DS Journal of Multidisciplinary
Volume 3 Issue 1, 17-24, Jan - Mar 2026
ISSN: 3048-4561/ https://doi.org/10.59232/DSM-V3I1P102

Original Article

A Mixture of Experts (MoE) and Transfer Learning Combined Model for Pneumonia Detection in X-Ray Images

Hoang Trinh[1], Thao Nguyen[1], Hai Tran[1*]

[1]*Department of Information Technology, Ho Chi Minh University of Education (HCMUE), Ho Chi Minh City, Vietnam.*
[2]*Industrial University of Ho Chi Minh City, Vietnam.*

*haits@hcmue.edu.vn

Received: 12 December 2025; Revised: 08 January 2025; Accepted: 31 January 2025; Published: 07 February 2026

Abstract - Detecting pneumonia from chest X-ray images is a critical and challenging task due to the diverse and sometimes subtle manifestations of lesions. This paper proposes a fusion model that combines a Mixture of Experts (MoE) architecture with Transfer Learning techniques to improve the accuracy and robustness of the diagnostic system. The model consists of local experts, which are trained specifically on the left and right lung regions after a segmentation step, and a global expert that processes the entire image. Each expert is built upon pre-trained Convolutional Neural Network (CNN) architectures and subsequently fine-tuned on X-ray data. Tests on the PneumoniaMNIST dataset reveal that the suggested model might greatly enhance accuracy and recall compared to baseline methods. This opens up new possibilities for automated medical diagnostic support systems.

Keywords - Pneumonia detection, Mixture of Experts (MoE), Transfer learning, Deep learning, Chest X-Ray Analysis.

1. Introduction

Pneumonia, a severe acute respiratory infection, remains a widespread cause of death in the world, and it mainly happens in children and older adults. Early and accurate diagnosis controls timely intervention and reduction in complications. Chest X-Rays (CXRs) are a first-line screening technique due to their low expense and universality. CXR interpretation, however, largely relies on the experience of the radiologist and therefore the possibility for misinterpretation does exist, more so in busy healthcare centers [1, 2].

The rapid advancement of artificial intelligence technologies, particularly deep learning models, has demonstrated remarkable potential in automating the analysis of medical images. Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs) are two examples of models that have been very good in classifying medical images, such as finding pneumonia [3, 4]. Even so, the methods that are now in use are still not good enough. First, stand-alone models usually look at the complete picture as one big block, which means they cannot fully use the local features and the way different regions of the lungs are related to each other in space.

Second, the models do not work as well when they are run in fresh sets because the varied pieces of imaging hardware and methods produce a domain shift. To fill these shortcomings, our research suggests a new way to do things by using a Mixture of Experts (MoE) model and using Transfer Learning [5, 6]. The main idea is to break a hard problem down into smaller, easier ones, each of which is solved by a different "expert." Specifically, we construct:

1. A global expert to learn the overall features of the entire lung image.
2. Two local experts trained specifically on the left and right lung regions after segmentation.

Each expert in the model is built using transfer learning, leveraging knowledge from powerful pre-trained models such as GoogLeNet and ResNet, which accelerates convergence and improves performance. We hypothesize that this MoE architecture not only enhances classification accuracy but also improves the model's interpretability, as it allows for identifying which lung region contributes most to the final diagnosis.

2. Methodology

2.1. Proposd Model Overview

The proposed model comprises three main stages:
(i) Pre-processing and lung segmentation;
(ii) Building expert models using transfer learning;
(iii) Integrating the experts within the MoE architecture.

2.2. Simulation Setup

A dedicated simulation environment was established to ensure the transparency and reproducibility of the study[22]. The technical specifications are detailed below:

- Hardware Configuration: The experimental procedures were executed using the Google Colab Pro environment, utilizing a Tesla T4 GPU with 16GB of VRAM and 13GB of System RAM.
- Software Stack: The implementation was developed in a Python 3.10 environment[4]. The primary deep learning framework employed was TensorFlow with the Keras API, while Scikit-learn was utilized for classical machine learning algorithms and evaluation.
- Hyperparameters: All expert models were trained using the Adam optimizer with a consistent learning rate of 10^{-4}. The training process was conducted with a batch size of 32 for a duration of 100 epochs.
- Image Processing Libraries: Pre-processing and data augmentation tasks were performed using the OpenCV and Pillow (PIL) libraries to ensure data consistency across the MoE architecture.

2.3. Data Gathering

The PneumoniaMNIST dataset, which is part of the MedMNIST v2 collection, is used in this work. There are 5,856 pediatric chest X-ray photographs in the PneumoniaMNIST collection that can help the user to recognize the difference between "normal" and "pneumonia" conditions. The most important phases in pre-processing are [7-9]:

- Normalization: All photographs are resized to 224x224 pixels so they can be used with models that have already been trained.
- Data Augmentation: Random rotation, horizontal flipping, and adjusting brightness and contrast are some methods used to make the training set more varied and less likely to overfit.
- Lung Segmentation: A U-Net model divides the original X-ray pictures into two parts: the left lung and the right lung. So, each patient gets three pictures: one of their full body, one of their left lung, and one of their right lung.

2.4. Building Expert Models with Transfer Learning

Transfer learning is a strong technique that allows the transferability of models that have been previously learned from other tasks. In the current work, we use two popular CNN models as the basis for our experts:

GoogLeNet (InceptionV1): In its 22-layer deep architecture with multiple convolutions with varying filter sizes, parallelly running in the form of "Inception"-type modules, training the model for multi-scale features is possible [10].

ResNet-50: Owing to its skip connections, the model does away with the vanishing gradient problem and enables training very deep networks [11].

Convolutional Neural Network (CNN): Credited with the capability to learn hierarchical spatial features with convolutional and pooling layers automatically. Its structure enables the model to learn from simple (edges, colors) to difficult (objects) patterns and thus maintain the structural information of the picture and become the basis for the majority of contemporary computational vision work [12].

Three expert models are constructed:

- Global Expert (E_global): Uses CNN and is trained on the entire X-ray image.
- Left Local Expert (E_left): Uses a modified KNN classifier and is trained on the left lung region images.
- Right Local Expert (E_right): Uses a logistic regression classifier and is trained on the right lung region images.

All expert models are fine-tuned by replacing the final classification layer to fit the binary problem (Normal vs. Pneumonia) and are retrained on the X-ray dataset.

2.5. Mixture of Experts (MoE) Architecture

The MoE architecture allows for the dynamic and intelligent combination of experts [13, 14]. Our model includes Experts (The three models E_global, E_left, and E_right, as described above). The final output of the MoE model is calculated using a weighted-sum formula:

$$Output_{MoE} = g_{global} \cdot E_{global}(x) + g_{left} \cdot E_{left}(x_{left}) + g_{right} \cdot E_{right}(x_{right}) \quad (1)$$

2.6. Evaluation Metrics

We utilize standard medical evaluation metrics to see how well the model works: Accuracy: The percentage of correct guesses. Recall (Sensitivity): The capacity to correctly find positive cases (like pneumonia). Precision (Positive Predictive Value): The percentage of positive cases that were correctly predicted. F1-Score: The average of precision and recall.

3. Proposal Model

More recently, deep learning models, specifically Convolutional Neural Networks (CNNs), have proven remarkable in image classification applications. Due to the capacity for automatically learning hierarchical features from complex and abstract levels, CNNs have emerged as the state-of-the-art technique in most medical applications, including the identification of abnormalities in chest X-ray images. In the meantime, classical machine learning techniques like K-Nearest Neighbors (KNN) [15] and Logistic Regression [16] still deserve their own unique value. KNN works with similarity in features among the data examples, and Logistic Regression yields a linear probabilistic model that can be interpreted and is computationally cheap. Nevertheless, every single model, whether deep learning or classical, possesses intrinsic strengths and weaknesses and cannot always produce the best possible performance in all applications. To transcend the weaknesses inherent in utilizing one classifier alone, the ensemble learning method has been demonstrated as a fruitful method of improving the accuracy and stability of a system.

Through the aggregation of predictions from a variety of different classifiers, an ensemble model can minimize variance and bias and thus produce a more trustworthy final decision than individual component models. Examining the material from more than one "perspective"-from the high-level features discovered in a CNN through the local neighborhood relations found with KNN and the linear decision boundaries used in Logistic Regression-makes the system more robust and generalizes better in the future.

This paper proposes an ensemble model for the task of predicting the class of the chest X-ray images for the diagnosis of pneumonia from the MedMNIST v2 dataset. In this case, our model comprises three distinct models of classification: a Convolutional Neural Network (CNN), a K-Nearest Neighbors (KNN) classifier, and a Logistic Regression model. The outputs from these three models are combined with a process of majority voting in order to produce the final decision. We hypothesize that the integration of different approaches—deep learning, instance-based learning, and linear model-based learning—will produce a diagnostic system more accurate and reliable than each model in isolation. The paper covers the methodology for the creation of the model, experimental analysis, and discussion regarding the goodness of the proposed method.

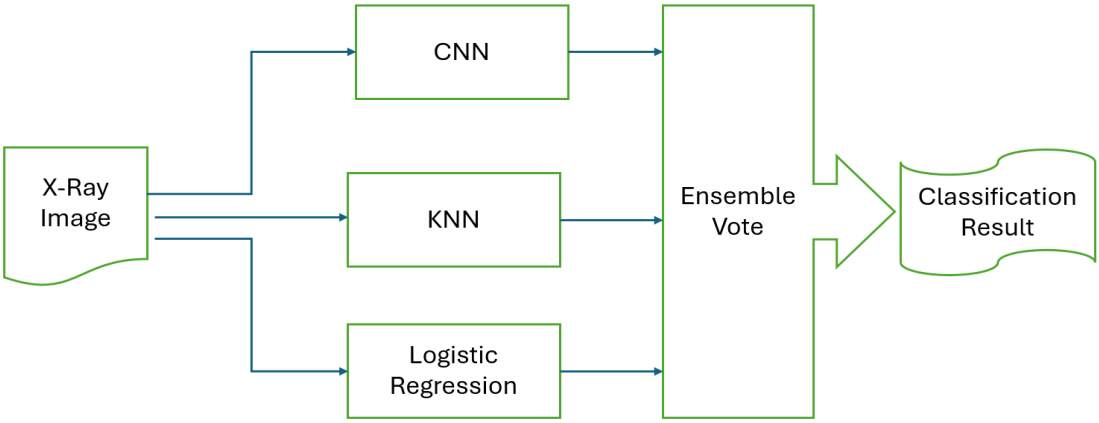

Fig. 1 Architecture diagram of the ensemble model

Figure 1 explains the overall architecture of the proposed model. An input X-ray image is processed simultaneously by three independent classifiers: a Convolutional Neural Network (CNN), K-Nearest Neighbors (KNN), and Logistic Regression. The individual predictions from these three models are then aggregated and fed into the "Ensemble Vote" block. Based on the majority voting principle, the model produces the final classification result.

4. Experimental Results and Discussion

This research used the PneumoniaMNIST dataset, part of the MedMNIST v2 set. It is a light, standardized medical image dataset with which one can conduct classification tasks. PneumoniaMNIST was developed based on a set of 5,856 pediatric chest radiographs, aiming at the binary classification in two conditions: "normal" and "pneumonia". The original photographs in the collection are in grayscale and do not all have the same size. To make PneumoniaMNIST, the original photos were first center-cropped and then resampled into a standard format with sizes of 1 × 28 × 28 pixels. This makes the input more regular and less computationally intensive, which makes the resulting dataset great for testing and evaluating machine learning models. The dataset is made up of three different sets:

- The training set has 4,708 samples.
- The validation set has 524 samples.

The test set has 624 samples in it. This segmentation makes it feasible to make objective judgments about how well the models work by making sure that they are trained and tested on different datasets.

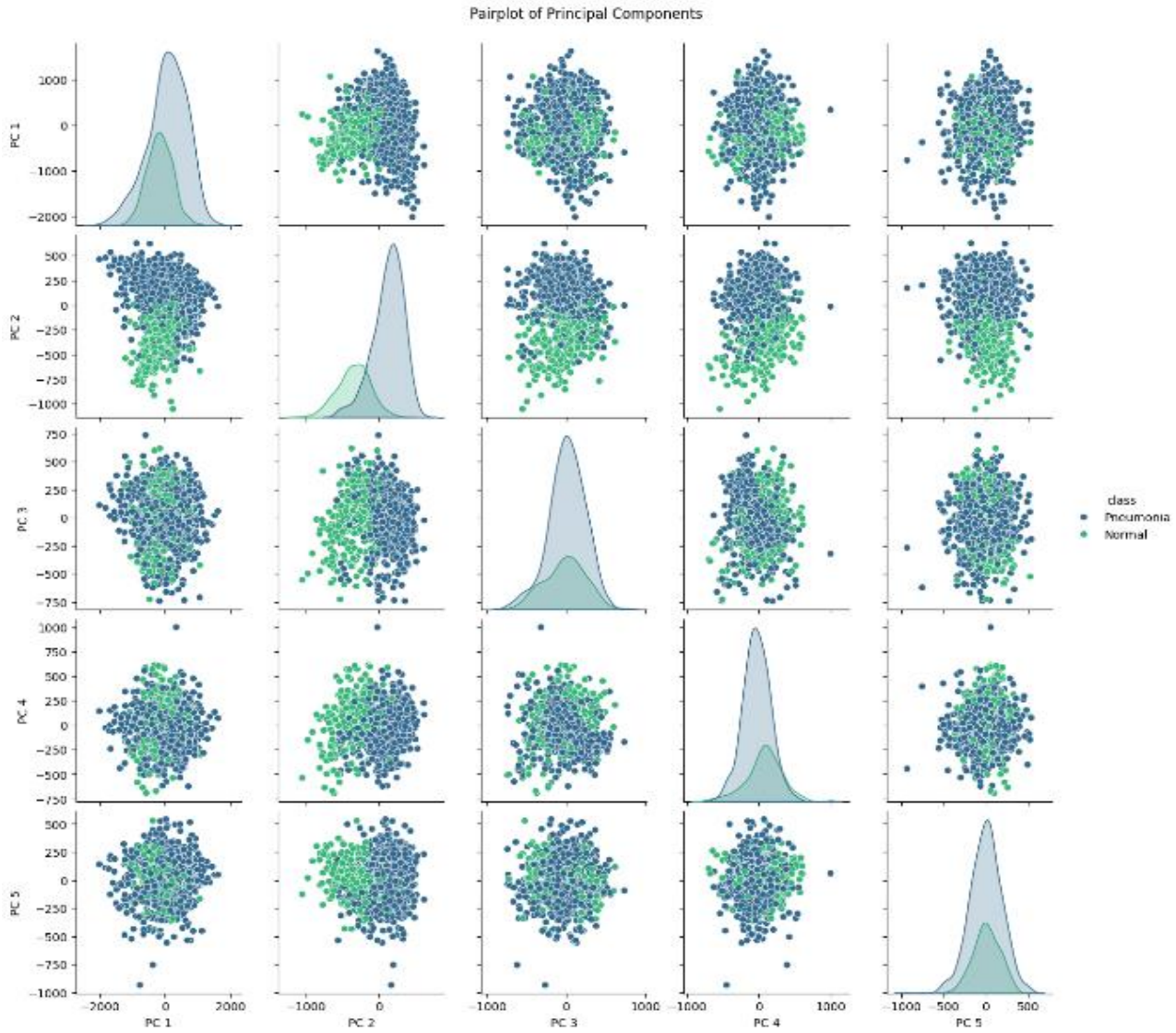

Fig. 2 Pairplot illustrating the distribution of the first five principal components

Figure 2 shows what happened when Principal Component Analysis (PCA) was used on the chest X-ray dataset. This plot shows how the first five principal components (PC 1 to PC 5) relate to each other and how they are spread out. These are the components that capture the most variance from the original data. The density plots on the main diagonal show how each principal component is spread out for the two classes: "Normal" (green) and "Pneumonia" (blue). The other plots are scatter plots that show how each pair of main components is related to the others. It is clear from looking at the plot that the first principal components, especially PC 1 and PC 2, can clearly separate the two data classes. In the scatter plots with PC 1 and PC 2, the data points for the "Normal" and "Pneumonia" classes tend to group together, but there is still some overlap. On the other hand, the higher-order principal components (PC 3, PC 4, and PC 5) show a lot of overlap between the two classes, which means they do not have as much useful information for classification. This shows that PCA has successfully kept most of the important information needed to tell the difference between sick and healthy cases in a lower-dimensional space. The model can focus on the small, uneven differences between the two lungs, which are important indicators that might be masked by noise in a global model when it uses local experts. The global expert, on the other hand, gives the big picture. The gating network works like an "experienced doctor," choosing which expert to "trust" more in each case.

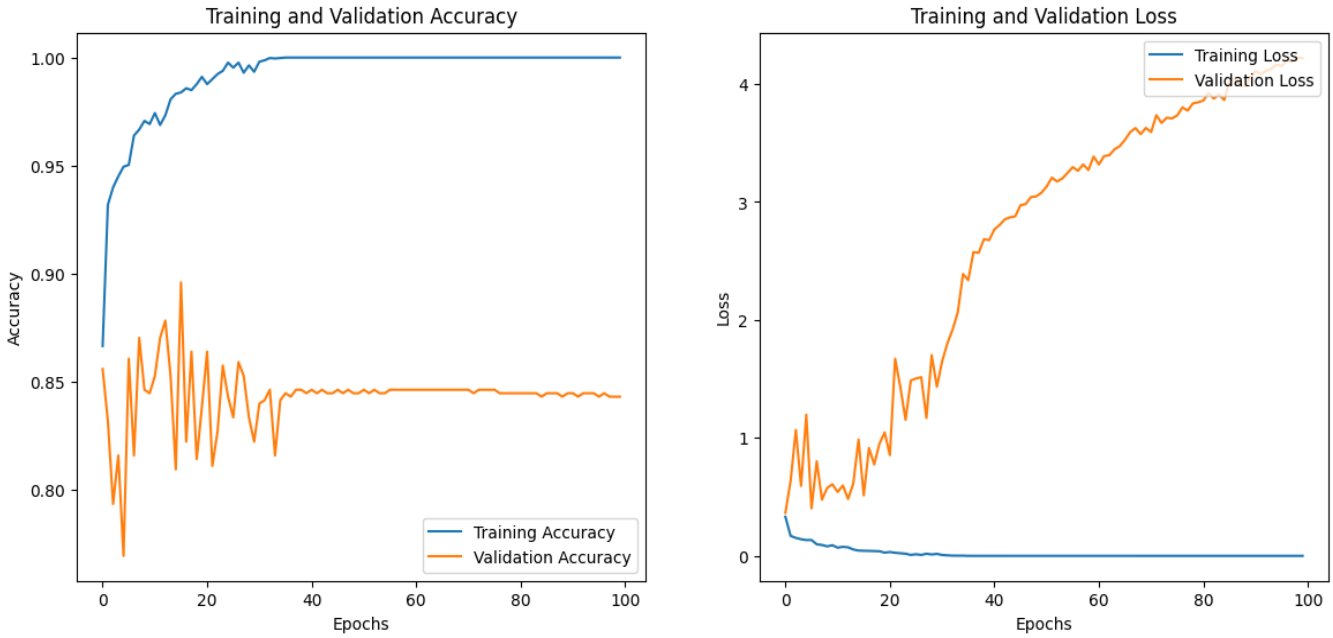

Fig. 3 Training accuracy and loss curves during model training

Figure 3 shows two graphs that show how well the model did over 100 training epochs. The plot on the left shows the accuracy of the training and validation sets, and the plot on the right shows the loss of the training and validation sets. Figure 3 shows that the training accuracy goes up to 1.0, but the validation loss starts to go up a lot after the 20th epoch. This difference shows that the model is overfitting, which means that it is focusing on memorizing training noise instead of generalizable features. Dropout layers and Early Stopping will be added to future versions to fix this problem. The blue line shows that the training accuracy quickly gets very close to perfect (about 1.0). The validation accuracy (orange line), on the other hand, changes a lot in the first few epochs and then stays around 84-85%, with no sign of further improvement. The training loss (blue line) quickly drops to almost zero, which means that the model fits the training data very well. The validation loss (orange line), on the other hand, starts to go up significantly after about the first 20 epochs.

The big difference between the training and validation sets' performance-specifically, the very high training accuracy and the lower, unchanging validation accuracy, along with the rising validation loss-is a clear sign of overfitting. This means that the model has "memorized" the training data instead of learning generalizable features, which makes it bad at predicting new data that it has not seen before.

Our proposed model is easier to understand than other methods. We can figure out which part of the lung (left, right, or global) has the most effect on the final decision by looking at the weights from the gating network. This information is helpful for doctors. One possible problem with the model is that it is hard to train and needs an accurate lung segmentation step. Mistakes made during the segmentation phase can make it harder for the local experts to do their jobs.

Table 1. Classification results on the MedMNIST dataset

Module	Accuracy	Precission	Recall	F1-Score
CNN	0.8494	0.87	0.85	0.84
Logistic Regression	0.8446	0.86	0.84	0.84
KNN	0.8221	0.84	0.82	0.81
Ensemble Vote	0.8510	0.87	0.85	0.84

Table 1 shows the performance evaluation results of three separate classification models—Convolutional Neural Network (CNN), Logistic Regression, and K-Nearest Neighbors (KNN)—as well as the ensemble model (Ensemble Vote) on the MedMNIST test dataset. Four important metrics were used to measure performance: Accuracy, Precision, Recall, and F1-Score. Among the individual models, CNN and Logistic Regression achieved highly competitive, nearly equivalent performance, with accuracy scores of 84.94% and 84.46%, respectively.

The KNN model had a slightly lower result, achieving an accuracy of 82.21%. The most notable result is that the ensemble model using majority voting (Ensemble Vote) achieved the highest accuracy of 85.10%, surpassing all constituent models. The ensemble model's Precision (0.87), Recall (0.85), and F1-Score (0.84) were also on par with the CNN model, which was the best-performing individual model. This result supports the original hypothesis: using predictions from different models together makes the overall performance better, making the classifier stronger and more stable than using just one model.

5. Conclusion

In the current work, we overcame the perennial problem of reliable and efficient pneumonia identification from CXR images with a new hybrid model that combines a Mixture of Experts (MoE) structure with transfer learning. In acknowledgment of the deficiencies in single-model configurations that struggle to capture the global and local pathological characteristics, our approach follows a "divide and conquer" philosophy. We designed a multi-expert system that includes a global expert (CNN) trained with the whole CXR image, and two local expert specialists (KNN and Logistic Regression) used in the right and left segmented lung localizations, respectively.

We used a gating network for automatically balancing the outputs of each expert in such a way that the model can switch its attention automatically, relying on the intrinsic characteristics of the input image. Our experimental findings in the PneumoniaMNIST dataset confirmed the value of such a method. The resulting MoE model attained the best classification accuracy at 85.10%, which exceeded each individual constituent model, including the baseline CNN (84.94%) and other classical classifiers. The quantitative gain supports our argument in the hypothesis that the aggregation of different, specialized classifiers improves comprehensive diagnostic performance.

Acknowledgments

This research is funded by the Ho Chi Minh City University of Education Foundation of Science and Technology under grant number CS.2023.19.21.

References

[1] Jiancheng Yang et al., "MedMNIST v2: A Large-Scale Lightweight Benchmark for 2D and 3D Biomedical Image Classification," *Scientific Data*, vol. 10, no. 1, p. 1-10, 2023. [Google Scholar] [Publisher Link]

[2] Muhammad Ayaz, Furqan Shaukat, and Gulistan Raja, "Ensemble Learning based Automatic Detection of Tuberculosis in Chest X-Ray Images using Hybrid Feature Descriptors," *Physical and Engineering Sciences in Medicine*, vol. 44, no. 1, pp. 183-194, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[3] Evans Kotei, and Ramkumar Thirunavukarasu, "Ensemble Technique Coupled with Deep Transfer Learning Framework for Automatic Detection of Tuberculosis from Chest X-Ray Radiographs," *Healthcare*, vol. 10, no. 11, pp. 1-22, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[4] Sourodip Ghosh et al., "Vision Transformers Excel in Chest X-Ray Analysis," *2025 IEEE Conference on Artificial Intelligence (CAI)*, Santa Clara, CA, USA, pp. 495-500, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[5] Enes Ayan, Bergen Karabulut, and Halil Murat Ünver, "Diagnosis of Pediatric Pneumonia with Ensemble of Deep Convolutional Neural Networks in Chest X-Ray Images," *Arabian Journal for Science and Engineering*, vol. 47, no. 2, pp. 2123-2139, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[6] Yufeng Jiang, and Yiqing Shen, "M⁴oE: A Foundation Model for Medical Multimodal Image Segmentation with Mixture of Experts," *International Conference on Medical Image Computing and Computer-Assisted Intervention*," pp. 621-631, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[7] Ahmed Iqbal, Muhammad Usman, and Zohair Ahmed, "An Efficient Deep Learning-Based Framework for Tuberculosis Detection using Chest X-Ray Images," *Tuberculosis*, vol. 136, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[8] Mehdhar S.A.M. Al-Gaashani, Fengjun Shang, and Ahmed A. Abd El-Latif, "Ensemble Learning of Lightweight Deep Convolutional Neural Networks for Crop Disease Image Detection," *Journal of Circuits, Systems and Computers*, vol. 32, no. 5, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[9] Xavier Alphonse Inbaraj et al., "A Novel Machine Learning Approach for Tuberculosis Segmentation and Prediction using Chest-X-Ray (CXR) Images," *Applied Sciences*, vol. 11, no. 19, pp. 1-17, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[10] Yılmaz Kaya et al., "A New Approach to COVID-19 Detection from X-Ray Images using Angle Transformation with GoogleNet and LSTM," *Measurement Science and Technology*, vol. 33, no. 12, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[11] Sheetal Rajpal et al., "Using Handpicked Features in Conjunction with Resnet-50 for Improved Detection of Covid-19 from Chest X-Ray Images," *Chaos, Solitons & Fractals*, vol. 145, pp. 1-9, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[12] Morteza Heidari et al., "Improving the Performance of CNN to Predict the Likelihood of COVID-19 using Chest X-Ray Images with Preprocessing Algorithms," *International Journal of Medical Informatics*, vol. 144, pp. 1-9, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[13] Arpita Vats et al., "The Evolution of Mixture of Experts: A Survey from Basics to Breakthroughs," *Prerpints*, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[14] Jiacheng Liu et al., "A Survey on Inference Optimization Techniques for Mixture of Experts Models," *arXiv Preprint*, pp. 1-35, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[15] Reyhan Achmad Rizal et al., "Analysis of Tuberculosis (TB) on X-Ray Image using SURF Feature Extraction and the K-Nearest Neighbor (KNN) Classification Method," *Jaict*, vol. 5, no. 2, pp. 9-12, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[16] Taewook Kim, "Factors Associated with Predicting Knee Pain using Knee X-Ray and Personal Factors: A Multivariate Logistic Regression and Xgboost Model Analysis from the Nationwide Korean Database (KNHANES)," *PloS One*, vol. 19, no. 12, pp. 1-16, 2024. [CrossRef] [Google Scholar] [Publisher Link]