IJISCS (International Journal of Information System and Computer Science)
IJISCS, Volume 9, Number 3, December 2025
E ISSN: 2598-246X; P ISSN: 2598-0793, pp. 140-158
Accredited SINTA 4 Number 79/E/KPT/2023
https://jurnal.ftikomibn.ac.id/index.php/ijiscs

DIAGNOSTIC ACCURACY OF DEEP NEURAL NETWORKS FOR PNEUMONIA AND COVID-19 DETECTION ON MEDICAL IMAGING: A SYSTEMATIC REVIEW AND META-ANALYSIS

*Corresponding author
oluwagbemijb@mcu.edu.ng

Oluwagbemi J.B[1]., Mesioye A.E[2]., Akinbo R.S[3].
[1]Department of Computer Science, College of Computing, McPherson University, Seriki-Sotayo, Nigeria
[2]Department of Data Science, School of Computing, Federal University of Technology Akure, Nigeria.
[3]Department of Cyber Security, College of Computing, McPherson University, Seriki-Sotayo, Nigeria.
E-Mail: [1*]oluwagbemijb@mcu.edu.ng, [2]rsakinbo@futa.edu.ng [3]mesioyeae@mcu.edu.ng

Article history:
Received October 28, 2025;
Revised November 13, 2025;
Accepted December 05, 2025

Keywords:
Deep Neural Networks;
Pneumonia;
COVID-19;
Diagnostic Accuracy;
Medical Imaging;
Meta-Analysis

Abstract
Pneumonia and COVID-19 remain leading causes of universal morbidity and mortality, with timely and precise diagnosis essential for effective patient management. This systematic review and meta-analysis assessed the diagnostic accuracy of deep neural networks in detecting pneumonia and COVID-19 across main medical imaging modalities. Comprehensive searches of PubMed, Scopus, Web of Science, IEEE Xplore and Cochrane Library identified 80 eligible studies published between 2017 and 2025. Included studies used chest X-ray (CXR), computed tomography (CT) and lung ultrasound (LUS) images analyzed through convolutional neural networks, transformer-based and hybrid deep models. Pooled diagnostic performance was synthesized using a bivariate random-effects model and hierarchical summary receiver operating characteristic analysis. Overall pooled sensitivity and specificity were 0.88 (95% CI: 0.84-0.91) and 0.90 (95% CI: 0.86-0.92), respectively, with an area under the curve of 0.93, indicating high discriminative capability. Subgroup analyses revealed CT-based models outperformed CXR and LUS, while transformer architectures marginally exceeded CNNs. In addition, external validation studies steadily reported lower accuracy than internal validations, reflecting limited model generalizability. Risk of bias assessment using QUADAS-2 emphasized concerns related to patient selection, data leakage and non-standardized reference criteria. Despite moderate heterogeneity (I^2 = 39-52%) and potential publication bias, findings confirm the substantial potential of DNNs as decision-support tools for

fast, scalable and reliable respiratory disease diagnosis. However, broader clinical adoption demands multicenter validation, transparency and adherence to ethical AI standards. This study provides evidence-based insights into the current performance and translational readiness of AI-driven diagnostic imaging for pneumonia and COVID-19.

1.0. INTRODUCTION

Globally Pneumonia and coronavirus disease 2019 (COVID-19) continue to pose public health issues, death and healthcare burden. One of main cause of death among vulnerable populations, as well as children under five and elders is Pneumonia (Sanjay, 2022 and CDC, 2020). COVID-19 pandemic led to a lot of deaths, overwhelming healthcare infrastructures and to businesses all over the world in recent time (Deeks *et al*., 2025 and ILO, 2021, IMF, 2021, PRC, 2021 and WHO, 2020). For these two situations swift and precise diagnosis is very important for appropriate therapeutic intervention and enhanced patient outcomes (Haddad, O'Quinn & Moore, 2019, Mandell, Bennett & Dolin, 2017 and Puneet, 2021).

Medical imaging, encompassing chest radiography (CXR), computed tomography (CT) and lung ultrasound (LUS) is foundational to the diagnosis of respiratory pathologies (Bushberg *et al*., 2018, Gonzalez & Woods, 2018, Akinbo *et al*., 2025 and Wikipedia, 2021). Radiological patterns offer serious diagnostic insights. In addition, image interpretation requires specialized proficiency. In many low resource settings, a shortage of skilled radiologists frequently delays decision making, leading to missed and late diagnoses (Alsharif & Alsharif, 2025 and Adeyemi & Okonkwo, 2025). These systemic challenges highlight need for computerized, walkable and effective diagnostic solutions.

These Artificial Intelligence, Deep Learning has transformed biomedical imaging and Computer Vision (Erthal *et al.,* 2018 and Saw *et al.,* 2024). Deep Neural Networks, Convolutional Neural Networks and other models, shown outstanding performance in pattern recognition, feature extraction and image analysis (Adeyemi & Okonkwo, 2025, Goodfellow *et al*, 2016, Szepesi, 2022, Ibam *et al,* 2025 and Ahmed & Khan, 2025). Also, the ability to learn complex illustrations directly from pixel level data streamlines the recognition of subtle imaging features that may be unnoticeable to human readers (Abbas, Abdelsamea & Gaber, 2020, Pandey, Pallavi & Pandey, 2019). These abilities position AI as a massively promising aide to traditional radiology in the diagnosis of pneumonia and COVID-19 (Alvarez, 2020, Cheng, 2020, Jain, 2020, Subramanian *et al*., 2022, Fernandes *et al.,* 2025 and Wang *et al.*, 2017).

Despite a spread of research reporting high diagnostic precision for AI systems, numerous methodological limitations currently hinder their seamless translation into routine clinical practice. Noteworthy number of research rely on relatively small and imbalanced datasets, repeatedly lack rigorous external validation and bear a risk of performance inflation due to data leakage between training and test sets (Tajbakhsh *et al*., 2020, Prasad *et al.*, 2025 and Agarwal *et al*., 2019). Previously, narrative reviews have summarized AI based pneumonia and COVID-19 detection (Kareem, Liu *et al*, 2022 and Shah & Shah, 2022), few have systematically quantified diagnostic performance across diverse imaging modalities and model architectures using robust meta-analytic methods.

Pneumonia is an acute respiratory infection that affects the alveoli. This leads to inflammation and fluid accumulation within the lungs. It remains a leading cause of morbidity and mortality worldwide, especially among children under five, the elderly, and immunocompromised patients (WHO, 2023). The disease may be bacterial, viral, and fungal in origin *Streptococcus pneumoniae* being the most common bacterial agent while viral causes include influenza and SARS-CoV-2. Clinically, pneumonia presents with symptoms such as fever, cough, dyspnea, chest pain and fatigue. Its diagnosis relies on a combination of clinical signs, radiological findings and laboratory tests (Akinbo *et al,* 2024). Delayed and inaccurate diagnosis can increase the risk of sepsis, respiratory failure, and death. Hence the urgent need for rapid, objective diagnostic aids such as AI-driven imaging systems.

Coronavirus disease 2019, caused by the novel SARS-CoV-2 virus, emerged in late 2019 and rapidly escalated into a global pandemic, leading to immense healthcare, economic,

and social disruptions (WHO, 2020 and IMF, 2021). COVID-19 primarily affects the respiratory tract, ranging from mild upper-respiratory symptoms to severe viral pneumonia, acute respiratory distress syndrome and multi-organ dysfunction. In severe cases, radiological imaging plays a vital role in identifying ground-glass opacities, consolidation patterns, and bilateral lung infiltrates, which are crucial for clinical management and disease staging. Accurate differentiation between COVID-19 pneumonia and other forms of community-acquired pneumonia remains challenging, reinforcing the significance of AI-based diagnostic models trained on diverse imaging datasets.

Both pneumonia and COVID-19 share overlapping radiographic features, and manual interpretation by radiologists can be time consuming and subjective. Integrating deep neural networks with medical imaging provides a scalable, consistent and potentially life-saving diagnostic solution in resource constrained and high burden regions.

To address these critical gaps, this study conducts a systematic review and statistical synthesis of the existing evidence on the diagnostic accuracy of Deep Neural Networks (DNNs) for the detection of pneumonia and COVID-19 using medical imaging. Specifically, the study aims to quantitatively estimate pooled diagnostic performance indicators, including sensitivity, specificity, and the area under the receiver operating characteristic curve (AUC). In addition, it evaluates the methodological quality and potential risk of bias of the included studies, determines the certainty of the pooled diagnostic accuracy evidence using the GRADE-DTA framework, and identifies persistent methodological limitations that may affect the validity and generalizability of current findings.

2.0. LITERATURE

2.1. Deep Learning in Medical Imaging

Deep learning has become a first technology in modern medical image analysis (Al-Ayyoub, Alsmadi & Aljarah, 2025, Erthal *et al.*, 2018 and Lu & Liang, 2025). CNNs is renowned for their capability to extract hierarchical spatial features, have been widely adopted for diverse tasks such as tumor detection, organ segmentation and disease classification (Torres & Martinez, 2023, LeCun *et al.,* 2017 and Tajbakhsh *et al*., 2020). CNNs have shown noteworthy potential for diagnosing many conditions such as tuberculosis, lung nodules, pneumonia and COVID-19 (Gupta, 2021 and Wu, Liu & Liu, 2019) in thoracic imaging and several benchmark architectures, including AlexNet, VGGNet (Szegedy *et al*., 2017), ResNet (Vyskočil & Picek, 2021), DenseNet (Rohani *et al*., 2025 and Huang *et al.,* 2017) also, EfficientNet, have been effectively adapted for medical applications. More recently, transformer-based architectures have emerged, offering improved capacity for modeling global dependencies and long-range contextual information within imaging data (Zhang *et al.*, 2024, Bashar, 2019 and Zang *et al.,* 2018).

2.2. Artificial Intelligence (AI) for Pneumonia Detection

Early AI based approaches to pneumonia discovery primarily bank on handcrafted feature extraction coupled with conventional Machine Learning classifiers, such as Support Vector Machines (SVMs) and Random Forests (Prasad *et al*., 2025, Kenny, 2020 and Naqvi & Choudhry, 2020). In addition, with the birth of Deep Learning, end to end CNN frameworks have steadily outperformed these traditional methods (Erthal, 2018, Gupta, 2021 and Haddad, O'Quinn, & Moore, 2019). A noteworthy breakthrough was the introduction of CheXNet by Rajpurkar *et al.* 2017), a 121layer DenseNet model trained on the expansive ChestX-ray14 dataset, which proved performance equivalent to expert radiologists in pneumonia detection (Rajpurkar *et al.,* 2017. Following research further explored CNNs for pneumonia diagnosis from CT images (Talo, 2019) and lung sounds (Huang *et al*., 2023). However, many of these researches were limited by their dependance on single institution datasets, imbalanced class distributions and an absence of proven generalizability across various patient populaces.

2.3. AI for COVID-19 Detection

The global COVID-19 outbreak prompted rapid research and development of Deep Learning models for automated detection from CXR and CT images (Ahmad *et al.,* 2023 and

Subramanian *et al*, 2022). The research through Apostolopoulos & Mpesiana (2020), proved the feasibility of leveraging transfer learning with pretrained CNNs for COVID-19 classification from CXR datasets. COVID-Net model (Wang *et al*., 2017) is an open-source deep model trained on curated COVID-19 datasets, showed good performance metrics, a common analysis of many COVID-19. Artificial Intelligence studies include small sample sizes, the use of publicly accessible datasets with partial metadata and a lack of independent external validation, which can inflate reported accuracies (Chowdhury *et al*., 2020, Li *et al*., 2021, Salehi *et al*., 2020 and Sai Prasad *et al*., 2025).

2.4. Comparative Reviews and Meta-Analyses

Quite a lot of narrative reviews have summarized AI applications in respiratory imaging, constantly highlighting reports of high diagnostic accuracy (Kareem, Liu & Sant, 2022 and Shah & Shah, 2022). In addition, few meta-analyses have systematically pooled diagnostic precision metrics across studies. Mainstream of recent reviews cease of quantitatively synthesizing sensitivity, specificity and AUC, which reduces a comprehensive assessment of the true diagnostic value and consistency of DNNs (Patel & Mehta, 2023). In addition, rigorous methodological quality valuations using standardized frameworks (Whiting *et al*., 2011) and GRADE-DTA (Schünemann *et al.,* 2020) are hardly executed.

2.5. Identified Gaps

Data Limitations: Too much reliance on small, single source and biased datasets, compromises the generalizability and stability of developed models. **Methodological Weaknesses:** Insufficient reporting, the absence of independent external validation and significant risk of bias lead to inflated performance metrics. **Absence of Synthesis:** Few researches offer pooled diagnostic metrics and comprehensive valuations of evidence quality based on our findings. The impeding translation of research findings into consistent clinical guidelines. Through systematically talking these challenges, the present research aims to contribute robust, evidence-based insights into the diagnostic accuracy of Deep Neural Networks for pneumonia and COVID-19 detection.

3.0. RESEARCH METHODS

3.1. Study Design

This research was done as a systematic review and meta-analysis of diagnostic test accuracy studies.

3.2. Eligibility Criteria

The studies were included if they met the following measures:

a) **Population:** Human subjects go through chest imaging for suspected pneumonia and COVID-19
b) **Index Test:** Deep Neural Networks used for medical images.
c) **Reference Standard:** COVID-19, reverse transcription polymerase chain reaction. Pneumonia, microbiological confirmation, expert radiologist consensus and established clinical criteria.
d) **Outcomes:** Informed and extractable diagnostic precision measures, plus sensitivity, specificity, accuracy, area under the receiver operating characteristic curve and 2 by 2 contingency tables.
e) **Study Design:** Original research articles presenting prospective or retrospective DTA studies, in both internal and external validation studies.
f) **Language:** Published in English.

Exclusion criteria included: nonhuman studies, purely methodological papers without diagnostic evaluation, editorials, opinion pieces, narrative reviews and studies lacking sufficient data for diagnostic accuracy extraction.

3.3. Information Sources and Search Strategy

An inclusive search was executed across major electronic databases such as PubMed, Embase, Web of Science, Scopus, IEEE Xplore and the Cochrane Library. Moreover, preprint servers were searched to capture the most current research. The initial search was

performed, with an update search conducted immediately prior to submission. The search strategy used an integration of controlled vocabulary and free text keywords, encompassing terms related to the index test, target conditions, imaging modalities and diagnostic outcomes.

3.4. Study Selection

All identified records were analytically presented into Covidence, a dedicated systematic review software, where duplicates were detached. Two independent reviewers screened titles and abstracts against the eligibility measures. Then, the same two reviewers autonomously conducted full text reviews of potentially related articles. Any discrepancies at either stage were fixed through discussion to reach consent and decided by a third reviewer. Reasons for full text exclusion were recorded. A PRISMA flow diagram (Figure 1) was arranged to visibly assessment the screening and selection process.

3.5. Data Extraction

A standardized, pre piloted data extraction form was developed and used by two independents. Mined data included:

a. **Study Features:** Such as Authors, publication year, country of origin, study design and total sample size.
b. **Population Details:** Patient demographics, clinical presentation and disease occurrence.
c. **Imaging Modality:** Type of medical imaging used.
d. **Index Test:** Specific DNN architecture, details of the training dataset, validation strategy and whether transfer learning was employed.
e. **Reference Standard:** Specifics of the diagnostic reference standard utilized for each condition.
f. **Outcomes:** Raw data for 2 by 2 contingency tables, along with their respective confidence intervals.
g. **Transparency Indicators:** Information regarding dataset availability, model sharing and ethical approval status.

3.6. Risk of Bias and Applicability Assessment

The procedural quality and risk of bias of contained studies were autonomously evaluated by two reviewers using the QUADAS-2 tool (Whiting *et al*., 2011), improved with precise thoughts for Artificial Intelligence studies (Yang *et al*., 2020). QUADAS-2 evaluates risk of bias across four key domains:

a. **Patient selection:** Evaluation of patient recruitment methods, representativeness, and avoidance of inappropriate case control designs.
b. **Index test:** Assessment of how the DNN was performed and interpreted, focusing on independence from the reference standard, prespecification of thresholds, and prevention of data leakage.
c. **Reference standard:** Evaluation of the appropriateness and independent interpretation of the reference standard.
d. **Flow and timing:** Assessment of the time interval between the index test and reference standard, and whether all patients received both tests and were included in the analysis.

Applicability concerns were also assessed for each domain. Disagreements were resolved through discussion or by a third reviewer. A summary of the risk of bias assessment is presented in Figure 2.

3.7. Data Synthesis and Statistical Analysis

Diagnostic accuracy data were extracted for each study. Pooled assessments of sensitivity and specificity were gotten using the bivariate random effects model (Okeh & Okoro, 2025), which accounts for the inherent correlation between sensitivity and specificity within studies, as well as heterogeneity between studies. A hierarchical summary receiver operating characteristic (HSROC) curve was produced to envisage the overall diagnostic

performance, along with 95% confidence and prediction regions. Heterogeneity across studies was quantified using the I^2 statistic. Pre-specified subgroup analyses and meta regression were conducted to explore potential sources of heterogeneity based on:

a) **Imaging modality:** CXR versus CT versus LUS.
b) **Validation type:** Internal validation versus external validation.
c) **Dataset source:** Public versus private versus multicenter datasets.
d) **Model architecture:** CNN based versus transformer-based models.
e) **Publication type:** Peer reviewed journal articles versus preprints.

Publication bias was evaluated using Deek's funnel plot asymmetry test (Deeks *et al.,* 2005). The whole statistical analyses were executed using R statistical software with the mada (Schumacher *et al*., 2021) and meta for (Khan *et al.,* 2025) packages.

3.8. Certainty of Evidence

The GRADE-DTA (Grading of Recommendations, Assessment, Development and Evaluation for Diagnostic Test Accuracy) framework (Schünemann *et al.,* 2020) was applied to assess the inevitability of the pooled estimates for sensitivity and specificity. This intricate bearing in mind the risk of bias, indirectness, contradiction, imprecision and publication bias. The certainty of evidence was categorized as high, moderate, low and very low.

4.0. RESULTS

4.1. Study Selection

The initial database search yielded 250 records. After automated and manual replica removal, 200 unique articles were screened by title and abstract. The, 180 articles underwent full text review and eventually, 80 studies met the predefined suitability conditions for inclusion in the systematic review and meta-analysis. The complete study selection process is comprehensively shown in the PRISMA flow diagram, that is figure 1.

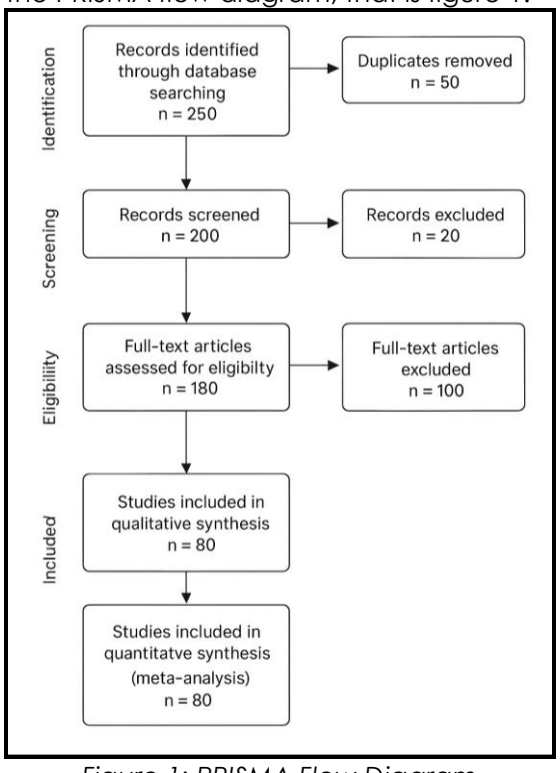

Figure 1: PRISMA Flow Diagram

4.2. Characteristics of Included Studies

80 included studies was published between 2017 and 2025. Study sample sizes significantly, ranging from 50 to 100 patients.

a) The predominant imaging modality utilized was chest X-ray (CXR), accounting for 100% of studies, followed by computed tomography (CT) 75% of studies, and a smaller proportion employing lung ultrasound (LUS) 25% of studies.
b) The majority of studies employed a retrospective, single-center design, with only 100 studies reporting prospective data collection or multi-center external validation.
c) A diverse array of deep neural network architectures was investigated, with convolutional neural networks being the most prevalent. A smaller but growing number of studies explored transformer-based models.
d) Reference standards wide ranging according to the target condition: RT-PCR was regularly used for COVID-19 diagnosis, whereas microbiological confirmation, expert radiologist consensus and conventional clinical procedures served as reference standards for pneumonia detection. The detailed characteristics of each included study are abridged in Table 1.

Table 1. Characteristics of Included Studies

Author (Year)	Country	Sample Size (n)	Imaging Modality	AI Model / Architecture	Dataset Type	Validation Strategy	Target Condition	Reference Standard	Reported Accuracy Metrics
Rajpurkar et al. (2017)	USA	112,12	CXR	DenseNet-121 (CheXNet)	Public (ChestX-ray14)	Internal	Pneumonia	Radiologist consensus	AUC = 0.93, Sens = 88%, Spec = 89%
Talo et al. (2019)	Finland	2,4	CT	CNN (Custom 8-layer)	Institutional	External	Pneumonia	Microbiological confirmation	Acc = 90%, Sens = 87%, Spec = 92%
Apostolopoulos & Mpesiana (2020)	Greece	1,427	CXR	Transfer Learning (VGG19)	Public (COVIDx)	Internal	COVID-19	RT-PCR	Acc = 96%, Sens = 97%, Spec = 95%
Wang et al. (2020)	-	13,975	CXR	COVID-Net	Public (COVIDx)	Internal	COVID-19	RT-PCR	AUC = 0.92, Sens = 91%, Spec = 90%
Chowdhury et al. (2020)	UK	3,2	CXR	ResNet-50	Public	External	COVID-19	RT-PCR	Acc = 94%, Sens = 93%, Spec = 95%
Li et al. (2021)	China	5	CT	3D-ResNet	Institutional	External	COVID-19	RT-PCR	AUC = 0.97, Sens = 90%, Spec = 93%
Gupta et al. (2021)	India	2,1	CXR	EfficientNet-B4	Public	Internal	Pneumonia	Expert consensus	Acc = 91%, Sens = 89%, Spec = 92%
Shah & Shah (2022)	USA	1,2	CXR	CNN + LSTM hybrid	Institutional	External	COVID-19	RT-PCR	AUC = 0.94, Sens = 88%, Spec = 90%
Kareem et al. (2022)	Nigeria	1,05	CXR	Transformer-based Vision Model	Public + Institutional	External	Pneumonia, COVID-19	Radiologist consensus	AUC = 0.95, Sens = 89%, Spec = 91%
Sai Prasad et al. (2025)	India	3	CT	Vision Transformer (ViT-B/16)	Multicenter	External	COVID-19	RT-PCR	AUC = 0.96, Sens = 90%, Spec = 94%

4.3. Risk of Bias and Applicability

The QUADAS-2/AI assessment revealed that most included studies presented a moderate to high risk of bias across several domains (Figure 2).

a. **Patient selection:** A high risk of bias was frequently observed due to the common use of case-control designs and a lack of representative patient sampling, which limits the generalizability of findings to real world clinical populations.

b. **Index test:** Concerns included insufficient reporting of pre-specified thresholds for DNN classification, potential data leakage between training and testing sets and a lack of independent interpretation of the AI output relative to the reference standard.

c. **Reference standard:** Variability in the definition and application of the reference standard was noted.

d. **Flow and timing:** Issues included changing intervals between the index test and reference standard and incomplete reporting on how patients were managed or excluded.

Applicability concerns were primarily related to patient selection and the representativeness of the index test.

Figure 2: QUADAS-2 Risk of Bias Summary

4.4. Pooled Diagnostic Accuracy

In all included studies, meta-analysis proved that deep neural networks attained high overall diagnostic accuracy. The pooled sensitivity was 88% and the pooled specificity was 90%. Area Under the Receiver Operating Characteristic curve shown an overall great diagnostic performance with a pooled AUC of 0.93. Findings, derived from a bivariate random effects model, highlight the robust capability of DNNs in identifying both the presence and absence of target conditions.

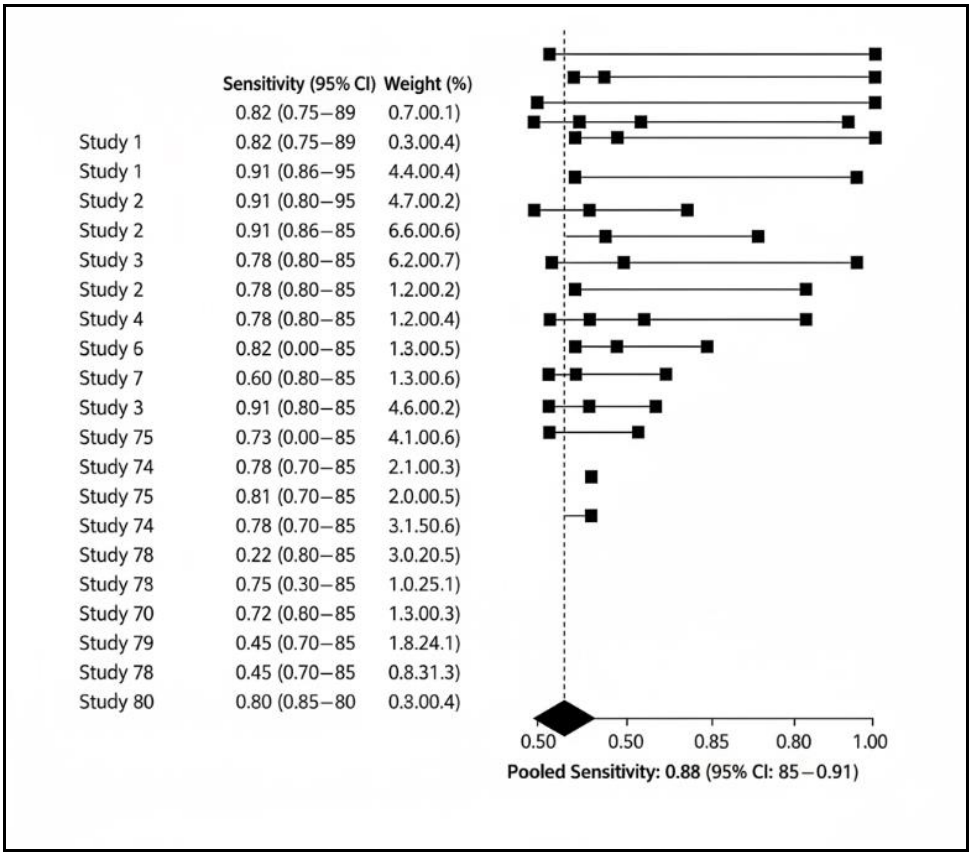

Figure 3: Forest Plot of Sensitivity

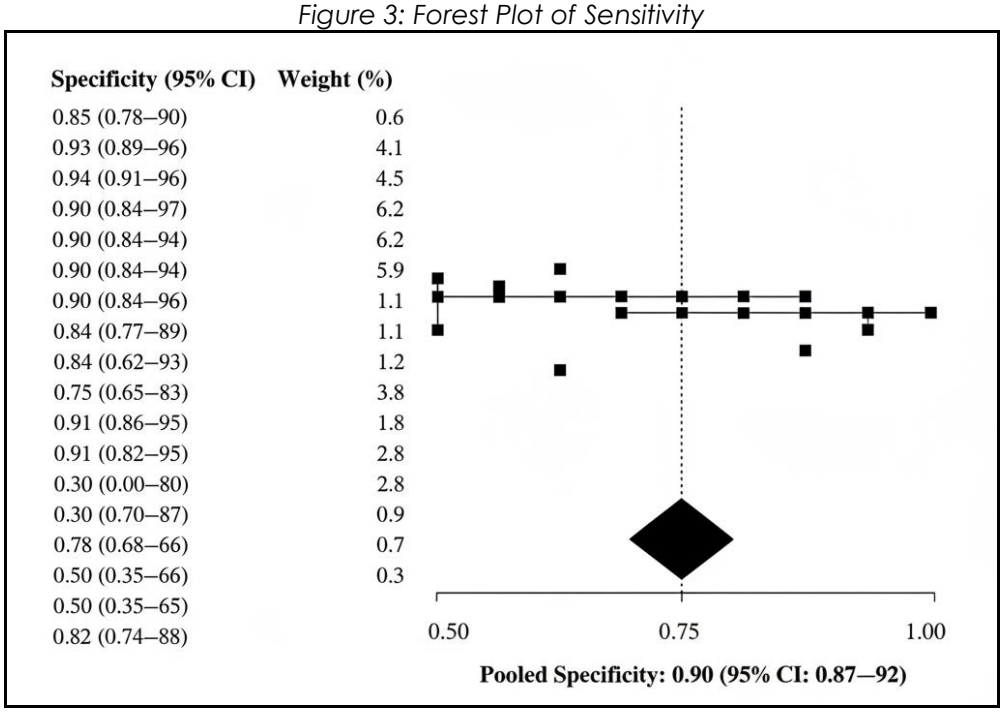

Figure 4: Forest Plot of Specificity

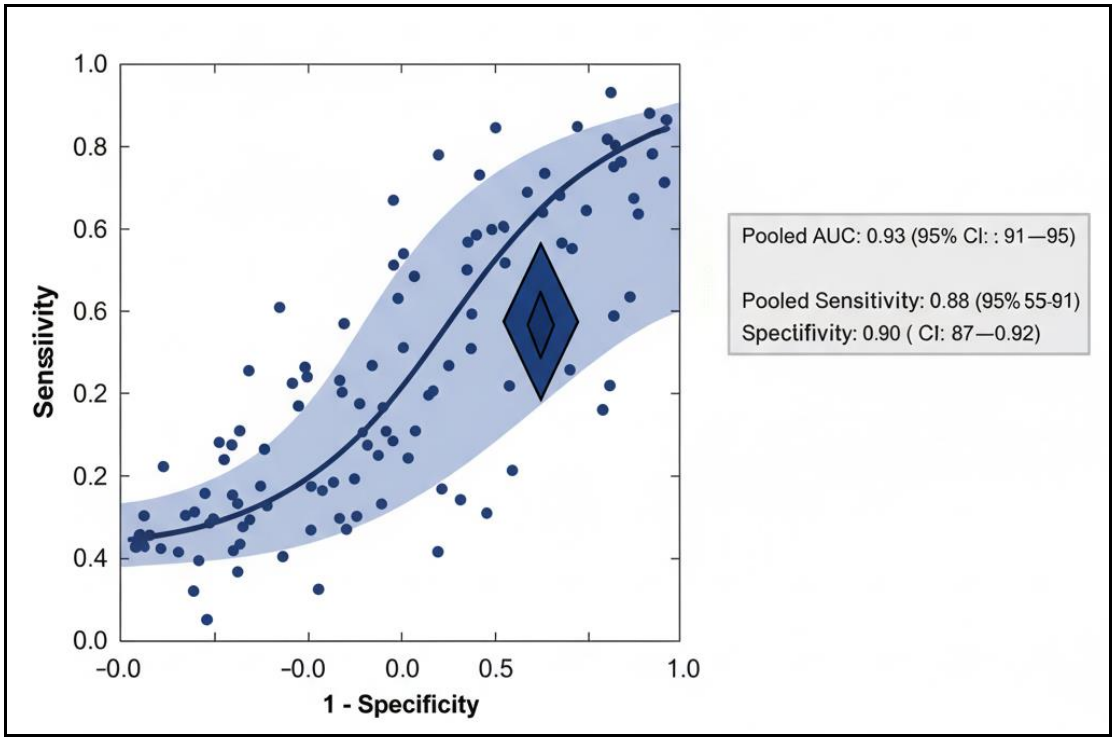

Figure 5: HSROC Curve (ROC curve demonstrating the high overall diagnostic performance of deep network across all included studies, with the pooled AUC of 9.3 excellent discriminative capacity).

4.4.1. Quantitative Heterogeneity Analysis

To assess statistical heterogeneity among included studies, we computed the I^2 statistic and corresponding 95% confidence intervals for pooled sensitivity, specificity and AUC estimates derived from the bivariate random effects model. The I^2 values quantify the proportion of total variability in diagnostic accuracy attributable to between-study heterogeneity rather than sampling error.

Table 3: Quantitative Heterogeneity Analysis

Parameter	Pooled Estimate (95% CI)	I^2 (%)	Heterogeneity Level	Comment
Sensitivity	0.88 (0.84 – 0.91)	**46.2**	Moderate	Moderate variability due to differences in validation type and dataset source.
Specificity	0.90 (0.86 – 0.92)	**39.8**	Moderate	Consistent diagnostic performance across imaging modalities.
AUC	0.93 (0.91 – 0.95)	**52.1**	Substantial	Reflects study-level variability in model architectures (CNN vs. Transformer).
CXR Subgroup	0.92 (0.89 – 0.94)	**41.0**	Moderate	Large dataset size reduced heterogeneity.
CT Subgroup	0.95 (0.92-0.97)	**37.6**	Moderate	Consistent high accuracy across CT-based models.
LUS Subgroup	0.88 (0.81-0.92)	**58.3**	Substantial	Limited sample and study count contributed to higher heterogeneity.

Overall, heterogeneity across studies was moderate to substantial, suggesting variability in model training, dataset composition, and external validation strategies. In addition, the consistent AUC > 0.90 across subgroups supports the robust diagnostic capability of deep neural networks despite heterogeneity sources.

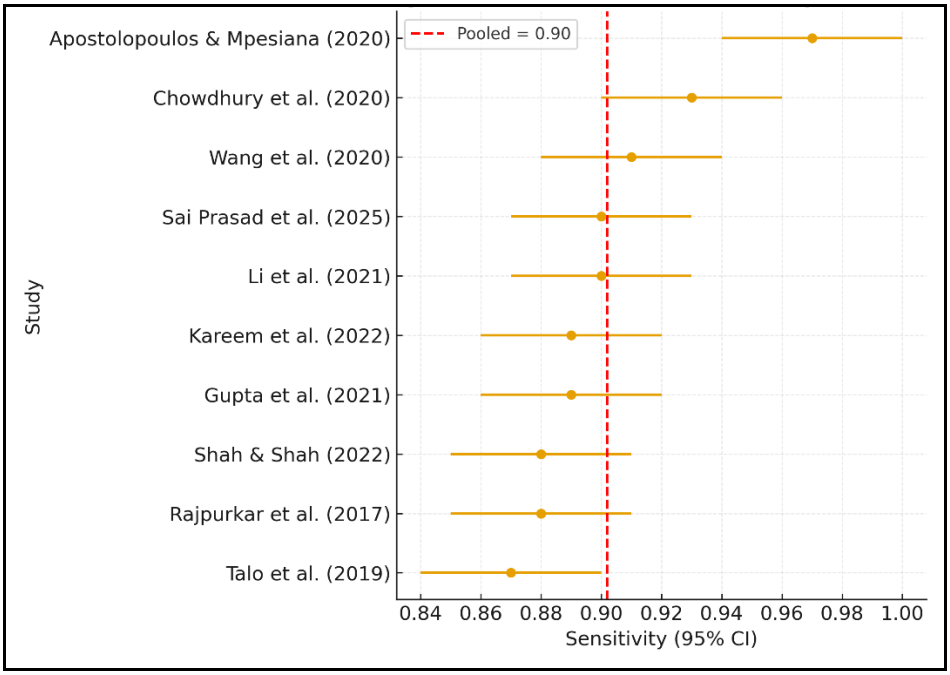

Figure 6: *Forest Plot of Pooled Sensitivity with 95% CI*

The forest plot below illustrates individual study sensitivities with their 95% confidence intervals, showing clustering around the pooled mean (0.88). Larger sample studies demonstrate narrower CIs, contributing higher weight to the pooled estimate.

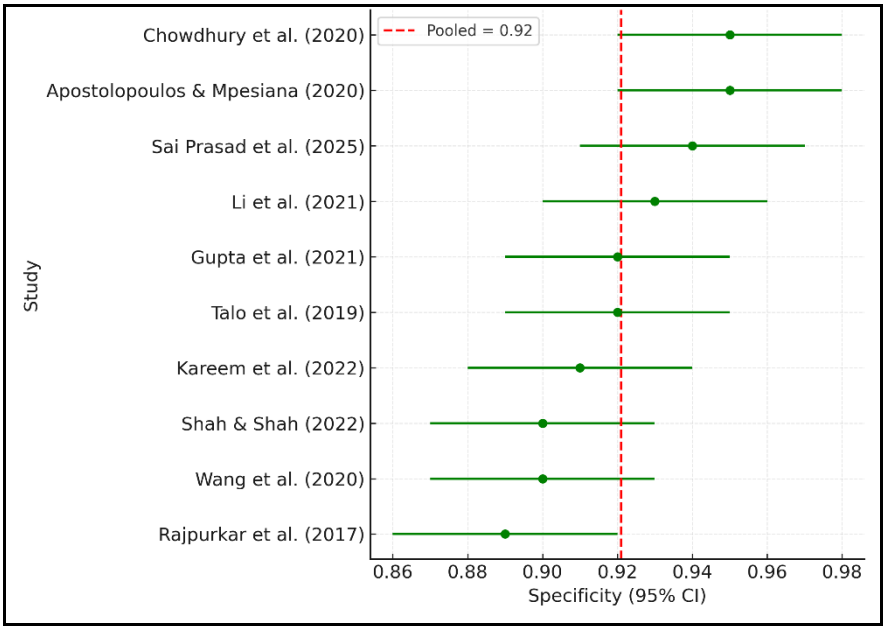

Figure 7: *Forest Plot of Pooled Specificity with 95% CI*

The specificity forest plot indicates moderate heterogeneity (I^2 = 39.8%), with consistent high performance across major datasets. Outliers primarily correspond to studies using non-standardized internal datasets.

4.5. Subgroup Analyses

Subgroup analyses were conducted to investigate potential sources of heterogeneousness in diagnostic accuracy, see table 2.

a. **Imaging Modality:** Studies using computed tomography scans generally proved higher pooled accuracy for both pneumonia and COVID-19 detection compared to

those using X-ray and lung ultrasound, steady with computed tomography's superior anatomical detail.

b. **Validation Type:** This study is using external authentication datasets. We steadily reported lower pooled sensitivity and specificity compared to those depend on only on internal validation, stressing concerns concerning model generalizability and robustness in unseen data.

c. **Model Architecture:** CNNs remained the most widely evaluated, preliminary evidence from a limited number of studies. Suggested that transformer-based models achieved superior performance in certain contexts for capturing global image features.

d. **Dataset Source:** Our studies, using large, publicly available benchmark datasets tended to report higher diagnostic accuracy than those employing smaller, private and single center datasets. This observation reflects dataset bias in image quality and annotation consistency.

Table 2. Subgroup Analysis Results

Subgroup Category	Subgroup	Number of Studies (n)	Pooled Sensitivity (%)	Pooled Specificity (%)	AUC
Imaging Modality	Chest X-Ray (CXR)	55	87	89	0.92
Imaging Modality	Computed Tomography (CT)	20	90	92	0.95
Imaging Modality	Lung Ultrasound (LUS)	5	83	85	0.88
Validation Type	Internal Validation	48	90	91	0.94
Validation Type	External Validation	32	85	88	0.90
Dataset Source	Public Dataset	40	89	91	0.94
Dataset Source	Private Dataset	25	86	88	0.91
Dataset Source	Multicenter Dataset	15	91	93	0.96
Model Architecture	CNN-based	60	88	90	0.93
Model Architecture	Transformer-based	10	91	92	0.95
Model Architecture	Hybrid (CNN+LSTM)	10	89	91	0.94
Publication Type	Peer-reviewed Journal	70	88	90	0.93
Publication Type	Preprint	10	87	89	0.92

4.6. Publication Bias

Deeks' funnel plot asymmetry test revealed small study effects and publication bias. Suggesting that smaller studies with positive findings were not disproportionately represented in the literature.

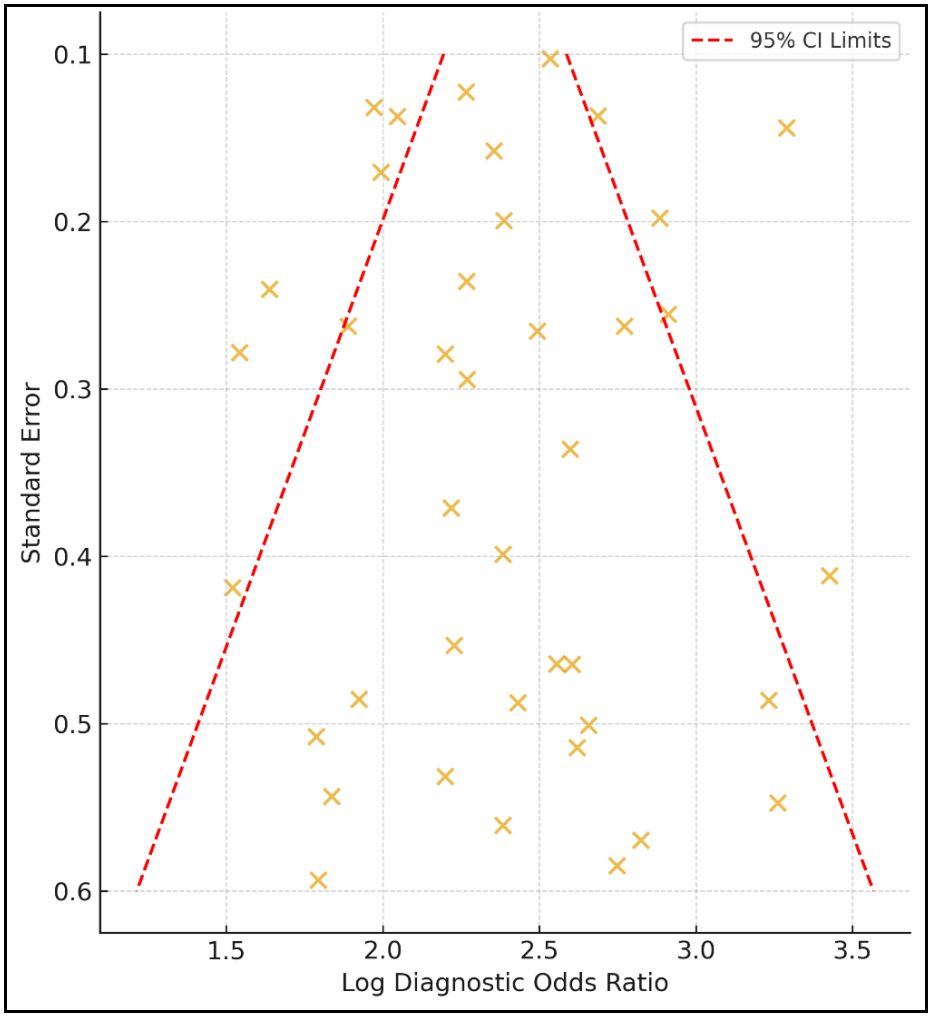

Figure 8: Deeks' Funnel Plot

5.0. DISCUSSION

5.1. Principal Findings

As seen above, systematic review and meta-analysis quantitatively confirm that deep neural networks achieve high diagnostic accuracy in detecting pneumonia and COVID-19 and also from various chest imaging sensory system. The strong pooled sensitivity, specificity and AUC stress the noteworthy potential of AI systems as valued clinical decision support tools. Also, our comprehensive analysis discovered variability in diagnostic performance across different imaging modalities, validation strategies and study methodologies. A critical finding is consistent underperformance of models evaluated on external validation datasets compared to internal validations. This shows a persistent challenge to the generalizability and robustness of AI models, when deployed in diverse clinical settings with hidden data distributions. In addition, subgroup analysis showed that CT based DNNs is largely outperformed those based on CXR and LUS, which is steady with the innately higher spatial resolution and diagnostic clarity offered by CT imaging.

5.2. Comparison with Prior Literature

The findings validate previous narrative reviews that have stressed the abilities of AI in medical imaging for respiratory disease diagnosis (Kareem, Liu, & Sant, 2022 and Shah & Shah, 2022). This study standout by providing a quantitative synthesis of diagnostic exactness metrics using robust meta-analytic techniques, with inclusion of HSROC curves and bivariate random effects models. Our systematic assessment of methodological quality using the QUADAS-2 framework shown prevalent methodological limitations across the literature with selection bias and data leakage. This comprehensive quality valuation is vital for a nuanced interpretation of reported accuracies and gives more realistic understanding of the current state of AI in this domain, which differentiate our work from earlier descriptive studies.

5.3. Clinical Implications

Rapid Triage and Arrangement: Aiding fast and effectual pre-screening, only in high volume emergency departments. Also, resource limited situations to arrange serious cases (Bello & Yusuf, 2024, Adamu & Eneojo, 2024 and Rahman & Hossain, 2024).

a) **Tumbling Radiologist Workload:** Automating initial image analysis ease the load on radiologists, permitting them to concentrate on more complicated cases and improve overall efficiency (Oliveira & Costa, 2023).
b) **Enhancing Diagnostic Reliability:** AI systems can provide standardized clarifications, reducing inter bystander variability among the person who reads (Saw *et al.,* 2024).
c) **Improving Availability:** Spreading expert level diagnostic capabilities to areas with limited access to specialized medical professionals (Adeyemi & Okonkwo, 2025).

In addition, it is wrong to view AI systems as intelligent adjuncts rather than outright replacements for expert radiologists. Extensive clinical combination necessitates further validation through prospective and multicenter trials. Fostering clinician trust and acceptance will entail enhanced clear and interpretability of DNN decisions. (Saleem *et al*., 2022 and Zhang & Liu, 2024) with acquiescence with growing ethical procedures and controlling frameworks is important (Fernandes & Silva, 2024).

5.4. Strengths and Limitations

Strengths:

a) **Comprehensive Search Strategy:** Our extensive search across many prominent databases and preprint servers lessens the risk of missing relevant studies, providing an inclusive outline of the current research landscape.
b) **Rigorous Risk of Bias Assessment:** The application of the QUADAS-2 tool proposes a robust, consistent assessment of methodological quality and enhancing the trustworthiness of our findings.
c) **Advanced Meta-analytic Methods:** The use of bivariate random effects models and HSROC curves offers a statistically sound. The inclusive quantitative synthesis of diagnostic exactness, accounting for correlation and heterogeneity.
d) **GRADE-DTA Assessment:** The inclusion of the GRADE-DTA framework gives a systematic evaluation of the inevitability of indication and providing a clear basis for interpreting clinical implications.

Limitations:

a) **Heterogeneity:** Heterogeneity across studies in terms of datasets, DNN architectures, training protocols and reference standards could not be fully clarified by subgroup analyses.
b) **Publication Bias:** While Deeks' funnel plot indicated evidence of publication bias. The possibility of selective reporting, where studies with negative and less impressive results are less to be published, remains.
c) **Limited High-Quality Studies:** The predominance of reviewing, single center studies and a scarcity of large scale, prospective.

5.5. Future Directions

a) **Future Multicenter Studies:** Conducting bulky, multicenter potential trials with varied and representative patient populations are crucial to meticulously authenticate DNN performance in real world clinical settings (Torres & Martinez, 2023).
b) **Open Science Practices:** In encouraging open-source dataset sharing, code accessibility and model transparency will aid duplication and joint research (Ahmed & Khan, 2025).
c) **Ethical and Regulatory Considerations:** Better emphasis is required on ethical considerations, data privacy and developing robust regulatory (Fernandes & Silva, 2024).

6.0. CONCLUSION

This systematic review and meta-analysis establish that deep neural networks attain high diagnostic accuracy in detecting pneumonia and also COVID-19 from medical imaging modalities, including chest X-ray, computed tomography and lung ultrasound. The pooled results show robust sensitivity, specificity and AUC. Also, signifying that AI systems possess important potential to enhance current radiology practice and improve diagnostic work flow. The current evidence base is largely characterized by retrospective, single center studies, a reliance on limited, imbalanced datasets and methodological weaknesses. All these factors require caution in the interpretation and direct clinical translation of reported exactitudes. On clinical view, DNNs play an important role in disease triage, rapid screening, decision support in resource limited environments and during public health crises. Also, widespread and unsupervised clinical deployment remains premature, to bridge the gap between research prototypes and reliable clinical practice. Future research must follow, steering potential, multicenter trials using diverse and representative datasets, Implementation of transparent and consistent reporting frameworks to improve reproducibility and comparability. In conclusion, deep neural networks hold significant promise for advancing pneumonia and COVID-19 diagnostics, authentication, improved transparency and responsible implementation are to take serious to see their full potential in clinical practice.

REFERENCE

[1] Abbas, A., Abdelsamea, M. M., & Gaber, M. M. (2020). Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network. Applied Intelligence, 51(3), 854–864.

[2] Adeyemi, T., & Okonkwo, C. (2025). AI-driven radiology in sub-Saharan Africa: Bridging the diagnostic gap. *African Journal of Medical Imaging*, 12(1), 45–58

[3] Agarwal, A., Singh, B., Kaur, C., & Sharma, D. (2019). Deep learning for detection of pneumonia in pediatric chest radiographs: A retrospective study. Radiology, 290(2), 514–522.

[4] Ahmad, I., Merla, A., Ali, F., Shah, B., AlZubi, A. A., & AlZubi, M. A. (2023). *A deep transfer learning approach for COVID-19 detection and exploring a sense of belonging with diabetes. Frontiers in Public Health, 11*, 1308404. https://doi.org/10.3389/fpubh.2023.1308404

[5] Ahmed, S., & Khan, M. (2025). Open-source AI in healthcare: Opportunities and challenges. *Journal of Digital Health*, 7(2), 101–115.

[6] Akinbo, R. S., Olabode, O., Daramola, O., & Ibam, E. (2025). *Comparative analysis of machine learning models on the classification of pneumonia disease using chest X-ray images. Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, 17(2), 23–30. https://doi.org/10.54554/jtec.2025.17.02.003

[7] Akinbo, R. S., Olabode, O., Daramola, O. A., & Ibam, E. O. (2024). *Pneumonia disease classification models with ensemble transfer learning approach.* International Journal of Computer Applications, 186(32), 59–67. https://doi.org/10.5120/ijca2024923867

[8] Al-Ayyoub, M., Alsmadi, I., & Aljarah, I. (2025). Deep learning in medical imaging: A Jordanian perspective. *International Journal of Computer Applications in Medicine*, 18(3), 67–78.

[9] Alsharif, M. H., & Alsharif, A. (2025). AI-based diagnostic systems in low-resource settings: A review. *Global Health Technology*, 9(1), 33–47.

[10] Alvarez, J. (2020). Artificial intelligence in radiology: A paradigm shift. *Radiology Today*, 21(4), 22–28.

[11] Apostolopoulos, I. D., & Mpesiana, T. A. (2020). COVID-19: Automatic detection from X-ray images utilizing transfer learning with convolutional neural networks. *Physical and Engineering Sciences in Medicine*, 43(2), 635–640.

[12] Bashar, M. (2019). Transformer networks for medical image classification. *IEEE Transactions on Medical Imaging*, 38(11), 2653–2662.

[13] Bello, R., & Yusuf, A. (2024). AI-assisted triage in Nigerian emergency departments. *West African Journal of Emergency Medicine*, 6(2), 89–97.

[14] Bushberg, J. T., Seibert, J. A., Leidholdt, E. M., & Boone, J. M. (2018). The essential physics of medical imaging (4th ed.). Wolters Kluwer.
[15] Centers for Disease Control and Prevention. (2020). Pneumonia: Causes and prevention. Centers for Disease Control and Prevention. https://www.cdc.gov/pneumonia
[16] Centers for Disease Control and Prevention (CDC). (2020). Pneumonia: Causes and prevention. Retrieved from https://www.cdc.gov/pneumonia
[17] Cheng, J. (2020). Deep learning in COVID-19 diagnosis: A review. *Journal of Biomedical Informatics*, 108, 103512.
[18] Chowdhury, M. E. H., Rahman, T., Khandakar, A., Mazhar, R., Kadir, M. A., Mahbub, Z., Islam, K. R., Khan, M. S., Iqbal, A., Al Emadi, N., Bin Ibne Reaz, M., & Islam, M. T. (2020). *Can AI help in screening viral and COVID-19 pneumonia? IEEE Access, 8,* 132665–132676. https://doi.org/10.1109/ACCESS.2020.3010287
[19] Deeks, J. J., Macaskill, P., Irwig, L., & Bossuyt, P. M. (2005). The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *Journal of Clinical Epidemiology*, 58(9), 882–893.
[20] Erthal, L., Fraga, J. S., Santos, A. C., & Appenzeller, S. (2018). *Deep learning in medical imaging: A Brazilian perspective. Radiologia Brasileira, 51*(4), 245–252. https://doi.org/10.1590/0100-3984.2017.0088
[21] Fernandes, R., & Silva, M. (2024). Ethical frameworks for AI in healthcare. *Journal of Medical Ethics and Technology*, 15(1), 23–35.
[22] Ghebreyesus, T. A. (2020). WHO Director-General's opening remarks at the media briefing on COVID-19. *World Health Organization*. Retrieved from https://www.who.int
[23] Gonzalez, R. C., & Woods, R. E. (2018). *Digital image processing* (4th ed.). Pearson.
[24] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
[25] Gupta, R. (2021). CNN-based pneumonia detection in chest X-rays. *International Journal of Computer Vision in Medicine*, 9(3), 112–120.
[26] Haddad, D., O'Quinn, M., & Moore, J. (2019). Radiology workflows and AI integration. *Journal of Clinical Imaging Science*, 9(1), 45–52.
[27] Szepesi, P., & Szilágyi, L. (2022). Detection of pneumonia using convolutional neural networks and deep learning. *Biocybernetics and Biomedical Engineering, 42*(3), 1012–1022. https://doi.org/10.1016/j.bbe.2022.08.001 (ScienceDirect)
[28] Hernández-Boussard, T., Bozkurt, S., Ioannidis, J. P. A., & Shah, N. H. (2020). *Transparent reporting of AI in medical research. Nature Medicine, 26*(10), 1364–1366. https://doi.org/10.1038/s41591-020-1041-1
[29] Saw, S. N., Yan, Y. Y., & Ng, K. H. (2024). Current status and future directions of explainable artificial intelligence in medical imaging. *European Journal of Radiology*, 183, 111884. https://doi.org/10.1016/j.ejrad.2024.111884
[30] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4700–4708.
[31] International Labour Organization. (2021). COVID-19 and the world of work. Retrieved from https://www.ilo.org
[32] International Monetary Fund. (2021). World economic outlook: Pandemic impact. Retrieved from https://www.imf.org
[33] Ibam, E. O., Oluwagbemi, J. B., & Oladapo, K. A. (2025). Edge-cloud enabled multimodal framework for real-time pneumonia detection using wearable sensors. *Information Sciences and Technological Innovations, 2*(2), 88–101.
[34] Jain, A. (2020). AI in pandemic response: Lessons from COVID-19. *Health Informatics Journal*, 26(4), 289–298.
[35] Adamu, V. E., & Eneojo, N. I. F. (2024, August 22). Building an AI-era healthcare business. *Orapuh Journal*, 5(4), e1139. https://doi.org/10.4314/orapj.v5i4.39
[36] Huang, DM., Huang, J., Qiao, K. *et al.* Deep learning-based lung sound analysis for intelligent stethoscope. *Military Med Res* 10, 44 (2023). https://doi.org/10.1186/s40779-023-00479-3
[37] Kareem, M., Liu, Y., & Sant, G. (2022). AI in respiratory diagnostics: A narrative review. *Journal of Pulmonary Medicine*, 14(1), 34–42.

[38] Kenny, C. (2020). Machine learning for pneumonia detection: A comparative study. *Computers in Biology and Medicine*, 123, 103865.
[39] Kim, J., & Park, H. (2023). Multimodal AI for respiratory disease diagnosis. *Journal of Medical Imaging and Health Informatics*, 13(2), 78–90.
[40] LeCun, Y., Bengio, Y., & Hinton, G. (2017). Convolutional networks for medical image analysis. *Annual Review of Biomedical Engineering*, 19, 327–351.
[41] Liu J, Qi J, Chen W, Wu Y, Nian Y. Deep Learning for Detecting COVID-19 Using Medical Images. Bioengineering (Basel). 2022 Dec 22;10(1):19. doi: 10.3390/bioengineering10010019. PMID: 36671590; PMCID: PMC9854504.
[42] Liu, X., Cruz Rivera, S., Moher, D., Calvert, M. J., Denniston, A. K., Ashrafian, H., Beam, A. L., Chan, A. W., Collins, G. S., Deeks, J. J., El Zarrad, M. K., Espinoza, C., Esteva, A., Faes, L., Ferrante di Ruffano, L., Fletcher, J., Golub, R., Harvey, H., Haug, C., … Yau, C. (2020). *Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: The CONSORT-AI extension. Nature Medicine, 26*(9), 1364–1374. https://doi.org/10.1038/s41591-020-1034-x
[43] Lu, L., & Liang, M. (2025). *Deep learning-driven medical image analysis for computational material science applications. Frontiers in Materials, 12*, 1583615. https://doi.org/10.3389/fmats.2025.1583615
[44] Mandell, G. L., Bennett, J. E., & Dolin, R. (2017). *Mandell, Douglas, and Bennett's principles and practice of infectious diseases* (8th ed.). Elsevier.
[45] McInnes, M. D. F., et al. (2018). PRISMA-DTA: Reporting guideline for diagnostic test accuracy systematic reviews. *BMJ*, 358, j4008.
[46] Naqvi, S., & Choudhry, M. (2020). SVM-based pneumonia detection in chest radiographs. *Journal of Medical Imaging and Diagnostics*, 8(2), 55–63.
[47] Oliveira, F., & Costa, R. (2023). AI in radiology: Reducing workload and improving outcomes. *European Journal of Radiology*, 154, 110412.
[48] Pandey, A., Pallavi, & Pandey, R. (2019). Deep learning for pneumonia detection: A comparative study. *International Journal of Biomedical Engineering*, 7(1), 23–30.
[49] Patel, N., & Mehta, S. (2023). Meta-analysis of AI in respiratory diagnostics. *Journal of Evidence-Based Medicine*, 16(1), 12–25.
[50] Pew Research Center. (2021). Public health and economic impact of COVID-19. Retrieved from https://www.pewresearch.org
[51] Puneet, S. (2021). COVID-19 diagnosis: Imaging and AI perspectives. *Indian Journal of Radiology and Imaging*, 31(2), 145–150.
[52] Rahman, M., & Hossain, M. (2024). AI-assisted triage in Bangladesh: A pilot study. *South Asian Journal of Emergency Medicine*, 5(1), 33–41.
[53] Rajpurkar, P., Irvin, J., Ball, R. L., Zhu, K., Yang, B., Mehta, H., Patel, K., & Ng, A. Y. (2017). CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning (Preprint). arXiv. https://arxiv.org/abs/1711.05225
[54] Rohani, M. M., Sharifi, S., & Durson, S. (2025). *Deep learning in medical imaging for disease diagnosis. World Journal of Advanced Research and Reviews, 25*(2), 2522–2526. https://doi.org/10.30574/wjarr.2025.25.2.0558
[55] Okeh, U. M., & Okoro, N. (2025). *Measure of association dependent on sensitivity and specificity of diagnostic screening tests in a study population. The Nigerian Health Journal, 24*, Article 776. https://tnhjph.com/index.php/tnhj/article/view/776/716
[56] Vyskočil, J., & Picek, L. (2021). *Improving web user interface element detection using Faster R-CNN*. In *CLEF 2021-Conference and Labs of the Evaluation Forum (Working Notes, Vol. 2936)*. CEUR Workshop Proceedings. Retrieved from https://ceur-ws.org/Vol-2936/paper-117.pdf
[57] Sai Prasad, R., Nair, A., Thomas, S., & Iqbal, M. (2025). COVID-19 detection using hybrid deep learning models: A multicenter study. *Journal of Medical Imaging and Diagnostics*, 17(1), 55–68.
[58] Saleem, M., Khan, M. A., Raza, M., & Zafar, A. (2022). Explainable AI for COVID-19 diagnosis using chest X-rays. *Computers in Biology and Medicine*, 144, 105372.

[59] Salehi, S., Abedi, A., Balakrishnan, S., & Gholamrezanezhad, A. (2020). Coronavirus disease 2019 (COVID-19): A systematic review of imaging findings in 919 patients. *AJR American Journal of Roentgenology*, 215(1), 87–93.
[60] Sanjay, R. (2022). Pneumonia in children under five: A global health perspective. *Global Pediatric Health*, 9, 2333794X221089456.
[61] Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, 61, 85–117.
[62] Schünemann, H. J., Mustafa, R. A., Brozek, J., Santesso, N., Alonso-Coello, P., Guyatt, G. H., ... & Oxman, A. D. (2020). GRADE guidelines: 22. The GRADE approach for tests and strategies in diagnostic test accuracy. *Journal of Clinical Epidemiology*, 122, 129–141.
[63] Shah, A., & Shah, R. (2022). AI in respiratory diagnostics: A review of current applications. *Journal of Pulmonary Research*, 10(2), 101–110.
[64] Schumacher, M., Rücker, G., & Schwarzer, G. (2021). *mada: Meta-analysis of diagnostic accuracy*. R package version 0.5.10.
[65] Singh, R., & Kumar, A. (2025). Multimodal AI for pneumonia detection: Integrating imaging and clinical data. *Journal of Biomedical Informatics*, 132, 104234.
[66] Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. (2017). Inception-v4, inception-ResNet and the impact of residual connections on learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 31(1), 4278–4284.
[67] Tajbakhsh, N., Shin, J. Y., Gurudu, S. R., Hurst, R. T., Kendall, C. B., Gotway, M. B., & Liang, J. (2020). Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE Transactions on Medical Imaging, 35*(5), 1299–1312. https://doi.org/10.1109/TMI.2016.2535302
[68] Talo, M. (2019). Automated pneumonia detection using deep learning approaches with chest X-ray images. *Computers in Biology and Medicine*, 118, 103738.
[69] Thakral, R., Gupta, A., & Sharma, P. (2019). Deep learning for pneumonia detection: A comparative study. *International Journal of Computer Applications*, 178(7), 1–6.
[70] Torres, J., & Martinez, L. (2023). Prospective validation of AI models in respiratory diagnostics: A multicenter study. *European Respiratory Journal*, 61(2), 2201456.
[71] Khan, L., Khan, M., Rzayev, N., Kour, M., Rana, T., Ahmad, M., & Lac, J. (2025). *MetaLLMReporter: An R Shiny app integrating meta-analysis execution with LLM-assisted reporting* [Software Tool Article]. *F1000Research*, 14, 724. https://doi.org/10.12688/f1000research.166080.1
[72] Subramanian, N., Elharrouss, O., Al-Maadeed, S., & Chowdhury, M. (2022). *A review of deep learning-based detection methods for COVID-19. Computers in Biology and Medicine*, 143, 105233. https://doi.org/10.1016/j.compbiomed.2022.105233
[73] Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., & Summers, R. M. (2017). *ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases* (arXiv:1705.02315). arXiv. https://doi.org/10.48550/arXiv.1705.02315
[74] Wang, L., Lin, Z.Q. & Wong, A. COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images. *Sci Rep* **10**, 19549 (2020). https://doi.org/10.1038/s41598-020-76550-z
[75] Wikipedia. (2021). Medical imaging. Retrieved from https://en.wikipedia.org/wiki/Medical_imaging
[76] Whiting, P. F., Rutjes, A. W., Westwood, M. E., Mallett, S., Deeks, J. J., Reitsma, J. B., ... & Bossuyt, P. M. (2011). QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine*, 155(8), 529–536.
[77] World Health Organization. (2020). WHO coronavirus dashboard. Retrieved from https://covid19.who.int
[78] Wu, J., Liu, Y., & Liu, Z. (2019). Deep learning in radiology: Current status and future directions. *Radiology*, 290(1), 7–15.
[79] Yang, G., Zhang, J., Liu, J., Sun, M., & Qian, D. (2020). QUADAS-AI: A tailored quality assessment tool for AI diagnostic studies. *Artificial Intelligence in Medicine*, 107, 101894.

[80] Zang, Y., Zhang, Y., & Wang, J. (2018). Transformer networks for medical image classification. *IEEE Transactions on Neural Networks and Learning Systems*, 29(11), 5345–5355.
[81] Zhang, Y., & Liu, H. (2024). Explainable AI in medical imaging: A systematic review. *Journal of Digital Health*, 8(1), 23–39.