(Received 2 July 2025, Accepted 28 July 2025, Scheduled 30 July 2025)

Classification of Severe Bacterial Pneumonia Based on CT Images and Deep Learning

Ke Cui[1, 3, #], Dawei Gong[1, 4, #], Xiaobo Chen[2, 5], Youzu Xu[6], Haiyan Li[6], Yefei Zhu[6], Julian Evans[2], Xin Gong[7], Zhenzhan Shi[5], Yinghe Xu[1, 3], and Sailing He[2, 1, *]

[1] *Zhejiang Engineering Research Center for Intelligent Medical Imaging, Sensing and Non-invasive Rapid Testing Taizhou Hospital of Zhejiang Province, Wenzhou Medical University, Linhai 317000, China*
[2] *Centre for Optical and Electromagnetic Research, National Engineering Research Center for Optical Instruments College of Optical Science and Engineering, Zhejiang University, Hangzhou 310058, China*
[3] *Department of Critical Care Medicine, Taizhou Hospital of Zhejiang Province Affiliated to Wenzhou Medical University Taizhou 318000, China*
[4] *MedEngInfo Collaborative Research Center, Taizhou Hospital of Zhejiang Province Affiliated to Wenzhou Medical University Taizhou 318000, China*
[5] *School of Opto-Electronic Engineering, Changchun University of Science and Technology, Changchun 130022, China*
[6] *Department of Respiratory and Critical Care Medicine Taizhou Hospital of Zhejiang Province Affiliated to Wenzhou Medical University, China*
[7] *Health Management Center, Taizhou Hospital of Zhejiang Province Affiliated to Wenzhou Medical University Taizhou 318000, China*
[#] *These authors contributed equally to this work.*

ABSTRACT: Severe bacterial pneumonia is a serious respiratory disease caused by bacteria, which is mainly transmitted through the respiratory tract. To achieve early recognition of severe pneumonia patients through images, this study collected the CT images of 180 patients diagnosed with bacterial infection in the lungs on the day of emergency admission to a large regional medical center (a Top-Tier (Grade 3 A) hospital). After classification by two deputy chief physicians of the respiratory department, 93 cases of severe bacterial infection were obtained and the rest 87 cases were identified as mild bacterial infection. The CT sequences were then preprocessed and annotated to obtain 599 images with annotated lung infection areas Together with 107 normal (non-infected) images, these bacterial infection images were randomly divided into a training set of 447 and a test set of 259. In the experiment, four deep learning methods, namely, FCN, PSPNet, deeplabv3, and deeplabv3plus, were used for training and three-class classification (severe bacterial infection, mild bacterial infection, and normal). Deeplabv3plus showed the best performance, with an overall accuracy of 96.91% (including a sensitivity of 95.25%, a specificity of 97.24%, an accuracy of 86.96%, a recall rate of 95.24%, and an F1 score of 90.91%) for severe bacterial infection. Using deep learning technology to diagnose severe pneumonia as early as possible can produce valuable treatment time for patients, thereby significantly reducing mortality and complication rates.

1. INTRODUCTION

Severe pneumonia refers to major respiratory dysfunction or systemic inflammatory response in patients with pneumonia. Patients usually present with symptoms such as high fever, cough, chest pain, and dyspnea. In severe cases, shock, multiple organ failure, and even death may occur. It is one of the most common critical illnesses in clinical practice, characterized by rapid onset, rapid progression, and high mortality. It is an important cause of death in ICU patients worldwide [1]. In the United States, community-acquired pneumonia (CAP) results in approximately 1.4 million emergency department visits, 740,000 hospitalizations, and 41,000 deaths each year [2]. Despite significant advances in diagnostic methods, anti-infective treatment, and intensive care in recent years, the clinical prognosis of severe pneumonia is still not ideal, with a mortality rate as high as 25%–50% [3]. Studies have shown that delayed diagnosis is one of the key factors affecting the prognosis [4].

To identify high-risk pneumonia patients early, the diagnostic criteria for severe pneumonia have undergone multiple revisions. The most influential of these are the American Thoracic Society (ATS) diagnostic criteria of 1993 and 2007. Clinical practice has shown that this diagnostic criteria is suitable for predicting whether a patient should be admitted to the ICU. Although it can significantly improve patient survival, it is not used for early diagnosis of severe pneumonia [5].

Conventional pulmonary imaging of severe pneumonia lacks specificity in early diagnosis. Advances in lung imaging technology have provided an important tool for early identification of severe pneumonia [6], as well as the rapid development of artificial intelligence (AI)-assisted image analysis technology [7, 10] AI technology uses deep learning algorithms to extract features and recognize patterns in massive amounts of medical imaging data, overcoming the limitations of traditional imaging diagnosis that relies on the physician's subjective experience and judgment and inconsistent diagnostic standards.

* Corresponding author: Sailing He (sailing@zju.edu.cn).

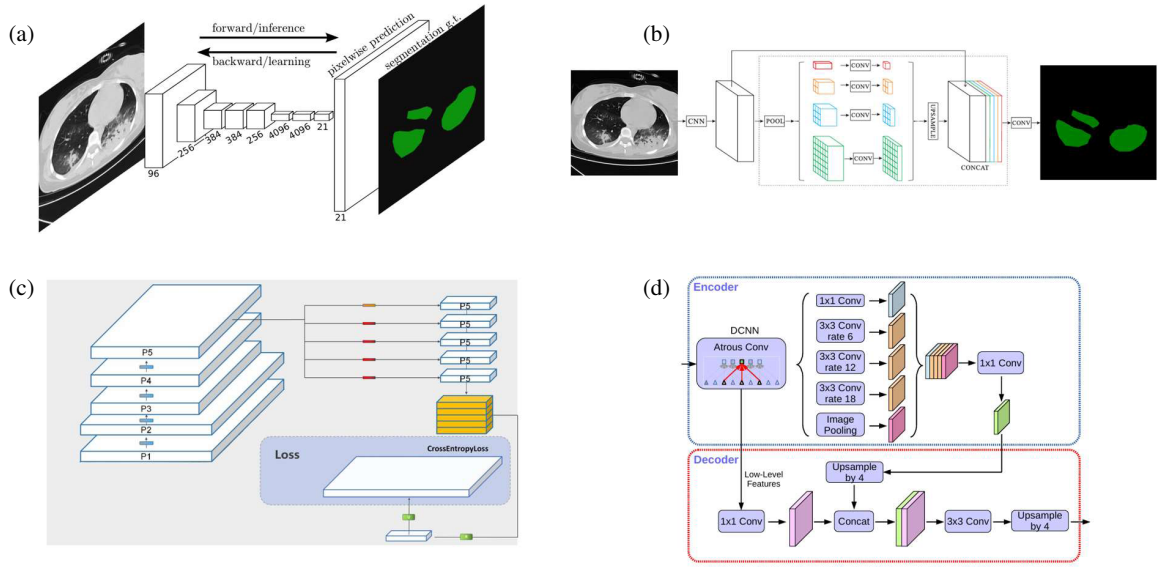

FIGURE 1. Deep learning network structures used here for bacterial pneumonia classification. (a) Fully Convolutional Networks (FCN). (b) Pyramid Scene Parsing Network (PSPNet). (c) Deeplabv3 network structure. (d) Deeplabv3plus network structure.

It achieves automated detection, quantitative analysis, severity assessment, and prognosis prediction of pneumonia lesions, significantly improving the accuracy and efficiency of early diagnosis [11].

The imaging features of the lungs will change accordingly with different infectious pathogens. If the pathogens can be identified early, anti-infection treatment is much more effective. Studies have shown that AI has advantages in distinguishing pathogen infections in lung CT images, and the differences in the incidence of various CT features in patients with fungal and bacterial infections are statistically significant [12] In this paper, we use deep learning methods to train and achieve three-class classifications of severe bacterial infection, mild bacterial infection and normal (no infection).

2. METHODS

In this paper, we use several deep learning network structures, namely, FCN, PSPNet, DeeplabV3, and DeeplabV3 plus, for bacterial pneumonia classification and compare the results. The fully convolutional network (FCN) [13] is the foundation of deep learning for semantic segmentation, as shown in Figure 1(a). The FCN model is relatively simple and its core principle is to use a convolutional neural network to extract image features. The last classification layer in the image classification network is discarded and replaced with a 1×1 convolution to obtain the feature map to be classified. The map is then restored to the size of the original image through upsampling, thereby achieving classification of each pixel and obtaining the segmentation result.

PSPNet (Pyramid Scene Parsing Network) [14] introduced the pyramid pooling module to capture contextual information at different scales, as shown in Figure 1(b). The pyramid pooling module can extract global and local contextual information at different scales, which helps to better understand the semantic content in the image and thus improve the segmentation performance. The pyramid pooling module fuses features at four different pyramid scales. The red mark represents global pooling, which generates a single bin output. The following pyramid levels divide the feature map into different sub-regions and form pooled representations at different locations. The outputs of different levels in the pyramid pooling module contain feature maps of different sizes. To maintain the weight of the global features, a 1×1 convolutional layer is used after each pyramid level to reduce the dimension of the context representation to $1/N$ of the original dimension (if the pyramid level size is N). The low-dimensional feature map is then upsampled, and bilinear interpolation is used to obtain features of the same size as the original feature map. Finally, the features at different levels are concatenated as the final global features of the pyramid pooling output.

DeepLabv3 [15] aims to improve segmentation accuracy by capturing multi-scale contextual information through atrous convolution and atrous spatial pyramid pooling (ASPP). Global average pooling is applied on the last feature map of the model, and the generated image-level features are fed into a 1×1 convolution with 256 filters, followed by bilinear upsampling of the features to the required spatial dimensions. Finally, the improved ASPP includes a 1×1 convolution and three 3×3 convolutions along with the image-level features. The feature results of all branches are then concatenated and passed through another 1×1 convolution, followed by a final 1×1 convolution, to generate the final logits.

DeepLabv3 plus [16] proposes an encoding and decoding structure to address the multi-scale problem of objects and the problem that multiple downsamplings will cause the resolution of feature maps to decrease, resulting in reduced prediction accuracy and loss of boundary information. The encoding part is a DeepLabV3 network. Since the low-level features account for

TABLE 1. Distribution of training and test sets of images for severe bacterial infection and mild bacterial infection.

	training set	test set	total
severe bacterial infection	133	46	179
mild bacterial infection	314	106	420
total	447	152	599

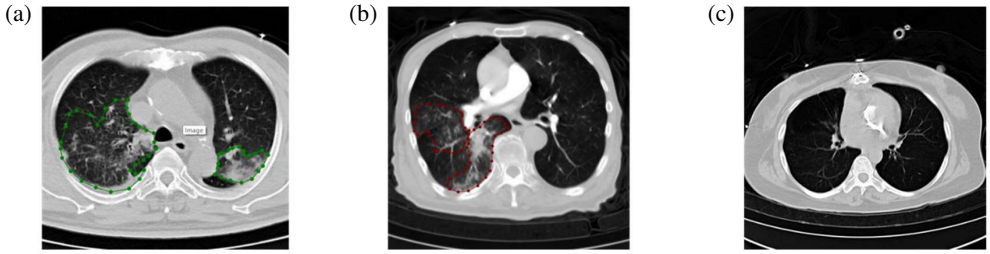

FIGURE 2. CT image samples with different levels of infection. (a) Severe bacterial infection. (b) Mild bacterial infection. (c) Normal.

a small proportion, 1×1 conv is used for channel compression. The features extracted by the encoder have richer information, and thus they account for a large proportion and are conducive to training. The encoder result is upsampled 4 times to be consistent with the underlying features. After the two feature maps are connected, they are refined through a 3×3 convolution, and finally upsampled 4 times to obtain pixel-level predictions.

3. CLASSIFICATION RESULTS

3.1. Dataset Introduction

After expert annotation and screening, a total of 706 CT images were obtained, including 179 images with severe bacterial infection, 420 images with mild bacterial infection, and 107 normal images. The images with severe and mild bacterial infection were randomly shuffled and divided into training and test sets, as shown in Table 1. In the training set, there are 133 images with severe bacterial infection and 314 images with mild bacterial infection. In the test set, there are 46 images with severe bacterial infection and 106 images with mild bacterial infection. Figure 2 displays CT image samples of different infection levels, where blue and red boundaries indicate severe and mild bacterial infections, respectively, based on manual annotation.

3.2. Evaluation Indicators

In the field of medical image classification, the confusion matrix is most commonly used to show the comparison between the model prediction results and the actual true results. For the classification of two classes, TP (True Positive) indicates the number of positive samples correctly predicted by the model, TN (True Negative) indicates the number of negative samples correctly predicted by the model, FP (False Positive) indicates that the model incorrectly predicts the negative class as the positive class and FN (False Negative) indicates that the model incorrectly predicts the positive class as the negative class. The quality of the classification results is usually evaluated by precision (P), recall (R), sensitivity (sens), specificity (spec), F1-score and overall accuracy. The calculation formulas are as follows

$$P = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \tag{1}$$

$$R = \mathrm{sens} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \tag{2}$$

$$\mathrm{spec} = \frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}} \tag{3}$$

$$\text{F1-score} = \frac{2 \times P \times R}{\mathrm{P}+\mathrm{R}} \tag{4}$$

$$\text{Overall accuracy} = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}} \tag{5}$$

Models of FCN, PSPNet, Deepabv3, and Deeplabv3 plus are trained using mmsegmentation [17] without using pre-trained models. The initial learning rate is set to 0.01, the gradient descent method uses stochastic gradient descent (SGD), the gradient update strategy uses PolyLR, the backbone uses ResNet 50, and the number of iterations is set to 20000. Figure 3 shows the convergence of the loss reduction for models FCN, PSPNet, Deeplabv3, and Deeplabv3 plus during the training process.

To achieve three-class classifications of severe bacterial infection, mild bacterial infection and normal (no infection), 107 normal images were added during the test.

From the experimental results in Figure 4, FCN has an overall accuracy of 93.82% in the 3-class classification of bacterial pneumonia (with a sensitivity of 81.25% for severe bacterial infection, a specificity of 96.68%, a precision of 84.78%, a recall of 81.25 %, and an F1 score of 82.98%). The conventional convolution operations in FCN models are limited to capturing local neighborhood information, making them ineffective for modeling long-range dependencies. While FCN's skip connections enable feature fusion at different depths, they

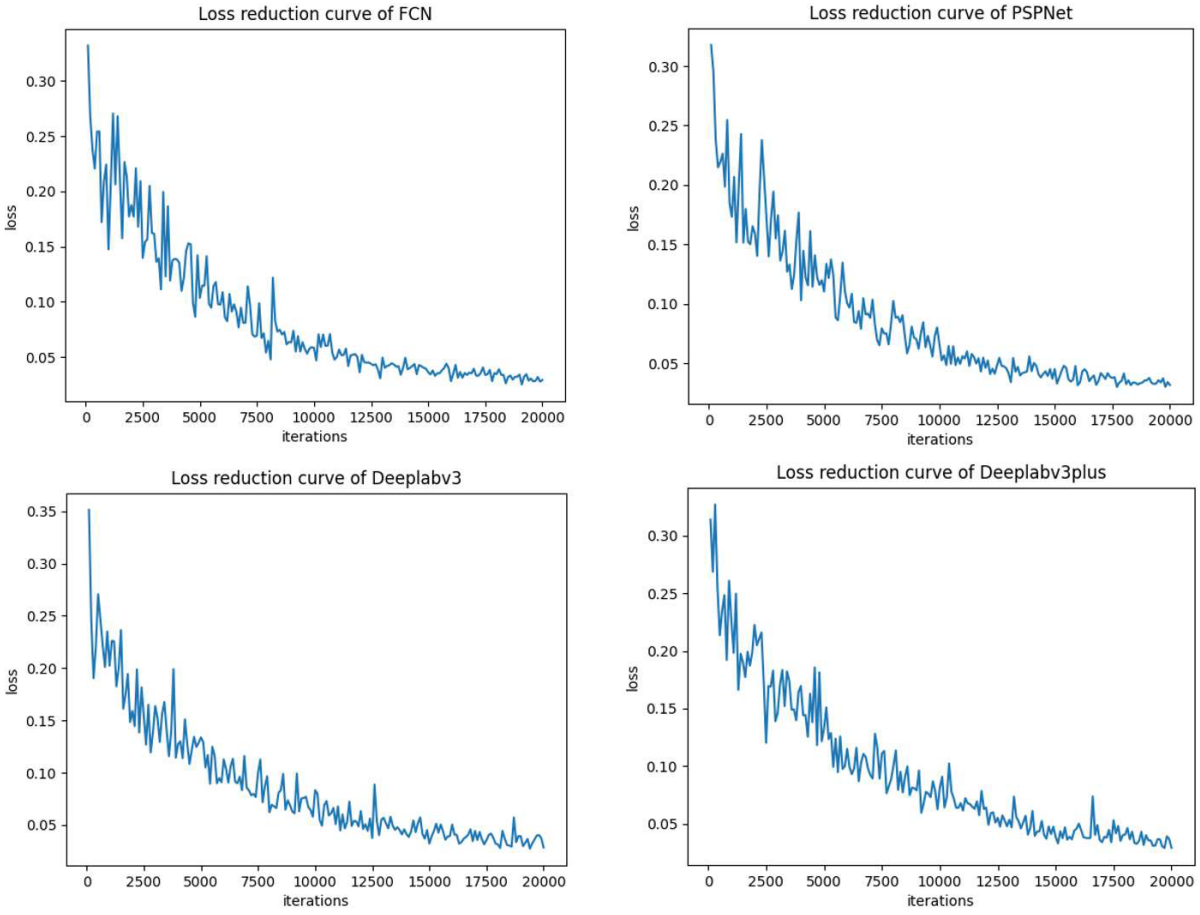

FIGURE 3. Loss reduction curve of models FCN, PSPNet, Deeplabv3, and Deeplabv3 plus during the training process.

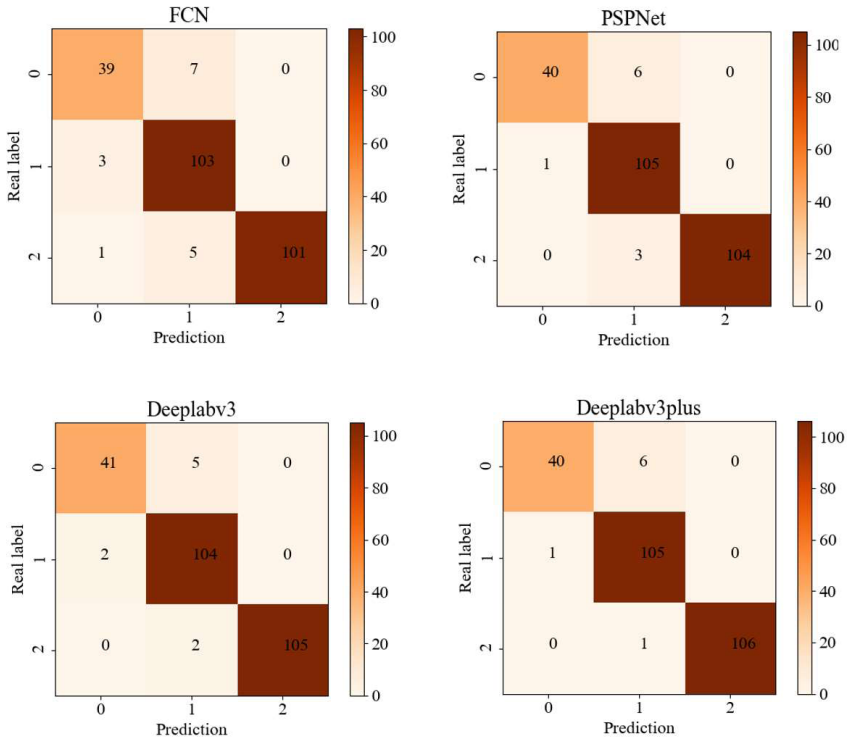

FIGURE 4. The confusion matrices of the 3-class classification results, with the coordinate axis 0 representing severe bacterial infection images, 1 representing mild bacterial infection images, and 2 representing normal images.

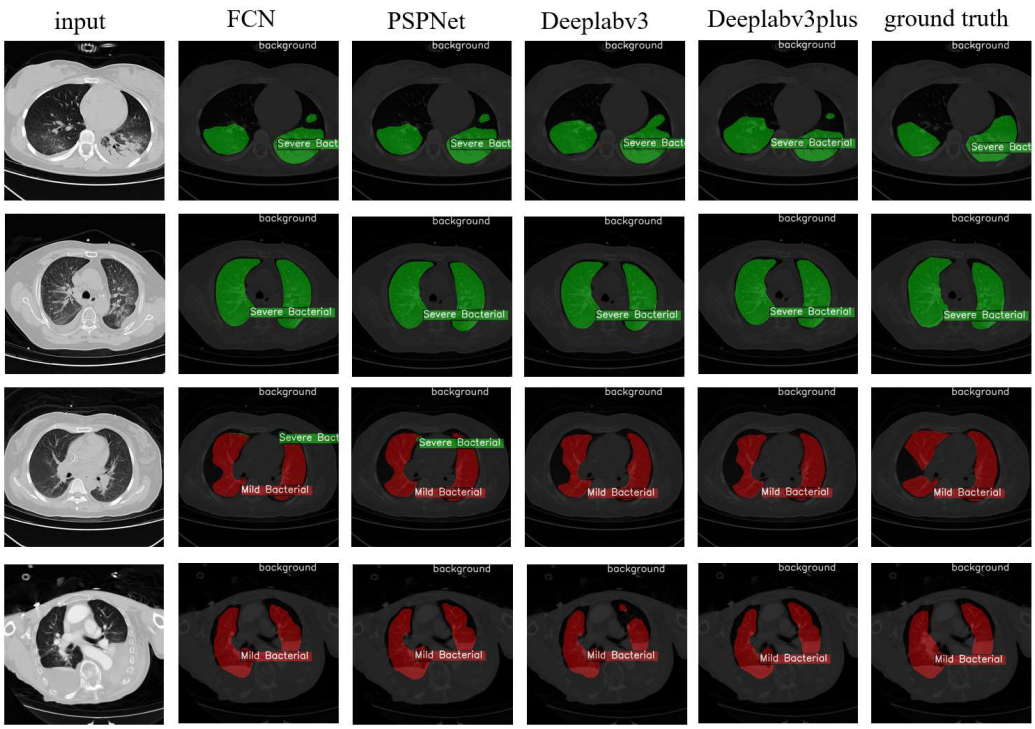

FIGURE 5. Visualization results: green color indicates severe bacterial infection, and red color indicates mild bacterial infection.

lack mechanisms for global context aggregation. This architectural limitation ultimately results in suboptimal F-score performance. PSPNet has an overall accuracy of 96.14% in the 3-class classification of bacterial pneumonia (with a sensitivity of 90.91% for severe bacterial infection, a specificity of 97.21%, a precision of 86.96%, a recall of 90.91%, and an F1 score of 88.89%). Deeplabv3 has an overall accuracy of 96.53% in the 3-class classification of bacterial pneumonia (with a sensitivity of 91.11% for severe bacterial infection, a specificity of 97.66%, a precision of 89.13%, a recall of 91.11%, and an F1 score of 90.11%). Deeplabv3 plus has the best overall accuracy of 96.91% in the 3-class classification of bacterial pneumonia (with a sensitivity of 95.24% for severe bacterial infection, a specificity of 97.24%, a precision of 86.96%, a recall of 95.24%, and an F1 score of 90.91%). Figure 5 visualizes some classification results using different models of FCN, PSPNet, Deeplabv3, and Deeplabv3 plus: green color indicates severe bacterial infection, and red color indicates mild bacterial infection. The consistent performance across training and test sets (accuracy difference $< 2\%$, demonstrated in DeepLabV3Plus) confirms the model's generalization capability, ruling out both overfitting and underfitting concerns. For the identification of severe pneumonia, we group mild and normal cases into the negative class, allowing the use of binary classification. As a result, conventional formulas for accuracy, sensitivity, and specificity are preserved.

4. DISCUSSION AND CONCLUSION

Based on a CT image dataset of 180 patients with bacterial pneumonia (706 annotated images in total), this study systematically evaluated the performance of four deep learning models, FCN, PSPNet, DeepLabv3, and DeepLabv3plus, in the task of three-class classification of bacterial pneumonia severity (severe bacterial infection, mild bacterial infection, and normal tissue). The study found that in terms of model performance, DeepLabv3plus showed the best classification ability, with an overall accuracy of 96.91% (259 test images). Its sensitivity for severe bacterial infection was 95.25% (95% confidence interval: 93.7–96.8%), specificity was 97.24% (96.1–98.3%), and F1 score was 90.91%, which was significantly better than the other models (McNemar test showed that it was significantly better than FCN, PSPNet, and DeepLabv3 models, $P < 0.05$). This quality is due to its multi-scale feature fusion mechanism: DeepLabv3+ employs an atrous spatial pyramid pooling (ASPP) module to capture multi-scale contextual information through parallel atrous convolutions and global average pooling. To improve computational efficiency while preserving multi-scale representation capabilities, the architecture incorporates depth-wise separable convolutions in the ASPP backbone. The encoder effectively extracts rich contextual features through ASPP, while the decoder progressively refines spatial details by sequentially upsampling and fusing these features with lower-level representations. The results show that deep learning technology has significant potential in the early diagnosis of severe pneumonia.

In addition, this study broke through the clinical applicability limitations of the traditional two-class classification model for the first time, achieved pixel-level three-class classification segmentation of the severity of pneumonia bacterial infection, and verified the universality of the multi-scale feature extraction mechanism for lung image analysis. DeepLabv3plus reduced the misjudgment rate of severe infection to only 4.76%

(7 false negatives) through hierarchical feature fusion, and no misdiagnosis of normal samples occurred (0 false positives). In the application scenario of the emergency department, early identification of severe pneumonia infection and the ability to distinguish the severity of different infections provide strong support for clinical decision-making. It enables rapid and accurate triage to avoid excessive occupation of ICU resources by non-infected patients and mildly infected patients. The AI diagnostic system established in this study is also of great significance for public health events, especially during the outbreak of respiratory infectious diseases (such as COVID-19) [18]. The 95.25% detection rate of severe infection can significantly shorten the diagnosis delay, identify critically ill patients early, help optimize the allocation of medical resources, and significantly improve work efficiency.

This study only used the first CT image of admission and did not track the changes in images during treatment. In the future, the introduction of a time series data analysis can improve the ability to predict prognosis. Multimodal data fusion can also be used to integrate clinical indicators (such as blood routine tests, inflammatory markers) and image features to establish a comprehensive prediction model, which is expected to promote the practical application of clinical decision-making support systems.

ACKNOWLEDGEMENT

This work was supported by "Pioneer" and "Leading Goose" R&D Program of Zhejiang Province (No. 2023C03083), the National Key Research and Development Program of China (2022YFC2010003), Ningbo Science and Technology Project (2023Z179), Science and Technology Plan Key Project of Taizhou City (24gyz01) and Science and Technology Plan Project of Luqiao (Taizhou) District(2024G2009).

REFERENCES

[1] Jain, S., W. H. Self, R. G. Wunderink, S. Fakhran, R. Balk, A. M. Bramley, C. Reed, C. G. Grijalva, E. J. Anderson, D. M. Courtney, *et al.*, "Community-acquired pneumonia requiring hospitalization among U.S. adults," *New England Journal of Medicine*, Vol. 373, No. 5, 415–427, 2015.

[2] Vaughn, V. M., R. P. Dickson, J. K. Horowitz, and S. A. Flanders, "Community-acquired pneumonia: A review," *Jama*, Vol. 332, No. 15, 1282–1295, 2024.

[3] Cillóniz, C., C. Dominedò, C. Garcia-Vidal, and A. Torres, "Community-acquired pneumonia as an emergency condition," *Current Opinion in Critical Care*, Vol. 24, No. 6, 531–539, 2018.

[4] Woodhead, M., C. A. Welch, D. A. Harrison, G. Bellingan, and J. G. Ayres, "Community-acquired pneumonia on the intensive care unit: Secondary analysis of 17,869 cases in the ICNARC Case Mix Programme Database," *Critical Care*, Vol. 10, S1, 2006.

[5] Marti, C., N. Garin, O. Grosgurin, A. Poncet, C. Combescure, S. Carballo, and A. Perrier, "Prediction of severe community-acquired pneumonia: A systematic review and meta-analysis," *Critical Care*, Vol. 16, No. 4, R141, Jul. 2012.

[6] Luo, N., H. Zhang, Y. Zhou, Z. Kong, W. Sun, N. Huang, and A. Zhang, "Utility of chest CT in diagnosis of COVID-19 pneumonia," *Diagnostic and Interventional Radiology*, Vol. 26, No. 5, 437–442, 2020.

[7] Kwon, T., S. P. Lee, D. Kim, J. Jang, M. Lee, S. U. Kang, H. Kim, K. Oh, J. On, Y. J. Kim, *et al.*, "Diagnostic performance of artificial intelligence model for pneumonia from chest radiography," *PLoS One*, Vol. 16, No. 4, e0249399, Apr. 2021.

[8] Xu, Z., Y. Jiang, J. Ji, E. Forsberg, Y. Li, and S. He, "Classification, identification, and growth stage estimation of microalgae based on transmission hyperspectral microscopic imaging and machine learning," *Optics Express*, Vol. 28, No. 21, 30 686–30 700, 2020.

[9] Zhu, H., J. Luo, J. Liao, and S. He, "High-accuracy rapid identification and classification of mixed bacteria using hyperspectral transmission microscopic imaging and machine learning," *Progress In Electromagnetics Research*, Vol. 178, 49–62, 2023.

[10] Wang, L. and S. He, "A 3-band iteration method to transfer knowledge learned in rgb pretrained models to hyperspectral domain," *Progress In Electromagnetics Research M*, Vol. 128, 1–9, 2024.

[11] Zhang, Y.-H., X.-F. Hu, J.-C. Ma, X.-Q. Wang, H.-R. Luo, Z.-F. Wu, S. Zhang, D.-J. Shi, Y.-Z. Yu, X.-M. Qiu, *et al.*, "Clinical applicable AI system based on deep learning algorithm for differentiation of pulmonary infectious disease," *Frontiers in Medicine*, Vol. 8, 753055, Dec. 2021.

[12] Huang, T., X. Zheng, L. He, and Z. Chen, "Diagnostic value of deep learning-based CT feature for severe pulmonary infection," *Journal of Healthcare Engineering*, Vol. 2021, No. 1, 5359084, Nov. 2021.

[13] Long, J., E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3431–3440, 2015.

[14] Zhao, H., J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2881–2890, 2017.

[15] Chen, L.-C., G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *Arxiv Preprint ArXiv:1706.05587*, 2017.

[16] Chen, L.-C., Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 801–818, 2018.

[17] MMSegmentation Contributors, "OpenMMLab semantic segmentation toolbox and benchmark," 2020.

[18] Hiremath, A., V. S. Viswanathan, K. Bera, R. Shiradkar, L. Yuan, K. Armitage, R. Gilkeson, M. Ji, P. Fu, A. Gupta, *et al.*, "Deep learning reveals lung shape differences on baseline chest CT between mild and severe COVID-19: A multi-site retrospective study," *Computers in Biology and Medicine*, Vol. 177, 108643, 2024.