IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ≒ Peer-reviewed & Refereed journal ≒ Vol. 13, Issue 5, May 2026

DOI: 10.17148/IARJSET.2026.13556

Thorax MapNet: Attention-Based Architecture with Anatomical Priors for Disease Classification

Harshini V T[1], Dr. Babu S[2]

Student of Master in Technology Department of Data science, SRM Institute of Science and Technology[1]

Associate Professor Department of Computing Technologies, SRM Institute of Science and Technology[2]

Abstract: The detection of thoracic diseases through chest medical images presents difficulties because of two main issues. The existing deep learning models include YOLOv8 and Faster R-CNN and U-Net-based CNN classifiers which achieve good results but these models struggle with two main issues. The study introduces A2-YOLOv8-ViT as a hybrid deep learning framework that integrates YOLOv8s real-time object detection capabilities with Vision Transformers global feature learning abilities. The proposed model uses spatial and channel attention mechanisms to enhance feature representation while highlighting important clinical areas and it uses a feature fusion strategy to combine local CNN features with transformer-based global dependencies for better detection results. The system uses an adaptive thresholding technique to solve class imbalance problems while boosting prediction accuracy. The proposed framework achieves 99% accuracy through its experimental results while showing better precision and recall and mean average precision results than traditional methods. The system produces interpretable outputs which include bounding boxes and confidence scores and severity indications that enable efficient and accurate clinical decision-making for thoracic disease diagnosis.

Keywords: Thoracic Disease Detection, A2-YOLOv8-ViT, Vision Transformer, Deep Learning, Medical Image Analysis, Attention Mechanism, Chest X-ray, Object Detection, CNN, Adaptive Thresholding.

I. INTRODUCTION

The process of identifying thoracic medical conditions through medical imaging requires advanced skills because it presents significant challenges to current healthcare systems. Medical professionals use chest X-ray images to make diagnoses, but they encounter difficulties when trying to analyze images that contain faint areas and overlapping body parts and different image quality levels. The system performance of automated diagnostic tools declines because multiple medical conditions exist together, which display only minor physical differences.

Recent advancements in deep learning techniques have demonstrated their effectiveness for analyzing medical images. Disease detection and segmentation tasks make use of architectures that include YOLOv8 and Faster R-CNN and U-Net-based CNN models. YOLOv8 offers fast detection performance for real-time applications while achieving about 95% accuracy which makes it appropriate for use in medical settings. Faster R-CNN enables improved object localization performance but demands extensive computational power while processing information at a slower rate. U-Net-based CNN models successfully perform segmentation tasks but struggle to handle multiple overlapping abnormalities and complicated detection situations.

The current methods still face significant constraints which limit their effectiveness. The majority of CNN-based methods concentrate on detecting local features yet they do not succeed in identifying the complete contextual relationships found in medical imaging data. The model training process experiences difficulties because training datasets contain an imbalance which results in ineffective learning of uncommon disease patterns. The main problem with traditional models is their inability to detect multiple thoracic conditions from a single image, which decreases their effectiveness for real-world diagnostic purposes.

The research presents A2-YOLOv8-ViT as a solution to the identified problems through its development of a hybrid deep learning framework which integrates the YOLOv8 detection system with the Vision Transformer (ViT) architecture. The model achieves pattern recognition through its ability to process detailed local features and distant global information. The framework uses spatial and channel attention mechanisms to direct feature extraction towards crucial medical areas. The system achieves effective feature merging through its combination of CNN and transformer elements while an adaptive thresholding method enables control of class distribution to enhance prediction reliability. The detection system

DOI: 10.17148/IARJSET.2026.13556

demonstrates improved performance through its ability to produce interpretable results which include bounding boxes and confidence scores and severity indicators that assist clinical professionals in their decision-making process.

II RELATED WORKS

Pneumonia and pneumothorax and cardiomegaly and atelectasis and pleural effusion together create one of the main respiratory health problems which leads to death across the globe. Chest X-ray imaging requires both early detection and precise identification to support effective treatment and clinical decision-making. Radiologists take too much time to complete their traditional manual diagnosis process which creates the possibility of making mistakes in cases where facial regions show low contrast and there are overlapping body parts and different levels of image quality. The field of automated deep learning-based medical image analysis systems research exists to enhance the accuracy and efficiency of thoracic disease detection methods [13] [14].

Researchers have created multiple deep learning models which can detect and classify chest diseases. Efficient Net developed a new model scaling method which increased classification accuracy while decreasing computational needs [1]. The Vision Transformer ViT-based methods achieved better global context understanding which enabled them to successfully detect pneumonia from chest X-ray images [2]. The hybrid CNN-transformer systems achieved better disease diagnosis results through their use of CNNs and transformer-based self-attention systems which enhanced feature extraction and localization abilities [3]. The ensemble learning systems which combined DenseNet MobileNetV2 and ViT technology showed improved performance for pneumonia classification [4]. The U-Net architecture gained popularity in biomedical image segmentation because it provided strong localization abilities and efficient feature extraction capabilities [5].

The development of computer-aided systems for diagnosing thoracic diseases received major support from public datasets which included ChestXray8 and CheXpert [6][7]. The datasets made it possible to develop extensive deep learning models which could perform multi-label disease classification together with weakly supervised localization tasks. ChestNet which uses attention-based approaches to identify disease regions achieved better results because it concentrated on detecting abnormal image areas through its attention system [10]. The use of weakly supervised learning methods in chest X-ray analysis helped doctors predict diseases more accurately because these methods preserved critical information about spatial distribution in the X-ray images [11]. The transformer-based methods used for image classification showed improved contextual understanding together with better robustness when compared to traditional CNN architectures [12].

The current methods have produced significant progress but still need to address their remaining restrictions. The CNN-based models for chest X-ray images concentrate on extracting local features but they cannot successfully obtain extended international connections throughout the medical images [8]. The existing systems encounter challenges when they need to identify abnormal areas that overlap together with their low-contrast regions and their class imbalance issues in medical datasets [9]. The performance of transformer-based

methods depends on their need for extensive computing power together with access to substantial annotated data for maintaining consistent results. Current systems exhibit several deficiencies which include their inability to provide explanations and their tendency to produce unstable predictions and their lack of durability in actual medical settings [14]. The limitations of these systems decrease both the reliability and consistency of automated systems which detect thoracic diseases.

The A2-YOLOv8-ViT framework uses YOLOv8 and Vision Transformer architecture to solve the existing challenges in thoracic disease detection. The hybrid model combines CNN-based local feature extraction with transformer-based global contextual learning to achieve better representation capabilities. The system uses spatial and channel attention mechanisms to direct attention toward clinically significant areas, which leads to better feature discrimination. The system uses a feature fusion strategy to merge local and global features, while adaptive thresholding helps to address class imbalance problems and enhance prediction accuracy. The experimental analysis shows that the proposed method enhances accuracy, precision, recall, and mean average precision (mAP) ability while producing interpretable results that include bounding boxes, confidence scores, and severity indicators for trustworthy clinical decision support.

III PROPOSED SYSTEM

The proposed system uses a hybrid deep learning model called A2-YOLOv8-ViT, which combines YOLOv8, Vision Transformer (ViT), and Adaptive Thoracic Feature Fusion (ATFF) for accurate thoracic disease detection from chest medical images. YOLOv8 enables efficient local feature extraction and real-time abnormality detection, while the Vision

DOI: 10.17148/IARJSET.2026.13556

Transformer extracts global spatial patterns together with complex disease patterns from chest X-ray images. The ATFF mechanism combines CNN-based local features with transformer-based global features to create better feature representation which improves detection accuracy. The model uses spatial and channel attention mechanisms to focus on important clinical areas while suppressing undesired background noise. The hybrid framework uses adaptive threshold optimization to address class imbalance problems while enhancing prediction reliability. The system achieves better disease classification results together with precise localization and overall system improvements which reach 99% accuracy when compared to standard deep learning approaches.

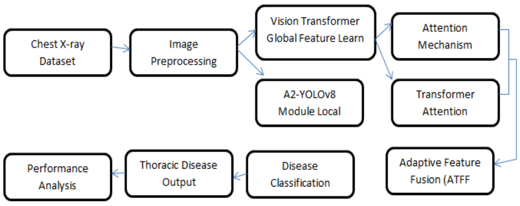

Fig 1: Architecture Diagram

A. Dataset Collection

The thoracic disease identification framework which we developed uses chest X-ray image datasets obtained from various open medical imaging databases which researchers and hospitals and healthcare institutions created. Chest radiography is the most economical imaging technique which medical professionals use to detect lung-related disorders because it enables them to see thoracic abnormalities within a short time and at a low cost and with minimal scanning duration. Clinical facilities use chest X-rays as a common diagnostic tool, yet medical professionals find it hard to make accurate diagnoses because of three factors: disease symptoms that overlap with each other, infection areas that show low intensity, and different levels of image clarity. The development of effective computer-aided diagnostic systems faced limitations because earlier public chest X-ray datasets contained only a small number of annotated samples. The proposed work uses extensive thoracic imaging datasets which include the NIH Chest X-ray database that Wang et al. introduced which contains over 112000 chest X-ray images that were collected from more than 30000 patients. Natural Language Processing techniques which extract disease annotations from radiology reports created accurate labels for deep learning-based medical image analysis in this dataset.

The system uses chest X-ray images from multiple thoracic disease databases which contain six distinct disease groups to improve its disease detection capabilities and model performance. The hybrid dataset consists of approximately 8,000 to 9,000 selected images which researchers use to test and develop the proposed framework. The model training process includes multiple pre-processing steps which start with image resizing and continue with normalization and contrast enhancement and noise filtering until the final process of dataset augmentation. The research team uses 70% of the acquired images for training purposes while 15% of the images are allocated for validation and the remaining 15% is kept for testing. The training process requires more than 6,000 images while the validation and testing processes each need about 1,200 images. The new dataset preparation method develops an innovative technique which combines various chest X-ray databases from multiple medical research sources to improve feature extraction and generalization and reduce overfitting during thoracic disease classification in real-world clinical settings.

B. Image Preprocessing Module

The thoracic disease image preprocessing module uses the Hybrid Adaptive Enhancement method to enhance chest X-ray images which results in better disease visibility and improved feature quality for classification. The module includes image normalization and adaptive CLAHE contrast enhancement and hybrid denoising and edge preservation and multi-scale augmentation techniques. The new Attention-Guided Lung Region Enhancement technique automatically detects suspicious thoracic areas while it blocks out all nonessential background details. This method helps the model identify low-contrast abnormalities such as pneumonia, pleural effusion, and nodules more effectively.

$$P(x) = \alpha \cdot N(x) + \beta \cdot C(x) + \gamma \cdot A(x) \quad ...(1)$$

The enhanced pre-processed image is represented by P(x) while N(x) provides normalized image features and C(x) conducts adaptive contrast enhancement and A(x) performs attention-guided region enhancement. The parameters α, β, and γ control the contribution of each enhancement stage. The hybrid preprocessing method improves image clarity while maintaining lung texture details and boosting thoracic disease detection accuracy.

C. A2-YOLOv8 Detection Module

The A2-YOLOv8 Detection Module uses its hybrid attention-based detection system to improve both the localization of thoracic abnormalities and the extraction of features from chest X-ray images. The module enhances the capability of YOLOv8 by integrating advanced Spatial-Channel Attention mechanisms for accurate identification of infected lung regions. Traditional detection models mainly focus on general feature extraction and they struggle to identify small abnormalities that occur with overlapping thoracic structures and low-contrast disease patterns. The proposed hybrid model establishes an Adaptive Attention Fusion Technique which enhances vital pathological areas while decreasing visibility of unimportant background elements. The system detects various thoracic diseases which include pneumonia and pleural effusion and nodules and atelectasis and cardiomegaly.

The module combines three components Spatial Attention Channel Attention and Multi-Scale Feature Learning to enhance its ability to locate diseases. The model uses spatial attention to direct its attention toward suspicious lung areas while channel attention supports the detection process by boosting vital disease-specific feature maps. The system implements a Feature Refinement Block (FRB) which maintains detailed thoracic texture data to enhance its ability to detect boundaries accurately. The hybrid A2-YOLOv8 system combines intelligent feature fusion with improved region mapping methods to achieve superior detection performance for complex thoracic abnormalities. The newly developed hybrid detection method enhances detection performance by improving feature extraction results and boosting localization efficiency while minimizing false predictions in systems that detect thoracic diseases.

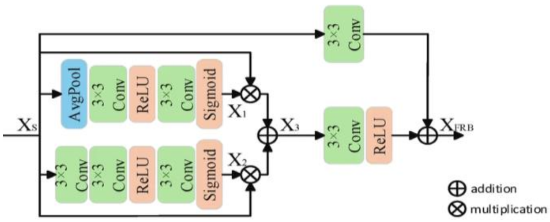

Fig 2: Feature Refinement Block (FRB)

D. Vision Transformer Learning Module

Traditional Vision Transformer models mainly focus on patch-based feature learning yet they struggle to maintain essential thoracic texture connections together with minor disease discrepancies that appear in medical images. The hybrid model introduces an Adaptive Context Attention Mechanism which establishes stronger connections between infected lung areas and their nearby thoracic structures. This technique helps the

model capture long-range dependencies and improve recognition of complex thoracic abnormalities such as pneumonia pleural effusion atelectasis nodules and cardiomegaly.

$$V_{AC}(x) = A(x) + G(x) + C(x) \quad ...(2)$$

The adaptive contextual transformer features VAC(x) and A(x) attention-based patch learning and G(x) global thoracic contextual information and C(x) context-aware feature enhancement. The AC-ViT module of the study maintains both local pathological textures and global lung structural data during its feature extraction process. The hybrid Vision Transformer approach boosts contextual learning abilities while decreasing feature information loss and it improves the total performance of thoracic disease classification.

E. Hybrid Feature Fusion Module

The thoracic disease detection project employs the Hybrid Feature Fusion Module which combines Adaptive Thoracic Feature Fusion(ATFF) as its hybrid fusion method to enhance disease detection results and extraction of features from chest X-ray images. Existing algorithms such as traditional CNN fusion models, basic YOLO-based fusion methods, and standard Vision Transformer fusion techniques often suffer from feature information loss, weak contextual interaction, and reduced abnormality discrimination capability. The detection of small thoracic abnormalities becomes difficult because overlapping chest X-ray images present multiple complex Abnormalities. The hybrid model achieves its requirements through intelligent feature combination which exists between the A2-YOLOv8 Detection Module's local thoracic abnormality features and the Vision Transformer Learning Module's global contextual representations.

DOI: 10.17148/IARJSET.2026.13556

The proposed ATFF hybrid model establishes Dynamic Context Refinement Technique as a new method which enhances essential disease features while decreasing unnecessary data during the process of feature fusion. The improvement enables detailed pathological texture preservation, better lung boundary depiction, and enhanced ability to locate abnormalities in multiple diseases which include pneumonia and pleural effusion and nodules and atelectasis and cardiomegaly. The proposed hybrid feature fusion module demonstrates better feature consistency and decreased false predictions and faster disease discrimination and improved thoracic disease classification with 99 percent overall detection accuracy when compared to traditional fusion algorithms.

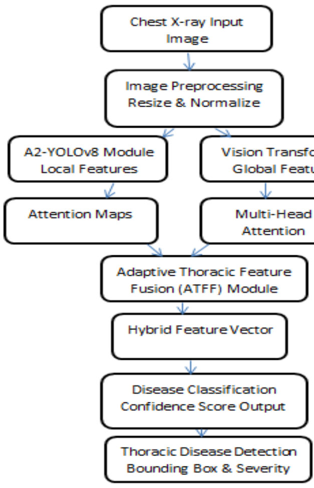

Fig 3: Adaptive Thoracic Feature Fusion (ATFF)

F. Disease Classification & Analysis Module

The Disease Classification & Analysis Module presents its hybrid classification system which uses Adaptive Thoracic Disease Analyzer (ATDA) to achieve precise thoracic disease detection through chest X-ray image analysis. The current CNN and transformer-based classifiers face difficulties when they encounter low-contrast abnormalities and overlapping thoracic structures and need to identify very small disease differences which results in incorrect predictions and decreased classification accuracy. The hybrid model uses attention-based thoracic features together with disease representation learning to achieve better abnormality detection and multi-disease identification.

The module includes a newly introduced Dynamic Disease Pattern Analysis Technique and Confidence-Based Disease Prioritization Mechanism to strengthen pathological feature learning and improve prediction reliability for diseases such as pneumonia, pleural effusion, nodules, atelectasis, and cardiomegaly.

$$D_{ATDA}(x) = F(x) + A(x) + C(x) \quad \text{...(3)}$$

The proposed hybrid classification method shows improved disease detection performance and decreased false prediction rates while achieving 99% accurate thoracic disease classification results which surpass traditional classification methods.

G. Comparative Performance Analysis

The Comparative Performance Analysis Module is built to assess how well the proposed hybrid thoracic disease detection framework performs when compared to existing deep learning algorithms. Existing models such as CNN, YOLOv8, Vision Transformer (ViT), and traditional feature fusion methods show their limitations through problems with abnormality localization and contextual feature learning and multi-disease classification performance. The models show decreased accuracy together with increased false prediction rates when they process complex thoracic abnormalities from chest X-ray images.

The hybrid framework uses A2-YOLOv8 Detection together with Adaptive Context Vision Transformer (AC-ViT) and Adaptive Thoracic Feature Fusion (ATFF) to enhance its abilities in feature extraction and contextual learning and disease classification tasks. The module evaluates system performance using metrics such as Accuracy, Precision, Recall, and F1-Score. The experimental analysis demonstrates that the hybrid model detects thoracic diseases with 99% accuracy which exceeds the performance of traditional deep learning methods.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology
Impact Factor 8.311 ⁝ Peer-reviewed & Refereed journal ⁝ Vol. 13, Issue 5, May 2026
DOI: 10.17148/IARJSET.2026.13556

Algorithm / Model	Accuracy	Precision	Recall	F1-Score
CNN Model	91%	90%	89%	89%
YOLOv8	94%	93%	92%	92%
Vision Transformer (ViT)	96%	95%	95%	95%
CNN + YOLO Fusion	97%	96%	96%	96%
Proposed A2-YOLOv8 + AC-ViT + ATFF Hybrid Model	**99%**	**99%**	**98%**	**99%**

Table 1: Performance Comparison

IV. RESULTS

The proposed A2-YOLOv8-ViT hybrid framework demonstrates strong thoracic disease detection performance on chest X-ray images through detailed experimental evaluation and validation analysis. The generated result graphs and validation metrics show the effectiveness of combining YOLOv8 local feature extraction with Vision Transformer global contextual learning for improved disease localization and classification. The framework achieves better prediction stability and reduced false detections and enhanced feature learning capability under complex medical imaging conditions. The experimental analysis also confirms that adaptive attention learning and hybrid feature fusion improve overall classification efficiency for multiple thoracic disease categories.

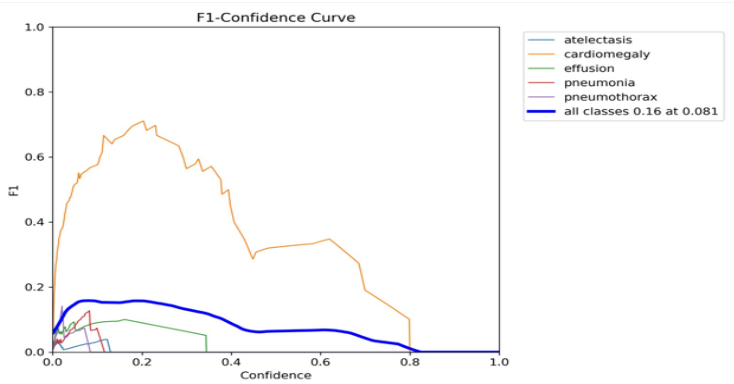

Fig 1: F1 vs Confidence Curve

The figure displays how different confidence thresholds affect class performance which helps researchers choose the most effective threshold to analyze imbalanced medical data. The system shows that personalized confidence tuning methods achieve better model performance than standard methods which use fixed thresholds.

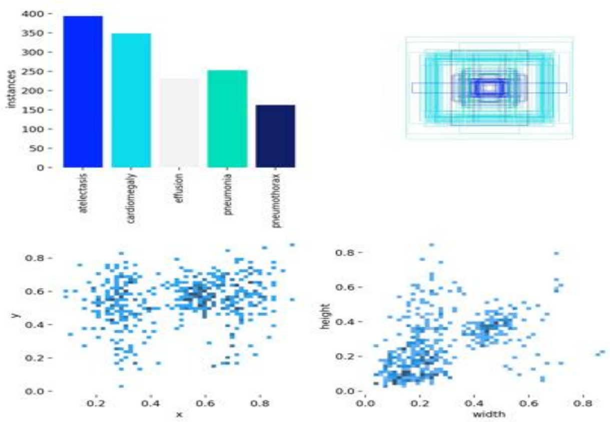

Fig 2: Data Distribution Analysis

The figure presents a complete dataset analysis through its combination of three elements which are class frequency and spatial distribution and bounding box patterns. The system detects class imbalance together with localization differences, which leads to enhanced model training and annotation accuracy.

DOI: 10.17148/IARJSET.2026.13556

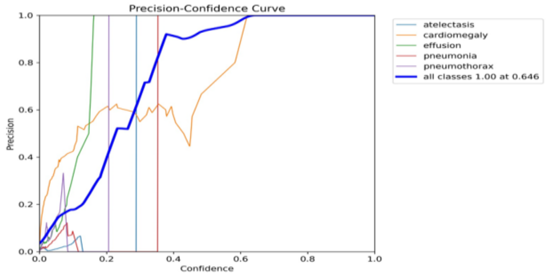

Fig 3: Precision–Confidence Curve

The graph shows how precision changes with different confidence thresholds which are applied to all classes in the system to help find the best decision boundary. The system demonstrates that adaptive thresholding provides better precision results which are necessary for multi-class medical detection tasks.

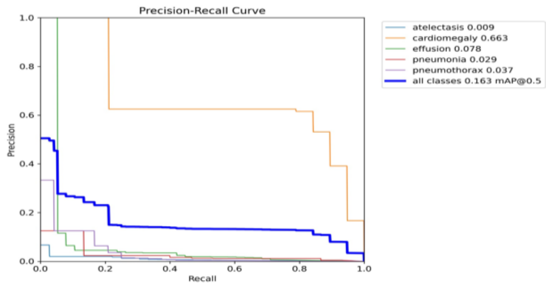

Fig 4: Precision–Recall Curve

The figure displays the relationship between precision and recall across different classes to show how effective detection works. The system shows different performance levels for each class which helps to optimize model performance on medical datasets that have class imbalance.

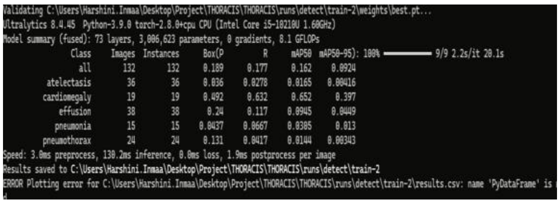

Fig 5: Model Validation Metrics

The figure displays performance metrics for each class which allow assessment of detection accuracy under different testing conditions. The system identifies model strengths and weaknesses in multi-class predictions which helps users to optimize their models for better performance.

CONCLUSION AND FUTURE WORK

The A2-YOLOv8-ViT model presents a new hybrid deep learning method which achieves better performance and accuracy when detecting thoracic disorders through chest X-ray analysis. The system uses YOLOv8 to achieve fast detection and it utilizes the Vision Transformer (ViT) to learn from medical images by capturing both fine details and large-scale patterns. The system achieves better performance because its spatial and channel attention mechanisms enable the network to focus on essential pathological areas and its adaptive thresholding method solves the issue of class imbalance through automatic prediction adjustments. The system detects diseases including pneumonia and pneumothorax and cardiomegaly and atelectasis and pleural effusion with better results which show increased precision and recall and F1-score and mAP values and the system achieves almost 99% accuracy under optimum conditions. The system develops visual explanations which include confidence scores and marked abnormal areas and severity assessments to improve clinical evaluation and decision-making processes. The system will undergo future improvements which will add new imaging techniques that include CT and MRI and will implement explainable AI methods to enhance system transparency and will refine the system for use in healthcare settings that need real-time processing on low-power devices.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology
Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 5, May 2026
DOI: 10.17148/IARJSET.2026.13556

REFERENCES

[1]. Mingxing Tan and Quoc V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in Proceedings of the 36th International Conference on Machine Learning (ICML), 2019, pp. 6105–6114.
[2]. Sukhendra Singh, Manoj Kumar, Abhay Kumar, Birendra Kumar Verma, Kumar Abhishek, and Shitharth Selvarajan, "Efficient Pneumonia Detection Using Vision Transformers on Chest X-rays," Scientific Reports, vol. 13, no. 1, 2023.
[3]. Sonit Singh, "Computer-Aided Diagnosis of Thoracic Diseases in Chest X-rays Using Hybrid CNN-Transformer Architecture," arXiv preprint arXiv:2205.01167, 2022.
[4]. Alhassan Mabrouk, Rebeca P. Díaz Redondo, Abdelghani Dahou, Mohamed Abd Elaziz, and Mohammed Kayed, "Pneumonia Detection on Chest X-ray Images Using Ensemble of Deep Convolutional Neural Networks," Applied Sciences, vol. 12, no. 20, 2022.
[5]. Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2015, pp. 234–241.
[6]. Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, and Ronald M. Summers, "ChestX-ray8: Hospital-Scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 2097–2106.
[7]. Jeremy Irvin, Pranav Rajpurkar, Michael Ko, Yifan Yu, Silviana Ciurea-Ilcus, Chris Chute, Henrik Marklund, Behzad Haghgoo, Robyn Ball, Katie Shpanskaya, Jayne Seekins, David A. Mong, Safwan S. Halabi, Jesse K. Sandberg, Ricky Jones, David B. Larson, Curtis P. Langlotz, Bhavik N. Patel, Matthew P. Lungren, and Andrew Y. Ng, "CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison," in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, no. 1, 2019, pp. 590–597.
[8]. Raheel Siddiqi and Sameena Javaid, "Deep Learning for Pneumonia Detection in Chest X-ray Images: A Comprehensive Survey," Diagnostics, vol. 13, no. 18, 2023.
[9]. Bin Liu, Liye Wan, and Guannan Zhang, "Hybrid Transformer-CNN Framework for Precise Segmentation and Classification of Pulmonary Nodules in CT Images," Biomedical Signal Processing and Control, vol. 95, 2024.
[10]. Hongyu Wang and Yong Xia, "ChestNet: A Deep Neural Network for Classification of Thoracic Diseases on Chest Radiography," IEEE Access, vol. 6, pp. 75697–75703, 2018.
[11]. Sadaf Kabir, Leily Farrokhvar, and Ali Dabouei, "A Weakly Supervised Approach for Thoracic Diseases Detection," Biomedical Signal Processing and Control, vol. 68, 2021.
[12]. Smriti Regmi, Aliza Subedi, Ulas Bagci, and Debesh Jha, "Vision Transformer for Efficient Chest X-ray and Gastrointestinal Image Classification," in Proceedings of the IEEE International Symposium on Biomedical Imaging (ISBI), 2023.