

عبدالعزیز ال عبود

Penumonia-related work version 4 final .docx

 Midocean University

Document Details

Submission ID

trn:oid::2945:378606951

Submission Date

Apr 20, 2026, 11:45 PM GMT+2

Download Date

Apr 21, 2026, 9:07 AM GMT+2

File Name

Penumonia-related work version 4 final .docx

File Size

714.1 KB

16 Pages**4,580 Words****28,999 Characters**

62% detected as AI

The percentage indicates the combined amount of likely AI-generated text as well as likely AI-generated text that was also likely AI-paraphrased.

Caution: Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.

Detection Groups



16 AI-generated only 58%

Likely AI-generated text from a large-language model.



1 AI-generated text that was AI-paraphrased 4%

Likely AI-generated text that was likely revised using an AI-paraphrase tool or word spinner.

Disclaimer

Our AI writing assessment is designed to help educators identify text that might be prepared by a generative AI tool. Our AI writing assessment may not always be accurate (i.e., our AI models may produce either false positive results or false negative results), so it should not be used as the sole basis for adverse actions against a student. It takes further scrutiny and human judgment in conjunction with an organization's application of its specific academic policies to determine whether any academic misconduct has occurred.

Frequently Asked Questions

How should I interpret Turnitin's AI writing percentage and false positives?

The percentage shown in the AI writing report is the amount of qualifying text within the submission that Turnitin's AI writing detection model determines was either likely AI-generated text from a large-language model or likely AI-generated text that was likely revised using an AI paraphrase tool or word spinner.

False positives (incorrectly flagging human-written text as AI-generated) are a possibility in AI models.

AI detection scores under 20%, which we do not surface in new reports, have a higher likelihood of false positives. To reduce the likelihood of misinterpretation, no score or highlights are attributed and are indicated with an asterisk in the report (*%).

The AI writing percentage should not be the sole basis to determine whether misconduct has occurred. The reviewer/instructor should use the percentage as a means to start a formative conversation with their student and/or use it to examine the submitted assignment in accordance with their school's policies.

What does 'qualifying text' mean?

Our model only processes qualifying text in the form of long-form writing. Long-form writing means individual sentences contained in paragraphs that make up a longer piece of written work, such as an essay, a dissertation, or an article, etc. Qualifying text that has been determined to be likely AI-generated will be highlighted in cyan in the submission, and likely AI-generated and then likely AI-paraphrased will be highlighted purple.

Non-qualifying text, such as bullet points, annotated bibliographies, etc., will not be processed and can create disparity between the submission highlights and the percentage shown.



Literature Review

Pneumonia Classification from Chest X-ray Images Using Convolutional Neural Networks

Pneumonia remains one of the leading causes of death worldwide, with young children and elderly patients carrying most of the burden, particularly in regions with limited access to specialist care. Chest X-ray imaging is the first-line diagnostic tool in most clinical settings because it is fast, inexpensive, and widely available. Interpreting these images, however, is not straightforward. Experienced radiologists frequently disagree on the same scan, and in many hospitals — especially in rural and low-resource areas — there are simply not enough trained readers to review every case in time. These limitations have driven growing interest in automated diagnostic tools based on deep learning, and in particular on Convolutional Neural Networks (CNNs), which learn visual patterns directly from large collections of labelled images. Over the past few years, a substantial body of work has applied CNNs and, more recently, Transformer-based architectures to chest X-ray classification, addressing both the binary task of detecting pneumonia and finer distinctions such as bacterial versus viral cases.

This chapter reviews recent studies that together represent the main directions currently being pursued in this area. The selected works span several methodological families, including pretrained CNN baselines, Vision Transformers, hybrid CNN–Transformer designs, feature-level ensembles, and federated learning approaches. Taken together, they offer a picture of where the field currently stands what works reliably, what remains difficult, and which problems are still open. The remainder of the chapter first presents the technical background needed to discuss these studies, then surveys the literature in detail, and finally identifies the gaps that motivate the work proposed in this thesis.

Introduction:

1.1 Problem Definition:

Pneumonia remains one of the most serious respiratory diseases worldwide, posing a significant threat to global health. According to global health reports, pneumonia is a leading cause of death among children under five years old and elderly populations, especially in regions with limited access to advanced medical care. Early and accurate diagnosis is critical for effective treatment; however, this remains a challenging task in clinical practice. Chest X-ray imaging is the most common diagnostic tool for pneumonia due to its availability and low cost. Although chest X-rays are widely used, interpreting them is often difficult and depends heavily on the radiologist's expertise. As a result, different specialists often disagree on diagnoses when they look at the same medical image, which can lead to diagnostic errors. Furthermore, healthcare systems, especially those in resource-limited settings, often grapple with a scarcity of seasoned radiologists, thereby exacerbating delays in both diagnosis and subsequent treatment. The core issue, at its most fundamental level, is the absence of diagnostic systems that are both efficient and precise, and that can scale effectively to analyze extensive datasets of medical imagery. [25]

Manual interpretation is not only labor-intensive but also susceptible to human error, particularly when distinguishing subtle visual discrepancies between healthy and pneumonia-affected lung tissue. Recent investigations have focused on the application of artificial intelligence, specifically deep learning methodologies like Convolutional Neural Networks (CNNs), for the automated detection of pneumonia. Previous research has shown that deep learning models have demonstrated considerable efficacy in the classification of chest X-ray images. For instance, CNN-based methodologies have achieved accuracies surpassing 98%, as evidenced by studies such as those conducted by Buriboev et al. In 2025 and Ahmed et al. (2023), sophisticated architectures such as DenseNet, ResNet, and EfficientNet have exhibited considerable efficacy in feature extraction from medical images, thereby enhancing diagnostic precision. The application of artificial intelligence in this domain is substantiated by its capacity to: automatically identify intricate patterns within medical images, diminish reliance on human interpretation, augment diagnostic consistency and accuracy, and furnish scalable solutions suitable for extensive healthcare infrastructures. Consequently, the central issue addressed by this research is the necessity for an intelligent, automated system capable of accurately identifying pneumonia from chest X-ray images, thereby circumventing the constraints inherent in conventional diagnostic methodologies. Through the utilization of deep learning techniques, this study endeavors to improve diagnostic performance, assist medical professionals, and contribute to a more efficient healthcare system. [25]

1.2 Thesis Objectives:

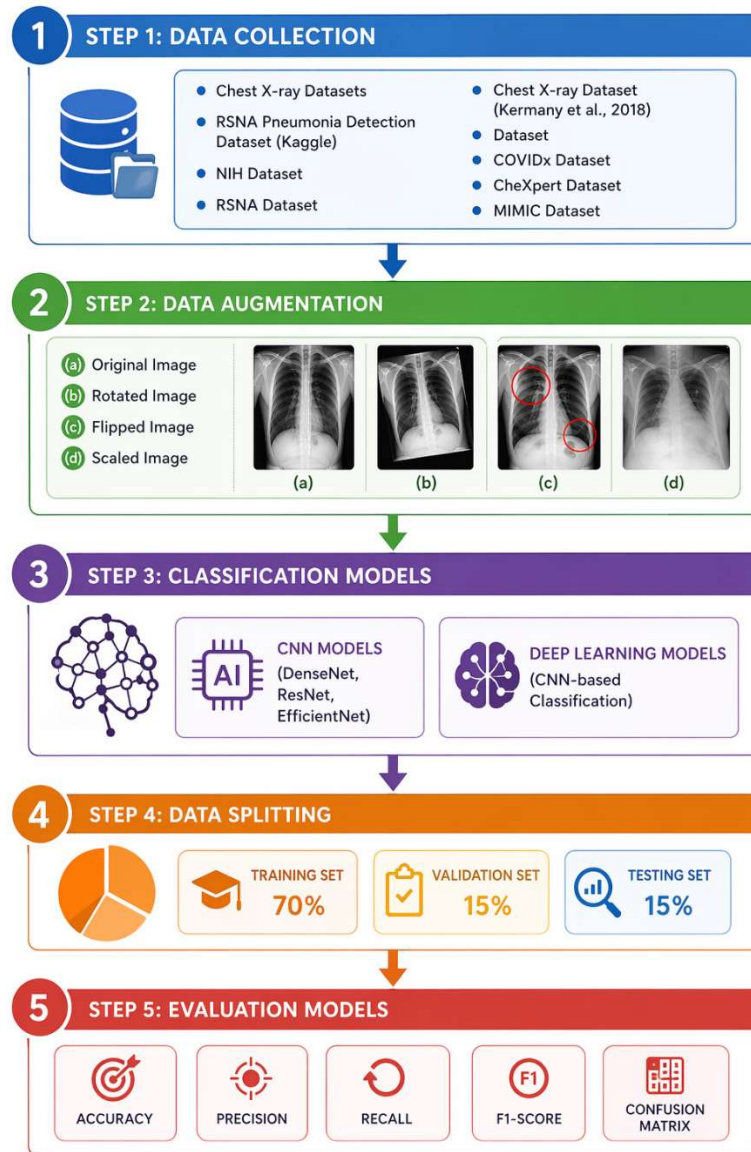
The primary goal of this dissertation is to create an automated pneumonia detection system, leveraging chest X-ray images and deep learning methodologies. This investigation , seeks to enhance diagnostic precision through the application of CNN models and transfer learning strategies. Furthermore, the research endeavors to refine model performance via data augmentation techniques and assess the efficacy of the proposed model using established evaluation metrics, including accuracy, precision, recall, and the F1-score.

1.3 Thesis Contributions:

1.4 Thesis Organization:

The current dissertation is organized into separate chapters. Chapter 1 introduces the main research question, outlines the specific goals, and provides an overview of how the study is structured. Chapter 2 offers the necessary background and reviews existing research on pneumonia detection, with a focus on deep learning methods and the overall methodological approach.

Proposed Framework:



Background:

1. Artificial intelligence (AI)

Artificial Intelligence (AI) constitutes a domain within computer science dedicated to the creation of systems capable of executing functions generally associated with human intelligence, encompassing learning, reasoning, problem-solving, perception, and decision-making processes. These AI systems are designed to simulate cognitive operations by employing algorithms and computational models, which facilitate the processing of large datasets, the identification of patterns, and the generation of well-informed predictions or decisions. [23]

Artificial intelligence in its broadest sense, encompasses several specialized areas, such as machine learning, deep learning, natural language processing, and computer vision.

2. Machine learning

Machine Learning (ML), a branch of artificial intelligence, centers on creating algorithms and models. These systems can identify patterns in data, enabling them to make predictions or decisions without explicit programming. Instead of adhering to fixed rules, machine learning models improve over time, learning from both their experiences and the data they analyze. [24]

3. Types of Machine Learning

3.1 Supervised Learning

Supervised learning, a subset of machine learning, involves training models on labeled datasets; these datasets pair each input with a corresponding output. The primary objective of supervised learning is to develop a mapping function capable of accurately forecasting outputs for novel, previously unobserved data. [24]

3.2 Unsupervised Learning

Unsupervised learning is a kind of machine learning where models are trained on data that doesn't have any labels. This means that there are no set output labels. The goal of unsupervised learning is to find hidden patterns, structures, or relationships in the data without knowing what the results. [24]

1 CNN:

A subset of deep learning models are engineered for the analysis of structured grid data, particularly images. Convolutional Neural Networks (CNNs), first proposed by Yann LeCun and his team, have become widely used in computer vision. This is because they can automatically extract hierarchical features from input images. [16]

The architecture of CNNs is characterized by convolutional layers, which are responsible for capturing spatial features; pooling layers, which serve to diminish dimensionality; and fully connected layers, which facilitate classification. This layered configuration empowers CNNs to discern intricate patterns, including edges, textures, and shapes, thereby rendering them exceptionally effective in medical image analysis, such as the detection of pneumonia from chest X-rays (LeCun et al., 1998). [16]

2 DenseNet

DenseNets a particular breed of deep neural networks, first appeared in 2017, thanks to Gao Huang and his colleagues. Their goal was to refine the flow of information and the way gradients traveled through these deep networks. The core concept of DenseNets hinges on the use of dense connections. In these connections, each layer gets input from all the layers before it. This encourages the reuse of features and reduces redundancy. As a result, this architecture helps to solve common problems in deep learning, like vanishing gradients, and it also makes better use of parameters. DenseNets have demonstrated considerable efficacy in image classification applications, proving particularly valuable in medical image analysis, including the identification of pneumonia within chest X-ray images. The DenseNet architecture is characterized by dense blocks, wherein numerous convolutional layers are interconnected, and transition layers, which serve to diminish the dimensionality of feature maps via pooling and convolutional operations. This design facilitates the alleviation of the vanishing gradient problem and enhances parameter efficiency. [25]

3 ResNet

ResNet a deep convolutional neural network architecture, They developed it to tackle the difficulties of training extremely deep CNNs. The defining characteristic of ResNet is its incorporation of residual connections, or skip connections. These connections allow the network to learn residual mappings, rather than attempting to learn the entire transformation. This architectural design choice helps gradients flow more easily during training, which reduces problems like vanishing gradients and performance issues in deep networks. ResNet facilitates the successful training of exceptionally deep models, thereby establishing itself as a crucial architecture within computer vision, particularly in the domain of medical image classification. [26]

The ResNet architecture is structured with convolutional layers succeeded by residual blocks; these blocks incorporate shortcut connections that bypass one or more layers, thereby enhancing gradient propagation. Subsequently, fully connected layers are employed for the purpose of classification. [26]

4 EfficientNet

Efficient Net is a family of CNN architectures designed to optimize accuracy while simultaneously reducing computational requirements. Unlike traditional models that adjust depth, width, or resolution independently, EfficientNet employs a compound scaling method, which methodically balances these three dimensions. [27]

Therefore, this method allows the model to perform better while using fewer parameters and requiring less computational power. EfficientNet has been shown to be very effective in different image classification tasks, including its use in medical imaging, such as detecting pneumonia. The EfficientNet architecture integrates convolutional layers, mobile inverted bottleneck blocks (MBConv), and squeeze-and-excitation (SE) layers, which together enhance feature representation. These elements are uniformly scaled to optimize both performance and efficiency. [27]

5 MobileNet:

MobileNet is a type of lightweight CNN architecture that is made to work well on devices with limited resources, like mobile and embedded systems. MobileNet is different from other CNNs because it uses depthwise separable convolutions, which break the standard convolution into two steps: depthwise convolution and pointwise convolution. [26]

4.6 Vision Transformer (ViT)

[12] ViT represent a deep learning architecture that employs transformer models for the purpose of image analysis. In contrast to convolutional neural networks, ViTs segment images into fixed-size patches, subsequently processing these patches as sequences. This approach facilitates the model's ability to comprehend global relationships within the image. The ViT architecture is characterized by several fundamental components. [20]

First, images are split into patches and converted into embeddings using a patch embedding layer. These embeddings are then processed by multiple transformer encoder layers, which include multi-head self-attention mechanisms and feed-forward neural networks. In addition, positional encoding is used to preserve spatial information in the image. Finally, a classification layer is applied to produce the output prediction, ViTs have demonstrated strong performance in computer vision tasks and are increasingly used in medical image analysis. [20]

4.7 Swin Transformer

The Swin Transformer is different from regular transformers because it uses a shifted window-based self-attention mechanism that limits attention computation to local windows while still allowing cross-window connections. This means that the Swin Transformer processes the image in multiple stages with progressively lower spatial resolution, which is similar to how CNN extract features while keeping the benefits of transformer-based models. The Swin Transformer is made up of several important parts. First, there is a patch partitioning layer, then a linear embedding layer. The model then uses multiple Swin Transformer blocks, which include window-based multi-head self-attention (W-MSA) and shifted window multi-head self-attention (SW-MSA), as well as feed-forward networks (MLP layers). [19]

5.1 Accuracy

Accuracy is an evaluation metric that measures the proportion of correctly classified instances out of the total number of instances in a dataset. [21]

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

5.2 Precision

Precision is a metric that measures the proportion of correctly predicted positive instances out of all instances predicted as positive by the model. [22]

$$\text{Precision} = \frac{TP}{TP + FP}$$

5.3 Recall

Recall is a metric that measures the ability of a model to correctly identify all actual positive instances in a dataset. [22]

$$\text{Recall} = \frac{TP}{TP + FN}$$

5.4 F1 Score

The F1-score is a metric that combines precision and recall into a single value by calculating their harmonic meaning, providing a balance between them. [21]

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

2.2 Related Work

Singh et al. [1] applied a ViT, specifically the DeiT_Base_Patch16_224 pretrained variant, along with several CNN-based baselines, including AlexNet, VGG, ResNet, DenseNet, and InceptionNet, all tested under the same conditions for a fair comparison. The dataset used is chest-xray-pneumonia from Kaggle a publicly available collection of 5,863 pediatric chest X-ray images gathered from the Guangzhou Women and Children's Medical Center, split into 80% training, 10% testing, and 10% validation. When the results came in, the ViT came out clearly ahead of every CNN baseline, reaching 97.61% accuracy, 95% sensitivity, and 98% specificity, with an AUC of 0.96 and an MCC of approximately 0.94. The ViT was the highest-scoring model in the study.

Mustapha et al. [2] applied a hybrid DL model that combines a CNN backbone with modified Swin Transformer blocks, tested alongside a standard CNN-only configuration and a pure Vision Transformer for comparison. The dataset was chest-xray-pneumonia from Kaggle, which is publicly available and widely used as a benchmark in this field. The hybrid CNN-ViT model achieved an accuracy of 98.72%, which was higher than both the CNN-only and the pure ViT configurations tested in the same experiment.

Shomsomi et al. [3] applied two DL models, MobileNetV2 and a ViT, and tested each one with two different training methods. In the first method, they kept the backbone frozen and only trained the classifier head. In the second method, they unfroze and updated some of the internal layers. The Chest X-ray Pneumonia dataset from Kaggle was utilized as a benchmark that is publicly available. When fine-tuning was used, the results showed a clear difference between the two architectures. MobileNetV2 reached 97.63% under feature extraction but actually dropped slightly to 97.41% after fine-tuning, suggesting the CNN backbone did not benefit much from further domain adaptation. On the other hand The ViT jumped to 98.59% accuracy and an F1-score of 0.99 with only minor fine-tuning. This was the model with the highest score in the study.

Kareem et al. [4] applied pretrained CNN models within a federated learning framework, where each participating healthcare institution trained a local model on its own data and shared only the model updates, never the raw images, with a central aggregation server. Rather than a single public benchmark, the dataset comprised real-time chest X-ray collections held across multiple institutions with different patient demographics and imaging equipment, making this a private federated setup rather than a publicly accessible dataset. The aggregated model reached an overall accuracy of 93%, which the authors acknowledge is lower than what centralized studies report on curated benchmarks.

Alghamdi et al. [5] applied three DL models, VGG16, ResNet50, and ViT, across three publicly available datasets to specifically study how class imbalance affects performance. The dataset they used were chest-xray-pneumonia from Kaggle, the BRAX dataset from Brazil, and the CheXpert dataset from Stanford University, which all are publicly available. Class weighting and data augmentation were applied to address the imbalance problem, and the Mean Teacher semi-supervised algorithm was also explored for making use of unlabeled images. On the Kaggle dataset, the ViT came out on top with 97.61% accuracy, 95% sensitivity, and 98% specificity. However, when the same models were tested on BRAX and CheXpert, performance dropped noticeably for all three architectures, and that drop is actually the most telling result in the paper.

Boukhe et al. [6] applied a novel hybrid DL architecture that brings together four components: standard CNN layers for spatial feature extraction, Gated Recurrent Units (GRUs) for modeling sequential dependencies across feature representations, a Spiking Neural Network (SNN) layer for noise-tolerant biologically inspired processing, and a spatial attention mechanism that highlights the most diagnostically relevant regions of the image. The dataset was the Chest X-ray Pneumonia dataset from Kaggle, which is publicly available. The only architecture evaluated in the study was the combined CNN-GRU-SNN-Attention model, which achieved 99.35% accuracy with strong precision and recall across the test set.

Ridouane et al. [7] applied a feature-level ensemble DL framework they named PELM, short for Pneumonia Ensemble Learning Model, which combines four pretrained architectures: InceptionV3, VGG16, ResNet50, and ViT. Rather than voting on a final class label at the output stage, PELM concatenates the feature vectors from all four models at the penultimate layer and passes the combined representation to a shared classification head, which is designed to capture relationships between the descriptors each architecture extracts. The author used training data from four publicly available sources: ChestX-ray14 from NIH, CheXpert from Stanford ML group, PadChest from BIMCV, and Chest X-ray Pneumonia collection from Keggles. The dataset contains 50,000 images. When running separately, PELM did better than every single component model examined in the study.

Devaux et al. [8] applied an EfficientNet-B4 model, pretrained on ImageNet and fine-tuned on a large multi-source chest X-ray collection, then validated it prospectively against two real patient cohorts where the reference diagnosis was established using a combination of expert radiologist review and CT scanning. The training data came from three publicly available datasets: Chexpert from Stanford ML group, ChestXray-14 from NIHCC, and MIMIC-CXR from Physio net, which contained 161,584 chest X-ray images. EfficientNet-B4 was the only model evaluated, and on the prospective clinical cohorts its diagnostic accuracy came out broadly comparable to senior radiologists reading the same films.

Buriboev et al. [9] proposed an enhanced pneumonia classification framework utilizing a CNN model trained using a supervised learning approach on datasets preprocessed with a modified CLAHE algorithm. The quality of chest X-ray images by introducing an adaptive contrast enhancement technique and applying preprocessing methods to enhance image features using 5,856 images from the Chest X-Ray Images (Pneumonia) dataset. The main contribution of integrating adaptive contrast enhancement with CNN-based classification. The results demonstrated the effectiveness of combining image enhancement techniques with deep learning models for pneumonia detection, with an accuracy of 98.7%.

Saber et al. [10] proposed a multi-scale, transformer-based method for detecting pneumonia using chest X-ray images. This classification approach was built on a deep learning framework. The transformer-based TransUNet model allowed for accurate lung segmentation, and ResNet models were used to extract multi-scale features relevant to pneumonia classification. The effectiveness of the methodology was evaluated using publicly available chest X-ray datasets, specifically the Kermany and Cohen datasets. The model's performance showed accuracies of 93.75% and 96.04% on the two datasets. The results suggest that using transformer models with deep learning methods for detecting pneumonia

Bundea et al. [11] investigated the use of deep learning techniques for pneumonia classification using chest X-ray images. The discussion analyzed chest X-ray images depicting both normal and pneumonia-affected lungs. To enhance dataset diversity, several preprocessing procedures were implemented, including resizing, normalization, and data augmentation methods like image flipping and rotation. The researchers utilized DenseNet, a deep CNN including DenseNet121, DenseNet169, and DenseNet201, to classify chest X-ray images into two categories: normal and pneumonia engineered to facilitate feature propagation and mitigate the vanishing gradient issue. The model's objective was to extract pertinent features from the chest X-ray images and subsequently categorize them as either pneumonia or normal. The results demonstrated the efficacy of DenseNet classification techniques with deep learning models for pneumonia ,achieving an accuracy of 97% for Pneumonia.

An et al. [12] proposed a deep learning approach for pneumonia detection using chest X-ray images. The research developed an ensemble model that EfficientNetB0 and DenseNet121 with attention mechanisms to improve feature and classification performance. The model classify chest X-ray images into normal, viral pneumonia, and bacterial pneumonia categories. The experimental results showed that the proposed model achieved an accuracy of about 98.4%, the effectiveness of ensemble deep learning models for pneumonia detection.

Usman et al. [13] proposed a deep learning model to classify pneumonia using chest X-ray images. Their research aimed to improve an automated system that could tell the difference between pneumonia and normal conditions in these images. The RSNA pneumonia dataset, comprising thousands of annotated chest X-ray images, served as the foundation for this investigation. To enhance the dataset's quality, several preprocessing procedures were implemented, encompassing image normalization and data augmentation strategies. Subsequently, the efficacy of two deep learning models was assessed: a convolutional neural network (CNN) and a transfer learning model, specifically ResNet-50. The empirical findings indicated that the ResNet-50 model attained an accuracy of approximately 96.2%, thereby illustrating the efficacy of deep learning models in the identification of pneumonia within medical imaging data.

Ahmed et al. [14] proposed a CNN model for the diagnosis of lung diseases using chest X-ray images. The model was trained to classify images, including pneumonia, and COVID-19, The experimental results that the proposed model achieved an accuracy of approximately 98.7%, demonstrating the potential of deep learning techniques for automated medical image analysis.

Al-Rashan et al.[15] proposed a deep learning method for classifying pneumonia using chest X-ray images. The study focused on developing to a system between chest X-ray scans showing pneumonia and normal. The researchers used datasets chest X-ray images of both healthy and pneumonia-affected lungs, several preprocessing techniques were applied, including image , normalization, and data augmentation methods such as rotation and flipping to increase dataset diversity and reduce. CNN architectures were then use to extract deep features from the chest X-ray images and perform the classification task. The experimental results showed that the proposed model demonstrated an accuracy of approximately 95%, demonstrating the effectiveness of deep learning approaches detecting pneumonia from chest X-ray images and supporting computer in medical imaging.

Comparative Summary

Study	CNN Models	Dataset(s)	Public / Private	Classification Task	Best Accuracy	Best Performer
Singh et al. (2024)	ViT (DeiT_Base_Patch16_224); CNN baselines: AlexNet, VGG, ResNet, DenseNet, InceptionNet	<i>Kaggle Chest X-ray Pneumonia</i> kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia	Public	Binary: normal vs. pneumonia	97.61%	Vision Transformer (ViT)
Mustapha et al. (2025)	Hybrid CNN + Swin Transformer; CNN-only; pure ViT (baselines)	<i>Kaggle Chest X-ray Pneumonia</i> kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia	Public	Binary: normal vs. pneumonia	98.72%	Hybrid CNN-ViT (Swin Transformer)
Shomsomi et al. (2026)	MobileNetV2; Vision Transformer (ViT) — both tested under feature extraction and partial fine-tuning	<i>Kaggle Chest X-ray Pneumonia</i> kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia	Public	Binary: normal vs. pneumonia	98.59% (ViT, fine-tuned)	ViT with partial fine-tuning
Kareem et al. (2024)	Pretrained CNNs under federated learning framework	Multi-institutional real-time CXR — federated setup, not publicly released	Private	Binary: normal vs. pneumonia	93%	Federated CNN (only model tested)
Alghamdi et al. (2024)	VGG16; ResNet50; Vision Transformer (ViT)	<i>Kaggle CXR Pneumonia (above)</i> BRAX: physionet.org/content/brax/1.1.0 CheXpert: stanfordmlgroup.github.io/competitions/chexpert	Public	Binary; class imbalance focus across 3 datasets	97.61% (ViT on Kaggle)	ViT — on Kaggle dataset
Boukhe et al. (2025)	Custom hybrid: CNN + GRU + SNN + Spatial Attention	<i>Kaggle Chest X-ray Pneumonia</i> kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia	Public	Binary: normal vs. pneumonia	99.35%	CNN-GRU-SNN-Attention (only model tested)

Ridouane et al. (2025)	PELM ensemble: InceptionV3 + VGG16 + ResNet50 + ViT (feature-level fusion); individual baselines: DenseNet121, DenseNet201, MobileNetV2 also tested	NIH ChestX-ray14: nihcc.app.box.com/v/ChestX-ray-NIHCC CheXpert: stanfordmlgroup.github.io/competitions/chexpert PadChest: bimev.cipf.es/bimev-projects/padchest Kaggle CXR (above) (50,000 images total)	Public	Binary: normal vs. pneumonia	Best among all tested (exact % not disclosed)	PELM ensemble
Devaux et al. (2024)	EfficientNet-B4 (pretrained on ImageNet, fine-tuned on 161,584 CXR images)	CheXpert: stanfordmlgroup.github.io/competitions/chexpert MIMIC-CXR: physionet.org/content/mimic-cxr/2.0.0 ChestXray-14: nihcc.app.box.com/v/ChestX-ray-NIHCC	Public	Binary; prospective clinical validation vs. radiologists and CT	Comparable to senior radiologists	EfficientNet-B4 (only model tested)
Buriboev et al. (2025)	CNN	RSNA Pneumonia Dataset	Public	Binary	Accuracy 97%	CNN
Saber et al. (2025)	ResNet, Attention	Chest X-ray Pneumonia Dataset	Public	Binary	Accuracy 97%	ResNet,
Bundea et al., 2024	DenseNet	Chest X-ray Pneumonia Dataset	Public	Binary (normal vs. pneumonia)	Accuracy 97%	DenseNet
An et al., 2024	Deep CNN, Transfer Learning	Chest X-ray Pneumonia Dataset	Public	Binary (normal vs. pneumonia)	Accuracy 97%	CNN
Usman et al., 2025	CNN, ResNet-50	RSNA Pneumonia Dataset	Public	Binary (normal vs. pneumonia)	Accuracy 98%	ResNet-50

Ahmed et al., 2023	Deep CNN	<u>Chest X-ray Pneumonia Dataset</u>	Public	Binary ,multi class	Accuracy 98.72%	CNN
Al-Reshan et al., 2023	CNN, MobileNet	<u>Chest X-ray Pneumonia Dataset</u>	Public	Binary (normal vs. pneumonia)	Accuracy 97%	MobileNet

References

- [1] Singh, S., Kumar, M., Kumar, A., Verma, B. K., Abhishek, K., & Selvarajan, S. (2024). Efficient pneumonia detection using Vision Transformers on chest X-rays. Scientific Reports, 14, 2487. <https://doi.org/10.1038/s41598-024-52703-2>
- [2] Mustapha, B., et al. (2025). Enhanced pneumonia detection in chest X-rays using hybrid convolutional and vision transformer networks. Current Medical Imaging. <https://doi.org/10.2174/0115734056326685>
- [3] Shomsomi, M., Triawan, W., & Purwadi. (2026). Classification of pneumonia using CNN and vision transformer. Journal of Artificial Intelligence and Engineering Applications (JAIEA), 5(2), 2495-2504. <https://doi.org/10.59934/jaiea.v5i2.1906>
- [4] Kareem, O., et al. (2024). Pneumonia detection using pretrained networks and federated learning on real-time datasets. Biomedical Engineering Letters.
- [5] Alghamdi, H. S., et al. (2024). Pneumonia detection from chest X-ray images using DL and transfer learning for imbalanced datasets. Biomedical Engineering Letters. <https://pmc.ncbi.nlm.nih.gov/articles/PMC12344030/>
- [6] Boukhe, M., et al. (2025). Trustworthy pneumonia detection in chest X-ray imaging through attention-guided DL. Scientific Reports. <https://doi.org/10.1038/s41598-025-23664-x>
- [7] Ridouane, F. L., et al. (2025). PELM: A DL model for early detection of pneumonia in chest radiography. Applied Sciences, 15(12), 6487. <https://doi.org/10.3390/app15126487>
- [8] Devaux, M., Lacombe, J.-N., Debray, M.-P., Claessens, Y.-E., Duval, X., et al. (2024). Validating the accuracy of DL for the diagnosis of pneumonia on chest X-ray against a robust multimodal reference diagnosis. European Radiology Experimental, 8, 20. <https://doi.org/10.1186/s41747-023-00416-y>

- [9] Buriboev, A. S., et al. (2025). Binary classification of pneumonia in chest X-ray images using convolutional neural networks. *Sensors*, 25(13), 3976.
<https://doi.org/10.3390/s25133976>
- [10] Saber, A., et al. (2025). Efficient and accurate pneumonia detection using deep learning. *Sensors*, 25(23), 7233.
<https://doi.org/10.3390/s25237233>
- [11] Bunde, M., et al. (2024). Pneumonia image classification using DenseNet. *Information*, 15(10), 611.
<https://doi.org/10.3390/info15100611>
- [12] An et al., (2024) Deep convolutional neural network for pneumonia detection using chest X-ray images. *Diagnostics*, 14(4), 390.
<https://doi.org/10.3390/diagnostics14040390>
- [13] Usman, C., et al. (2025). Pneumonia disease detection using chest X-rays and machine learning. *Algorithms*, 18(2), 82.
<https://doi.org/10.3390/a18020082>
- [14] Ahmed, M. S., et al. (2023). Joint diagnosis of pneumonia and other lung diseases using deep learning from chest X-ray images. *Diagnostics*, 13(15), 2562.
<https://doi.org/10.3390/diagnostics13152562>
- [15] Al-Reshan, M. S. A., et al. (2023). Detection of pneumonia from chest X-ray images utilizing MobileNet model. *Healthcare*, 11(11), 1561.
<https://doi.org/10.3390/healthcare11111561>
- [16] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). *Gradient-based learning applied to document recognition*. Proceedings of the IEEE.
- [17] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). *Densely Connected Convolutional Networks*. CVPR.
- [18] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. CVPR.
- [19] Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. ICML.

[20] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2020). *An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale*. arXiv preprint arXiv:2010.11929.

[21] Sokolova, M., & Lapalme, G. (2009). *A systematic analysis of performance measures for classification tasks*. Information Processing & Management.

[22] Powers, D. M. W. (2011). *Evaluation: From Precision, Recall and F-measure to ROC, Informed Ness, Markedness & Correlation*. Journal of Machine Learning Technologies.

[23] Russell, S., & Norvig, P. (2021). *Artificial Intelligence: A Modern Approach* (4th ed.). Pearson.

[24] Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill.

[25] World Health Organization (WHO). (2023). *Pneumonia*.
<https://www.who.int/news-room/fact-sheets/detail/pneumonia>

[26] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). *MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications*.

[27] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., & Guo, B. (2021). *Swin Transformer: Hierarchical Vision Transformer using Shifted Windows*. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV).