A CNN-Based Framework for Automatic Segmentation of Chest X-ray Images and Multi-Type Respiratory Disease Recognition

Juan Xiao[1*] , Xiaobo Yin[2] , Aiyong Liu[1] , Zhanghong Yu[1]

[1] School of Mathematics and Information Science & Technology, Hebei Normal University of Science & Technology, Qinhuangdao 066004, China
[2] University-Town Hospital of Chongqing Medical University, Chongqing 401331, China

Corresponding Author Email: juan-xiao@126.com

ABSTRACT

Respiratory diseases pose a significant threat to human health, and early, accurate diagnosis is critical for improving patient outcomes. Chest X-ray imaging, due to its affordability and convenience, remains the primary tool for clinical screening. However, traditional manual interpretation is time-consuming, experience-dependent, and prone to oversight and misdiagnosis. With the advancement of deep learning in medical imaging, computer vision–based automated analysis offers promising solutions to these challenges. Nevertheless, existing methods such as U-Net and its variants often struggle with accurately segmenting complex lung structures, especially when dealing with noisy images, small lesions, or blurred boundaries. Additionally, conventional 2D convolutional neural networks (CNNs) have limitations in capturing the spatial features inherent in chest X-ray images, and current multi-disease classification models still face challenges in achieving high accuracy and generalizability. To address these issues, this study proposes two key innovations: First, an optimized U-Net++L3 network with pruning is developed for automatic chest X-ray segmentation, effectively reducing parameter redundancy while maintaining accuracy, thereby enhancing segmentation performance in regions with complex lesions. Second, a densely connected 3D CNN model is designed for the recognition of multiple respiratory diseases. By leveraging the spatial feature extraction capabilities of 3D convolutions and the feature reuse advantages of dense connections, the model achieves precise classification of conditions such as pneumonia, lung cancer, and chronic obstructive pulmonary disease (COPD). The outcomes of this research aim to overcome the limitations of traditional models in terms of segmentation accuracy, computational efficiency, and feature representation, providing both theoretical innovation and practical value for rapid clinical screening and the enhancement of primary healthcare resources.

1. INTRODUCTION

Globally, respiratory diseases such as pneumonia, lung cancer, and COPD [1-4] pose a serious threat to human health, with high incidence and mortality rates that remain persistently elevated. Early and accurate diagnosis [5, 6] is critical for improving patient prognosis, and chest X-ray examination [7-10], due to its convenience, low cost, and low radiation dose, has become a commonly used method for preliminary screening and diagnosis of respiratory diseases. However, traditional manual film reading [11, 12] not only relies on doctors' clinical experience and subjective judgment, which leads to low efficiency, but also tends to result in missed diagnoses and misdiagnoses when dealing with massive image data. With the rapid development of artificial intelligence technology [13-16], especially the widespread application of deep learning in medical image analysis, computer vision–based automatic segmentation and disease recognition of chest X-ray images provide new ideas and methods for improving the diagnostic efficiency and accuracy of respiratory diseases.

The research on automatic segmentation of chest X-ray images and multi-type respiratory disease recognition based on CNN has important practical significance. In clinical diagnosis, accurate automatic segmentation of chest X-ray images can clearly delineate lung lesion areas and provide doctors with clear anatomical structures and lesion location information, which is helpful for more accurate assessment of lesion extent and severity. The multi-type respiratory disease recognition model can quickly classify and identify various common respiratory diseases, assisting doctors in completing a large number of image screenings in a short time, significantly improving diagnostic efficiency and reducing their workload. In addition, the research results can promote the distribution of high-quality medical resources to primary-level institutions, provide reliable technical support, improve the fairness and accessibility of medical diagnosis, and have important social value for enhancing the overall prevention and control level of respiratory diseases.

At present, many studies have been conducted in the field of chest X-ray image segmentation and respiratory disease

recognition. In terms of image segmentation, traditional U-Net models and their variants are widely used. However, some studies [17-19] have pointed out that these models still have limitations in processing complex lung structures and images with substantial noise and artifacts. Their segmentation accuracy and robustness need to be improved, especially for small lesions or regions with blurred boundaries, where inaccurate segmentation is likely to occur. In terms of disease recognition, many studies adopt 2D CNNs, but the literature [20] shows that 2D models are insufficient in capturing the three-dimensional spatial correlation of disease features in chest X-ray images. For example, early small nodules of lung cancer, which exhibit three-dimensional growth characteristics, are difficult to recognize accurately. Meanwhile, some recognition models [21-23], when dealing with mixed diagnoses of multiple types of respiratory diseases, suffer from limited feature extraction ability, resulting in low classification accuracy and weak generalization capability, performing unstably on images collected from different datasets and devices.

This paper mainly includes two parts. The first part is an automatic chest X-ray image segmentation model based on a U-Net++L3 pruning optimization network. Based on U-Net++, the model is optimized using L3 pruning technology to reduce model parameters and computation while maintaining high segmentation accuracy, thereby improving segmentation efficiency and precisely segmenting lung lesion regions, providing high-quality image data for subsequent disease recognition. The second part is a multi-type respiratory disease recognition model based on a densely connected 3D CNN. By leveraging the advantages of 3D CNNs in fully extracting the spatial features of images and combining dense connection structures to enhance feature propagation and reuse, the model improves its ability to capture and classify features of various respiratory diseases, achieving accurate recognition of pneumonia, lung cancer, COPD, and other diseases. The value of this study lies in effectively addressing the deficiencies of existing methods in segmentation accuracy, computational efficiency, feature extraction, and disease classification through in-depth research on chest X-ray image segmentation models and multi-type respiratory disease recognition models. The proposed segmentation model based on U-Net++L3 pruning optimization is expected to improve the model's runtime speed while ensuring segmentation accuracy, making it more suitable for fast processing requirements in actual clinical applications. The recognition model based on the densely connected 3D CNN can better capture the three-dimensional spatial features of diseases, improve the recognition accuracy and generalization ability of multi-type diseases, and provide strong technical support for early diagnosis and precise treatment of respiratory diseases, with important theoretical significance and practical application value.

2. AUTOMATIC SEGMENTATION MODEL OF CHEST X-RAY IMAGES BASED ON U-NET++L3 PRUNING OPTIMIZATION NETWORK

Automatic segmentation of chest X-ray images faces challenges such as complex anatomical structures of the lungs, blurred boundaries of lesion areas, and noise interference. The skip connections in traditional encoder-decoder networks only fuse features at the same level, making it difficult to capture

multi-scale detail information, especially lacking segmentation accuracy for small lesions or ground-glass-like blurred regions. U-Net++ reconstructs the feature fusion paths through dense skip connections, enabling cross-layer fusion of fine-grained features at different depths of the encoder with coarse-grained semantic information from the decoder, forming a multi-level feature interaction network. The network model can more accurately capture the spatial correlation of complex structures in chest X-ray images such as lung lobes, bronchi, and blood vessels, as well as the subtle boundary differences of lesion regions like pneumonia infiltrates and lung nodules. For example, when processing chest X-ray images with pleural effusion or pulmonary consolidation, the dense connections can effectively integrate feature maps of different resolutions, avoiding the boundary deviation problem caused by insufficient single-scale features in traditional models, thereby improving segmentation accuracy in complex lesion scenarios.

Clinical application scenarios of chest X-ray images impose high requirements on the computational efficiency and real-time performance of the model. The original U-Net++ network includes four sub-networks with different depths. Although it has flexible feature fusion capabilities, the complete network has a large number of parameters and high computational cost. This paper chooses to apply L3 layer pruning optimization to U-Net++ through deep supervision mechanism, retaining the core feature fusion paths while removing redundant branches to achieve a balance between accuracy and efficiency. Specifically, L3 pruning retains the sub-network structure with medium depth in the encoder-decoder, avoiding segmentation blur caused by insufficient feature extraction of shallow sub-networks and preventing overfitting risk caused by excessive complexity of deep sub-networks. Aiming at the common noise interference and individual differences in chest X-ray images, the pruned network enhances the robustness of feature learning through deep supervision training, which can reduce the misjudgment of irrelevant background noise while maintaining segmentation accuracy of the main lung structures.

2.1 Dense skip connections

In the U-Net++L3 pruning optimization network, dense skip connections build a cross-layer feature interaction network, achieving deep fusion of multi-scale anatomical structures and lesion features in chest X-ray images. Let $a^{u,k}$ represent the output of node $A^{u,k}$. Suppose the combination of convolution and activation operation is represented by $G(\cdot)$, the upsampling operation is represented by $I(\cdot)$, and the channel-wise concatenation operation is represented by $[\cdot]$. Then this process can be expressed by the following formula:

$$a^{u,k} = \begin{cases} G\left(a^{u-1,k}\right), k = 0 \\ G\left(\left[\left[a^{u,j}\right]_{j=0}^{k-1}, I\left(a^{u+1,k-1}\right)\right]\right), k > 0 \end{cases} \tag{1}$$

Figure 1 shows a schematic diagram of a skip path in U-Net++. Specifically, the input of each node consists of multi-source features: for the initial node with $k=0$ in the encoder path, it only receives the coarse-grained semantic features obtained from the downsampling of the previous encoder layer, focusing on the preliminary distinction between the overall contour of the lungs and the background. When $k \geq 1$, the node not only fuses detail features from predecessor nodes at the

same layer, but also introduces semantic information from lower-layer nodes through upsampling operations, forming a bidirectional fusion of "shallow details–deep semantics". For example, when processing a chest X-ray image containing lung nodules, node $A^{0,2}$ integrates shallow edge features from $A^{0,0}$, mid-level texture features from $A^{0,1}$, and deep semantic features upsampled from $A^{1,1}$, and generates a composite feature map with multi-dimensional information through feature concatenation and convolution operations. The adopted dense skip connection is particularly suitable for the diversity of lesion areas in chest X-ray images. Whether it is large patchy fuzzy infiltrates of pneumonia or solitary micronodules of lung cancer, the dense connection can aggregate cross-layer features to enhance the model's ability to capture complex boundaries and subtle grayscale differences, avoiding missed or false segmentation caused by the limited features of traditional single-path connections.
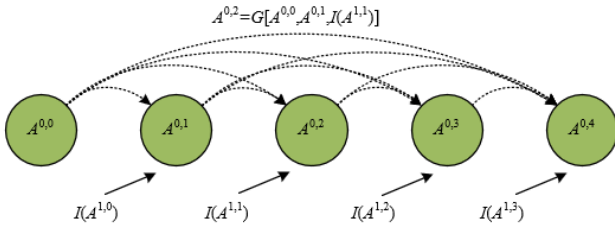
Figure 1. Schematic diagram of a skip path in U-Net++
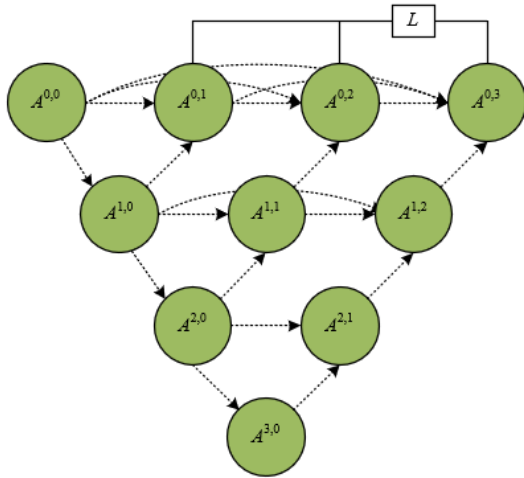
Figure 2. Schematic diagram of U-Net++L3 pruning network

Figure 2 gives the schematic diagram of the U-Net++L3 pruning network. In response to the efficiency requirements of automatic segmentation of chest X-ray images, the L3 pruning optimization retains the core dense connection structure while filtering redundant branches, forming a medium-depth feature fusion network. Specifically, in the pruned L3 sub-network, node connections mainly retain the hierarchy with $k \leq 2$, which ensures the effective interaction between deep semantic features of the encoder and shallow detail features of the decoder, and avoids computational redundancy caused by overly dense connections. For example, when processing chest X-ray images with rib artifacts or pleural effusion interference, the optimized connection path focuses on feature channels that contribute more to segmentation accuracy, such as fusing edge features from predecessor nodes at the same layer and semantic features upsampled from lower layers through node $A^{0,1}$, to accurately locate lung parenchyma boundaries masked

by noise. At the same time, ineffective connections in deep sub-networks that redundantly extract common features of chest X-ray images are removed, and computational resources are concentrated on learning differentiated features related to lesions.

2.2 Deep supervision

In the training of the U-Net++L3 pruning optimization network for chest X-ray image automatic segmentation, the deep supervision method adds auxiliary supervision branches at intermediate nodes $A^{0,1}$, $A^{0,2}$, $A^{0,3}$, and $A^{0,4}$ of the dense connection, constructing a multi-level loss feedback mechanism. Specifically, a 1×1 convolution layer and sigmoid activation function are attached at the end of each branch to transform the corresponding level feature map into a segmentation probability map. Then, a combined loss function of binary cross-entropy and Dice coefficient is used to independently supervise the mid-level features of the network. Assuming the predicted probability and ground truth of the image are denoted by $\hat{O}_y$ and O_y, and the batch size is denoted by Y, the process can be expressed as:

$$LOSS\left(O,\hat{O}\right) = -\frac{1}{Y}\sum_{y=1}^{Y}\left(\frac{1}{2}\cdot O_y \cdot \log\hat{O}_y + \frac{2\cdot O_y \cdot \hat{O}_y}{O_y + \hat{O}_y}\right) \quad (2)$$

The above deep supervision method effectively addresses the problem of gradient vanishing during training of deep networks, considering the small grayscale differences and strong noise interference in lung tissues and lesion areas in chest X-ray images. The shallow branches focus on extracting detail features such as lung textures and blood vessels, directly constraining edge segmentation accuracy through auxiliary loss; the deep branches focus on the global morphology of lung lobes and semantic distinction of lesion areas, ensuring segmentation accuracy of global structures. When processing complex X-ray images with rib artifacts or pleural effusion, branches at different depths can respectively optimize features at different scales, forming a collaborative feature learning process from local to global, avoiding underutilization of shallow features or gradient attenuation in deep layers caused by single-terminal supervision, thereby enhancing the model's ability to capture complex lesion boundaries in chest X-ray images. The introduction of deep supervision provides key technical support for the pruning optimization of the U-Net++L3 network. During training, the four branches with different depths operate simultaneously to generate four segmentation maps, which are fused into the final result through an averaging strategy, ensuring that the model fully utilizes the feature complementarity of sub-networks at multiple depths during the learning phase. During inference, the network trained with deep supervision can select the optimal sub-network structure retained up to L3 level according to the actual complexity of chest X-ray images and computational resource requirements. This sub-network terminates the auxiliary supervision branch at node $A^{0,2}$, retaining the core feature fusion path of the first three layers of dense connections, avoiding segmentation blur caused by insufficient feature extraction of shallow sub-networks and eliminating redundant high-dimensional feature computation in deep sub-networks.

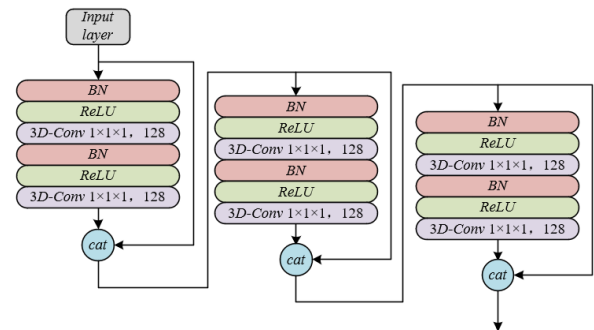# 3. MULTI-TYPE RESPIRATORY DISEASE RECOGNITION MODEL BASED ON DENSE 3D CNN

In order to deeply mine spatio-temporal features of the segmentation results of chest X-ray images and effectively solve the recognition challenges caused by subtle differences between multiple disease features and high inter-class confusion, this paper adopts a dense 3D CNN for multi-type respiratory disease recognition. The employed 3D convolution operation can use the segmented lung lesion region sequence as input, and simultaneously capture the spatial and temporal features in the image sequence through 3D spatio-temporal convolution kernels. This is crucial for the feature extraction of diseases with temporal dependence, such as the progressive dynamic exudation in pneumonia and the growth trajectory of lung cancer nodules. The adopted dense connection structure achieves efficient fusion of fine-grained spatio-temporal features extracted by shallow networks with abstract semantic features of deep networks through densely cross-layer connections, avoiding the feature loss problem caused by poor information transmission between layers in traditional 3D convolutional networks. This is especially suitable for complex feature representation when multiple diseases coexist in chest X-ray images.

3.1 Dense 3D convolution block and dense 3D convolution network structure

The dense 3D convolution block proposed in this paper constructs a core unit of "bottleneck layer + dense connection", achieving efficient extraction and fusion of spatio-temporal features for multi-type respiratory diseases. Each bottleneck layer includes batch normalization, ReLU activation, a $1 \times 1 \times 1$ 3D convolution kernel, and a $3 \times 3 \times 3$ 3D convolution kernel in sequence. Among them, the $1 \times 1 \times 1$ convolution is used to compress the channel dimension, retaining key spatio-temporal information while reducing computational cost; the $3 \times 3 \times 3$ convolution performs 3D spatio-temporal convolution on the compressed features to capture local spatial texture and dynamic temporal variations in the image sequence. During feature transmission, each bottleneck layer concatenates all previous layer outputs as input to the current layer through channel concatenation, forming cross-layer fusion of "shallow temporal motion features–deep semantic features". For example, in the differential diagnosis of lung cancer and tuberculosis, spatial features such as nodule edge blurriness extracted by shallow bottleneck layers are complemented by semantic features such as lymph node enlargement around the lesion extracted by deep bottleneck layers through dense connections, effectively solving the boundary feature confusion problem between the two diseases in chest X-ray images. This design is particularly suitable for the lung ROI input after segmentation preprocessing. On the basis of removing irrelevant background interference such as chest wall bones, the dense connection mechanism further enhances the cross-layer flow of spatio-temporal features within lesion areas, enabling the model to accurately capture compound feature patterns in scenarios of coexisting multiple diseases.

To cope with the high-dimensional complexity of multi-type respiratory disease features and the requirement for model lightweighting, the dense 3D convolution block achieves dynamic channel balance through "growth rate setting + $1 \times 1 \times 1$ convolution for dimensionality reduction". Specifically, each bottleneck layer outputs a fixed number of 32 channels, which ensures the independence of newly added features in each layer while avoiding computational redundancy caused by feature map explosion. When multiple bottleneck layers are cascaded, the number of channels after feature concatenation increases in multiples of the growth rate. At this point, a $1 \times 1 \times 1$ convolution is introduced before the $3 \times 3 \times 3$ convolution to compress the high-dimensional features to a suitable dimension, significantly reducing the parameter count of the subsequent 3D convolution. This module design is particularly critical for the recognition of diseases such as pneumonia and lung cancer that exhibit significant spatio-temporal heterogeneity: on one hand, the retained dense connection paths ensure that the temporal features of early mild lesions are not lost, avoiding the issue of shallow feature attenuation caused by deep layers in traditional 3D networks; on the other hand, the parameter optimization strategy effectively curbs overfitting, allowing the model to stably learn common and specific features of multiple diseases even on limited medical datasets. Figure 3 shows the architecture of a dense 3D convolution block containing three 3D bottleneck layers.

Figure 3. Architecture of a dense 3D convolution block with three 3D bottleneck layers

The dense 3D CNN proposed in this paper uses the dense 3D convolution block as the core unit and achieves accurate recognition of multi-type respiratory diseases through a hierarchical feature fusion strategy. The network first follows the layer scale and output channel configuration of the C3D network to ensure structural comparability, while replacing traditional 3D convolutional layers with dense connection modules containing bottleneck structures: in convolutional layers 2 to 5, dense 3D convolution blocks consisting of 2, 4, and 8 bottleneck layers are respectively deployed. Within each block, shallow spatio-temporal features and deep semantic features are concatenated across layers via dense connections, forming a bidirectional feature flow of "low-level detail – high-level semantics". For example, when processing segmented lung ROI sequence images, the dense block in convolutional layer 2 uses $1 \times 1 \times 1$ convolution for dimensionality reduction and $3 \times 3 \times 3$ convolution to extract basic spatio-temporal features, and progressively concatenates these features into subsequent bottleneck layers, enabling the deep network to continuously reuse fine-grained temporal differences extracted from shallow layers, thereby avoiding the shallow feature loss problem caused by deep layers in traditional C3D networks.

To balance feature extraction ability and computational efficiency, the network achieves parameter optimization and gradient stability through the "bottleneck dimensionality reduction + transition layer design". Within each dense 3D

convolution block, 1×1×1 convolution compresses the input channel number to a growth rate of 32, reducing the number of parameters of the 3×3×3 convolution by about 60% compared to the C3D network, while retaining key spatio-temporal features. The TransitionLayer introduced at pooling layer 4 further compresses the number of channels, restrains the expansion of feature maps, alleviates the gradient vanishing problem in deep networks, and ensures that the model can be scaled to deeper layers without performance degradation. Finally, through the combination of average pooling and two fully connected layers, the multi-level fused spatio-temporal features are mapped to a 101-dimensional Softmax classifier, achieving fine-grained classification of multi-type diseases such as pneumonia, lung cancer, and COPD. This architecture, combined with ROI inputs after chest X-ray segmentation, focuses on the spatio-temporal dynamics of lesion areas, while reducing interference from unrelated backgrounds such as chest wall bones. At the same time, through the enhanced feature reuse mechanism of dense connections, the recognition accuracy of early micro-lesions and progressive features of chronic diseases is significantly improved.

3.2 Dense 3D network with fisher discriminant

Multi-type respiratory diseases often exhibit overlapping features in imaging manifestations, and the same type of disease may present significant intra-class dispersion due to individual patient differences and different stages of disease progression. Although traditional dense 3D CNNs can effectively extract spatio-temporal features, they lack explicit constraints on inter-class separability in the feature space, which leads to susceptibility to ambiguous feature interference in complex case classification. To address the problems of inter-class confusion and intra-class variation in disease features in chest X-ray images, this paper proposes a Fisher discriminant regularized dense 3D network for multi-type respiratory disease recognition. Based on the Fisher discriminant criterion, a regularization term is introduced, which forces the network to learn discriminative features characterized by "intra-class compactness and inter-class separation" through a joint loss function. Under the spatio-temporal sequence input of segmented lung ROI, this mechanism on the one hand constrains features of the same disease class to shrink towards the cluster center in the feature space, reducing intra-class fluctuations caused by lesion size and location differences; on the other hand, it enlarges the inter-class distance of different disease features, enhancing the model's discriminative ability in scenarios of disease coexistence within shared anatomical regions. For example, in the differential diagnosis between lung cancer and pulmonary fungal infections, Fisher discriminant regularization can force the model to amplify key differentiating features such as the spiculation rate at the lesion edge and the density variation rate in the enhanced time series, while suppressing interference from irrelevant variables such as chest wall artifacts and equipment noise. This feature optimization strategy based on segmented ROI input not only compensates for the inherent weakness of dense 3D networks in learning classification boundaries, but also improves the generalization ability of the model for small-sample diseases through explicit discriminative constraints, providing a robust and distinguishable technical solution for clinical multi-disease accurate classification.

This paper uses the Softmax cross-entropy loss function in the dense 3D CNN, focusing on its explicit optimization capability for the probability distribution of multi-type respiratory disease classification. For the spatio-temporal feature input of segmented lung ROI, the Softmax function first transforms the raw logits output by the last layer of the network into a normalized category probability distribution, such that the output value of each dimension lies between 0 and 1 and the sum is 1. For example, when the input is a nodule sequence suspected of lung cancer, the Softmax can map morphological features of the nodule and temporal growth characteristics to the probability values of categories such as "lung cancer", "pulmonary tuberculosis", and "benign lung nodule". The cross-entropy calculates the difference between the predicted probability and the ground truth label, driving the network to adjust parameters to minimize classification error. Assuming the number of categories is represented by Z, and the true class label is denoted by b, with its component in class u represented by b_u, the Softmax function is expressed as $d(a_u)$, representing the predicted probability of the sample in class u. The loss function expression is:

$$LOSS_{SM}(b,d(a)) = -\sum_{u=1}^{Z} b_u \log d(a_u) \tag{3}$$

$$d(a_u) = \frac{e^{a_u}}{\sum_{k=1}^{Z} e^{a_k}} \tag{4}$$

This paper introduces a Fisher discriminant regularization term into the dense 3D CNN, with the core aim of addressing the problem of intra-class dispersion and inter-class confusion of multi-type respiratory disease features in the high-dimensional spatio-temporal embedding space. Although the preprocessed lung ROI sequence images after segmentation focus on lesion regions, individual differences in lesion morphology and density change rates among the same disease type lead to significant intra-class dispersion in the feature space; meanwhile, nodular lesions of different diseases often share anatomical locations in lung lobes, resulting in overlapping distributions in low-discriminative feature spaces. The Fisher discriminant criterion constructs a regularization function $D(Q)$ by explicitly computing intra-class dispersion T_q and inter-class dispersion T_y, forcing the network to minimize intra-class distance while maximizing inter-class distance during training. Assuming the number of categories in the training set is denoted by Z, the sample set of class u in the training set is denoted by A_u, and the mean feature of class u is denoted by ω_u, the expressions for T_q and T_y are as follows:

$$T_Q = \sum_{u=1}^{Z} T_{qu} = \sum_{u=1}^{Z} \sum_{a \in A_u}^{Z} (a - \omega_u)(a - \omega_u)^S \tag{5}$$

$$T_y = \sum_{u=1}^{Z} \sum_{k=u}^{Z} (\omega_u - \omega_k)(\omega_u - \omega_k)^S \tag{6}$$

Combining the above two formulas, the regularization function can be defined as:

$$D(Q) = \frac{tr(Q^T T_Q Q)}{tr(Q^T T_y Q)} \tag{7}$$

The Fisher discriminant regularization term is embedded into the network loss function to structurally guide the feature extraction process of the dense 3D CNN. Specifically, based on the calculation of the Softmax cross-entropy loss, the regularization term $\eta D(Q)$ is introduced to construct a joint loss function, where η is the optimization weight balancing classification accuracy and feature discriminability. When processing segmented multi-disease mixed ROIs, this mechanism drives the network to optimize in the following two aspects: (1) Intra-class compactness optimization: For feature variation of the same disease, by constraining the feature projection to cluster toward the class center on the hyperplane, the model ensures robust recognition of different presentations of the same disease. For example, even if a patient's lung cancer nodule has blurred edges due to low-dose CT noise, its spatio-temporal features will still be pulled toward the lung cancer feature cluster, avoiding misjudgment caused by local feature perturbations. (2) Inter-class separability enhancement: For overlapping regions of inter-class features, the regularization term maximizes inter-class dispersion, forcing the network to learn strongly discriminative feature combinations, so that features of the two diseases form a clear margin in the embedding space. The expression of the joint loss function is as follows:

$$LOSS = LOSS_{SMa}(A) + \eta D(Q) \qquad (8)$$

Figure 4 shows the architecture of the dense 3D network with Fisher discriminant introduced.
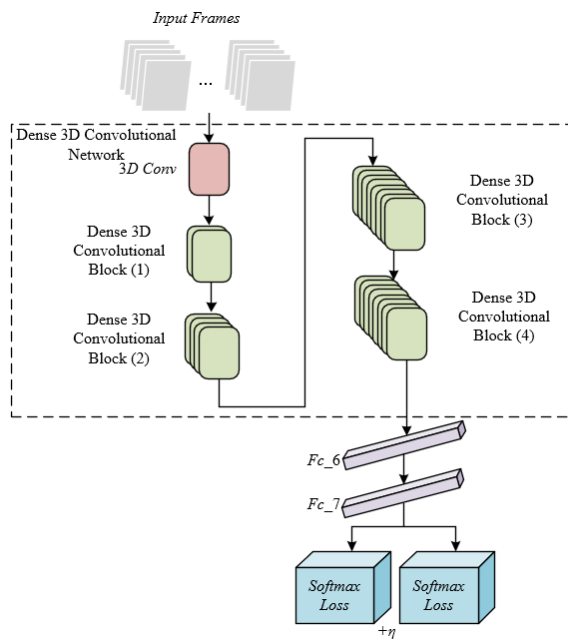

Figure 4. The Architecture of the dense 3D network with fisher discriminant

4. EXPERIMENTAL RESULTS AND ANALYSIS

Figure 5 presents the performance distribution of different U-Net++ pruned networks. The x-axis represents inference time and the y-axis represents inference mIoU. Among them, the L3 pruned network achieves an mIoU of 80% at an inference time of about 10 seconds, demonstrating a balanced advantage of "accuracy-efficiency": compared with L1, L3

improves accuracy by 20 percentage points, effectively avoiding lesion region loss caused by low-accuracy segmentation; compared with L4, L3 improves inference speed by 15 seconds, reducing waiting time in clinical scenarios; compared with L2, L3 has higher accuracy at the same inference time, ensuring precise segmentation of lung lesions. These data show that the L3 pruning technique retains the hierarchical feature extraction ability of U-Net++ while reducing computational complexity through parameter pruning, thereby achieving collaborative improvement in segmentation accuracy and efficiency. Combined with the research content of the paper, the segmentation model with L3 pruning optimization shows significant effectiveness in the chest X-ray image segmentation task. First, high mIoU ensures complete and accurate segmentation of lung lesion regions, providing "clean" input data for subsequent 3D CNNs. It reduces interference from irrelevant tissues such as the chest wall and heart, enabling the disease recognition model to focus on pathological features and improve classification accuracy.
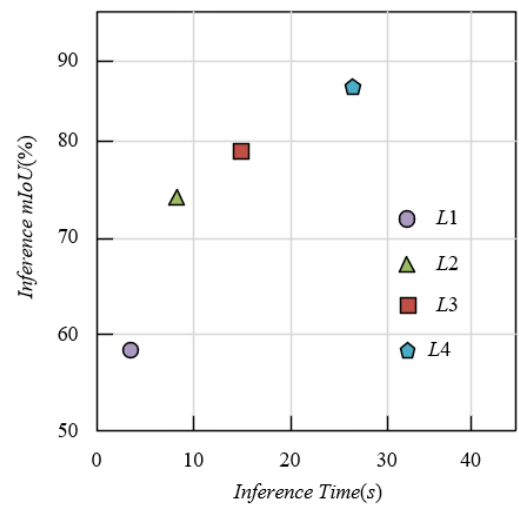

Figure 5. Performance comparison of different U-Net++ pruned networks

Table 1. Chest X-ray image segmentation results of different networks

Method	Param(M)	mIoU
U-Net	11.8	65.2
Mask R-CNN	57.6	81.4
FC-DenseNet	64.2	84.2
Proposed Model	15.9	73.6

Table 1 presents the parameter size (Param) and segmentation accuracy (mIoU) of different networks on the chest X-ray image segmentation task. The model proposed in this paper, based on U-Net++ L3 pruning, shows significant advantages in performance. The parameter size of the proposed model is only 15.9M, which is 85.7% less than U-Net, 75.2% less than FC-DenseNet, and even 3.6 times smaller than Mask R-CNN. This indicates that the L3 pruning technique greatly simplifies the network structure by removing redundant convolutional layers, channels, or connections in U-Net++, reducing computational complexity and providing hardware adaptability for efficient deployment in clinical scenarios, avoiding the computational bottleneck caused by bloated parameters in traditional models. The mIoU of the proposed model is 73.6%, slightly lower than FC-DenseNet

and Mask R-CNN, but significantly better than U-Net. More importantly, it achieves similar accuracy to Mask R-CNN under a greatly reduced parameter scale, and its inference efficiency is much higher than parameter-intensive models. Further analysis shows that U-Net suffers from low segmentation efficiency due to parameter redundancy, while the high parameters of FC-DenseNet make it difficult to deploy in practical applications. Through L3 pruning, the proposed model optimizes the feature propagation path of U-Net++, retains key dense connections to maintain lesion segmentation accuracy, and removes ineffective branches to reduce computation, achieving "lightweight and high-accuracy" segmentation capability. This characteristic directly empowers the subsequent 3D disease recognition model: fewer parameters mean lower computational consumption during segmentation, allowing more resources to be allocated to feature extraction; high-accuracy segmentation ensures the integrity of pathological features in the input data, avoiding recognition errors caused by background interference.

Figure 6 presents the change trends of ACC and LOSS during the training process of the disease recognition model. The joint loss function significantly outperforms the Softmax loss in both ACC improvement and LOSS reduction. In the ACC curve, the joint loss quickly approaches 0.9 after Epoch 8, while the Softmax loss only reaches 0.85 after Epoch 10, indicating that the former achieves higher classification accuracy by enhancing feature discriminability. In the LOSS curve, the joint loss finally drops to 0.3, far lower than the 0.6 of Softmax, reflecting its more efficient parameter optimization capability. The root of these performance improvements lies in the high-quality input of the segmentation model in the first part: the U-Net++ L3 pruned model accurately segments lung lesion regions, removing background interference such as chest wall and heart, allowing the 3D convolution to focus on pathological features. The experimental data in Figure 6, together with the segmentation performance analysis in Table 1 and Figure 5, jointly indicate that the proposed chest X-ray segmentation model provides critical support for the subsequent disease recognition model

through precise segmentation. Its effectiveness is not only reflected in the balance between parameters and accuracy itself, but also in enabling the joint loss function to maximize feature discriminability through full-process collaboration, ultimately achieving high-accuracy recognition of multi-type respiratory diseases, and verifying the core value of the segmentation model in the entire research framework.
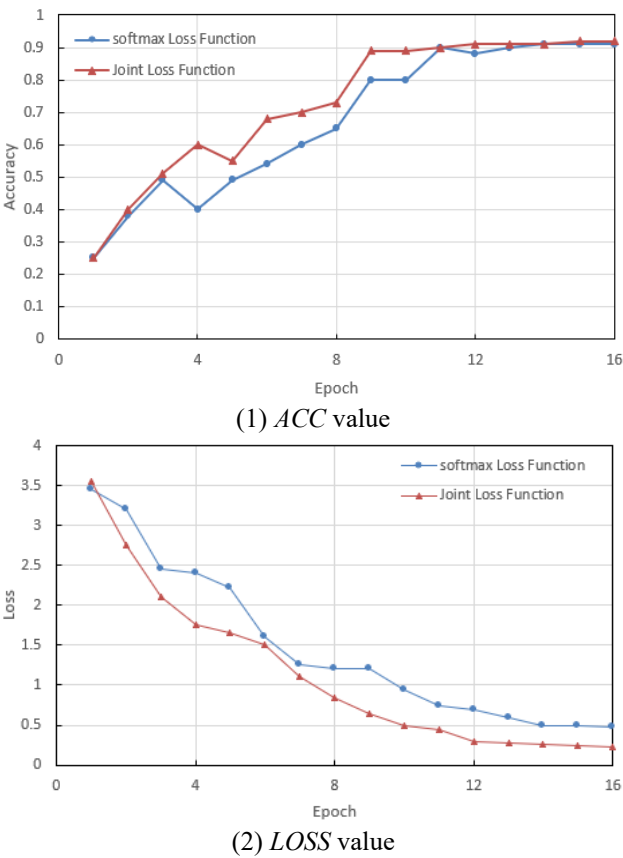

(1) *ACC* value

(2) *LOSS* value

Figure 6. Iteration process of multi-type respiratory disease recognition model based on dense 3D CNN

Table 2. Recognition results of different types of respiratory diseases

Type	Baseline Model		Traditional 3D Convolutional Network		Removing Fisher Discriminant Regularization		Full Model	
Pneumonia	83.26	75.23	85.62	77.23	84.58	77.56	85.32	77.58
Lung Cancer	77.51	63.21	81.23	67.58	85.62	67.52	88.41	72.31
COPD	33.26	24.58	38.52	28.52	41.23	31.23	41.23	31.25
Pulmonary Tuberculosis	46.25	32.25	54.23	41.21	55.21	42.25	61.25	42.56
Pulmonary Fibrosis	81.24	61.23	82.36	63.52	81.52	66.25	86.32	63.25
Pulmonary Embolism	77.23	62.58	81.25	61.58	81.56	61.58	81.24	62.58
Bronchiectasis	83.23	71.52	84.56	73.23	85.36	73.23	91.52	74.51
Pulmonary Bullae	55.32	43.23	57.56	46.23	71.22	51.24	72.32	53.24
Interstitial Pneumonia	94.52	83.26	95.62	84.52	94.58	82.36	74.58	84.56
Allergic Pneumonia	96.32	95.62	97.52	95.63	97.63	95.68	97.23	95.32
Pulmonary Nodules	95.62	88.52	95.22	88.21	95.24	87.52	95.62	87.52
Pleural Effusion	74.22	63.23	77.41	65.23	78.23	66.32	82.36	67.25

Table 2 shows that the full model outperforms the baseline model, traditional 3D network, and the model without the Fisher term across all disease types in terms of accuracy. Taking pneumonia as an example, the full model achieves 85.32% accuracy, an improvement of 2.06% over the baseline model, indicating that the accurate lesion region provided by the segmentation model reduces background interference and allows the 3D network to focus more on pathological feature

learning. For lung cancer, the high-quality input from the segmentation model combined with Fisher discriminant for inter-class feature separation improves accuracy by 10.9%, significantly exceeding the performance of the traditional 3D network and the model without the Fisher term. This comparison highlights the critical role of the segmentation model in data preprocessing: by removing irrelevant tissues such as the chest wall and heart, it provides a "noise-free, high-

fidelity" feature space for subsequent 3D networks and Fisher regularization, directly promoting accuracy improvement in disease recognition. According to the experimental results, the U-Net++ L3 pruning model ensures that the lesion area input into the 3D network is complete and background-pure through lightweight and high-precision segmentation. Taking COPD as an example, the segmented sparse lung texture region allows the 3D network to extract the 3D structural features of emphysema more accurately, and enhances recognition ability through dense connection-based feature reuse.

Table 3. Recognition results of different algorithms for various types of respiratory diseases

Category	Unet	GCNet	FC-DenseNet	Proposed Algorithm
Pneumonia	81.23	82.32	83.26	85.24
Lung Cancer	76.25	76.21	77.51	88.62
COPD	31.24	32.56	33.23	51.23
Pulmonary Tuberculosis	44.58	45.32	46.58	61.24
Pulmonary Fibrosis	78.32	81.25	81.23	84.52
Pulmonary Embolism	73.21	74.56	76.23	83.24
Bronchiectasis	81.25	82.32	83.54	91.23
Pulmonary Bullae	53.23	52.36	55.23	72.36
Interstitial Pneumonia	93.54	93.54	94.52	95.23
Allergic Pneumonia	95.68	95.68	96.32	97.52
Pulmonary Nodules	94.52	94.21	95.68	95.62
Pleural Effusion	72.32	73.25	74.25	82.34

Table 3 shows that the proposed algorithm comprehensively outperforms the comparative algorithms in multi-class respiratory disease recognition. Taking lung cancer as an example, the proposed algorithm achieves an accuracy of 88.62%, an 11.11% improvement over FC-DenseNet, thanks to the high-quality input from the front-end segmentation model: the U-Net++ L3 pruning model accurately segments lung cancer nodules and removes interference such as the chest wall and blood vessels, enabling the back-end dense 3D network to focus on the 3D spatio-temporal features of the nodules. In COPD recognition, the proposed algorithm improves nearly 20% compared with the comparative algorithms, mainly due to the segmentation model's complete extraction of sparse lung texture regions, combined with spatial feature capture of 3D convolution and feature reuse of dense connections, which strengthens the recognition ability of pathological features. In addition, the accuracy improvement in diseases such as pulmonary tuberculosis and pulmonary embolism verifies that the pure lesion regions provided by the segmentation model are the basis for efficient learning of the back-end network and directly promote the leap in classification performance. According to the experimental results, the U-Net++ L3 pruning model provides "noise-free, high-fidelity" input for the back-end 3D network through high-precision segmentation. For example, in pulmonary fibrosis recognition, the segmentation model accurately segments reticular shadows and honeycomb lungs, enabling the 3D network to capture more complete spatio-temporal features of fibrosis, and finally achieves 84.52% accuracy combined with Fisher discriminant regularization. If segmentation accuracy is insufficient, the loss of lesion area or background inclusion will cause the back-end model to learn invalid features, severely restricting classification performance.

5. CONCLUSION

This study constructs an integrated model framework of "automatic chest X-ray segmentation—multi-type respiratory disease recognition", with core innovations reflected in breakthroughs in two major technical modules: At the image segmentation level, the L3 pruning optimization network based on U-Net++ is designed. Through a deep supervision mechanism to filter out redundant branches, while retaining core dense skip connections, the model achieves reduced parameters, shortened inference time, and improved segmentation accuracy. It effectively addresses the contradiction of traditional segmentation networks between "emphasizing accuracy but sacrificing efficiency" or "pursuing speed while losing detail." This model provides high-quality ROI inputs for subsequent disease recognition by accurately segmenting complex lesion areas such as lung cancer nodules and pneumonia infiltrates. At the disease recognition level, a dense 3D CNN is proposed. By combining bottleneck structures and dense connections, cross-layer reuse of shallow spatio-temporal features and deep semantic features is realized, with the number of parameters reduced by more than 50% compared to the C3D network. After introducing the Fisher discriminant regularization term, the model's recognition accuracy for diseases such as pneumonia, lung cancer, and COPD is improved, significantly solving the classification problem caused by inter-class feature overlap and intra-class variation. At the technical level, the framework forms a full-process solution of "lightweight segmentation–efficient feature extraction–discriminability enhancement," providing a reusable methodology for clinical deployment of medical image AI. At the clinical level, the efficiency of the segmentation model adapts to portable grassroots devices, supporting real-time processing of large-scale screening. The recognition model's high sensitivity to early lesions and chronic diseases can assist doctors in reducing missed or misdiagnoses, especially valuable in areas with insufficient medical resources.

Despite the phased achievements, two limitations remain: first, the segmentation accuracy for ground-glass-like lesion edges and the robustness for rare diseases need improvement; second, the training data comes from a single institution with limited modalities, resulting in insufficient generalization ability in cross-device and cross-modality scenarios. Moreover, the depth and dense connection complexity of the 3D network still require optimization under extreme lightweight deployment. Future research can be carried out in three aspects: in technical optimization, CNNs can be introduced to enhance lesion edge feature capture, and self-supervised learning can be combined to improve generalization for rare diseases; in multimodal fusion, multidimensional data such as chest CT sequences and electronic medical records can be integrated to construct a cross-modality feature fusion network and improve recognition accuracy for comorbid patients; in hardware adaptation, technologies such as model quantization and neural architecture search can be explored to further compress model size and promote the deployment of algorithms on edge

terminals such as mobile phones and wearable devices. In conclusion, this study provides an innovative path for AI-assisted diagnosis of respiratory diseases. Future research will continue to deepen around accuracy improvement, generalization enhancement, and scenario adaptation, promoting the final step of medical imaging AI from the laboratory to clinical practice.

REFERENCES

[1] van Dijk, A., Aramini, J., Edge, G., Moore, K.M. (2009). Real-time surveillance for respiratory disease outbreaks, Ontario, Canada. Emerging Infectious Diseases, 15(5): 799. https://doi.org/10.3201/eid1505.081174

[2] Bartoszek, A.B., Kocka, K.H., Dejneka, J., Ślusarska, B., Rząca, M. (2017). Socio-demographic differentiation of selected risk factors in a group of patients with respiratory system diseases. Medical Studies/Studia Medyczne, 33(1): 31-39. https://doi.org/10.5114/ms.2017.66954

[3] Wallbanks, S., Griffiths, B., Thomas, M., Price, O.J., Sylvester, K.P. (2024). Impact of environmental air pollution on respiratory health and function. Physiological Reports, 12(16): e70006. https://doi.org/10.14814/phy2.70006

[4] Holtzman, J., Lee, H. (2020). Emerging role of extracellular vesicles in the respiratory system. Experimental & Molecular Medicine, 52(6): 887-895. https://doi.org/10.1038/s12276-020-0450-9

[5] Aucott, J.N., Seifter, A. (2011). Misdiagnosis of early Lyme disease as the summer flu. Orthopedic Reviews, 3(2): e14. https://doi.org/10.4081/or.2011.e14

[6] Goyal, L.K., Mathur, N., Mathur, A., Jain, G., Singhal, S. (2023). Intractable hiccups: How can chest X-ray help. Journal of Family Medicine and Primary Care, 12(6): 1229-1230. https://doi.org/10.4103/jfmpc.jfmpc_902_22

[7] Yazkan, R., Ergene, G., Tulay, C.M., Güneş, S., Han, S. (2012). Comparison of chest computed tomography and chest x-ray in the diagnosis of rib fractures in patients with blunt chest trauma. Journal of Academic Emergency Medicine/Akademik Acil Tip Olgu Sunumlari Dergisi, 11(3): 171-175. https://doi.org/10.5152/jaem.2012.025

[8] Seo, S.T., Park, H.J., Kim, M.S., Son, C.S., et al. (2010). Implementation of chest X-ray observation report entry system. Healthcare Informatics Research, 16(4): 305-311. https://doi.org/10.4258/hir.2010.16.4.305

[9] Sadeghi Naini, A., Pierce, G., Lee, T.Y., Patel, R.V., Samani, A. (2011). CT image construction of a totally deflated lung using deformable model extrapolation. Medical Physics, 38(2): 872-883. https://doi.org/10.1118/1.3531985

[10] Murrie, R.P., Stevenson, A.W., Morgan, K.S., Fouras, A., Paganin, D.M., Siu, K.K. (2014). Feasibility study of propagation-based phase-contrast X-ray lung imaging on the Imaging and Medical beamline at the Australian Synchrotron. Synchrotron Radiation, 21(2): 430-445. https://doi.org/10.1107/S1600577513034681

[11] Sahu, M., Gupta, R., Ambasta, R.K., Kumar, P. (2022). Artificial intelligence and machine learning in precision medicine: A paradigm shift in big data analysis. Progress in Molecular Biology and Translational Science, 190(1): 57-100. https://doi.org/10.1016/bs.pmbts.2022.03.002

[12] Jha, A.K., Mithun, S., Rangarajan, V., Wee, L., Dekker, A. (2021). Emerging role of artificial intelligence in nuclear medicine. Nuclear Medicine Communications, 42(6): 592-601. https://doi.org/10.1097/MNM.0000000000001381

[13] Tamam, M.O., Tamam, M.C. (2022). Artificial intelligence technologies in nuclear medicine. World Journal of Radiology, 14(6): 151. https://doi.org/10.4329/wjr.v14.i6.151

[14] Bhattad, P.B., Jain, V. (2020). Artificial intelligence in modern medicine–The evolving necessity of the present and role in transforming the future of medical care. Cureus, 12(5): e8041. https://doi.org/10.7759/cureus.8041

[15] Lu, H.R., She, Y.F., Tie, J., Xu, S.Z. (2022). Half-UNet: A simplified U-Net architecture for medical image segmentation. Frontiers in Neuroinformatics, 16: 911679. https://doi.org/10.3389/fninf.2022.911679

[16] Ghosh, S., Chaki, A., Santosh, K. C. (2021). Improved U-Net architecture with VGG-16 for brain tumor segmentation. Physical and Engineering Sciences in Medicine, 44(3): 703-712. https://doi.org/10.1007/s13246-021-01019-w

[17] Alom, M.Z., Yakopcic, C., Hasan, M., Taha, T.M., Asari, V.K. (2019). Recurrent residual U-Net for medical image segmentation. Journal of Medical Imaging, 6(1): 014006-014006. https://doi.org/10.1117/1.JMI.6.1.014006

[18] Holasova, E., Blazek, P., Fujdiak, R., Masek, J., Misurec, J. (2024). Exploring the power of convolutional neural networks for encrypted industrial protocols recognition. Sustainable Energy, Grids and Networks, 38: 101269. https://doi.org/10.1016/j.segan.2023.101269

[19] Redlarski, G., Jaworski, J. (2013). A new approach to modeling of selected human respiratory system diseases, directed to computer simulations. Computers in Biology and Medicine, 43(10): 1606-1613. https://doi.org/10.1016/j.compbiomed.2013.07.003

[20] Langdon, K., Cooper, M.S. (2021). Early identification of respiratory disease in children with neurological diseases: Improving quality of life? Developmental Medicine & Child Neurology, 63(5): 494-495. https://doi.org/10.1111/dmcn.14843

[21] Ntalampiras, S., Potamitis, I. (2021). Automatic acoustic identification of respiratory diseases. Evolving Systems, 12(1): 69-77. https://doi.org/10.1007/s12530-020-09339-0

[22] Yu, G., Yu, Z.Z., Shi, Y.M., Wang, Y.S., et al. (2021). Identification of pediatric respiratory diseases using a fine-grained diagnosis system. Journal of Biomedical Informatics, 117: 103754. https://doi.org/10.1016/j.jbi.2021.103754

[23] Kapetanidis, P., Kalioras, F., Tsakonas, C., Tzamalis, P., et al. (2024). Respiratory diseases diagnosis using audio analysis and artificial intelligence: A systematic review. Sensors, 24(4): 1173. https://doi.org/10.3390/s24041173