Proceedings of ICBioMed 2025 Symposium: AI for Healthcare: Advanced Medical Data Analytics and Smart Rehabilitation
DOI: 10.54254/2753-8818/113/2025.AU23577

Research on an Intelligent Pneumonia Diagnosis Model Based on Improved Convolutional Neural Networks Using Chest X-ray Images

Xingrong Li

Shandong University, Weihai, China
li53243322@gmail.com

Abstract: Pneumonia, a common health concern today, requires early and accurate diagnosis. Chest X-ray examinations play a critical role in the early detection of pneumonia. To enhance diagnostic accuracy, this study utilizes a deep learning-based convolutional neural network (CNN) model, trained on a dataset of 5,216 chest X-ray images obtained from pediatric patients aged 1-5 years at Guangzhou Women and Children's Medical Center. Among these, 3,875 images show signs of pneumonia and 1,341 images are normal, serving as the training and testing data for the model. By incorporating Dropout techniques and Batch Normalization methods, the model's robustness and generalization ability were significantly improved. Experimental results demonstrate that the model achieves a diagnostic accuracy of 97.83%, which will effectively alleviate physicians' workload and holds substantial clinical application value.

Keywords: Medical Image Classification, Convolutional Neural Network, Chest X-ray Pneumonia Diagnosis

1. Introduction

Pneumonia, a respiratory disease that poses a significant threat to public health globally, necessitates early and accurate diagnosis to reduce patient mortality. According to research published in The Lancet, pneumonia accounts for over 250 million medical visits annually. Chest X-ray imaging, due to its accessibility and low cost, has become a primary diagnostic tool for pneumonia. However, two major challenges persist: (1) Pneumonia lesions exhibit diverse morphologies and high heterogeneity in distribution; (2) There is an imbalance in medical resource allocation—primary healthcare institutions lack specialized physicians, while tertiary hospitals must process thousands of images daily. Traditional manual diagnostic approaches are increasingly inadequate for meeting the growing demand for image analysis.

In recent years, convolutional neural networks (CNNs) based on deep learning have demonstrated remarkable potential in medical image analysis. Classical models such as ResNet and DenseNet have achieved over 90% accuracy in pulmonary disease detection tasks. This study constructs a CNN model to classify chest X-ray images, thereby improving the efficiency and accuracy of image processing and diagnosis, and providing a more efficient and convenient auxiliary diagnostic tool for primary healthcare settings.

DOI: 10.54254/2753-8818/113/2025.AU23577

This paper is organized into six sections. Section Two reviews relevant classical literature and discusses the development of deep learning combined with image recognition in recent years. Section Three details the feedforward structure of the convolutional neural network employed in this study. Section Four elaborates on the research process and presents the research results. Section Five provides a discussion, analyzing the strengths of the research methodology and interpreting the results. Finally, Section Six concludes the paper, discussing the advantages, limitations, and future research directions of this study.

2. Literature review

In recent years, deep learning technology has demonstrated broad application prospects in the field of medical image analysis. Researchers have developed automatic detection, classification, and risk prediction models targeting specific diseases based on various deep learning frameworks, significantly improving the efficiency and accuracy of medical image diagnosis.

In 2023, a study employing a deep learning (DL) model investigated the performance of coronary computed tomography angiography (CTA) in diagnosing coronary artery stenosis. Using coronary CTA data from 89 patients with suspected coronary artery disease at Peking University Shougang Hospital and taking invasive coronary angiography as the reference standard, the study found that the DL model achieved high diagnostic accuracy for obstructive coronary artery stenosis, with an AUC of 0.92, sensitivity of 86.2%, and specificity of 87.6%. Moreover, its performance in diagnosing stenosis caused by non-calcified plaques, mixed plaques, and segmental plaques surpassed that of physicians [1].

In early applications, convolutional neural networks (CNNs) were widely used for the automatic detection and classification of specific organs or lesions. In 2017, a study based on brain CT image classification used a CNN model for preliminary screening of Alzheimer's disease (AD), utilizing brain CT data from 342 subjects to distinguish among AD, organic lesions, and normal aging states, achieving an average classification accuracy of 82.3% [2]. In 2020, Zhou et al. employed a CNN model using CT images from 1,024 adult rib fracture patients from three hospitals to accurately classify fresh fractures, healing fractures, and old fractures, demonstrating robust performance comparable to that of radiologists and reducing the average detection time by 132.07 seconds [3].

In 2021, Gu et al. utilized the GISTNet model to conduct differential diagnosis between gastrointestinal stromal tumors (GISTs) with diameters ≤5 cm and other gastric submucosal tumors (SMTs) based on enhanced CT images. Using data from 168 patients (107 with GIST-related SMTs), the GISTNet model achieved an AUC of 0.900 on the test set, with a sensitivity of 100%, specificity of 67%, and accuracy of 83%, outperforming traditional imaging models and junior radiologists [4].

With the advancement of deep learning technology, some studies have integrated traditional imaging feature extraction methods with radiomics techniques to further enhance diagnostic accuracy and generalization ability. In 2018, a radiomics-based study using MRI images from 294 patients extracted key features and combined them with a support vector machine (SVM) algorithm to classify mass-forming cholangiocarcinoma, hepatocellular carcinoma, and combined hepatocellular-cholangiocarcinoma, achieving a maximum classification accuracy of 85.3% [5]. In 2019, Fu et al. adopted traditional machine learning and deep learning methods to study the automatic classification of liver fibrosis progression in patients with chronic hepatitis B. Based on grayscale ultrasound and elastography images from 354 patients, classification models were built using SVM, sparse representation classifiers, and LeNet-5 neural networks. The binary classification accuracies were 89.8%, 91.8%, and 90.7% respectively, while the multiclass classification accuracies ranged from 75.3% to 87.2% [6]. In 2022, another study combined MRI radiomics and machine learning techniques, using DWI and PWI imaging data from 214 acute stroke patients to extract 792 radiomics

DOI: 10.54254/2753-8818/113/2025.AU23577

features, and constructed an SVM model to predict hemorrhagic transformation in acute stroke, achieving an AUC of 0.921 on the test set and demonstrating excellent performance [7].

Deep convolutional neural network (DCNN) technology has shown significant advantages in tumor detection tasks and has gradually been applied to more complex multimodal image fusion analyses. In 2024, a DCNN-based study achieved a tumor detection accuracy of 91.80% on liver CT images, significantly outperforming traditional watershed and connected component algorithms, effectively reducing misdiagnosis and missed diagnosis rates [8]. Furthermore, in 2022, a study based on the Faster R-CNN model used 2,354 CT images from 32 hepatocellular carcinoma patients (16 with bile duct tumor thrombus) and employed the Faster R-CNN network to identify dilated bile ducts, achieving an outstanding AUC of 0.94 for preoperative diagnosis of bile duct tumor thrombus in liver cancer patients [9].

In 2020, another study on deep learning and multimodal medical image fusion applied CNNs to identify patients with subthreshold depression (StD). Using MRI and fMRI data from 56 StD patients and 70 healthy individuals, the study trained and analyzed models through MRI/fMRI data fusion, improving the classification accuracy of subthreshold depression to 78.57%, an increase of 5.55% compared to unimodal data, significantly enhancing classification performance [10].

From early single-modality image classification studies to more recent complex tasks integrating radiomics and multimodal data, deep learning has made remarkable progress in the field of medical image analysis. In the future, improving model interpretability, enhancing multicenter data generalization ability, and optimizing clinical applications will become important research directions in this field to better serve practical healthcare needs.

3. Methodology

Convolutional Neural Networks (CNNs) are a type of feedforward neural network specifically designed to handle image data, primarily used for image recognition and processing. Their core concept is to extract multi-level feature representations from data through local connections, weight sharing, and spatial downsampling, thereby effectively reducing the number of parameters while preserving spatial information.

A convolutional neural network is composed of an input layer, hidden layers (including convolutional layers, activation functions, pooling layers, and fully connected layers), and an output layer.

The convolutional layer generates a set of parallel feature maps by sliding different convolution kernels (filters) across the input image to perform local feature extraction and calculating the weighted sum at each position. The formula for the convolution operation is:

$$Y(i,j) = (X \times W)(i,j) + b \tag{1}$$

Where, X represents the input image or feature map, W is the convolution kernel (typically of size $k \times k$), b is the bias term, and $Y(i,j)$ is the output value of the feature map.

The activation function introduces nonlinearity into the model. A commonly used activation function is ReLU (Rectified Linear Unit), which retains the input value when positive and outputs zero when negative. In this study, LeakyReLU is used as the activation function, allowing a small, nonzero gradient when the input is negative. LeakyReLU is defined as:

$$\text{LeakyReLu}(x) = \begin{cases} x, x \geq 0 \\ \alpha x, x < 0 \end{cases} \tag{2}$$

Where, α is a small positive constant, typically set to 1e−2 by default. In this study, $\alpha = 0.2$.

DOI: 10.54254/2753-8818/113/2025.AU23577

The pooling layer is used for downsampling feature maps, reducing the number of parameters and enhancing translation invariance. The most commonly used pooling operations are max pooling and average pooling. The formula for max pooling is:

$$Y(i,j) = \max(X(i,j), X(i+1,j), X(i,j+1), X(i+1,j+1)) \quad (3)$$

Where, $Y(i,j)$ is the output value of the feature map.

The fully connected layer follows the convolution and pooling operations and is used for classification or regression tasks. Each neuron in the fully connected layer connects to all neurons in the previous layer, mapping high-level features to the classification space:

$$z = Wx + b \quad (4)$$

Where, x is the input vector, W is the weight matrix, and b is the bias term.

The common architectural pattern of convolutional neural networks is:

$$INPUT -> [[CONV] * N -> POOL] * M -> [FC] * K -> OUTPUT \quad (5)$$

Where, $INPUT$ denotes the input layer; $CONY$ denotes the convolutional layer; $POOL$ denotes the pooling layer; FC denotes the fully connected layer; $OUTPUT$ denotes the output layer; and N, M, K represent positive integers. (Pooling layers may be omitted.)

The structure adopted in this study is:

$$INPUT -> [CONV] * 4 -> [FC] * 2 -> OUTPUT \quad (6)$$

The hyperparameter list for this study is shown below (Table 1).

Table 1: Hyperparameter list

Layer	Output shape	Param number
InputLayer	(None, 256, 256, 3)	0
Conv2D	(None, 128, 128, 32)	896
BatchNormalization	(None, 128, 128, 32)	128
LeakyReLu	(None, 128, 128, 32)	0
Dropout	(None, 128, 128, 32)	0
Conv2D	(None, 64, 64, 64)	18496
BatchNormalization	(None, 64, 64, 64)	256
LeakyReLu	(None, 64, 64, 64)	0
Dropout	(None, 64, 64, 64)	0
Conv2D	(None, 32, 32, 128)	73856
BatchNormalization	(None, 32, 32, 128)	512
LeakyReLu	(None, 32, 32, 128)	0
Dropout	(None, 32, 32, 128)	0
Conv2D	(None, 16, 16, 256)	295168
BatchNormalization	(None, 16, 16, 256)	1024
LeakyReLu	(None, 16, 16, 256)	0
Dropout	(None, 16, 16, 256)	0
Flatten	(None, 65536)	0
Dense	(None, 100)	6553700
BatchNormalization	(None, 100)	400
LeakyReLu	(None, 100)	0
Dropout	(None, 100)	0
Dense	(None, 2)	202

DOI: 10.54254/2753-8818/113/2025.AU23577

4. Results

In this study, we focused on determining whether patients had pneumonia based on chest X-ray images. The research data were collected from pediatric patients aged one to five at the Guangzhou Women and Children's Medical Center, comprising a total of 5,216 chest X-ray images, including 3,875 images of pneumonia cases and 1,341 normal images. The dataset was randomly divided into a training set (4,731 cases) and a test set (485 cases) using the hold-out method.

During the data preprocessing stage, all images were normalized. Specifically, the 5,216 images were proportionally stretched or compressed to a uniform size of 256×256 pixels, and the color mode of the images was standardized to RGB.

For model construction, a convolutional neural network composed of four convolutional layers and two fully connected layers was built to perform binary classification (NORMAL/PNEUMONIA) on the input images. The specific configurations of the convolutional layers were as follows: the numbers of convolutional kernels (filters) in each layer were 32, 64, 128, and 256, respectively, with a kernel size of 3×3 and a stride of 2. Batch normalization was applied to each batch of input data. The activation function used was the LeakyReLU function, with a parameter $\alpha = 0.2$. Additionally, dropout was introduced with $\text{dropout rate} = 0.2$. In the fully connected layers, the first contained 100 neurons, and the second, serving as the output layer, contained 2 neurons, employing softmax as the activation function.

Regarding training parameters, the learning rate was set to 0.0005, determining the step size for parameter updates during training. The loss function used was categorical cross-entropy ("categorical_crossentropy"), and accuracy was monitored as the key metric throughout the training and evaluation processes. Moreover, the ModelCheckpoint callback function was utilized to save model checkpoints during training, specifically preserving only the model that performed best on the test set to avoid overfitting or poor performance.

Figure 1 and Figure 2 display images of the NORMAL category, while Figure 3 and Figure 4 show images of the PNEUMONIA category.

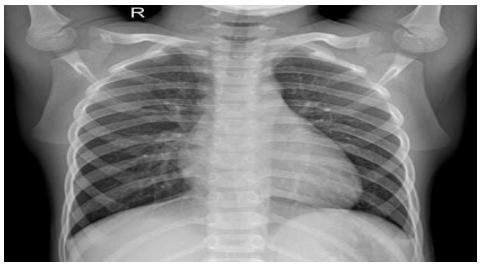

Figure 1: Image 1 of normal

Figure 2: Image 2 of normal

Figure 3: Image 1 of pneumonia

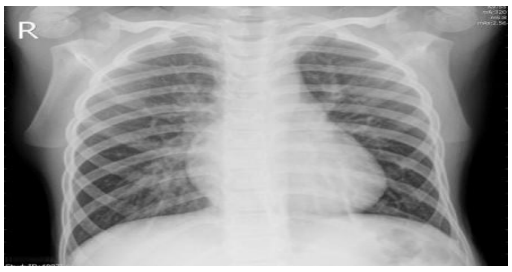

Figure 4: Image 2 of pneumonia

The experiment was conducted over 30 epochs, with the training set achieving a final accuracy of 0.9985 and a loss value of 0.0051. On the test set, the highest accuracy obtained was 0.9793, with a

DOI: 10.54254/2753-8818/113/2025.AU23577

corresponding test loss of 0.0821. This best result occurred during the 29th epoch. Detailed accuracy and loss values for both the training and test sets over the 30 epochs are presented in [Table 2]. Figure 5 and Figure 6 illustrate the changes in loss (Training and Test Loss) and accuracy (Training and Test Accuracy) throughout the training and testing processes, respectively.

Table 2: Training results of each epoch

Epoch	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss
1	0.917564988	0.234142438	0.872164965	0.325331271
2	0.968716979	0.093950987	0.942268014	0.172519907
3	0.979919672	0.060765516	0.950515449	0.098253578
4	0.987317681	0.041428987	0.925773203	0.213178158
5	0.990699649	0.030258378	0.962886572	0.106838636
6	0.989008665	0.030921673	0.969072163	0.084634304
7	0.993236125	0.020276889	0.954639196	0.114069179
8	0.993447483	0.019337399	0.868041217	0.405696213
9	0.995349824	0.015466789	0.95670104	0.145721182
10	0.994504333	0.018674087	0.960824728	0.113141328
11	0.995349824	0.012712151	0.973195851	0.091583982
12	0.996195316	0.013728292	0.967010319	0.100921676
13	0.995561182	0.011852826	0.954639196	0.135485202
14	0.994927049	0.014774959	0.967010319	0.102580681
15	0.996406674	0.011312119	0.969072163	0.106226236
16	0.996195316	0.010879521	0.964948475	0.095662303
17	0.9978863	0.006715842	0.971134007	0.095564947
18	0.997252166	0.011268928	0.954639196	0.12324208
19	0.997674882	0.00834303	0.967010319	0.131089255
20	0.996618032	0.009587267	0.960824728	0.116533756
21	0.997463524	0.007445859	0.962886572	0.104139052
22	0.997040808	0.009834411	0.964948475	0.097293124
23	0.998097658	0.006101265	0.960824728	0.133533746
24	0.998731792	0.00561474	0.975257754	0.087317802
25	0.998309016	0.006626317	0.971134007	0.108881526
26	0.99682945	0.010370547	0.964948475	0.114163302
27	0.998309016	0.006423338	0.919587612	0.33523199
28	0.997463524	0.007406317	0.964948475	0.104180545
29	0.99682945	0.00815128	0.979381442	0.082091108
30	0.998520374	0.005115778	0.946391761	0.193137333

DOI: 10.54254/2753-8818/113/2025.AU23577

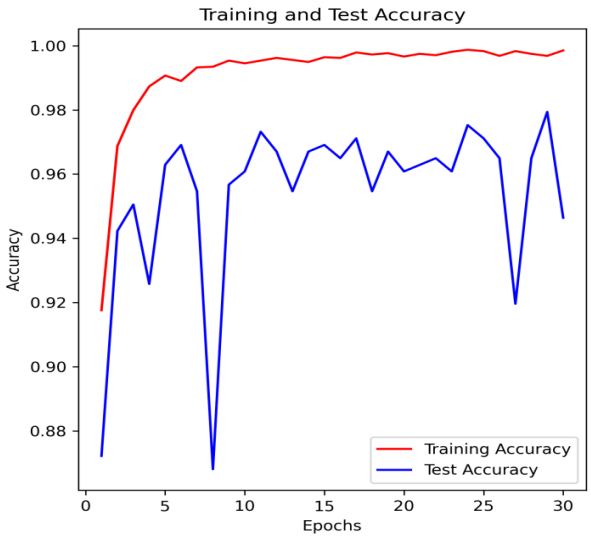

Figure 5: Training and test accuracy

Figure 6: Training and test loss

5. Discussion

This study selected a total of 5,216 chest X-ray images from pediatric patients aged one to five years at Guangzhou Women and Children's Medical Center. The large sample size provides sufficient data support for the research; the distribution ratio between positive and negative samples is relatively balanced, avoiding the issue of dataset imbalance, which helps to enhance the model's training effectiveness and generalization ability. The training set comprises 4,731 images, while the test set consists of 485 images, following a common data splitting standard that ensures the reliability of the model during the training and validation phases. The original images have a high resolution and a clear level of visualization, providing a solid foundation for subsequent image analysis and diagnosis, thus ensuring the accuracy and reliability of the research results.

The original 5,216 images were stretched or compressed to unify the image size to 256×256, mainly because the structure of convolutional neural networks is fixed and cannot directly process input images of inconsistent sizes. Furthermore, standardizing the image size ensures that all input data share a consistent feature space before entering the network, thereby improving training efficiency and the model's generalization ability.

The activation function layer adopts the LeakyReLU function, which offers high computational efficiency and fast training speed. LeakyReLU introduces a non-zero slope α when the input is less than zero, allowing neurons to output non-zero values even for negative inputs, thus avoiding the "dying ReLU" problem and preventing gradient vanishing. This ensures that gradients can continuously update parameters during backpropagation. It also enhances the network's responsiveness to negative inputs, helping to retain information when encountering noise and outliers, thereby improving the model's robustness and generalization ability.

The output layer uses softmax as the activation function to convert the raw outputs into a probability distribution, representing the likelihood that an image belongs to either the NORMAL or PNEUMONIA category.

Batch normalization is applied to each batch of input data to reduce internal covariate shift, alleviate issues such as vanishing and exploding gradients, and ensure that the mean and variance of input data are maintained at 0 and 1, respectively.

DOI: 10.54254/2753-8818/113/2025.AU23577

Dropout technology is introduced to reduce inter-neuron dependency by randomly "dropping out" a portion of the neuron outputs during training, with $\text{dropout rate} = 0.2$, meaning that each neuron has a 20% probability of being dropped during the training process.

The learning rate is set at 0.0005, which is relatively small to ensure that parameter updates are smooth and the model is optimized stably in the direction of the loss function descent. At the same time, the learning speed is not overly slow, allowing the model to achieve good performance within a reasonable timeframe. Moreover, with sufficient sample data, the model can learn features effectively, avoiding underfitting and enabling good overall performance.

The loss function adopts categorical cross-entropy, which measures the divergence between the predicted probability distribution and the true labels. The softmax activation function at the output layer transforms the raw outputs into a probability distribution, and categorical cross-entropy directly optimizes these probabilities. During backpropagation, the combined gradient expression of softmax and categorical cross-entropy is concise, making model training more efficient. Furthermore, with softmax outputs, the loss function remains convex, ensuring a global optimal solution.

Accuracy is used as the key metric monitored during training and evaluation, directly reflecting the model's classification or prediction accuracy for various samples and facilitating the comparison and selection of models.

5.1. Training and test accuracy analysis

The training accuracy curve (red line) rises rapidly at the initial stage, quickly approaching 1.00 and stabilizing, indicating that the model learns well and improves steadily on the training set. The curve is smooth in the early stages and shows minor oscillations later, demonstrating good convergence overall.

The test accuracy curve (blue line) fluctuates significantly at the beginning. Although the overall trend is upward, it does not reach the stability level of the training accuracy. The oscillations are relatively large, suggesting some instability in the model's generalization ability on the test set, possibly due to differences in data distribution and the relatively small size of the test set.

5.2. Training and test loss analysis

The training loss curve (red line) drops sharply in the early stages of training, then levels off and stabilizes near a low value, indicating good model fitting and stable loss convergence on the training set.

The test loss curve (blue line) declines initially but fluctuates greatly, with some apparent peaks at certain epochs, and does not exhibit a continuously stable downward trend. This phenomenon may result from the small sample size of the test set, leading to insufficient generalization and unstable error rates.

This model significantly reduces the workload of manual diagnosis through automated analysis, providing an efficient and convenient auxiliary diagnostic tool for primary healthcare institutions, thus holding important clinical application value.

6. Conclusion

This study, based on convolutional neural networks in deep learning, investigated the diagnosis of pneumonia through chest X-ray images. A total of 5,216 chest X-ray images of pediatric patients aged one to five years from Guangzhou Women and Children's Medical Center were used in the experiments. The results achieved were a training accuracy of 0.9985 with a loss of 0.0051, and a test accuracy of 0.9793 with a loss of 0.0821. These outcomes demonstrate that the CNN model possesses high accuracy and reliability in pneumonia diagnosis and holds significant clinical application value.

DOI: 10.54254/2753-8818/113/2025.AU23577

It can effectively alleviate the workload of physicians and plays an important role in addressing the shortage and relative underdevelopment of rural primary healthcare resources. However, the current study still has certain limitations. The sample source is relatively homogeneous, which restricts the model's generalizability. In future research, we will further expand the sample size and actively broaden the channels for sample collection, aiming to gather chest X-ray data from diverse regions and age groups. This continuous optimization of the model will enhance its generalization ability and provide more robust and reliable auxiliary support for the accurate diagnosis of pneumonia.

References

[1] *Geng, J., Chang, Y., Zhang, B., Wang, S., & Zhang, F. (2023). Diagnostic performance of a deep learning model based on coronary CT angiography for coronary heart disease. Chinese Journal of Medical Imaging Technology, 31(7), 706–712.*

[2] *Hui, R., Gao, X., & Tian, Z. (2017). A preliminary screening method for Alzheimer's disease using CT brain image classification based on deep learning. China Medical Equipment, 32(12), 5.*

[3] *Zhou, Q., Wang, J., Tang, W., & Zhang, R. (2020). Application of convolutional neural network in automatic detection and classification of adult rib fractures on CT images. Imaging Diagnosis and Interventional Radiology, 29(1), 5.*

[4] *Gu, J., Shi, H., Yang, L., Shen, Y., Wang, Z., Feng, Q., et al. (2021). Clinical significance of a deep learning algorithm based on enhanced CT in differentiating gastrointestinal stromal tumors of the stomach with diameter ≤5 cm. Chinese Journal of Gastrointestinal Surgery, 24(9), 8.*

[5] *Zhang, J., Chen, F., Xue, X., Zhang, S., Yao, L., Wang, X., et al. (2018). Value of MRI radiomics based on support vector machine in differentiating pathological types of primary liver cancer. Chinese Journal of Radiology, 52(5), 5.*

[6] *Fu, T., Yao, Z., Ding, H., Xu, Z., Yang, M., & Yu, J., et al. (2019). Value analysis of computer-aided diagnosis in monitoring liver fibrosis progression in chronic hepatitis B patients. National Medical Journal of China, 99(7), 491–495.*

[7] *Miao, L., Peng, M., Wang, T., Chen, G., Yin, X., & Wu, G. (2022). Prediction of hemorrhagic transformation after acute ischemic stroke based on MRI radiomics and machine learning. Magnetic Resonance Imaging, 13(3), 5.*

[8] *Huang, X., & Ma, J. (2024). Research on liver tumor detection algorithms based on deep convolutional neural networks. Journal of Ningxia Normal University, 45(7), 84–91.*

[9] *Liu, J., Wu, J., Liu, A., Bai, Y., Zhang, H., & Yan, M. (2022). Preoperative diagnosis of hepatocellular carcinoma with biliary tumor thrombus based on deep learning methods. Journal of University of Science and Technology of China, 52(12), 47–57.*

[10] *Yin, X., Li, D., Tu, Y., & Shan, B. (2020). Recognition of subthreshold depression patients based on deep learning and multimodal medical image fusion. Chinese Journal of Medical Imaging Technology, 36(8), 5.*