Progress of Deep Learning-Based Pulmonary Disease Identification from Chest X-Rays

Jiale Zhou

School of Electrical Engineering and Computer Science, The University of Queensland, Brisbane, 4072, Australia
jiale.zhou@student.uq.edu.au

Abstract. With the increasing prevalence of epidemic respiratory diseases and lung cancer in recent years, intelligent diagnostic technologies based on chest X-ray (CXR) image recognition have become a research focus. The breakthrough development of deep learning has provided a new paradigm for automatic CXR image analysis. This paper systematically reviews the evolution and limitations of convolutional neural networks (CNNs) and their derivative architectures in pulmonary disease detection over the past 5-10 years, comparing their performance and exploring advancement in applications for diseases like pneumonia. The study examines the enormous potential of emerging technologies, such as Transformer models and multimodal learning, in transforming the automatic diagnosis of pulmonary diseases from CXR images. Finally, this paper introduces commonly used CXR image datasets, focusing on the challenges faced at the data level and analyzing corresponding solution strategies. The review comprehensively analyzes the current applications, challenges, and future development paths of deep learning technologies and datasets applied to pulmonary disease recognition in CXR imaging, offering valuable insights for developing more accurate, rapid, portable, and interpretable AI-powered diagnostic systems for pulmonary diseases, potentially improving early patient outcomes globally.

Keywords: Deep Learning, Chest X-Ray, Medical Image Recognition

1 Introduction

Pulmonary diseases have become one of the major challenges in global public health, with their high prevalence, diverse aetiology, and harmful effects imposing a heavy medical burden on countries worldwide. Diseases such as pneumonia and lung cancer continue to threaten human health, whilst the sudden global pandemic in 2019 further intensified the strain on medical resource allocation. The large-scale spread of COVID-19 in 2020 elevated it to the third leading cause of death in the United States [1]. By the end of 2024, confirmed cases had appeared in no fewer than 200 countries and regions, with cumulative cases exceeding 400 million [2]. These phenomena indicate the urgent global need for efficient pulmonary disease identification methods to

J. Iqbal (ed.), *Proceedings of the 2025 2nd International Conference on Mechanics, Electronics Engineering and Automation (ICMEEA 2025)*, Advances in Engineering Research 272,
https://doi.org/10.2991/978-94-6463-821-9_56

enhance public health emergency response capabilities and reduce disease pressure on both people and medical resources.

CXR imaging, as an efficient and economical examination method, can help physicians observe multiple vital organs in the human body, including the lungs. This examination method has become a key diagnostic tool for detecting lesions and evaluating treatment effectiveness in clinical practice due to its advantages of reasonable cost, non-invasiveness, and painlessness for patients. In current medical practice, image diagnosis primarily relies on visual interpretation by physicians, a traditional method facing severe challenges. On one hand, the annual growth rate of medical imaging data significantly exceeds the expansion rate of radiologist teams [3]. On the other hand, the quality of imaging diagnosis is often affected by subjective factors such as visual fatigue and individual experience levels, making it difficult to reduce the risk of missed or incorrect diagnoses and to quantify disease severity [3]. This experience-driven model is both time-consuming and labour-intensive, struggling to meet the current growing demand for rapid diagnosis of pulmonary diseases.

With the rapid development of deep learning technology, intelligent diagnostic systems based on CXR images have provided new possibilities for addressing the challenges in medical practice. Deep learning methods centred on convolutional neural networks, through automatic analysis and feature learning of large-scale medical imaging data, can rapidly identify potential pathological features, providing objective quantitative references for clinical decision-making, significantly improving diagnostic efficiency, and reducing human error. This paper analyses the current applications, technological evolution paths, and limitations of convolutional neural networks in this field and explores improvement methods of other derivative models in addressing difficulties such as insufficient generalisation ability and computational efficiency. The paper then further summarises the application prospects of Transformer models, multimodal learning, and large language models in improving automatic diagnostic accuracy and system performance. Last, the review enumerates commonly used training datasets and current data issues, discusses preprocessing strategies for CXR images and solutions for medical dataset defects, and concludes with prospects to promote progress and breakthroughs in the field of automatic medical image diagnosis.

2 Applications of CNN and Their Derivative Architectures in CXR Image Recognition

As an important branch of deep neural networks, CNNs possess unique advantages in medical image processing. They can receive medical image data, such as chest X-rays, and automatically detect and extract significant features from various regions of the image through the configuration of diverse convolutional kernels. The standard CNN model building process is shown in Figure 1. Academic research on CNN framework improvements primarily focuses on the following directions.

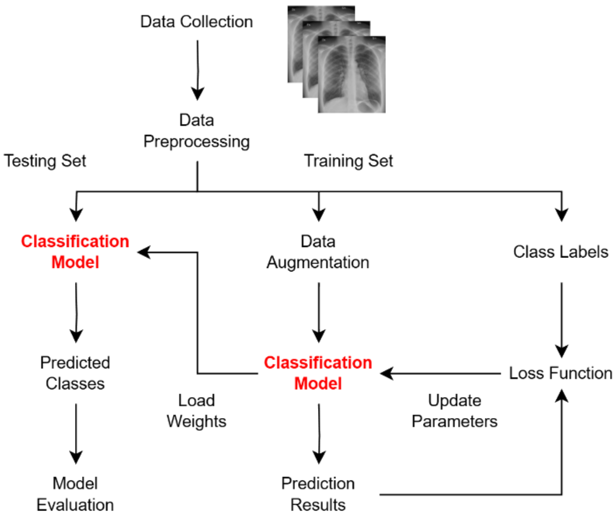

Fig. 1. CNN model building process [3].

2.1 CNN-based Improvement Methods

Improvements to traditional CNN models primarily focus on network architecture optimization and integration. Researchers enhance network performance by adjusting convolutional kernels and network depth, introducing skip connection structures, and implementing ensemble learning. Although these strategies increase model complexity, they have achieved remarkable results in medical image feature recognition tasks, significantly improving inadequate recognition rates and addressing gradient vanishing problems.

Alshmran et al. proposed an improved deep learning architecture combining VGG19 and CNN for multi-class chest disease diagnosis (five-class classification). This model uses a pre-trained VGG19 as the base architecture with three additional CNN modules for feature extraction, each containing convolutional layers with ReLU activation functions, batch normalization layers, max pooling layers, and dropout layers, receiving 224×224×3 CXR images as input data [4]. In the classification phase, features are transformed into one-dimensional vectors through a flattening layer, then processed by three fully connected layers (containing 512, 256, and 128 neurons respectively) and output through a SoftMax layer with 6 neurons, ultimately achieving 96.48% accuracy, 99.82% AUC value, and 93.75% recall [4], demonstrating the enormous potential of redesigning network structure depth and adjusting feature channel quantities to improve CNN model accuracy for chest X-ray image recognition.

To address the difficulties of traditional networks in capturing global context and information loss, Liu et al. introduced a Class Residual Attention (CRA) module and proposed a multi-scale feature fusion framework [5], adding a multi-scale convolutional network to ResNet's convolutional layers and laterally connecting it with top-down feature pyramid layers to achieve feature fusion, while utilizing 1×1 convolutional layers and SoftMax normalization layers to construct attention modules that assign weight scores (range 0-1) to feature spaces of different categories, achieving enhancement of lesion area features and suppression of irrelevant features [5]. The introduction of the CRA mechanism combined with ResNet increased COVID-19 detection accuracy from 96.87% to 97.25% (an improvement of 0.38%); with the further introduction of the multi-scale feature fusion module, accuracy improvement could reach 0.84%; successfully preserving the semantic information of COVID-19 minute pathological features in deep networks and alleviating the problem of information loss for small-sized lesions caused by convolutional behaviour [5].

Compared to single CNN models, ensemble learning methods integrate the advantages of multiple independent networks through multi-model fusion strategies, significantly improving model generalization ability, achieving more comprehensive image feature representation, and thereby enhancing classification performance [6]. Kumar et al. proposed an Enhanced Multi-model Deep Learning method (EMDL), combining five advanced deep learning models (VGG16, VGG19, ResNet, AlexNet, and GoogleNet), processing images using histogram equalization and contrast enhancement techniques, and constructing a multi-stage pipeline based on PCA, SelectKBest, Binary Particle Swarm Optimization (BPSO), and Binary Grey Wolf Optimizer (BGWO) for feature selection to reduce computation time while retaining key information, ultimately achieving improved diagnostic accuracy and robust training on two independent datasets [7]. Compared to the optimal base learner VGG19 with 94.23% accuracy, the EMDL method achieved 98.88% peak accuracy; and compared to high-precision models based on Transformer architecture (Swin Transformer, Vision Transformer), the EMDL method significantly reduced computation time (65 seconds) while maintaining similar accuracy [7].

Structural depth improvements, connection structures, feature representation and fusion, and multi-model ensemble methods based on traditional CNNs have significantly enhanced the performance of automated diagnosis of chest diseases. However, complex CNN-based models still face problems such as poor interpretability, gradient explosion on large-scale datasets, and high computational resource requirements, necessitating appropriate image preprocessing and limiting their range of use in medical diagnosis.

2.2 Lightweight CNN Models

MobileNet and EfficientNet, through innovative architectural designs (such as depthwise separable convolutions and compound scaling), significantly outperform traditional CNNs in terms of computational efficiency, model compactness, and real-time performance. They are particularly suitable for operation on resource-constrained edge computing devices while maintaining high accuracy and are widely applied in mobile devices, robotics, and Internet of Things domains.

MobileNet (V1 version) is a lightweight architecture specifically designed for mobile and embedded vision applications, constructed based on bottleneck residual blocks and employing depthwise separable convolutions to replace traditional convolution operations [8]. The core concept of MobileNet is to avoid using conventional 3×3 convolution filters, instead dividing the operation into depthwise separable 3×3 convolution filters followed by 1×1 pointwise convolutions, thereby reducing the number of operations and parameters while achieving the same filtering and combination processes as regular convolutions [8]. Souid et al. proposed an improved MobileNet V2 model based on the Keras framework to classify and predict pulmonary pathologies using the ChestX-Ray14 dataset (14 classes of chest abnormalities) comprising approximately 112,000 CXR images. Compared to the V1 version, V2 incorporates inverted residual structures (high-dimensional features mapped to low dimensions through depthwise separable convolutions, then restored to high dimensions after 1×1 convolutions with linear activation) and residual connections to minimize information loss and facilitate gradient flow [8]. The model achieved AUC values exceeding 85% for five categories, including emphysema, effusion, edema, pneumothorax, and cardiomegaly (0.891%, 0.876%, 0.884%, 0.880%, 0.885%), while maintaining a small size and low latency [8].

Compared to MobileNet, EfficientNet places greater emphasis on the overall balance between performance and efficiency, pursuing higher accuracy under the same computational cost. EfficientNet avoids the limitations of traditional CNN's single-dimension scaling by using compound scaling to simultaneously optimize three dimensions: network depth, width, and resolution; and compared to MobileNet's inverted residual structure, EfficientNet employs improved Mobile inverted Bottleneck Convolution (MBConv) blocks to maintain flexible adaptation to tasks of different scales [9]. Diallo et al. utilized an EfficientNet model extended with progressive scaling (K-EfficientNet) and applied transfer learning on ImageNet, ultimately achieving 100% and 89% PPV (Positive Predictive Value) for COVID-19 and pneumonia cases respectively, with an average accuracy of 97.3%, surpassing the accuracy of complex models such as VGG19 while being 8.4 times smaller and 6.1 times faster than ConvNet [9].

The models, as representatives of lightweight CNNs, are highly suitable for automated diagnostic scenarios with resource constraints, enabling deployment on edge devices. MobileNet places greater emphasis on low latency suitable for real-time inference, but its extremely lightweight design may sacrifice some feature expression capability; EfficientNet focuses on balancing accuracy and efficiency but increases model structural complexity. Lightweight models require a balance between local lesion feature extraction capability and computational overhead; if training data is insufficient or lesions are subtle, performance may be affected due to limited model capacity, necessitating supplementation with transfer learning or attention mechanisms.

2.3 Construction, Preprocessing, and Feature Extraction of CXR Data

High-quality medical image datasets constitute the cornerstone for deep learning model development, with large-scale standardized data being one of the key factors enabling these models to reach or even surpass specialist-level performance in specific tasks. Despite significant advances in automated detection across various diseases, medical data still faces challenges in constructing high-quality, high-volume, diverse, and class-balanced datasets. The preprocessing and feature extraction stages similarly have decisive impacts on overall model performance. These include noise and artifact removal, exclusion of irrelevant anatomical structures such as the spine and heart, and identification and extraction of multi-scale lesion features to optimize model performance.

2.3.1 Commonly used CXR datasets, annotation, and augmentation.

Ideal training datasets should encompass various disease manifestations, undergo review by professional physicians, and maintain balanced class samples. Currently, automated precise annotation and data augmentation of large-scale medical data represent focal areas of research in constructing high-quality datasets. Jaeger et al. proposed two datasets suitable for automatic tuberculosis diagnosis: the Montgomery County dataset from the Department of Health and Human Services of Montgomery County, Maryland, USA, containing 138 posteroanterior chest radiographs (80 normal, 58 abnormal with tuberculosis), in PNG format with 12-bit grayscale and 4892×4020 resolution, supervised by radiologists who annotated lung segmentation labels and left-right lung masks following anatomical landmarks and the Shenzhen dataset from Shenzhen No. 3 People's Hospital comprising 662 posteroanterior chest radiographs (326 normal, 336 abnormal with tuberculosis) with 3000×3000 resolution but only simple annotations [10]. Although both are small in scale and focus solely on tuberculosis, they are highly suitable for validating model details and serving as performance benchmarks. Wang et al. introduced ChestX-Ray8, a novel CXR image database containing 108,948 frontal-view X-ray images (1024×1024) from 32,717 patients, automatically extracting eight disease labels from radiology reports using natural language tools (DNorm and MetaMap) and employing weakly-supervised multi-label classification methods to localize lesions, successfully achieving automated efficient annotation of large-scale medical data [11]. ChestX-Ray14 expanded the pulmonary disease labels to 14 categories while maintaining a similar scale (112,120 images) and resolution to ChestX-Ray8. The COVID-ChestXray-15k and CoronaHack databases focus on CXR images of COVID-19 cases, pneumonia, and normal images, including 15,000 and 5,933 relatively balanced samples across 3 and 4 categories, providing models with more precise data understanding of COVID-19 [12].

Sufficient data can directly enhance model performance; however, medical images are costly, require manual annotation by professional radiologists, and involve privacy concerns such as data leakage, making large-scale acquisition challenging. Chen et al. summarized two technical approaches for medical image dataset augmentation: one based on geometric transformations of pixel grayscale or lesion region deformation, which offers limited improvement in data diversity; the other leveraging Generative Adversarial Networks (GANs) to produce synthetic images with good differentiation based on the pathological feature distribution of original images, yielding considerable accuracy improvements for models trained on such datasets [13]. Among existing technical pathways, small-scale fine-grained annotation is constrained by disease singularity and data volume, while large-scale weakly supervised annotation, though complementary, may introduce noisy labels that undermine model credibility. Regarding data augmentation, GAN-based generation methods can significantly enrich image diversity but require attention to synthetic artifact risks. Manual correction and data balancing remain crucial for model pre-training and subsequent transfer learning.

2.3.2 Preprocessing and feature extraction methods for CXR data.

Traditional digital image processing employs conversion operations, cropping, exposure adjustment, and contrast enhancement to systematically modify the fundamental elements of CXR images, improving overall image quality and highlighting key information features. Ganesan et al. utilized a pre-trained VGG16 model to test the differences between no augmentation (NA) and traditional augmentation (TA); performing real-time TA on each mini-batch, including Gaussian smoothing to reduce random noise, sharpening masks to enhance high-frequency components and strengthen edge features, and minimum filtering to find minimum pixel values within local regions to remove positive outlier noise [14]. This ultimately achieved a 4.93% accuracy improvement (79.32% vs. 84.25%), a 4.67% AUC value increase (86.03% vs. 90.7%), and a 0.09 improvement in MCC (Matthews Correlation Coefficient) (0.55 vs. 0.64) after TA was applied to all samples compared to NA [14]. Deep learning-based pre-feature extraction methods can identify and extract the lung field portions in CXR images, attenuating and eliminating other irrelevant anatomical structures, effectively reducing various interfering elements and highlighting core lesion areas. Maity et al. proposed a Deep Convolutional Neural Network (DCNN) model that employs end-to-end training to achieve fully automated lung region segmentation for anteroposterior (AP, X-rays projected from the patient's anterior chest toward the posterior back) or posteroanterior (PA, X-rays projected from the patient's posterior back toward the anterior chest) CXR images [15]. By learning to differentiate between parenchymal and non-parenchymal regions, it predicts binary masks for CXR images and adopts techniques such as Top-Hat Bottom-Hat Transform and Contrast Limited Adaptive Histogram Equalization (CLAHE), achieving Dice similarity coefficients of 0.982±0.018 and Jaccard similarity coefficients of 0.967±0.015 on the JSRT, NLM-MC, and Shenzhen datasets while successfully operating in a lightweight manner on standard GPUs [15].

Traditional preprocessing/enhancement methods rely on computer graphics, requiring no complex algorithms while maintaining strong explainability and processing efficiency, highly applicable in medical practice, but necessitating certain expertise and unable to address complex or marginal cases. Deep learning-based feature extraction significantly enhances lesion recognition capabilities under complex anatomical interference through automated segmentation and feature focusing yet faces challenges of strong data dependency and insufficient model interpretability.

3 Applications of Transformer Models, Multimodal Learning, and MLLMs

In recent years, the emergence of Transformers has marked a significant advancement in the field of deep learning, abandoning convolutional structures and breaking through the bottlenecks of traditional attention mechanisms. The novel self-attention mechanism represents a qualitative leap in model design philosophy, offering new possibilities for medical vision tasks. Based on Transformer-based multimodal learning, MLLMs (Multimodal Large Language Models) aim to integrate information from different modalities including images, text, and speech, enabling models to comprehensively understand and process complex diagnostic tasks. Figure 2 illustrates the overall development trajectory of the model.

Transformers possess a broader receptive field, successfully overcoming limitations in accessing global information. This allows the model to freely gather information between different positions in the input sequence, effectively establishing long-range dependencies. Duong et al. proposed a hybrid approach utilizing EfficientNet and Vision Transformer (ViT) models combined with various transfer learning techniques (such as ImageNet, AdvProp, and Noisy Student), achieving 97.72% accuracy and 100% AUC for tuberculosis identification on datasets including Montgomery County (MC) and Shenzhen [16]. Their predictive performance and generalization capabilities surpass individual models while maintaining excellent computational efficiency and robustness in practical scenarios, overcoming the limitations of previous methods that were often only sensitive to small-scale or non-diverse datasets [16].

Beyond imaging, medical practice contains abundant diverse information, including reports and blood test results. How to comprehensively utilize these to improve automatic diagnosis is currently a mainstream research direction in medical deep learning. Zhou et al. proposed IRENE, a Transformer-based multimodal representation learning model that processes and holistically represents unstructured medical images, patient condition descriptions, medical history, structured clinical data (height, weight, gender, age, etc.), and laboratory test information [17]. This achieved a 12% AUC improvement over image-only models and 9% over traditional non-unified methods in lung disease identification. For COVID-19 patient clinical outcome prediction, it demonstrated a 29% AUPRC improvement over traditional methods, avoiding the problem of independent processing and fusion of different modal features while simplifying cumbersome clinical text structuring steps [17].

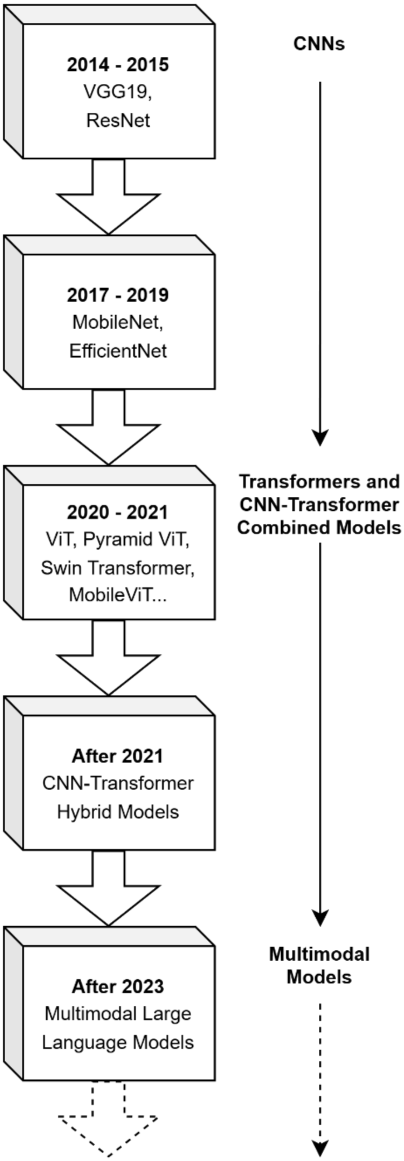

Fig. 2. Development of the model used in CXR image recognition. (Picture credit: Original)

Lee et al. introduced CXR-LLAVA, a LLAMA-2-based multimodal large language model that integrates a ViT architecture visual encoder for CXR image interpretation [18]. Through joint fine-tuning of CXR images and radiological reports, the model demonstrated the ability to generate detailed reports, including accurate descriptions of lesion locations and providing differential diagnoses and other key clinical information. CXR-LLAVA achieved an average F1 score of 0.81 in diagnosing six pulmonary pathologies, outperforming mainstream MLLMs such as GPT-4-Vision and Gemini-Pro-Vision [18]. On external test sets, the model's report success rate was 72.7%; although lower than the 84.0% ground truth benchmark, it demonstrated tremendous potential value for clinical applications [18].

The Transformer architecture shows disruptive potential in the future of medical AI, redefining clinical decision-making paradigms through deep semantic mining of images and multimodal fusion capabilities. However, numerous challenges remain in practical applications. The computational complexity of self-attention mechanisms increases quadratically with image size, requiring downsampling to save computational costs. Medical diagnosis demands high precision, requiring more detailed alignment to address hallucination risks in large models. Additionally, the unique nature of medical ethics requires models to have traceable reasoning chains, while existing black-box structures struggle to meet clinical credibility requirements.

4 Conclusion

Deep learning-based CXR image analysis technology has developed rapidly, significantly enhancing the automated diagnostic capabilities for pulmonary diseases and greatly contributing to future medical resource conservation and response to public health emergencies. Models centred on convolutional neural networks achieve a balance between accuracy and efficiency in detecting pneumonia, tuberculosis, and COVID-19 through architectural optimization, lightweight design, and integrated approaches. Large-scale weakly supervised annotation and augmentation synthesis provide models with substantial data of observable quality, while preprocessing further ensures correct learning from the data. The new mechanisms introduced by Transformer architecture and multimodal learning have initiated the next wave of medical automated diagnosis, offering clinicians more comprehensive and efficient auxiliary tools.

CXR images and other medical imaging data originate from diverse equipment and various academic and medical institutions. This diversity necessitates good domain adaptation capabilities for data, features, models, and even labels, including the introduction of image style transfer, feature alignment and adversarial training, and model federated learning techniques.

The multimodal cognitive paradigm remains in its nascent stage with substantial room for deepening; simultaneously, simple classification can no longer meet evolving medical requirements, necessitating expansion into complex tasks such as treatment plan evaluation and disease causal inference. Current MLLMs (Multimodal Large Language Models) have demonstrated potential for cross-modal reasoning, though semantic mining remains insufficient. Future work requires the construction of medically knowledge-guided attention mechanisms, incorporating clinical guidelines, disease progression temporality, and even audio, video, and other examination data into model architectures to achieve broader multimodal, multi-task learning.

The requirements are difficult to achieve without more complex models and algorithms with greater computational overhead, which conflicts with the real-time nature of medical practice and the limited computational resources of portable, small medical devices. To apply theoretically high-performing models from academic research to practical applications, future work needs to focus more on model miniaturization and resource-friendly optimization strategies, such as distillation, early stopping, and sparsity regularization.

Finally, the black-box characteristics of models contradict the evidence-based pursuit of medicine, with trustworthy and interpretable deep learning remaining one of the barriers to clinical translation. Current research predominantly focuses on performance metrics, lacking systematic evaluation of model biases and hallucinations, necessitating the establishment of multidimensional fairness tests and clinical utility benchmarks. Regarding model interpretability, future efforts should focus on constructing transparent inference chains and human-machine collaborative diagnostic paradigms, including hierarchical interpretable attention mechanism design, lesion heatmaps, and causal association analysis.

In summary, this paper comprehensively reviews the development of deep learning technology in the automated diagnosis of CXR images, from traditional CNN models, derived optimization, integrated and lightweight architectures to visual and multimodal models based on the novel Transformer architecture, analyzing the advantages and limitations of various models in pulmonary disease detection from multiple perspectives. This paper also focuses on recent changes in the CXR image data domain, emphasizing commonly used datasets and data augmentation, annotation, and image preprocessing and pre-extraction, identifying key issues related to quality, quantity, and corresponding countermeasures. Finally, this paper envisions future directions in medical image analysis, supporting its greater role in reducing patient suffering and physician fatigue, optimizing service quality and medical resources, and even responding to the next public health emergency.

References

1. Murphy S. L., Kochanek K. D., Xu J., and Arias E.: 'Mortality in the United States, 2020,' NCHS Data Brief, 2021, 427, pp. 1–8.
2. Ai T. et al.: 'Correlation of Chest CT and RT-PCR Testing for Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases,' Radiology, 2020, 296, (2), pp. 32–40.
3. Li J. L., Liu X., Lin Z. Y., Long X. R., Jiang Z. H., and Huo Y. Y.: 'The current application status of deep learning in Chest X-ray screening for lung diseases,' Chinese Journal of Antituberculosis, 2024, 46, p. 12.
4. Alshmrani G. M. M., Ni Q., Jiang R., Pervaiz H., and Elshennawy N. M.: 'A deep learning architecture for multi-class lung diseases classification using chest X-ray (CXR) images,' Alexandria Engineering Journal, 2023, 64, pp. 923–935.
5. Liu S., Cai T., Tang X., Zhang Y., and Wang C.: 'COVID-19 diagnosis via chest X-ray image classification based on multiscale class residual attention,' Computers in Biology and Medicine, 2022, 149, p. 106065.

6. Huang H., Li Y., Wu R., Li Z., and Zhang J.: 'Benign-malignant classification of pulmonary nodule with deep feature optimization framework,' Biomedical Signal Processing and Control, 2022, 76, p. 103701.
7. Kumar R., Pan C.-T., Lin Y.-M., Yow-Ling S., Chung T.-S., and Janesha U. G. S.: 'Enhanced Multi-Model Deep Learning for Rapid and Precise Diagnosis of Pulmonary Diseases Using Chest X-Ray Imaging,' Diagnostics, 2025, 15, 3, 3.
8. Souid A., Sakli N., and Sakli H.: 'Classification and Predictions of Lung Diseases from Chest X-rays Using MobileNet V2,' Applied Sciences, 2021, 11, 6.
9. Diallo P. A. K. K. and Ju Y.: 'Accurate Detection of COVID-19 Using K-EfficientNet Deep Learning Image Classifier and K-COVID Chest X-Ray Images Dataset,' in 2020 IEEE 6th International Conference on Computer and Communications (ICCC), 2020, pp. 1527–1531.
10. Jaeger S., Candemir S., Antani S., Wáng Y.-X. J., Lu P.-X., and Thoma G.: 'Two public chest X-ray datasets for computer-aided screening of pulmonary diseases,' Quant Imaging Med Surg, 2014, 4, 6, pp. 475–477.
11. Wang X., Peng Y., Lu L., Lu Z., Bagheri M., and Summers R. M.: 'ChestX-Ray8: Hospital-Scale Chest X-Ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases,' in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 3462–3471.
12. Chowdhury M. E. H. et al.: 'Can AI Help in Screening Viral and COVID-19 Pneumonia?,' IEEE Access, 2020, 8, pp. 132665-132676.
13. Chen Y. et al.: 'Research progress on medical image dataset expansion methods,' Sheng Wu Yi Xue Gong Cheng Xue Za Zhi, 2023, 40, 1, pp. 185–192.
14. Ganesan P., Rajaraman S., Long R., Ghoraani B., and Antani S.: 'Assessment of Data Augmentation Strategies Toward Performance Improvement of Abnormality Classification in Chest Radiographs,' in 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2019, pp. 841–844.
15. Maity A., Nair T. R., Mehta S., and Prakasam P.: 'Automatic lung parenchyma segmentation using a deep convolutional neural network from chest X-rays,' Biomedical Signal Processing and Control, 2022, 73, p. 103398.
16. Duong L. T., Le N. H., Tran T. B., Ngo V. M., and Nguyen P. T.: 'Detection of tuberculosis from chest X-ray images: Boosting the performance with vision transformer and transfer learning,' Expert Systems with Applications, 2021, 184, p. 115519.
17. Zhou H.-Y. et al.: 'A Transformer-based representation-learning model with unified processing of multimodal input for clinical diagnostics,' 2023, arXiv: arXiv:2306.00864.
18. Lee S., Youn J., Kim H., Kim M., and Yoon S. H.: 'CXR-LLAVA: a multimodal large language model for interpreting chest X-ray images,' 2024, arXiv: arXiv:2310.18341.