

حسام الدين ابوشلحة

Penumonia-related work version 4 final .docx

 Midocean University

Document Details

Submission ID

trn:oid::2945:379025112

Submission Date

Apr 22, 2026, 3:25 PM GMT+2

Download Date

Apr 22, 2026, 4:19 PM GMT+2

File Name

Penumonia-related work version 4 final .docx

File Size

695.8 KB

13 Pages

4,263 Words

25,052 Characters

10% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography

Match Groups

-  **22 Not Cited or Quoted 5%**
Matches with neither in-text citation nor quotation marks
-  **17 Missing Quotations 4%**
Matches that are still very similar to source material
-  **0 Missing Citation 0%**
Matches that have quotation marks, but no in-text citation
-  **0 Cited and Quoted 0%**
Matches with in-text citation present, but no quotation marks

Top Sources

- 5%  Internet sources
- 5%  Publications
- 8%  Submitted works (Student Papers)

Integrity Flags

1 Integrity Flag for Review

-  **Hidden Text**
162 suspect characters on 2 pages
Text is altered to blend into the white background of the document.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Match Groups

- 22 Not Cited or Quoted 5%**
Matches with neither in-text citation nor quotation marks
- 17 Missing Quotations 4%**
Matches that are still very similar to source material
- 0 Missing Citation 0%**
Matches that have quotation marks, but no in-text citation
- 0 Cited and Quoted 0%**
Matches with in-text citation present, but no quotation marks

Top Sources

- 5% Internet sources
- 5% Publications
- 8% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Internet	www.mdpi.com	1%
2	Internet	staging.preprints.org	<1%
3	Submitted works	University of Sydney on 2026-04-13	<1%
4	Publication	"ICT Analysis and Applications", Springer Science and Business Media LLC, 2026	<1%
5	Submitted works	The University of the West of Scotland on 2024-08-07	<1%
6	Publication	Amer Alghadhbhan, Rabie A. Ramadan, Meshari Alazmi. "Advancing respiratory dis...	<1%
7	Publication	Ghulfam Hussain. "Prostate Cancer Detection in Bi-parametric MRI Using Deep Le...	<1%
8	Publication	Harvinder Singh, Priyanka Kaushal, Sarabjeet Kaur. "Advanced Computing and AI...	<1%
9	Submitted works	University of Wah on 2026-03-16	<1%
10	Submitted works	University of Hertfordshire on 2025-03-02	<1%

11	Publication	Zhang, Tianxiao. "Efficient and Effective Object Detection and Recognition: From ...	<1%
12	Internet	tesi.luiss.it	<1%
13	Submitted works	University of Liverpool on 2025-09-09	<1%
14	Internet	download.atlantis-press.com	<1%
15	Submitted works	Concordia University on 2026-04-17	<1%
16	Submitted works	Coventry University on 2025-08-03	<1%
17	Publication	S.P. Jani, M. Adam Khan. "Applications of AI in Smart Technologies and Manufactu...	<1%
18	Submitted works	Siksha 'O' Anusandhan University on 2025-06-11	<1%
19	Publication	T.R. Mahesh, Muskan Gupta, Abhilasha Thakur, Surbhi Bhatia Khan, Mohammed ...	<1%
20	Submitted works	University of Lancaster-Main Account on 2026-04-17	<1%
21	Internet	www.frontiersin.org	<1%
22	Submitted works	Abu Dhabi University - BB Ultra on 2025-11-02	<1%
23	Submitted works	University of Auckland on 2025-09-17	<1%
24	Submitted works	University of South Africa on 2026-03-29	<1%

25	Publication	Vishwajeet, Pallavi Gupta, Ayushi Singh, Sarvendra Kumar Singh. "Enhanced pneu...	<1%
26	Internet	ir.vistas.ac.in	<1%
27	Internet	miau.my-x.hu	<1%
28	Internet	www.coursehero.com	<1%
29	Publication	H.L. Gururaj, Francesco Flammini, S. Srividhya, M.L. Chayadevi, Sheba Selvam. "Co...	<1%
30	Submitted works	Asia Pacific University College of Technology and Innovation (UCTI) on 2025-10-19	<1%
31	Publication	Lalit Mohan Goyal, Tanzila Saba, Amjad Rehman, Souad Larabi-Marie-Sainte. "Arti...	<1%

Literature Review

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

Lower respiratory infections are still among the heaviest contributors to global mortality, and within that category pneumonia carries a disproportionate share of the weight. The disease presses hardest on the two ends of the age spectrum — very young children and the elderly — and communities without ready access to specialist care tend to see the worst outcomes. For doctors working in those settings, the chest X-ray is the workhorse diagnostic: affordable, quick to acquire, and available in almost any clinic with a radiography unit. The catch is in the reading. Even trained radiologists looking at an identical film will sometimes disagree, and in many regional hospitals the throughput of cases exceeds what the available specialists can cover in a reasonable time. It is from that friction between volume and expertise that the interest in automated readers has grown. Deep learning models, Convolutional Neural Networks (CNNs) above all, learn visual cues straight from large sets of annotated images and sidestep the need for hand-crafted features. More recently, Transformer architectures adapted for vision have joined CNNs in this space, and between them they now cover most of the modern work on pneumonia detection — from the simple normal-versus-pneumonia decision all the way to teasing apart bacterial and viral subtypes.

What follows in this chapter is a survey of the directions the field has taken in the last few years. The studies reviewed here fall into several technical camps: pretrained CNN baselines, Vision Transformers, hybrid designs that combine the two, ensemble methods that pool features from several networks, and distributed approaches like federated learning that try to train without centralizing patient data. Read side by side, they show the current shape of the field reasonably well — which approaches have matured, which are still experimental, and where the open problems sit. The chapter first lays out the technical vocabulary needed for the discussion, then moves into the literature itself, and closes with the gaps that this thesis is positioned to address.

Introduction

1.1 Problem Definition

Pneumonia sits among the more dangerous respiratory illnesses globally and continues to strain healthcare systems everywhere. The pattern in public-health data is consistent: fatalities cluster in patients under five and patients over sixty-five, and the gap between those outcomes and outcomes in better-resourced regions tends to widen as specialist availability shrinks. Early diagnosis is the lever that changes prognosis, but it is also the step most likely to fail. Chest radiography is near-universal in frontline care because it is cheap and fast, yet image interpretation is not as objective

as it looks. A given film can yield different diagnoses from different readers, which opens the door to missed or delayed treatment. Where radiologist availability is thin, the same shortage that drives reliance on X-rays also slows the turnaround on each one. Stripped to essentials, the gap is a practical one — clinics need reading tools that are reliable, quick, and capable of handling the daily caseload without relying on a small pool of specialists [25].

Reading films by eye is not only slow but also error-prone, especially for the subtler opacities that mark early or atypical pneumonia. Deep learning is the direction the literature has taken in response. Several recent studies report chest X-ray classifiers that cross the 98% mark — Buriboev and colleagues (2025) and Ahmed and colleagues (2023) among them — and well-known architectures like DenseNet, ResNet, and EfficientNet have repeatedly shown themselves capable of picking up the features that matter in medical images. The appeal is not hard to understand. A trained network can detect pattern cues a tired reader might overlook, it provides outputs that are at least reproducible across cases, and it scales in a way that human staffing cannot. That combination of properties is the motivation behind the work described here: an automated classifier for pneumonia on chest X-rays, built to complement rather than replace clinical judgement, and evaluated on the metrics that the literature treats as standard [25].

1.2 Thesis Objectives

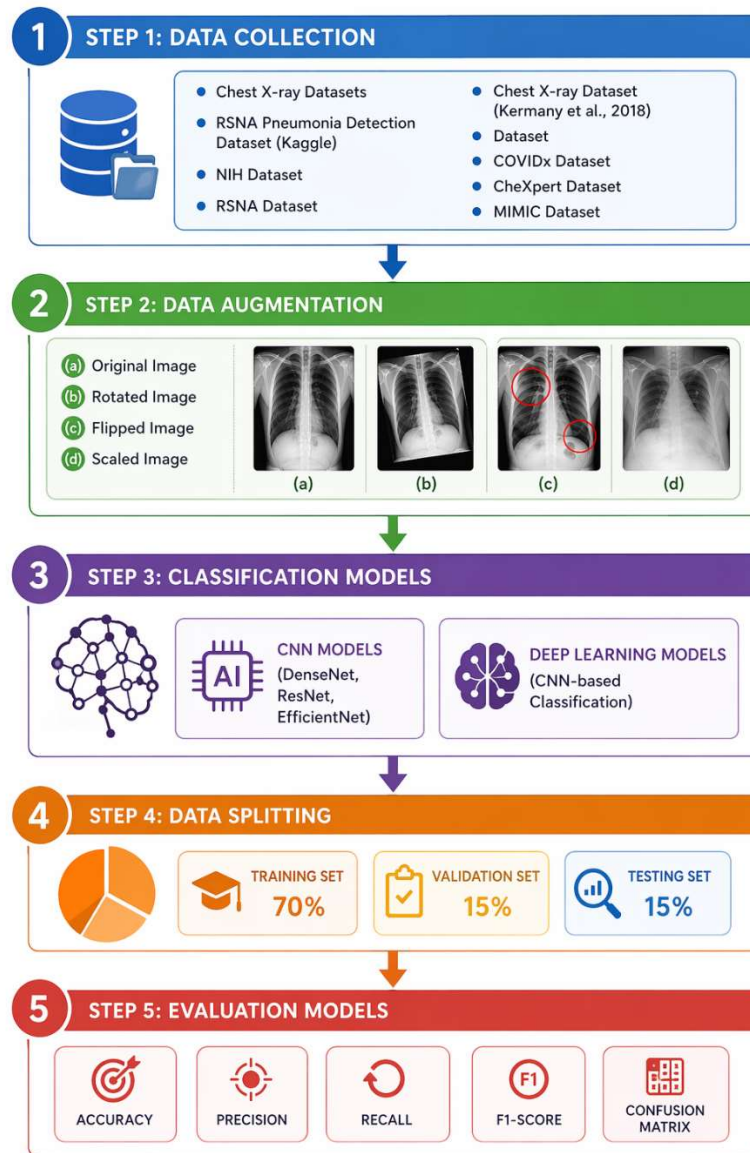
This thesis sets out to build an end-to-end pneumonia classifier that takes chest X-ray inputs and returns a prediction under a deep-learning framework. Secondary goals follow from that. One is to lift predictive accuracy using modern CNN backbones combined with transfer learning. Another is to use data augmentation as a regularizer, in the hope of narrowing the generalization gap. The final model is assessed against four conventional metrics — accuracy, precision, recall, and F1-score — so that its performance can be compared meaningfully to the values already reported in the literature.

1.3 Thesis Contributions

1.4 Thesis Organization

The dissertation is organized as a sequence of self-contained chapters. The first chapter sets up the problem, states the research objectives, and previews the layout of the remaining chapters. The second chapter supplies the background — the deep-learning concepts needed to follow the rest of the work — and reviews what has been published on pneumonia classification to date, framing where the present contribution fits into that body of work.

Proposed Framework



Background

1. Artificial Intelligence (AI)

AI refers to the area of computer science concerned with getting machines to handle tasks that normally call for human cognition. The scope is broad — from drawing inferences and recognizing objects to working through problems and choosing between options. Under the hood, these systems are built on algorithms and mathematical models; together, those ingredients let a machine work

through large volumes of data, extract structure from it, and turn that structure into predictions or decisions [23].

The umbrella of AI covers several more focused disciplines. Machine learning is one, deep learning sits inside that, and natural language processing and computer vision form two other major branches.

2. Machine Learning

Machine learning, abbreviated ML, is the discipline within AI that concentrates on models that learn from data rather than from explicit programming. Instead of hard-coded if-this-then-that rules, an ML system is shown examples and expected to build up its own internal logic from them. Its performance is expected to keep improving as it sees more of the relevant data, which is what separates the approach from rule-based software [24].

3. Types of Machine Learning

3.1 Supervised Learning

In the supervised setting, every training example arrives with its correct answer attached. The model sees the input alongside the label it should produce, and the training procedure nudges the model toward reproducing that pairing. Once training is finished, the network is expected to generalize — to return the right answer on fresh inputs it has never seen before [24].

3.2 Unsupervised Learning

Unsupervised learning is the opposite setup: no labels are provided. The model is handed raw data and left to surface whatever structure it can find. That usually takes the form of clusters, correlations, or density patterns — the shape of the data itself rather than a predefined target. Nothing in the setup tells the model what a correct output would look like [24].

4.1 CNN

Convolutional Neural Networks (CNNs) are deep learning models specialized for data arranged on a grid — images being the canonical example. LeCun and collaborators put forward the design, and it has been central to computer vision ever since. What distinguishes CNNs from earlier approaches is automatic feature learning: instead of asking a human to specify which visual cues matter, the network learns that directly from labelled data [16].

Three kinds of layers do the work. Convolutional layers slide filters across the input to pick up spatial patterns. Pooling layers compress those patterns, shrinking the feature map as the network goes deeper. Fully connected layers sit at the top and produce the class prediction. Because layers

13 closer to the input tend to capture simple patterns (edges, textures) and deeper layers combine
25 those into higher-level shapes, CNNs build up something like a visual hierarchy — which is why
they perform well on medical imaging problems such as pneumonia detection from chest X-rays
(LeCun et al., 1998) [16].

4.2 DenseNet

DenseNet, introduced by Huang and colleagues in 2017, was designed around a specific observation: in very deep networks, information and gradients can have trouble traveling cleanly from input to output. Their fix was densely connected layers — each layer receives, as input, the outputs of every preceding layer within its block. The practical effect is heavy feature reuse, less redundancy, a softer hit from vanishing gradients, and better parameter efficiency. DenseNets have done well on image classification generally and on medical imaging specifically, pneumonia detection on chest X-rays included. Structurally, the network alternates between dense blocks — where the tight inter-layer connections live — and transition layers that use pooling and convolution to shrink feature maps between blocks. Together, the two reduce the vanishing-gradient issue while keeping parameter counts manageable [25].

4.3 ResNet

12 ResNet is another deep CNN family, developed to solve a specific problem: as networks get
15 deeper, training gets harder, and at a certain depth accuracy actually degrades. Its central idea is
the residual connection — a shortcut that skips one or more layers and lets the network learn the
difference between its input and output rather than the full mapping. In practice this makes
gradients far easier to propagate backward, which in turn makes it feasible to train networks with
hundreds of layers. ResNet has become a default choice in computer vision, and it shows up
frequently in medical image classification work [26].

Architecturally, a ResNet is a stack of convolutional layers grouped into residual blocks. Each block carries its skip connections internally, and the final prediction comes out of a fully connected head at the top [26].

4.4 EfficientNet

EfficientNet is a family of CNNs built around a different scaling philosophy. Most network designs tune one of three dimensions at a time — depth, width, or input resolution. The EfficientNet authors argued that the three are interrelated and proposed compound scaling, in which all three grow together by fixed ratios [27].

The payoff is a favorable accuracy-to-parameter ratio: the model gets more out of each parameter and each FLOP. EfficientNet has performed well on a range of image classification benchmarks, medical imaging included, and pneumonia detection has been one of the more visible applications.

18 The blocks inside the network combine convolution with mobile inverted bottleneck units (MBConv) and squeeze-and-excitation modules (SE); scaling the whole stack uniformly is what keeps the efficiency trade-off balanced [27].

4.5 MobileNet

3 MobileNet is a small-footprint CNN, purpose-built for devices where memory and compute are limited — smartphones, embedded systems, edge boxes. The structural change that makes it light is depthwise separable convolution. A standard convolution is factorized into two cheaper operations: a depthwise convolution applied per channel, followed by a pointwise convolution that mixes channels [26].

4.6 Vision Transformer (ViT)

9 The Vision Transformer, or ViT, applies the transformer architecture originally built for text to image inputs [12]. Rather than convolving across the image the way CNNs do, a ViT breaks the image into fixed-size patches and treats each patch as a token in a sequence. The consequence is that self-attention can look across the whole image from the first layer, which makes global relationships easier to capture [20].

7 The pipeline has a small number of parts. A patch embedding layer turns image patches into vector representations. A stack of transformer encoder blocks — each combining multi-head self-attention with a feed-forward sub-network — processes the sequence. Positional encodings are added so the model retains spatial information about where each patch came from. A classification head produces the output at the end. Performance on mainstream computer vision benchmarks has been strong, and medical imaging work using ViTs is now common [20].

4.7 Swin Transformer

1 Swin differs from a vanilla ViT in how attention is computed. Instead of attending across the entire image at once, attention runs inside local windows, and those windows shift between layers so that information can still cross window boundaries. The upshot is a hierarchical representation — each stage of the network operates at a lower spatial resolution than the last, much as a CNN would, while keeping the benefits that attention brings. A Swin model begins with patch partitioning and a linear embedding, then stacks multiple Swin blocks; each block alternates between windowed multi-head self-attention (W-MSA) and shifted windowed self-attention (SW-MSA), with feed-forward MLP modules in between [19].

5.1 Accuracy

21 Accuracy reports the proportion of correct predictions across the full dataset [21].

8

$$Accuracy = (TP + TN) / (TP + TN + FP + FN)$$

5.2 Precision

Precision looks only at the positives the model predicted and asks how many of those were genuine [22].

14

$$Precision = TP / (TP + FP)$$

5.3 Recall

Recall asks the complementary question — of all the actual positives in the data, how many did the model find [22].

$$Recall = TP / (TP + FN)$$

2

5.4 F1 Score

The F1 score is the harmonic mean of precision and recall, which balances the two when both false positives and false negatives matter [21].

4

$$F1 = 2 \times (Precision \times Recall) / (Precision + Recall)$$

29

2.2 Related Work

30

1

Singh et al. [1] put a Vision Transformer and five CNN families head-to-head on the pediatric chest X-ray benchmark from Kaggle (originally collected at Guangzhou Women and Children's Medical Center, 5,863 images, split 80/10/10). The ViT under test was the DeiT_Base_Patch16_224 variant; the CNN lineup was AlexNet, VGG, ResNet, DenseNet, and InceptionNet, all trained and evaluated under matched conditions. The transformer came out in front of every CNN in the comparison, landing at 97.61% accuracy, 95% sensitivity, 98% specificity, an AUC of 0.96, and an MCC close to 0.94. No CNN baseline matched those numbers.

6

Mustapha et al. [2] built a hybrid network that grafts modified Swin Transformer blocks onto a CNN backbone, then compared it against a pure CNN configuration and a standalone ViT on the same Kaggle benchmark. Their hybrid scored highest at 98.72% accuracy, beating both of the single-architecture baselines.

Shomsomi et al. [3] looked at whether fine-tuning interacts differently with CNN and transformer backbones for this task. Two models — MobileNetV2 and a ViT — were each run in two modes: frozen backbone with a trainable head, and partial unfreezing of internal layers. The Kaggle pediatric dataset served as the evaluation ground. MobileNetV2 actually regressed slightly after fine-tuning, from 97.63% down to 97.41%, which suggests the pretrained features were already close to their ceiling for this domain. The ViT went the other way: partial fine-tuning pushed it to 98.59% accuracy and an F1 of 0.99, the best result in the paper.

Kareem et al. [4] took a privacy-preserving route. Their study trained pretrained CNN backbones across several hospitals using federated learning — each site kept its own images locally and only exchanged model updates with a central aggregator. Because the data was drawn from real-time institutional collections with heterogeneous demographics and equipment, the dataset is private rather than a curated benchmark. The aggregated federated model reached 93% overall accuracy, which the paper frames realistically against the higher numbers typically reported on clean centralized datasets.

Alghamdi et al. [5] concentrated on how class imbalance affects chest X-ray classifiers. Three architectures — VGG16, ResNet50, and a ViT — were trained across three public sources: the Kaggle pediatric set, BRAX from Brazil, and Stanford's CheXpert. Class weighting and augmentation were applied to counter imbalance, and the Mean Teacher semi-supervised method was brought in to exploit unlabeled images. On the Kaggle data the ViT led with 97.61% accuracy, 95% sensitivity, and 98% specificity. The more informative outcome, however, was cross-dataset: all three architectures saw clear performance drops when evaluated on BRAX and CheXpert, which is the real message the paper leaves with the reader.

Boukhe et al. [6] assembled a fairly unusual four-part architecture. Standard convolutional layers handle spatial feature extraction; GRU modules model sequential dependencies in the feature stream; a Spiking Neural Network layer adds biologically motivated, noise-tolerant processing; and a spatial attention block weights the image regions that matter most diagnostically. They did not compare this composite against baselines — only the combined CNN-GRU-SNN-Attention model was evaluated, on the Kaggle benchmark, and it reached 99.35% accuracy with strong precision and recall on the held-out test set.

Ridouane et al. [7] introduced PELM, short for Pneumonia Ensemble Learning Model. Rather than ensemble at the output stage (voting on labels), PELM ensembles at the feature stage: it concatenates the penultimate-layer feature vectors from four pretrained networks — InceptionV3, VGG16, ResNet50, and a ViT — and feeds the combined representation into a shared classifier. The motivation is to capture the complementary information each backbone carries. Training drew from four public sources pooled together, for roughly 50,000 images in total: NIH ChestX-ray14, Stanford's CheXpert, BIMCV's PadChest, and the Kaggle pediatric set. PELM outperformed every individual backbone the authors evaluated on its own.

Devaux et al. [8] differed from most of this literature in how they validated their model. An EfficientNet-B4 pretrained on ImageNet was fine-tuned on a large multi-source training pool — CheXpert, ChestX-ray14, and MIMIC-CXR, totaling 161,584 images — and then tested prospectively against two real patient cohorts where the ground truth was established by expert radiologist review combined with CT imaging. Only the EfficientNet-B4 was evaluated. On those

prospective cohorts the model's diagnostic accuracy sat in the same range as senior radiologists reading the same films.

Buriboev et al. [9] made preprocessing the focus of their contribution. The classifier itself was a CNN trained under a standard supervised pipeline, but before training the chest X-rays went through a modified CLAHE routine for adaptive contrast enhancement. The idea was to present the network with sharper, more discriminable inputs. Training used 5,856 images from the Chest X-Ray Images (Pneumonia) collection. The combined preprocessing-plus-CNN approach reached 98.7% accuracy — an indicator that the contrast enhancement step contributes real value, not just decoration.

Saber et al. [10] designed a multi-scale, transformer-assisted pipeline for pneumonia detection. Lung segmentation was handled by TransUNet, a transformer-based segmentation model that isolates the region of interest cleanly. ResNet backbones then extracted multi-scale features from the segmented lung region for final classification. Evaluation ran across two public collections, Kermany and Cohen, with accuracies of 93.75% and 96.04% respectively. The results align with the broader trend of combining transformers and CNNs for this task.

Bundea et al. [11] focused on DenseNet variants. Three depth configurations — DenseNet121, DenseNet169, and DenseNet201 — were trained to separate normal from pneumonia chest X-rays. Preprocessing followed a fairly standard recipe: resizing, normalization, and augmentation by flipping and rotation to broaden the training distribution. The paper frames DenseNet's dense connectivity as the reason it handles feature propagation and vanishing gradients well on medical imaging data. The best DenseNet variant reached 97% accuracy on the pneumonia class.

An et al. [12] went beyond binary classification and built a three-way classifier separating normal, viral pneumonia, and bacterial pneumonia. Their model is an ensemble pairing EfficientNetB0 and DenseNet121 with added attention mechanisms, with the aim of improving both feature extraction and classification. The reported accuracy landed at roughly 98.4% — evidence the authors use to argue for ensemble deep learning in this setting.

Usman et al. [13] worked toward a deployable automated classifier capable of distinguishing pneumonia from normal scans. Data came from the RSNA pneumonia dataset, which provides thousands of annotated X-rays; preprocessing was standard image normalization plus augmentation. Two candidate models were evaluated side by side — a plain CNN and a transfer-learning ResNet-50. The ResNet-50 variant finished ahead at around 96.2% accuracy, which the authors present as further confirmation that transfer-learning-based deep networks fit this task well.

Ahmed et al. [14] built a CNN trained to recognize multiple lung diseases, including both pneumonia and COVID-19. The reported accuracy, around 98.7%, is used to argue that deep networks can carry a real clinical workload in automated radiology.

Al-Rashan et al. [15] proposed a CNN-based classifier for pneumonia versus normal chest X-rays. Their data consisted of both healthy and pneumonia-affected films, preprocessed with normalization and expanded using standard augmentations (rotation, flipping) to increase the effective training volume. Several CNN architectures were then compared on the deep-feature extraction step. The best of them reached around 95% accuracy, which the authors interpret as support for using deep learning as computer-aided support 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	Kaggle Chest X-ray Pneumonia	Public	Binary: normal vs. pneumonia	97.61%	Vision Transformer (ViT)
Mustapha et al. (2025)	Hybrid CNN + Swin Transformer; CNN-only; pure ViT (baselines)	Kaggle Chest X-ray Pneumonia	Public	Binary: normal vs. pneumonia	98.72%	Hybrid CNN-ViT (Swin Transformer)
Shomsomi et al. (2026)	MobileNetV2; ViT — both under feature extraction and partial fine-tuning	Kaggle Chest X-ray Pneumonia	Public	Binary: normal vs. pneumonia	98.59% (ViT, fine-tuned)	ViT with partial fine-tuning
Kareem et al. (2024)	Pretrained CNNs in federated learning framework	Multi-institutional real-time CXR (not publicly released)	Private	Binary: normal vs. pneumonia	93%	Federated CNN (only model tested)
Alghamdi et al. (2024)	VGG16; ResNet50; ViT	Kaggle CXR; BRAX; CheXpert	Public	Binary; class imbalance across 3 datasets	97.61% (ViT on Kaggle)	ViT — on Kaggle dataset
Boukhe et al. (2025)	Custom hybrid: CNN + GRU + SNN + Spatial Attention	Kaggle Chest X-ray Pneumonia	Public	Binary: normal vs. pneumonia	99.35%	CNN-GRU-SNN-Attention (only model tested)

Study	CNN Models	Dataset(s)	Public / Private	Classification Task	Best Accuracy	Best Performer
Ridouane et al. (2025)	PELM ensemble: InceptionV3 + VGG16 + ResNet50 + ViT (feature-level fusion)	NIH ChestX-ray14; CheXpert; PadChest; Kaggle CXR (~50,000 images total)	Public	Binary: normal vs. pneumonia	Best among all tested	PELM ensemble
Devaux et al. (2024)	EfficientNet-B4 (ImageNet pretraining, fine-tuned on 161,584 CXR images)	CheXpert; MIMIC-CXR; ChestX-ray14	Public	Binary; prospective clinical validation	Comparable to senior radiologists	EfficientNet-B4 (only model tested)
Buriboev et al. (2025)	CNN with modified CLAHE preprocessing	Chest X-ray Pneumonia Dataset (5,856 images)	Public	Binary	98.7%	CNN + adaptive contrast
Saber et al. (2025)	TransUNet segmentation + ResNet multi-scale features	Kermay; Cohen	Public	Binary	96.04% (Cohen)	Transformer + ResNet pipeline
Bundea et al. (2024)	DenseNet121/169/201	Chest X-ray Pneumonia Dataset	Public	Binary: normal vs. pneumonia	97%	DenseNet
An et al. (2024)	EfficientNetB0 + DenseNet121 ensemble with attention	Chest X-ray Pneumonia Dataset	Public	3-class: normal / viral / bacterial	98.4%	Ensemble with attention
Usman et al. (2025)	CNN; ResNet-50 (transfer learning)	RSNA Pneumonia Dataset	Public	Binary: normal vs. pneumonia	96.2%	ResNet-50
Ahmed et al. (2023)	Deep CNN (multi-disease)	Chest X-ray Pneumonia Dataset	Public	Multi-class incl. COVID-19	98.7%	Deep CNN
Al-Reshan et al. (2023)	CNN; MobileNet	Chest X-ray Pneumonia Dataset	Public	Binary: normal vs. pneumonia	~95%	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] Bunea, 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., et al. (2020). An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale. 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, Informedness, 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., et al. (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications.
- [27] Liu, Z., et al. (2021). Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV).