

Article

Not peer-reviewed version

Pulmonary Medical Image Recognition Based on Deep Learning

Linden Fairbairn , Greer Tolland ^{*} , [Mingxuan Xiao](#)

Posted Date: 17 November 2025

doi: 10.20944/preprints202511.1133.v1

Keywords: medical imaging; deep learning; transfer learning; GoogLeNet



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Pulmonary Medical Image Recognition Based on Deep Learning

Linden Fairbairn, Greer Tolland and Mingxuan Xiao *

University of Regina, Regina, Saskatchewan, Canada

* Correspondence: xiaomingxuan7788@outlook.com

Abstract

Pulmonary nodules are important indicators for the early diagnosis of lung cancer, and their identification and classification are of great significance. At present, there exist significant domain discrepancies between source and target datasets when using transfer learning-based recognition algorithms, which leads to poor feature extraction of pulmonary nodules and thus unsatisfactory results. Therefore, this paper proposes an improved convolutional neural network model based on a modified neural network architecture. The model integrates features extracted by the pre-trained GoogLeNet Inception V3 network to enhance its ability to extract relevant features. In order to determine the optimal combination, different groups were tested using accuracy as the evaluation metric. Experiments were conducted on the LUNA16 lung nodule dataset. The results from cross-validation testing show that the improved network achieves an accuracy of 88.80% and a sensitivity of 87.15%. In terms of recognition accuracy and sensitivity, the proposed method outperforms GoogLeNet Inception V3, with improvements of 2.72 and 2.19 percentage points, respectively. Even when tested on small-scale datasets, the model demonstrates better generalization capability. This method can provide objective reference indicators for clinical diagnosis.

Keywords: medical imaging; deep learning; transfer learning; GoogLeNet

1. Introduction

Currently, the incidence and mortality rates of lung cancer are increasing year by year. According to data from the American Cancer Society, approximately 14% of newly diagnosed malignant tumors are lung cancer. Among all malignant tumors, lung cancer ranks highest in incidence and mortality. As pulmonary nodules are closely related to lung cancer and may be an early feature of lung cancer, their detection and diagnosis are of great significance for early lung cancer diagnosis and can serve as an important reference [1,2].

Computed Tomography (CT) is commonly used to evaluate pulmonary nodules, requiring physicians to have specialized medical imaging knowledge and clinical experience. However, due to the differences among individual physicians, Computer-Aided Diagnosis (CAD) technology has been widely applied in recent years as an auxiliary medical tool, providing relatively objective evaluation metrics to assist doctors in making diagnoses [3].

Early research methods often relied on feature extraction techniques, which involved identifying Regions of Interest (ROI) in the images and analyzing the characteristics of the nodules based on shape, grayscale, texture, etc. Manually extracting features is time-consuming and labor-intensive [4]. With the rapid development of machine learning, deep learning algorithms represented by convolutional neural networks have been widely adopted. One approach is the two-stage detection algorithm based on candidate regions, such as RCNN; another is the one-stage detection algorithm, such as the YOLO series [5,6]. One method uses the YOLO detection framework to detect CT nodules and predicts multiple bounding boxes using a Convolutional Neural Network (CNN), followed by evaluation on the LIDC-IDRI dataset (Lung Image Database Consortium), achieving a sensitivity of 89%. Currently,

CNN, YOLO, FCN, R-CNN [7–9], and other networks are often used for training on different datasets to extract features from target datasets.

Transfer learning methods used at present can achieve relatively good results, but due to the large domain gap between the source and target datasets, there is still room for improvement in feature extraction. Therefore, this paper proposes an improved deep neural network model based on GoogLeNet Inception V3 to recognize pulmonary nodules.

The main work of this paper is to design a feature fusion layer and optimize related parameters to better extract features from the target dataset.

2. Related Work

Recent advancements in deep learning have significantly enhanced medical image analysis. Works that emphasize the integration of clinical and imaging data to enhance diagnostic performance. For instance, joint modeling of medical images and clinical text facilitates early detection of chronic conditions like diabetes [10]. Similarly, leveraging Transformer-CNN fusion for skin lesion segmentation [11] and temporal graph attention for EEG anomaly detection [12] shows the potential of deep neural architectures in specialized diagnostic contexts.

Research has also emphasized predictive modeling in clinical settings, particularly for chronic disease progression and health risk assessment. Structure-aware temporal models offer accurate progression prediction [13], while federated frameworks are enabling privacy-preserving risk analysis on electronic health records [14]. Furthermore, the integration of deep learning and natural language processing (NLP) for structuring electronic medical records enhances clinical workflow efficiency [15].

In the broader context of cloud-based AI applications, reinforcement learning plays a vital role in dynamic resource management and anomaly detection. Multi-agent reinforcement learning strategies [16] and temporal anomaly detection models [17] have demonstrated robust scalability. Personalized modeling for federated anomaly detection [18] and collaborative data mining across distributed systems [19] further contribute to intelligent cloud system design. Moreover, cross-domain transfer learning techniques [20] and privacy-aware model fine-tuning mechanisms [21] highlight efforts to bridge domain gaps and ensure model generalizability and security.

Concurrently, the integration of large language models (LLMs) with retrieval-augmented mechanisms and knowledge graphs has gained momentum. Innovations in question answering systems [22], adapter-based knowledge injection [23], and knowledge graph-infused reasoning models [24] exemplify this trend. Additionally, uncertainty-aware summarization techniques [25], prompt engineering for controllable abstraction [26], and contrastive learning for secure alignment [27] underscore the growing intersection of language modeling and deep learning robustness.

In terms of enterprise and infrastructure intelligence, anomaly detection in ETL workflows using autoencoders [28] and knowledge-driven threat identification frameworks in IoT environments [29] expand the applicability of deep learning techniques. Reinforcement learning has also been harnessed for optimizing microservice orchestration, including scheduling [30], adaptive HCI strategies [31], graph-based service routing [32], and rate limiting in distributed systems [33].

These diverse yet interconnected works collectively inform the development of robust pulmonary image recognition systems. Methodologies such as transfer learning, federated modeling, multimodal fusion, and reinforcement optimization are especially relevant for enhancing feature extraction and generalization across heterogeneous datasets like LUNA16, as adopted in our proposed RGIV3 model.

3. Dataset

This paper uses the LUNA16 dataset. As a publicly available lung nodule dataset and a subset of LIDC-IDRI [34], LUNA16 includes CT images from 888 patients, divided into 10 subsets for storage. Each image set consists of mhd and raw files, where the mhd file provides basic image information, and the raw file stores pixel data. An example CT image is shown in Figure 1.

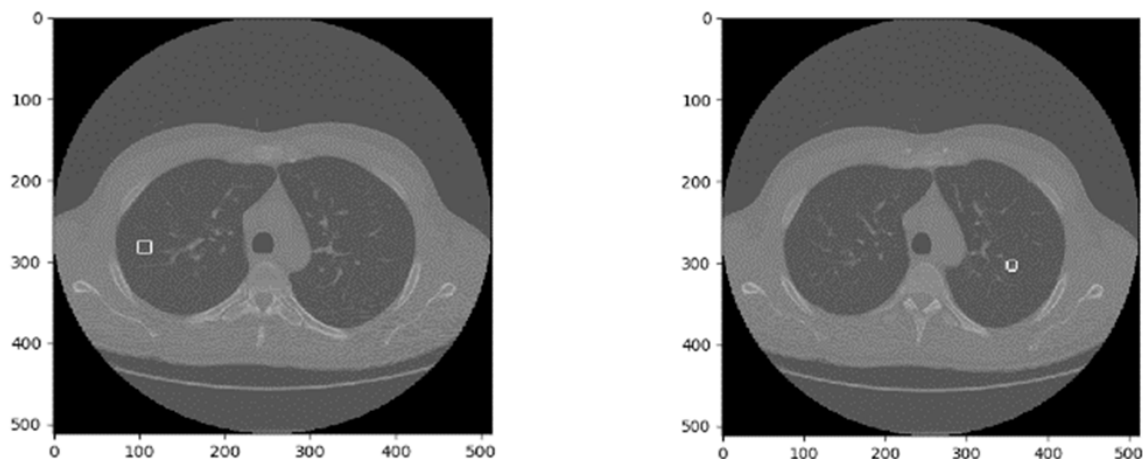


Figure 1. LUNA16 image examples.

4. Improved Pulmonary Nodule Recognition Model

4.1. GooLeNet Network Algorithm

In 2012, the GooLeNet network was proposed [35]. Compared with networks such as AlexNet and VGG, GooLeNet features a deeper network structure and fewer parameters (see Table 1). By adjusting the network structure (see Figure 2) and designing Inception layers, GooLeNet significantly reduces the number of parameters, thereby improving training efficiency.

Table 1. Comparison of Common Neural Network Parameters.

Model	TOP-1 Accuracy (%)	Parameters	Depth
Inception V3	78.8	23,851,784	159
ResNet50	75.9	25,636,712	168
VGG16	71.5	138,357,544	23
VGG19	72.7	143,667,240	26

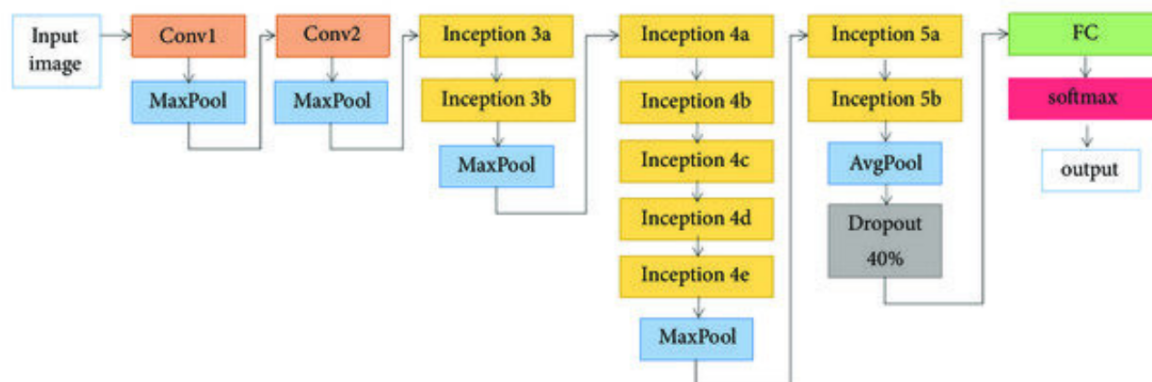


Figure 2. GooLeNet network architecture.

4.2. Improved RGIV3 Network Structure Design

GooLeNet Inception V3 [36,37] is a model trained on the ImageNet dataset. However, due to the large domain gap between ImageNet and the LUNA16 dataset used in this paper, directly applying it yields poor results. Therefore, a fusion layer is designed to reduce the negative impact of the domain gap between the two datasets. This fusion layer combines feature information learned by the network in a non-linear way, improving the learning of image features.

The fusion layer is composed of three fully connected layers and one dropout layer, as shown in Figure 3. This layer is then combined with the GooLeNet Inception V3 network, where the original GooLeNet Inception V3 is referred to as GIV3, and the combined network is named RGIV3.

According to experience, the number of neurons is chosen from 1024, 512, 256, and 128, and the dropout rate is selected from 0.4, 0.5, and 0.6. To determine the optimal configuration, this paper conducts experiments on different combinations of neuron numbers and dropout rates—a total of 30 combinations—using accuracy as the selection criterion. The experimental results are shown in Table 2. Based on the results, the selected configuration is 1024 and 512 neurons with a dropout rate of 0.5.

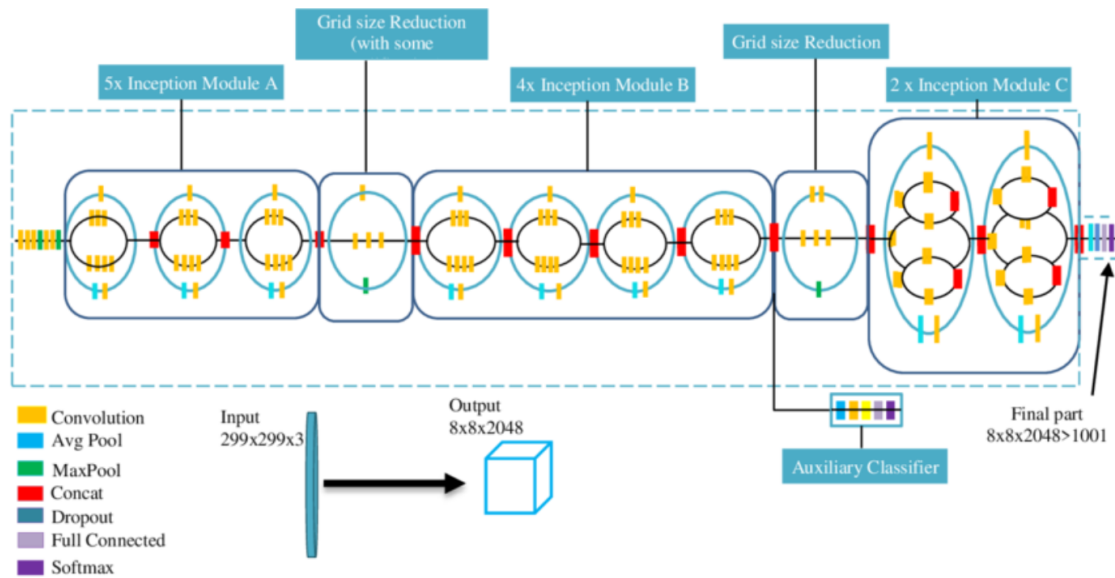


Figure 3. Architecture of the improved RGIV3 network.

Table 2. Experimental Results of Neuron Number and Dropout Rate Combinations.

Neuron Combination	Dropout = 0.3	Dropout = 0.4	Dropout = 0.5
1024, 1024	86.35	87.66	88.64
1024, 512	87.71	88.46	88.80
1024, 256	86.56	87.61	88.58
1024, 128	87.28	86.83	88.34
512, 512	86.28	84.43	87.34
512, 256	86.28	87.15	88.12
512, 128	86.78	87.12	86.85
256, 256	81.43	87.48	88.23
256, 128	83.05	84.53	87.28
128, 128	86.28	84.45	87.24

Due to the limited number of samples in the target dataset, newly trained models are prone to overfitting. In the experiment, pretrained weights from large-scale datasets such as ImageNet were used as initialization parameters, and fine-tuning was conducted with a relatively low learning rate.

4.3. Network Weight Update Method

The Stochastic Gradient Descent (SGD) method is used to update the weights by minimizing the loss function along the direction of the negative gradient of the learning rate [38]. The update formula is as follows:

$$W_{i+1} = W_i - \Delta h_{is_{i+1}} \quad (1)$$

In this formula, W_{i+1} represents the weight after the $(i + 1)$ -th iteration; W_i is the weight after the i -th iteration; Δhis_{i+1} is the accumulated gradient. The formula is:

$$\Delta his_{i+1} = momentum \times \Delta his_i + L \times \frac{\partial loss}{\partial w} \quad (2)$$

Here, Δhis_{i+1} denotes the accumulated gradient after the $(i + 1)$ -th iteration; Δhis_i is the accumulated gradient after the i -th iteration; *momentum* is the momentum coefficient; L represents the current learning rate; $\frac{\partial loss}{\partial w}$ is the gradient of the loss function.

5. Experiments and Results Analysis

5.1. Dataset Preprocessing

The original LUNA16 data consists of image data (.raw) and annotation data (.mhd). Since neural networks cannot directly read this image format, the mhd-format images are converted into the commonly used BMP format. Through this conversion, each CT slice corresponds to a BMP image with the same slice count as RGB images, and each image displays the full CT scan content. From the annotation files, the position and properties of the nodules can be obtained. With the position and slice depth, the exact coordinates and slices can be quickly located [38]. When patients undergo CT scans, different body postures may cause image flipping, so the images need to be adjusted based on annotation data to align properly, and the coordinate annotations also need to be corrected.

Due to the small dataset size, this study expands the dataset through data augmentation. Supervised data augmentation techniques are used to enhance the BMP-format images, and a low learning rate is adopted to prevent overfitting.

Pulmonary nodules typically appear at the ends of bronchial and vascular branches. In CT images, these are the black circular regions representing air in the lungs. This study aims to identify pulmonary nodules, so it is necessary to segment the actual lung region and exclude the outer lung fields, which may introduce noise into the network [39].

Figure 4 shows a slice from the LUNA16 dataset. After binarization, corrosion, and dilation operations, the lung parenchyma and outer lung field can be separated for further processing [40].

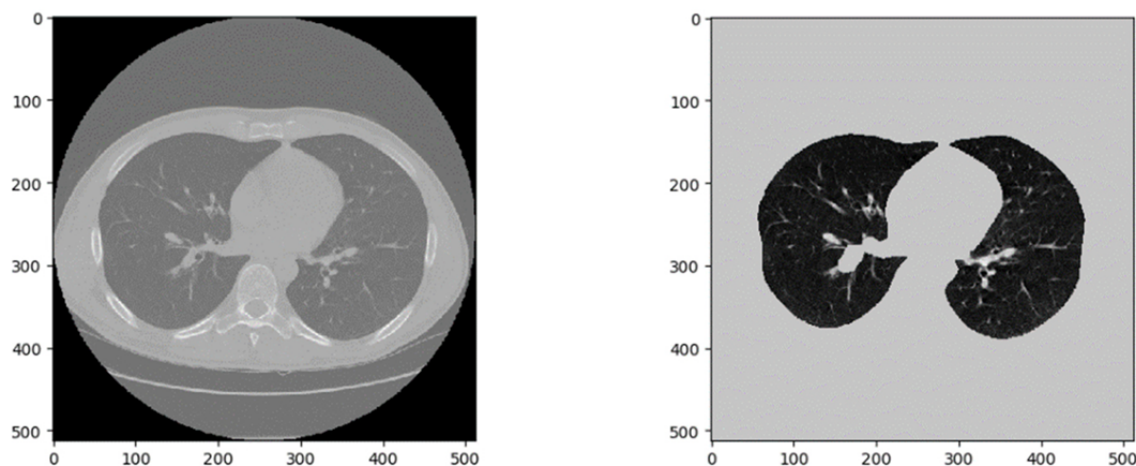


Figure 4. Original image and processed lung parenchyma image.

5.2. Experimental Procedure

The experiments were conducted under Windows 10 system using an Intel Core i5-6200U @ 2.30GHz CPU. TensorFlow framework and the slim module were adopted, with Python version 3.6.

In the experimental process, 750 patients' CT data were used for training, and 150 patients' data were used for testing. A total of 2000 pulmonary nodule images and 2000 healthy tissue images were extracted. After data augmentation, the dataset expanded to 5000 nodule images and 5000 healthy images. All images were resized to 330×330 . The data were split as follows: 80% training, 20% testing;

70% training, 30% testing; and 60% training, 40% testing. Each group was trained with the same loss function for fair comparison. Among them, the 80% group was trained using the slim model for 2776 steps, reducing the loss to 0.38 and achieving an accuracy of 88.8%. Other group results are shown in Table 4.

In the experiment, the designed RGIV3 network was trained on the training set and validated on the test set. To avoid falling into local minima, a gradually decreasing learning rate was used, reducing the rate every 200 iterations.

5.3. Experimental Results and Analysis

According to the experimental results, True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) were used to evaluate model performance [14-16]. Three evaluation metrics were selected: Accuracy (Acc), Sensitivity (Sen), and Specificity (Spe), calculated as follows:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

$$Sen = \frac{TP}{TP + FN} \quad (4)$$

$$Spe = \frac{TN}{TN + FP} \quad (5)$$

Accuracy reflects the consistency between predictions and ground truth. Sensitivity reflects the model's ability to detect nodules, while specificity indicates the model's ability to distinguish non-nodule images. The training processes of GIV3 and RGIV3 are shown in Figure 5. According to Figure 5, RGIV3 has a faster convergence speed and more stable performance. Combined with the accuracy metric, the improved RGIV3 model achieves better results on the dataset.

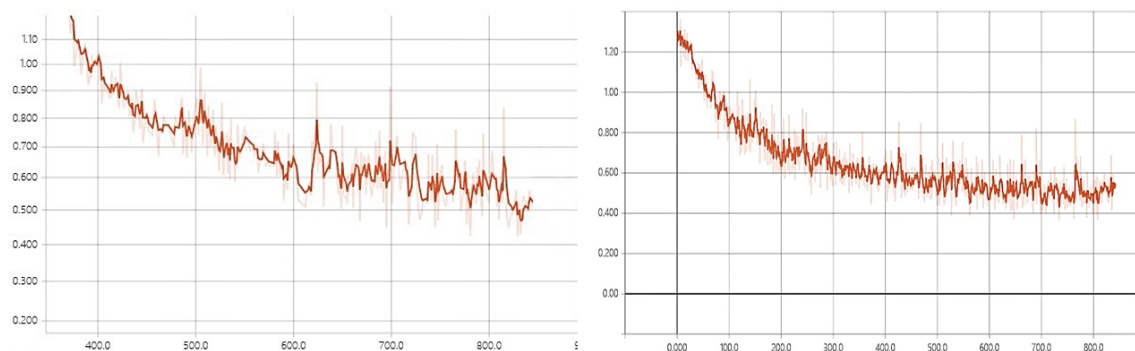


Figure 5. Training process comparison: (a) GIV3, (b) RGIV3.

When using 80% of the data as the training set and 20% as the test set, the performance comparison of different models is shown in Table 3.

Table 3. Performance Comparison of Models (%).

Experiment No.	Recognition Model	Accuracy	Sensitivity	Specificity
1	AlexNet	85.37	87.70	73.0
2	YOLO V3	84.07	88.72	82.3
3	GIV3	86.08	84.96	80.6
4	RGIV3	88.80	87.15	80.6

As shown in Table 3, the improved model achieves an accuracy of 88.80% and a sensitivity of 87.15%, which are 2.72% and 2.19% higher than those of the traditional GIV model, respectively. This indicates a relatively better recognition performance.

To comprehensively evaluate the performance of the improved RGIV3 model, experiments were conducted under different ratios of training and testing sets. The results are shown in Table 4.

Table 4. Performance Comparison Under Different Data Ratios.

Training/Test Set Ratio	Accuracy %	
	GIV3	RGIV3
8-2	86.08	88.80
7-3	84.21	87.58
6-4	84.85	87.23

According to Table 4 and the training process analysis, the improved RGIV3 model outperforms the original GIV model in terms of recognition performance. Under the 80%, 70%, and 60% data ratios, RGIV3 consistently achieves better recognition results than GIV3, with faster convergence and more stable performance.

6. Conclusion

This paper adopts an improved neural network model to detect pulmonary nodules in part of the LUNA16 dataset. By combining the GIV3 network with the designed feature fusion layer, the network's feature extraction capability is enhanced, which to some extent alleviates the poor transfer learning performance caused by large domain differences between source and target datasets.

In comparative experiments, the proposed model demonstrates better recognition performance compared to other existing models. Due to the relatively small scale of the dataset and the existence of overfitting, data augmentation and fine-tuning using low learning rates were employed, which effectively mitigated the problem.

In future research, computational features can be associated with semantic features to provide quantitative evaluation for pulmonary nodules and offer diagnostic references for clinicians.

References

1. S. H. Liang, M. Hu, Y. Ma, L. Yang, J. Chen, L. Lou, C. Chen and Y. Xiao, "Performance of deep-learning solutions on lung nodule malignancy classification: a systematic review," *Life*, vol. 13, no. 9, article 1911, 2023.
2. I. Marinakis, K. Karampidis and G. Papadourakis, "Pulmonary nodule detection, segmentation and classification using deep learning: a comprehensive literature review," *BioMedInformatics*, vol. 4, no. 3, pp. 2043–2106, 2024.
3. A. Mahajan, U. Agarwal, R. Agrawal, A. Venkatesh, S. Shukla, K. S. S. Bharadwaj, M. L. V. Apparao, V. Pawar and V. Poonia, "Deep learning-based pulmonary nodule screening: a narrative review," *[Journal Name]*, online ahead of print, 2024.
4. "3D deep learning for detecting pulmonary nodules in CT scans," *PMC*, vol. ..., 2018.
5. "Deep learning in pulmonary nodule detection and segmentation," *[Journal]*, 2023.
6. "Deep learning for the detection of benign and malignant pulmonary nodules," *Nature Communications*, 2023.
7. G. Francis and J. P. Ko, "Pulmonary nodules: detection, assessment, and CAD," *American Journal of Roentgenology*, vol. 191, no. 4, pp. 1057–1069, 2008.
8. D. Liu, S. Zhu, B. Liu, et al., "Improvement of CT target scanning quality for pulmonary nodules by PDCA management method," *Mathematical Problems in Engineering*, vol. 2021, no. 6, pp. 1–9, 2021.
9. G. Litjens, T. Kooi, B. Ehteshami Bejnordi, A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken and C. I. Sánchez, "A survey on deep learning in medical image analysis," arXiv:1702.05747, 2017.
10. Y. Zi and X. Deng, "Joint modeling of medical images and clinical text for early diabetes risk detection," *Journal of Computer Technology and Software*, vol. 4, no. 7, 2025.

11. X. Wang, X. Zhang and X. Wang, "Deep skin lesion segmentation with Transformer-CNN fusion: toward intelligent skin cancer analysis," arXiv:2508.14509, 2025.
12. X. Zhang and Q. Wang, "EEG anomaly detection using temporal graph attention for clinical applications," *Journal of Computer Technology and Software*, vol. 4, no. 7, 2025.
13. J. Hu, B. Zhang, T. Xu, H. Yang and M. Gao, "Structure-aware temporal modeling for chronic disease progression prediction," arXiv:2508.14942, 2025.
14. R. Hao, W. C. Chang, J. Hu and M. Gao, "Federated learning-driven health risk prediction on electronic health records under privacy constraints," 2025.
15. N. Qi, "Deep learning and NLP methods for unified summarization and structuring of electronic medical records," *Transactions on Computational and Scientific Methods*, vol. 4, no. 3, 2024.
16. B. Fang and D. Gao, "Collaborative multi-agent reinforcement learning approach for elastic cloud resource scaling," arXiv:2507.00550, 2025.
17. L. Lian, Y. Li, S. Han, R. Meng, S. Wang and M. Wang, "Artificial intelligence-based multiscale temporal modeling for anomaly detection in cloud services," arXiv:2508.14503, 2025.
18. Y. Wang, H. Liu, N. Long and G. Yao, "Federated anomaly detection for multi-tenant cloud platforms with personalized modeling," arXiv:2508.10255, 2025.
19. M. Wang, T. Kang, L. Dai, H. Yang, J. Du and C. Liu, "Scalable multi-party collaborative data mining based on federated learning," 2025.
20. Y. Zhou, "Self-supervised transfer learning with shared encoders for cross-domain cloud optimization," 2025.
21. Y. Li, "Task-aware differential privacy and modular structural perturbation for secure fine-tuning of large language models," *Transactions on Computational and Scientific Methods*, vol. 4, no. 7, 2024.
22. Y. Sun, R. Zhang, R. Meng, L. Lian, H. Wang and X. Quan, "Fusion-based retrieval-augmented generation for complex question answering with LLMs," in *Proc. CISAT*, pp. 116–120, July 2025.
23. H. Zheng, L. Zhu, W. Cui, R. Pan, X. Yan and Y. Xing, "Selective knowledge injection via adapter modules in large-scale language models," 2025.
24. W. Zhang, Y. Tian, X. Meng, M. Wang and J. Du, "Knowledge graph-infused fine-tuning for structured reasoning in large language models," arXiv:2508.14427, 2025.
25. S. Pan and D. Wu, "Trustworthy summarization via uncertainty quantification and risk awareness in large language models," 2025.
26. X. Song, Y. Liu, Y. Luan, J. Guo and X. Guo, "Controllable abstraction in summary generation for large language models via prompt engineering," arXiv:2510.15436, 2025.
27. J. Zheng, H. Zhang, X. Yan, R. Hao and C. Peng, "Contrastive knowledge transfer and robust optimization for secure alignment of large language models," arXiv:2510.27077, 2025.
28. X. Chen, S. U. Gadgil, K. Gao, Y. Hu and C. Nie, "Deep learning approach to anomaly detection in enterprise ETL processes with autoencoders," arXiv:2511.00462, 2025.
29. L. Yan, Q. Wang and C. Liu, "Semantic knowledge graph framework for intelligent threat identification in IoT," 2025.
30. R. Zhang, "AI-driven multi-agent scheduling and service quality optimization in microservice systems," *Transactions on Computational and Scientific Methods*, vol. 5, no. 8, 2025.
31. R. Liu, Y. Zhuang and R. Zhang, "Adaptive human-computer interaction strategies through reinforcement learning in complex environments," arXiv:2510.27058, 2025.
32. C. Hu, Z. Cheng, D. Wu, Y. Wang, F. Liu and Z. Qiu, "Structural generalization for microservice routing using graph neural networks," arXiv:2510.15210, 2025.
33. N. Lyu, Y. Wang, Z. Cheng, Q. Zhang and F. Chen, "Multi-objective adaptive rate limiting in microservices using deep reinforcement learning," arXiv:2511.03279, 2025.
34. R. Siddiqi, "Deep Learning for Pneumonia Detection in Chest X-ray Images," *Images*, vol. 10, no. 8, p. 176, 2024.
35. C. Szegedy, W. Liu, Y. Jia, *et al.*, "Going deeper with convolutions," in *IEEE Computer Society*, pp. 1–9, 2015.
36. Y. Ohno, M. Nishio, H. Koyama, *et al.*, "Dynamic contrast-enhanced CT and MRI for pulmonary nodule assessment," *American Journal of Roentgenology*, vol. 202, no. 3, pp. 515–529, 2014.
37. M. Souto, P. G. Tahoces, J. J. Suarez Cuenca, *et al.*, "Automatic detection of pulmonary nodules on computed tomography: a preliminary study," *Radiologia*, vol. 50, no. 5, pp. 387–392, 2008.

38. C. Zhang, J. Li, J. Huang, *et al.*, "Computed tomography image under convolutional neural network deep learning algorithm in pulmonary nodule detection and lung function examination," *Journal of Healthcare Engineering*, Article ID: 3417285, 2021.
39. C. H. Lee and J. S. Jwo, "Automatic segmentation for pulmonary nodules in CT images based on multifractal analysis," *IET Image Processing*, vol. 14, no. 7, pp. 1347–1353, 2020.
40. E. Itouji, M. Kono, S. Adachi, *et al.*, "The role of CT and MR imaging in the diagnosis of lung cancer," *Gan to Kagaku Ryoho Cancer & Chemotherapy*, vol. 24, suppl. 3, pp. 353–358, 1997.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.