such as histogram features, texture features, and shape features. However, the large number of extractable features often results in high-dimensional data challenges [7]. To address this, researchers frequently employ machine learning techniques for dimensionality reduction, focusing on the most relevant features.

To overcome some limitations of traditional convolutional neural networks (CNNs), several novel network architectures have been introduced in recent years. These architectures, including ResNet, Inception networks [8], and DenseNet, have demonstrated effectiveness in learning target features from CT images with varying parameters. Nevertheless, existing methods still face challenges that limit their clinical application and require significant involvement from medical professionals [9]. Currently, lung cancer datasets are derived from various imaging techniques, including CT, positron emission tomography (PET), and X-rays. Among these, PET/CT is widely regarded as the standard imaging technology for evaluating lung cancer patients [10].

Lung cancer is broadly classified into two types: non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC) [11]. Adenocarcinoma, a major subtype of NSCLC, is characterized by various molecular mutations, some of which have significant clinical implications [12]. Figure 1 compares adenocarcinoma slices with other types of lung cancer [13]. Medical experts believe that regular examination of a large number of CT images from patients can effectively reduce the risk of disease [14]. However, the complex nodule information contained in CT scans makes accurate diagnosis increasingly difficult for doctors as the volume of images grows [15].

Zhao et al. [16] provided a comprehensive review of the application of two-dimensional convolutional neural networks (2D-CNNs), three-dimensional convolutional neural networks (3D-CNNs), and Faster Region-Based Convolutional Neural Networks (Faster R-CNNs) in lung nodule detection, emphasizing the immense potential of deep learning techniques in enhancing diagnostic accuracy.

Tan et al. [17] analyzed the characteristics of CT imaging and lung nodules, discussing the challenges and recent advancements in deep learning-based detection models. They highlighted the strengths and weaknesses of prominent achievements and proposed future directions for application and improvement.

Ilse [18] introduced an enhanced ResNeXt network, incorporating novel convolutional structures to improve the classification accuracy of pathological lung cancer images, achieving an overall accuracy of 99.47

Chi [19] proposed a lung nodule detection and classification method based on convolutional neural networks (CNNs). By optimizing the Faster R-CNN model, they significantly improved detection accuracy and developed a malignancy classification model using a Dual Path Network (DPN) with multi-scale decision fusion, achieving exceptional performance on the LIDC-IDRI dataset.

Shu et al. [20] proposed a malignancy classification network that integrates multi-scale feature extraction with global and local attention mechanisms, effectively improving classification accuracy through the incorporation of multi-scale features and attention mechanisms. Although these methods perform exceptionally well in cancer image processing applications, they also have some drawbacks. Many models lack sufficient feature processing capabilities, resulting in suboptimal quality of critical information in the images.

Based on the current state of research, we propose a learning framework that optimizes feature extraction through a multi-model fusion algorithm. The aim of this study is to develop a deep learning framework that leverages chest CT scan data for the classification of lung cancer subtypes. Each patient's 3D CT volume is regarded as an integrated whole composed of multiple slices. The model extracts instance-level features enriched with deep spatial information from individual slices and consolidates them into a unified representation. Additionally, an attention mechanism is employed to capture inter-slice correlations, further enhancing the robustness of feature learning and enabling accurate adenocarcinoma classification. Experimental results on publicly available datasets demonstrate that the proposed model outperforms existing cancer classification approaches, exhibiting superior accuracy and reliability.

The proposed framework integrates state-of-the-art neural networks, including MobileNetV2, EfficientNet-B3, and ResNeXt50, to extract complementary features from CT slices, capturing both global structural insights and fine-grained details. Attention-driven feature optimization is achieved through the Convolutional Block Attention Module (CBAM), refining the extracted features to ensure the model prioritizes the most informative regions. Simultaneously, a weighted feature fusion mechanism effectively combines features from multiple networks, enhancing robustness while mitigating the risk of overfitting.

The proposed model framework is adaptable to 3D CT scans with varying slice counts. Extensive experimental validation on lung CT images from the Cancer Imaging Archive (TCIA) public dataset confirms the effectiveness of the proposed method, achieving exceptional performance across metrics such as accuracy, thus demonstrating its superiority and reliability.

II. Related works

In recent years, deep neural networks have achieved remarkable success in various computer vision tasks, demonstrating significant potential in image feature learning. By increasing the depth and width of networks, researchers aim to capture more complex and abstract feature representations, thereby enhancing task performance through the preservation of relevant information. To address diverse task requirements and objectives, researchers continuously refine network architectures and develop classification algorithms that offer improved performance and generalization capabilities, thereby aiding medical professionals in making accurate diagnoses. Aligned with the goals of this study, this section provides an overview of related work on lung cancer classification and summarizes the design approaches employed in previous research. Figure 1 shows an image of lung cancer subtypes.

A. Dataset preprocessing

This study included 347 lung cancer patients (191A, 29B, 34G), with imaging data obtained from the TCIA database. Each case consisted of a CT volume, a PET volume, and fused PET/CT images. The CT images had a resolution of 512 × 512 pixels with a voxel size of 1 mm × 1 mm, while the PET images had a resolution of 200 × 200 pixels with a voxel size of 4.07 mm × 4.07 mm. Both modalities had a slice thickness and spacing of 1 mm, and the volumes were reconstructed using the same number of slices. Three-dimensional emission and transmission scans were acquired from the skull base to the mid-thigh. PET images were reconstructed using the TrueX TOF method, ensuring a consistent slice thickness of 1 mm.

The original CT images consisted of anisotropic voxels with varying in-plane resolutions. Due to differences in scanners and acquisition protocols, the voxel spacing of the resulting CT dataset varied. To facilitate training, all medical images were resampled based on the voxel spacing provided in the DICOM files, and the resolution was standardized. Slice thickness ranged from 0.625 mm to 5 mm, with scan modes including plain, contrast, and 3D reconstruction. Images were analyzed using a window width of 1050 HU and a window level of -475 HU. Reconstructions were performed in a lung window with a slice thickness of 2 mm.

The location of each tumor was annotated by five academic chest radiologists specializing in lung cancer, making the dataset a valuable resource for developing medical diagnostic algorithms. Two radiologists had over 15 years of experience, while the others had over 5 years. After one radiologist annotated each subject, the remaining four radiologists performed verification, ensuring that all annotations in the dataset were reviewed. Annotations were captured using LabelImg and saved in PASCAL VOC format as XML files.

B. Image Feature Extraction

The evolution of image feature extraction techniques has progressed from traditional hand-crafted methods to deep learning-driven automated approaches. Traditional methods, such as edge detection, texture analysis, and scale-invariant feature extraction, have proven effective in simpler scenarios but exhibit limitations when dealing with complex datasets. With the advent of deep learning, convolutional neural networks (CNNs) and their derivatives (e.g., Inception and ResNet) have significantly enhanced feature representation through multi-layered and multi-scale feature extraction. The integration of attention mechanisms, such

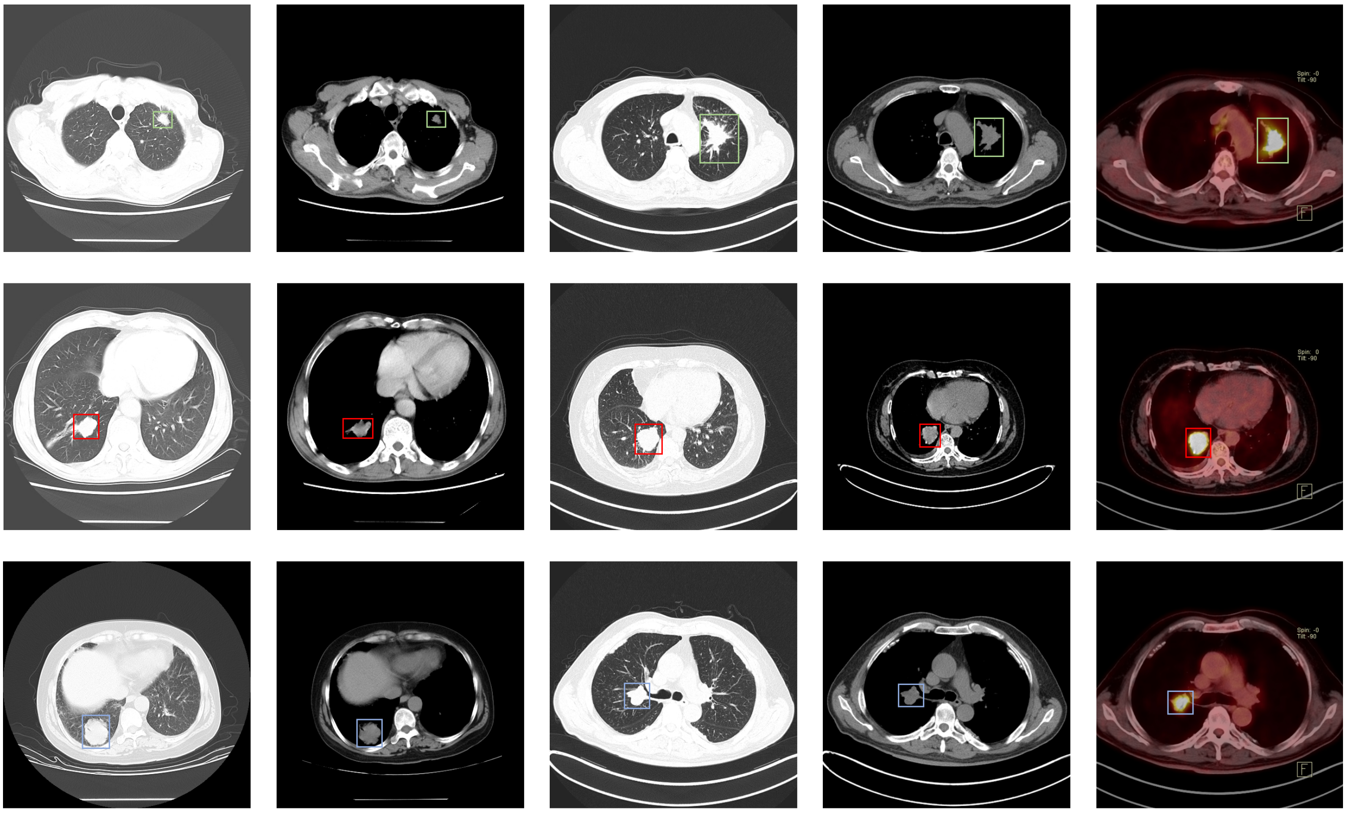

Fig. 1 Lung cancer subtypes image

as the Convolutional Block Attention Module (CBAM) and the Squeeze-and-Excitation (SE) block, alongside feature fusion strategies, allows models to focus more effectively on critical regions, thus strengthening feature representation. Furthermore, the introduction of self-attention mechanisms and Transformer-based architectures, such as the Vision Transformer (ViT) and Swin Transformer, has enabled global dependency modeling, further optimizing feature extraction capabilities. In medical image analysis, the application of 3D feature extraction, attention mechanisms, and multi-model fusion has significantly enhanced the analytical precision of high-dimensional data, such as lung CT scans. Shallow neural networks often struggle to handle complex image features, even with sufficient data, as they are unable to effectively focus on the distribution of key features. To address this, the present study employs a hybrid deep feature extraction model, composed of fine-tuned MobileNetV2, EfficientNet-B3, and ResNeXt50 networks, where the final fully connected layers output the extracted features. Multi-model fusion represents an advanced methodological framework designed to integrate the feature extraction capabilities of multiple models, effectively addressing complex learning tasks and high-dimensional data challenges. This study proposes a multi-model fusion framework that leverages the complementary strengths of pre-trained models such as MobileNetV2, EfficientNet-B3, and ResNeXt50. Through three distinct stages, the framework achieves precise feature extraction and optimized classification.

C. Multi-Model Feature Extraction and Fusion

This study employs three advanced deep learning architectures—MobileNetV2, EfficientNet-B3, and ResNeXt50—to extract multi-scale and multi-dimensional features. MobileNetV2 focuses on capturing global structural information with its lightweight and efficient design, enabling the rapid processing of large-scale medical images. EfficientNet-B3 excels in extracting fine-grained texture and morphological features, making it particularly adept at identifying subtle tumor characteristics with exceptional precision. ResNeXt50, utilizing grouped convolutions, enhances the extraction of critical regional features and effectively processes high-resolution input data. These models are integrated through a dynamic weighted fusion strategy, where adaptive weights are assigned to each model's contribution, optimized based on the specific task requirements. The fused features are then mapped into a tensor representation, ensuring seamless integration and consistency. This multi-model fusion approach balances local feature sensitivity with global contextual understanding, effectively combining global insights and local precision to enhance the robustness and comprehensiveness of feature representation.

D. Feature Optimization and Selection

Swarm intelligence optimization algorithms, inspired by the collective behavior observed in nature, represent a class of intelligent optimization techniques widely used in feature selection, particularly in high-dimensional and complex datasets. These algorithms encode feature subsets as individuals or particles and iteratively optimize them within the search space to identify the optimal feature combination, thereby reducing redundancy and improving model performance. Notable swarm intelligence algorithms include Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO), Firefly Algorithm (FA), Genetic Algorithm (GA), and Bat Algorithm (BA). For example, PSO performs rapid global searches through iterative updates of individual velocity and position; ACO evaluates feature importance using pheromone-guided search; FA optimizes feature subsets based on light intensity; GA evolves feature subsets through genetic operations; and BA balances search efficiency and accuracy by integrating local and global search strategies. These algorithms exhibit strong global search capabilities, high flexibility, and compatibility with various classifiers, effectively eliminating irrelevant or redundant features while enhancing model accuracy and computational efficiency.

In this study, an enhanced Firefly Algorithm (FA) is employed to refine the extracted features. The algorithm uses a fitness function to balance classification accuracy and feature sparsity, ensuring effective feature selection while minimizing redundancy. Additionally, an improved search strategy is incorporated, integrating LSAMV flight dynamics into the FA for a more refined optimization process. This enhancement ensures an effective balance between global exploration and local exploitation, leading to superior feature optimization performance.

E. Attention Mechanism for Classification

In this study, the model utilizes an attention mechanism to aggregate the optimized features, assigning dynamic importance to key slices for bag-level prediction. The Convolutional Block Attention Module (CBAM) enhances feature relevance through both channel and spatial attention mechanisms, allowing the model to focus on high-information regions and amplify the influence of critical instances, thereby improving classification accuracy and interpretability. Simultaneously, the attention mechanism preserves the spatial correlations between CT slices during feature aggregation, ensuring the integrity and fidelity of the bag-level representation. This effectively filters and emphasizes key features, providing strong support for bag-level classification.

III. Methods

In the field of medical image classification, feature extraction and selection are pivotal components that significantly impact the overall performance of a model. Multi-model fusion frameworks leverage outputs from multiple models or feature spaces, integrating the unique structures and mechanisms of each model to extract complementary features. This approach effectively addresses the challenges posed by multi-scale and multi-dimensional features in complex datasets. However, the feature fusion process is often hindered by redundant or conflicting information, which increases the complexity of optimization and selection. Furthermore, existing methods generally lack the ability to dynamically evaluate feature importance across models, leading to the selection of features that may not meet the specific requirements of the task at hand. Additionally, the stochastic nature of cross-model feature extraction further exacerbates challenges related to global optimization and local refinement. Therefore, optimizing strategies for selecting fused features is critical for improving classification performance. This section outlines the overall workflow of the proposed framework for multi-model fusion and feature optimization, with subsequent sections providing a detailed design and development of each module and model. Figure 2 presents the architecture diagram of the model.

A. Overall Network Architecture

This paper proposes an image classification framework based on multi-model fusion and algorithmic optimization, designed to efficiently extract multi-scale and multi-dimensional features, thereby achieving high-precision classification. The process begins with preprocessing raw CT images, including resolution adjustment, voxel spacing standardization, and resampling, to ensure data quality and consistency that meet the model's training requirements. Subsequently, a dynamic weighted fusion mechanism integrates feature outputs from multiple deep learning networks by assigning adaptive weights to the feature vectors of each model and mapping the fused features into a unified high-dimensional tensor representation.

In the feature optimization stage, an improved Firefly Algorithm (FFA) is employed, which combines global search and local refinement strategies. This effectively reduces redundant features while preserving spatial correlations between CT slices, ensuring the integrity and fidelity of the feature representations. Additionally, channel and spatial attention mechanisms are incorporated to dynamically aggregate the optimized features, focusing on high-information regions and further enhancing the semantic representation of critical features.

Finally, the aggregated features are fed into the classifier for prediction. By synergistically combining multi-model fusion, algorithmic optimization, and attention mechanisms, the proposed framework achieves a balance between global semantic understanding and local detail sensitivity. This provides an efficient and robust solution for precise lung cancer subtype classification and offers a broadly applicable paradigm for high-resolution medical image classification tasks.

B. Optimization of Image Features Using the FFA

The Improved Firefly Algorithm (FFA) is a swarm intelligence-based optimization method specifically designed to address the problem of feature selection in image classification. This algorithm incorporates multiple enhancements over the classical Firefly Algorithm, including the introduction of the Lévy flight mechanism to enhance global search capabilities, as well as dynamic attractiveness adjustment and feature sparsity constraints to effectively extract key features while eliminating redundancies, thereby significantly improving classification performance. Compared to the original algorithm, the proposed improvements focus on three critical aspects: Prioritizing features that make substantial contributions to classification accuracy; Introducing sparsity constraints to reduce redundant features, thereby lowering computational complexity and enhancing model generalization; Leveraging the Lévy flight mechanism to avoid local optima while dynamically adjusting attractiveness to strengthen local search capabilities.

Furthermore, the classical Firefly Algorithm simulates the attraction between fireflies to identify the global optimum. In the improved version, additional optimizations are implemented in the initialization phase. Each firefly is initialized

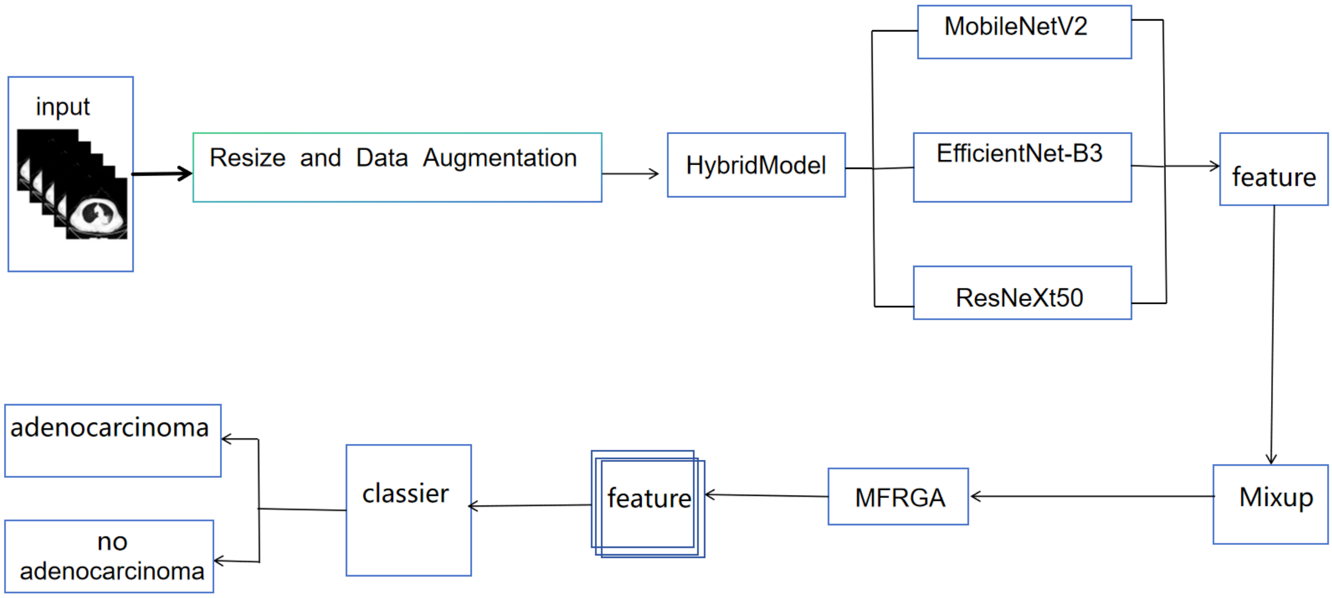

Fig. 2 Overview of the framework

as a binary feature selection vector (e.g., [1, 0, 1, ...], where 1 denotes selected features and 0 denotes excluded ones), ensuring that the initial population comprehensively covers the solution space. This enhancement strengthens the algorithm's global search ability and improves optimization efficiency. These advancements make the proposed algorithm highly effective and broadly applicable for feature selection tasks in complex, high-dimensional datasets.

Generate a random population, where each individual (firefly) is represented by a position vector $x_i = [x_1, x_2, \ldots, x_n]$, with each element $x_j \in [0, 1]$. Ensure the population is well-distributed across the solution space to enhance global exploration capability.

The fitness function balances classification accuracy and the number of features. It is defined as follows:

$$F(X) = F(X \cdot Ex\left(\sum Sq(X)\right)) \tag{1}$$

Calculated using a simple classifier such as KNN on training and validation data. Represents the proportion of unselected features:

$$FeatureFactor = \frac{TotalFeatures - SelectedFeatures}{TotalFeatures} \tag{2}$$

Equation 3 updates the position of firefly *i* by utilizing the position difference between fireflies *j* and *i* ($X_j - X_i$) and the random jumps of Levy flight. Equation 4 calculates the Euclidean distance between two fireflies in a d-dimensional space.

$$X_i = X_i + \beta \cdot (X_j - X_i) + \alpha \cdot LevyFlight \tag{3}$$

$$r_{ij} = \sqrt{\sum_{k=1}^{d} (x_{j,k} - x_{i,k})^2} \tag{4}$$

Where X_i, X_j represent the positions of firefly individuals (feature selection vectors). β represents the attractiveness factor. *r* represents the distance between two fireflies, where x_i and x_j are the position vectors of the two fireflies in a d-dimensional space. Equation 5 represents the relationship between the attraction β of fireflies and the distance r between them:

$$\beta = \beta_0 \cdot e^{-\gamma r^2} \tag{5}$$

Where γ represents the controls the decay rate of attraction. r represents the distance between firefly individuals. Lévy flight is a form of random walk characterized by step lengths that follow a power-law distribution, seamlessly integrating local exploitation with long-range exploration. This behavior makes it a core mechanism in numerous optimization algorithms and search strategies. Its mathematical representation is as follows:

$$LevyFlight \sim \frac{u}{|\nu|^{1/\lambda}} \tag{6}$$

Here, u and ν are random variables sampled from specific distributions, such as Gaussian or uniform distributions, while λ lambda serves as the scaling exponent of the power-law distribution, governing the step-length characteristics. By balancing fine-grained local search with broad global exploration, Lévy flight significantly enhances the efficiency of global optimization. Feature Optimization Process.

In conclusion, the Firefly Algorithm (FFA) is an efficient feature selection optimization technique that enhances global search capabilities through dynamic attractiveness adjustment and Lévy flight, while incorporating feature sparsity constraints to extract critical features and eliminate redundant ones. This approach not only improves classification performance but also significantly reduces computational complexity.

The algorithm represents feature selection vectors using the positions of firefly individuals and initializes the population to ensure comprehensive coverage of the global solution space. A fitness function is used to balance classification accuracy with the minimization of the number of selected features. During each iteration, the algorithm evaluates the fitness of each individual, updates the positions of weaker

Algorithm 1 Firefly Optimization Process

Input: n: Population size (number of fireflies), d: Dimensionality of the solution space (number of features), α: Randomization parameter, β_0: Maximum attractiveness, γ: Light absorption coefficient, λ: Lévy flight distribution parameter, $f(x)$:Fitness function, max_iter:Maximum number of iterations.

Output: x^*: Best solution found.

for $t = 1$ to max_iter **do**
  for $i = 1$ to n **do**
    for $j = 1$ to n **do**
      if $L_i > L_j$ **then**
        $r_j \leftarrow$ eq. (4) ▷ Equation (4) for r_j
        $B_j \leftarrow$ eq. (5) ▷ Equation (5) for B_j
        $step \leftarrow$ arrow ▷ Use arrow-based update step
        $\lambda_i \leftarrow \lambda_i + step$ ▷ Update λ_i
        $x_i \leftarrow \text{clip}(x_i, 0, 1)$ ▷ Clip to valid bounds
      end if
    end for
  end for
  $L_i \leftarrow f(\lambda_i), i = 1, 2, \ldots, n$
  if $\max(L_i) > f(x^*)$ **then**
    $x^* \leftarrow \arg\max_i(L_i), f(x^*) = \max(L_i)$
  end if
end for

individuals using a dynamic attractiveness mechanism, and employs random Lévy flights to explore a broader solution space, effectively avoiding local optima. The continuous position vectors are then binarized to generate the final feature selection vectors.

By incorporating feature sparsity constraints, the algorithm prioritizes the enhancement of classification performance, ensuring that the selected feature subset demonstrates strong discriminative power and robust generalization capabilities. The IFA strikes a delicate balance between global exploration and local exploitation during the feature selection process, with convergence curves documenting the optimization trajectory. Ultimately, the algorithm outputs the globally optimal feature selection vector along with its corresponding fitness value, providing a robust and efficient solution to high-dimensional feature extraction challenges.

C. Multi-Model Fusion

This section presents an improved deep learning framework based on multi-model feature fusion, referred to as HybridMode-J. The framework leverages the feature extraction capabilities of advanced pre-trained models, namely MobileNetV2, EfficientNet-B3, and ResNeXt-50. By harnessing the strengths of these models in capturing diverse feature representations across various task scenarios, HybridMode-J achieves collaborative feature fusion across multiple architectures.

To address the limitations of fixed-weight designs, the framework incorporates a dynamic weight adjustment strategy, allowing the model to adaptively balance the contributions of each pre-trained model during feature fusion. Furthermore, to resolve the issue of inconsistent feature dimensions produced by different pre-trained models, a unified dimensionality reduction method is introduced. This method maps the outputs of multiple models into a consistent feature space, reducing the computational complexity of feature fusion, enhancing the efficiency of the process, and improving the model's overall stability.

HybridMode-J demonstrates advanced capabilities for multi-model feature extraction by leveraging the pre-trained feature representations of MobileNetV2, EfficientNet-B3, and ResNeXt-50. Specifically, MobileNetV2 offers lightweight, high-resolution feature representations, making it ideal for efficiently capturing key spatial features. EfficientNet-B3 employs compound scaling to achieve an optimal balance between precision and computational efficiency, excelling at extracting fine-grained details. Meanwhile, ResNeXt-50 capitalizes on its multi-path feature aggregation architecture, providing robust and diverse high-level semantic representations.

By integrating the outputs from these three models into a unified representation, the framework effectively captures a rich spectrum of semantic information, combining the strengths of multiple architectures to deliver enhanced feature expressiveness and task performance.

To achieve adaptive multi-model feature fusion, HybridMode-J incorporates a learnable dynamic weighting mechanism. Through weight optimization during training, the model dynamically adjusts the contribution of MobileNetV2, EfficientNet-B3, and ResNeXt-50 to the feature fusion process based on the input data distribution. Specifically, the framework ensures that the sum of the weights is constrained to 1 via a normalization operation and optimizes the weight parameters through a regularization term embedded in the loss function. This design allows the model to flexibly adapt the feature contributions of different architectures across varying scenarios, significantly enhancing its adaptability and generalization capability.

The varying feature dimensions generated by different models pose challenges for direct fusion, leading to potential dimensional mismatches or excessive computational complexity. To mitigate this, HybridMode-J employs an independent linear dimensionality reduction strategy, transforming the feature outputs of each model into a unified 512-dimensional space. This approach ensures the retention of the original semantic richness while substantially reducing computational overhead, establishing a more efficient and scalable foundation for feature fusion.

In summary, HybridMode-J seamlessly integrates the feature extraction capabilities of multiple pre-trained models, utilizing a dynamic weighting mechanism and a dimensionality reduction-based fusion strategy to achieve highly compact and efficient feature representation learning. The model exhibits exceptional performance across diverse tasks, providing a novel and robust solution for high-dimensional feature fusion in complex and challenging scenarios.

D. Integrating the FFA with classifiers

The Improved Firefly Optimization Algorithm (FFA) employs a global search strategy based on the LSAMV flight mechanism, enabling dynamic and efficient optimization of the feature selection process while effectively avoiding local optima. This approach excels at identifying the features most relevant to the classification task, substantially reducing feature redundancy and mitigating the influence of noise, thereby enhancing the classifier's overall performance.

To further refine classification accuracy, the classifier is designed to focus on identifying key features while minimizing the impact of irrelevant or noisy inputs. To address the challenge of label imbalance within datasets, the fitness function incorporates a compactness evaluation metric, allowing the algorithm to prioritize fewer but more discriminative features. This not only reduces computational complexity but also enhances the model's generalization capabilities. Additionally, the classifier integrates global pooling and attention mechanisms, enabling it to capture complex interrelationships between features. When paired with the optimized feature set, this approach significantly improves classification outcomes.

By combining the global optimization capabilities of the improved FFA with the deep embedding and representation learning strengths of the classifier, the proposed framework adapts seamlessly to high-dimensional feature spaces and complex data distributions, such as those encountered in CT slice analysis. The enhanced attraction formula and feature sparsity constraints in the improved FFA ensure rapid convergence while minimizing the effects of noise and redundant features, resulting in a highly robust and efficient model.

The integration of the improved FFA with the classifier fully harnesses the optimization algorithm's ability for precise and efficient feature selection and the classifier's superior feature representation and classification performance. This synergistic framework not only enhances the accuracy and efficiency of feature selection but also significantly improves classification performance on complex and challenging datasets. The approach is particularly suited for real-world applications involving high-dimensional and imbalanced data, such as the classification of adenocarcinoma CT images.

E. Attention-Based Multi-Model Fusion

This study proposes an attention-based multi-model fusion approach to address the complex feature distributions and sequential relationships within CT slices.This method effectively integrates features from multiple models and optimizes them using the Convolutional Block Attention Module (CBAM), improving the model's focus on critical features and enhancing classification performance.

Multi-model feature extraction preprocessed CT images are input into three deep learning models: MobileNetV2, EfficientNet-B3, and ResNeXt50. Each model extracts features at different levels, capturing both global structural information and fine-grained details. The outputs from these models are unified into low-dimensional embeddings through a dimensionality reduction layer:

$$Feature_i = Reduce(f_i(x)) \tag{7}$$

Where $f_i(x)$ denotes the output of the *i-th* model for input *x*, and Reduce is the dimensionality reduction operation mapping high-dimensional outputs to low-dimensional feature embeddings.

Dynamic attention-weighted fusion to dynamically adjust each model's contribution to feature fusion, an attention mechanism assigns weights to the extracted features. These weights are generated by a multi-layer perceptron (MLP) and normalized using a Softmax operation:

$$\omega_i = softmax(MLP(Feature_i)) \tag{8}$$

The features from each model are then weighted and fused as follows:

$$Fusion_i = \omega_i \cdot \mathrm{Feature}_i \tag{9}$$

Finally, the weighted features from all models are summed to produce a unified fusion representation:

$$Fusion = \sum_{i=1}^{N} Fusion_i \tag{10}$$

Where N is the number of models being fused.

Channel and spatial attention optimization the fused features are further optimized using the Convolutional Block Attention Module (CBAM), which enhances features along both the channel and spatial dimensions. The CBAM operations are defined as follows.

Channel Attention:

$$M_c(F) = \sigma\left(MLP\left(AvgPool(F)\right) + MLP\left(MaxPool(F)\right)\right) \tag{11}$$

Spatial Attention:

$$M_s(F) = \sigma\left(Conv2D\left([AvgPool(F)], MaxPool(F)\right)\right) \tag{12}$$

Here, $M_c(F)$and $M_s(F)$ represent the channel and spatial attention maps, respectively, F is the input feature, and σ sigma is the Sigmoid activation function. The CBAM-generated attention maps adjust feature weights, enhancing the model's focus on critical regions.

The attention mechanism dynamically integrates the features of three models, effectively combining multi-level feature information. Simultaneously, CBAM optimizes the fused features, enabling the model to focus on critical regions and information within CT images, thereby enhancing robustness and accuracy. This attention-based multi-model fusion approach strengthens the model's ability to extract key features from CT images, achieving efficient and accurate classification in lung cancer detection tasks.

TABLE I
A DESCRIPTION OF THE LUNG CT IMAGE DATASET

Datasets	Classes	Train	Test
TCIA	A	8844	3791
	B	151	65
	E	141	60
	G	3597	1542
	Total	14034	6015

TABLE II
THE DIFFERENT HYPERPARAMETER SETS OF VARIOUS OPTIMIZATION

Optimization Algorithm	Parameter	Value
Crow Search Algorithm (CSA)	Awareness_Prob	0.1
	Flight_Length	1.0
Particle Swarm Optimization (PSO)	W	0.7
	C1	1.5
	C2	1.5
Atom Search Optimization (ASO)	Alpha	50
	Beta	0.2
	Mass_Min	1.0
	Mass_Max	10.0
	Obj_Function	Compute_Fitness
Mayfly Algorithm (MA)	Alpha	0.7
	Beta	1.5
	Gamma	0.2
	Mutation_Rate	0.1
Ant Lion Optimizer (ALO)	Pop_Size	20
	Max_Iter	50
	Obj_Function	Compute_Fitness
Generalized Normal DistributionOptimization (GNDO)	Num_Agents	20
	Max_Iterations	50

IV. EXPERIMENTS

A. Datasets

In this study, the publicly available lung CT dataset TCIA was selected for experimentation and used to evaluate the proposed model. This section provides a detailed description of the dataset, as outlined in Table 1. The TCIA dataset consists of four categories: ”A” for adenocarcinoma, ”B” for small-cell carcinoma, ”E” for large-cell carcinoma, and ”G” for squamous cell carcinoma, totaling 20,049 slices. The ratio for training and validation was set at 4:1.

For this experiment, all images in the dataset were stored in tensor format, with each slice resized to a resolution of 256 × 256. Compared to the original slices, this downscaling reduces the size to a quarter of the original, significantly minimizing the overall dataset file size. Additionally, this approach substantially accelerates training time and reduces the demands on GPU memory. After testing, it was observed that the resized data exhibited negligible differences from the original data.

B. Experimental Details

All experiments were conducted on a local workstation equipped with an Intel(R) Core(TM) i7-12700H processor and an NVIDIA GeForce RTX 3090 Laptop GPU. Extensive experiments were then carried out to evaluate the model parameters. Specifically, based on empirical experience, the model’s initial learning rate was set to 0.0005, and the maximum number of iterations (Tmax) was configured to 60. The model was trained using the Adam optimization algorithm, with the number of epochs set to 64. The proposed model integrates deep learning methodologies, offering advantages that are applicable to both small and large datasets. Compared to similar approaches, this method demonstrates strengths such as weak supervision, eliminating the need for lesion segmentation.

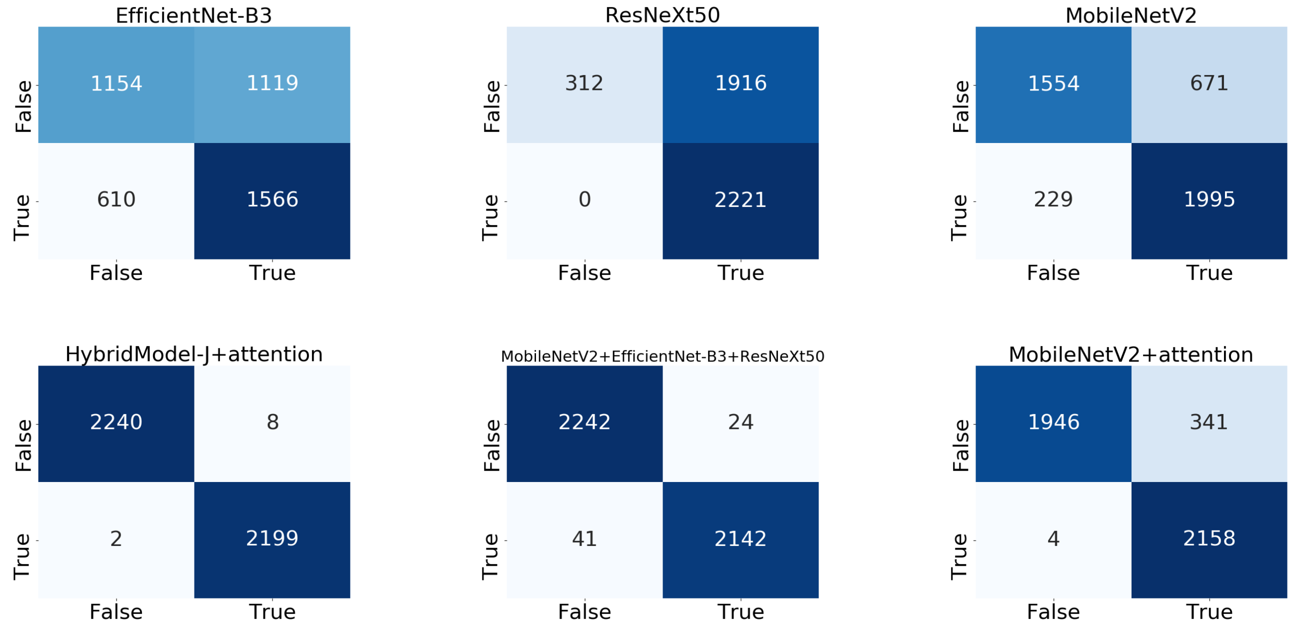

Fig. 3 Confusion matrix for the proposed and comparable models

Standard classification performance metrics, namely accuracy (ACC) and the F1 score, were used to evaluate the model's performance. In this context, 0 and 1 represent adenocarcinoma and non-adenocarcinoma, respectively. Additionally, recall corresponds to sensitivity (SEN). The F1 score helps mitigate the impact of imbalanced data. Ablation experiments were then conducted to validate each module within the model. This study integrates the FFA module and the HybridModel-J module. The performance of the two modules will be assessed based on their respective metrics and significance. Tables 2 and 3 present the detailed evaluation criteria for these two modules.

C. Comparison with Other Feature Selection Algorithms

We evaluated the proposed enhanced Firefly Algorithm (FFA) against 11 widely recognized metaheuristic-based feature selection techniques. These methods include the Mayfly Algorithm (MA), Particle Swarm Optimization (PSO), Atom Search Optimization (ASO), Ant Lion Optimizer (ALO), and the Crow Search Algorithm (CSA). The comparative results demonstrate that FFA outperforms these approaches in both the quality and quantity of the selected features.

The significant performance enhancement of FFA can be primarily attributed to the incorporation of the random flight variable mechanism. This innovative mechanism substantially improves the algorithm's learning capability, increases the exploration potential of the search space, and establishes a balanced trade-off between exploration and exploitation through a systematic fitness evaluation and position update strategy.

We also analyzed the convergence characteristics of these algorithms. Based on the convergence results, FFA exhibited exceptional robustness, accelerated convergence rates, and superior optimization efficacy.

Through this comprehensive comparative analysis, it can be concluded that FFA demonstrates outstanding performance across datasets. As a result, FFA excels in performing feature selection tasks with remarkable consistency and reliably achieves its intended objectives.

TABLE III
THE DIFFERENT HYPERPARAMETER SETS OF VARIOUS OPTIMIZATION

Method	Acc
CSA	98.79
PSO	99.01
ASO	98.85
MA	98.55
ALO	99.15
GNDO	98.11
FFA	99.99

D. Experimnts Based on the HybridModel—J Module

In this section, the experiments are divided into four fundamental models based on the proposed architecture, as outlined in Table 4. The primary feature extraction modules utilized are EfficientNet-B3t, MobileNetV2t, and ResNeXt50, upon which the CBAM module is incrementally integrated to evaluate performance enhancements. As shown in Figure 4, the left side displays the change in training and validation loss over the training epochs, while the right side illustrates the change in training and validation accuracy over the training epochs.

As depicted in Table 4, the configuration of HybridModel-J + CBAM (attention) demonstrates superior performance at

TABLE IV
COMPARISON OF MODEL PERFORMANCE FOR SEVERAL DIFFERENT MODULE COMBINATIONS

Model	ACC	F1 score	Precision
MobileNetV2	0.57	0.47	0.77
EfficientNet-B3	0.61	0.61	0.62
ResNeXt50	0.62	0.62	0.62
MobileNetV2+EfficientNet-B3+ResNeXt50	0.92	0.92	0.93
MobileNetV2+attention	0.81	0.81	0.81
HybridModel-J+attention	0.99	0.99	0.99

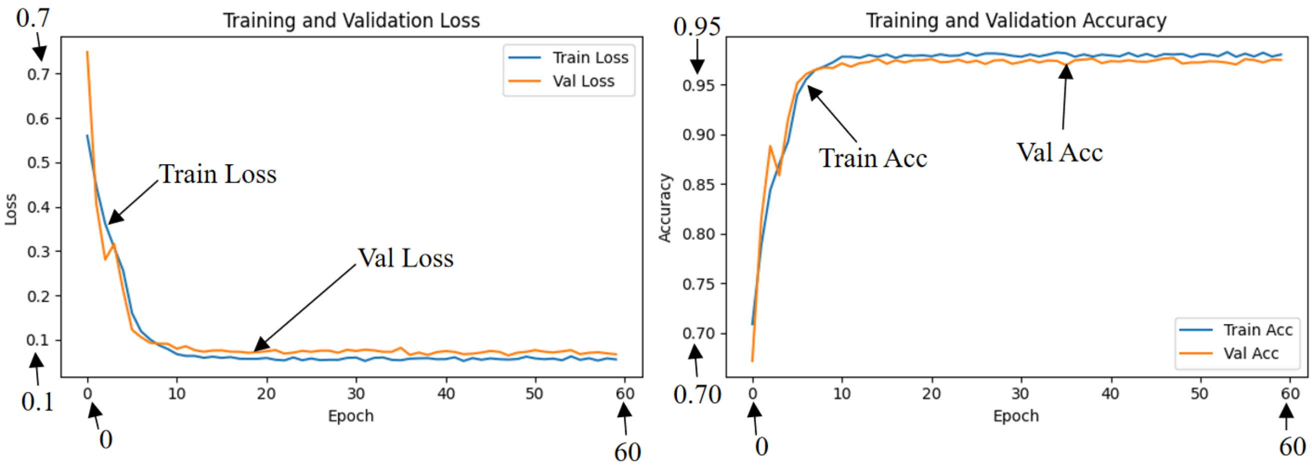

Fig.4 Training and validation results

the patient level, achieving an accuracy (ACC) of 0.99 and an F1 score of 0.99. The corresponding confusion matrix is shown in Figure 3.

V. CONCLUSION

This paper proposes a multi-model fusion-based approach for extracting and characterizing deep features of lung cancer lesions in CT images. In this approach, deep learning is utilized as a feature extractor, and a hybrid model, created by fusing multiple models, is employed for feature extraction. After feature optimization using the FFA algorithm, a classifier is applied for classification. The model is capable of learning and fusing 2D and 3D feature maps from an arbitrary number of CT images. Experimental results demonstrate that the proposed method effectively aggregates features by leveraging both explicit diagnostic characteristics and latent deep information, significantly improving classification accuracy. Given the validated effectiveness of this approach in the domain of lung cancer, it is expected to perform exceptionally well in other medical image analysis tasks as well.

REFERENCES

[1] R. Manafi-Farid, E. Askari, and I. Shiri, "[^{18}f]fdg-pet/ct radiomics and artificial intelligence in lung cancer: Technical aspects and potential clinical applications," *Seminars in Nuclear Medicine*, vol. 52, pp. 759–780, 2022.

[2] G. Chassagnon, C. Margerie-Mellon De, and M. Vakalopoulou, "Artificial intelligence in lung cancer: current applications and perspectives," *Japanese Journal of Radiology*, vol. 41, pp. 235–244, 2023.

[3] A. Gopinath, P. Gowthaman, and L. Gopal, "Enhanced lung cancer classification and prediction based on hybrid neural network approach," in *2023 8th International Conference on Communication and Electronics Systems (ICCES)*, vol. 2023, 2023, pp. 933–938.

[4] S. Tomassini, N. Falcionelli, and P. Sernani, "Cloud-ylung for non-small cell lung cancer histology classification from 3d computed tomography whole-lung scans," in *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, vol. 2022, 2022, pp. 1556–1560.

[5] Y. Wang, Z. Liu, and X. Zhang, "Lung cancer subtype classification using deep convolutional neural networks and radiomics," *Medical Physics*, vol. 51, no. 3, pp. 981–992, 2024.

[6] L. Chen, Q. Li, and Z. Yang, "Deep learning for early detection of lung cancer using low-dose ct scans," *IEEE Transactions on Biomedical Engineering*, vol. 71, no. 4, pp. 870–880, 2024.

[7] T. Wang, Q. Fan, H. Cai, and B. Zhang, "Application of machine learning for tracing the origin of metastatic lung cancer tissues," *IAENG International Journal of Computer Science*, vol. 50, no. 2, pp. 359–367, 2023.

[8] T. Liu, Z. Wang, and X. Zhang, "Multi-feature fusion for lung cancer metastasis detection using pet/ct and mri images," *IEEE Transactions on Biomedical Engineering*, vol. 71, no. 7, pp. 1974–1986, 2024.

[9] M. Zhou, J. Wang, and J. Chen, "Deep multi-modal fusion for colorectal cancer diagnosis using histopathology and radiological images," *Journal of Clinical Imaging*, vol. 42, no. 5, pp. 1579–1589, 2024.

[10] S. K. Thakur, D. P. Singh, and J. Choudhary, "Lung cancer identifica-

tion: a review on detection and classification," *Cancer and Metastasis Reviews*, vol. 39, pp. 989–998, 2020.

[11] Z. Liu and Y. Liu, "Multi-modal fusion learning for diabetic retinopathy detection using fundus and oct images," *IEEE Transactions on Biomedical Engineering*, vol. 68, no. 4, pp. 1187–1195, 2021.

[12] V. Patel, P. Patel, and S. Patel, "Fusion of clinical and radiological data for predicting breast cancer using a hybrid machine learning model," *Journal of Clinical Oncology*, vol. 41, no. 6, pp. 457–468, 2023.

[13] X. Zhang, Y. Li, and Z. Wang, "Multi-modal fusion for lung cancer detection using deep learning," *IEEE Transactions on Medical Imaging*, vol. 43, no. 2, pp. 180–192, 2024.

[14] W. Li, Z. Wang, and F. Zhang, "A hybrid machine learning model for prostate cancer detection using multi-modal mri and ultrasound imaging," *Journal of Clinical Imaging*, vol. 42, pp. 123–135, 2024.

[15] Y. Xia, Z. Yang, and X. Liu, "Multimodal data fusion using deep learning for predicting alzheimer's disease," *IEEE Transactions on Medical Imaging*, vol. 40, no. 12, pp. 3165–3174, 2021.

[16] J. Zhao, Y. Zhang, L. Liu *et al.*, "A comprehensive review of lung nodule detection using 2d-cnns, 3d-cnns, and faster r-cnns," *Journal of Medical Imaging*, vol. 6, no. 2, pp. 151–165, 2019.

[17] Y. Tan, Z. Li, X. Zhang *et al.*, "Deep learning for lung nodule detection: Challenges and advances in ct imaging," *IEEE Transactions on Medical Imaging*, vol. 39, no. 5, pp. 1214–1234, 2020.

[18] M. Ilse, J. Munoz, J. Zwart *et al.*, "Improving lung cancer image classification using enhanced resnext network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 352–358.

[19] M. Chi, H. Zhang, T. Yu *et al.*, "Lung nodule detection and classification based on optimized faster r-cnn and dual path network (dpn)," *IEEE Transactions on Medical Imaging*, vol. 40, no. 9, pp. 2879–2889, 2021.

[20] X. Shu, H. Zhang, X. Liu *et al.*, "Malignancy classification of lung cancer using multi-scale feature extraction and attention mechanisms," *Journal of Medical Imaging and Health Informatics*, vol. 11, no. 6, pp. 1472–1481, 2021.