



Enhancing Diagnostic Accuracy in Nigerian Hospitals: Analysis of Deep Learning Applications in Medical Imaging

*¹Yahya Umar Muhammad, ²Abdulrahman Nasiru Sada and ³Umar Faruk Muktar

¹Department of General Studies, Katsina State College of Health Sciences and Technology, Katsina State, Nigeria.

²Department of Information Technology, Federal University Dutsinma, Katsina State, Nigeria.

³Department of Information Technology, Federal University Dutsinma, Katsina State, Nigeria.

*Corresponding authors' email: yumarnine@gmail.com

ABSTRACT

Nigeria faces a critical shortage of radiologists and other medical specialists, creating significant challenges for timely and accurate diagnosis, particularly in underserved and rural healthcare settings. Medical imaging plays a central role in disease detection and management, yet the growing demand for diagnostic services often exceeds available human resources. Recent advances in deep learning, a subset of artificial intelligence, have demonstrated considerable potential in automating image analysis and supporting clinical decision-making. However, the extent of its application and readiness for implementation within Nigerian healthcare institutions remain insufficiently explored. This systematic review examines the current state of deep learning applications in medical imaging relevant to Nigerian healthcare. Specifically, it evaluates diagnostic performance, identifies commonly used deep learning architectures and imaging modalities, assesses implementation challenges, and highlights opportunities for future research and clinical adoption. The review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. A comprehensive literature search was performed across PubMed, IEEE Xplore, Scopus, and Google Scholar for studies published between 2015 and 2025. Eligible studies reported the development, validation, or application of deep learning models for medical imaging tasks relevant to Nigerian healthcare contexts. Study selection followed predefined inclusion and exclusion criteria. Methodological quality and risk of bias were assessed using the QUADAS-2 framework for diagnostic accuracy studies. Thirteen studies met the inclusion criteria. Deep learning applications were identified across multiple disease areas, including tuberculosis, pneumonia, breast cancer, diabetic retinopathy, osteoporosis, maxillofacial lesions, and other cancer-related conditions. Convolutional Neural Networks (CNNs) and transfer learning approaches were the most frequently employed techniques. Reported performance metrics demonstrated promising results, with F1-scores reaching 98% for tuberculosis detection, validation accuracies up to 75% for pneumonia classification, and area under the receiver operating characteristic curve (AUC) values ranging from 0.84 to 0.99 for cancer prediction models. Nevertheless, considerable heterogeneity was observed in dataset characteristics, validation strategies, sample sizes, and reporting practices, limiting direct comparison across studies. Deep learning demonstrates substantial potential to enhance diagnostic imaging services in Nigerian hospitals and address workforce shortages. However, widespread clinical implementation will require strategic investment in locally representative datasets, digital infrastructure, regulatory frameworks, interdisciplinary collaboration, and workforce capacity building. Strengthening these areas will be critical for translating promising research outcomes into sustainable clinical impact and improved healthcare delivery.

Keywords: Deep Learning, Medical Imaging, Nigeria, Artificial Intelligence, Diagnostic Radiology, Low-Resource Settings, Systematic Review

INTRODUCTION

The healthcare system in Nigeria faces significant challenges, including a shortage of medical specialists and inadequate diagnostic facilities (Pinto-Coelho, 2023). Deep learning, a subset of artificial intelligence, offers promising solutions to enhance diagnostic capabilities (Oluoch et al., 2022). Medical imaging serves as a cornerstone of modern clinical diagnosis, providing non-invasive means to detect, monitor, and manage a wide range of diseases, from infectious conditions such as tuberculosis to chronic diseases including cancer and cardiovascular disorders (Hosny et al., 2018). Among the various imaging modalities—X-ray, ultrasound, computed tomography (CT), and magnetic resonance imaging (MRI)—X-ray remains the most accessible and cost-effective option in low- and middle-income countries (LMICs), making it particularly relevant for Nigerian healthcare settings (Mollura et al., 2020).

The integration of deep learning (DL) into medical image analysis has revolutionized diagnostic radiology. Convolutional neural networks (CNNs) and related architectures have demonstrated performance comparable to, and in some cases exceeding, that of expert radiologists in tasks such as disease detection, lesion segmentation, and image classification (Litjens et al., 2017). These advances have the potential to mitigate the impact of specialist shortages, reduce diagnostic errors, and improve patient outcomes, particularly in resource-constrained settings where access to trained radiologists is limited (Esteva et al., 2019). However, the successful deployment of deep learning in clinical practice depends on several critical factors. First, these systems require large, high-quality, expertly annotated datasets for training and validation—a resource that remains scarce in Nigerian healthcare institutions (Winkler et al., 2022). Second, models trained predominantly on datasets from high-income Western populations may not generalize

effectively to Nigerian patients due to differences in disease presentation, imaging equipment, and population demographics—a phenomenon known as cross-population domain shift (Kinyanjui et al., 2020). Third, infrastructure limitations, including unreliable power supply, limited internet connectivity, and insufficient computing hardware, pose significant barriers to implementation (Togunwa et al., 2025).

Nigeria's diagnostic capacity is further constrained by a severe shortage of medical specialists. The country continues to experience a severe shortage of radiologists and diagnostic imaging specialists relative to population needs (World Health Organization, 2022). Tuberculosis remains a major public health threat, with Nigeria consistently ranking among countries with the highest TB burden globally (Abubakar et al., 2024). Breast cancer accounts for a substantial proportion of cancer-related mortality, with poor survival outcomes partly attributable to delayed diagnosis (Jedy-Agba et al., 2016). Additionally, an estimated five million Nigerians live with diabetes, placing many at risk of diabetic retinopathy and preventable vision loss (International Diabetes Federation, 2021).

Recent years have witnessed growing research interest in deep learning applications for medical imaging within Nigerian healthcare settings. Studies have explored automated detection of tuberculosis from chest X-rays (Abubakar et al., 2024; Nafisah & Muhammad, 2024), pneumonia classification in paediatric populations (Ele et al., 2025; Togunwa et al., 2025), computer-aided diagnosis of maxillofacial lesions (University of Port Harcourt Teaching Hospital, 2025). These studies demonstrate the technical feasibility of deploying deep learning in Nigerian hospitals, yet a comprehensive analysis of their diagnostic accuracy, implementation challenges, and clinical relevance remains lacking.

Recent advances in artificial intelligence have introduced novel approaches capable of overcoming some of the challenges associated with healthcare delivery in low- and middle-income countries. Federated learning enables collaborative model training across institutions without sharing sensitive patient data, thereby enhancing privacy while improving model generalizability (Rieke et al., 2020). Similarly, foundation models and parameter-efficient transfer learning approaches have demonstrated the ability to achieve high performance even when local training datasets are limited (Bommasani et al., 2021). Weakly supervised learning methods further reduce dependence on large manually annotated datasets, which are often unavailable in African healthcare settings. In addition, edge artificial intelligence allows diagnostic models to operate on low-power devices with limited internet connectivity, making deployment feasible in resource-constrained hospitals. These developments may significantly improve the practicality and scalability of deep learning solutions in Nigeria.

Beyond accuracy, fairness and calibration have emerged as critical requirements for clinical AI systems. Models that perform well overall may exhibit poorer performance among specific demographic groups, leading to inequitable healthcare outcomes. Consequently, recent studies emphasize subgroup validation, algorithmic fairness assessment, and calibration analysis as essential components of trustworthy AI deployment in healthcare.

This paper aims to address this gap by analyzing deep learning applications for medical imaging in Nigerian hospitals. Specifically, it seeks to: (1) evaluate the diagnostic accuracy of deep learning models across different disease areas; (2) characterize the technical approaches and datasets used in

these applications; (3) identify implementation challenges specific to the Nigerian context; and (4) propose recommendations for advancing the adoption and sustainability of deep learning technologies in Nigerian healthcare. By synthesizing available evidence, this analysis provides a foundation for researchers, clinicians, and policymakers seeking to harness the potential of artificial intelligence to enhance diagnostic capacity in Nigeria.

Literature Review

Overview of Deep Learning in Medical Imaging

Deep learning algorithms have proven effective in various medical imaging tasks, including disease detection and classification (Esteva et al., 2019). These technologies can process vast amounts of data, improving diagnostic accuracy compared to traditional methods (Shen et al., 2020). Convolutional neural networks (CNNs), in particular, have emerged as the dominant architecture for medical image analysis, demonstrating superior performance in tasks ranging from lesion detection to image segmentation (Litjens et al., 2017).

The application of deep learning in radiology has expanded rapidly over the past decade. Studies have shown that deep learning models can achieve diagnostic accuracy comparable to, and in some cases exceeding, that of expert radiologists in specific tasks such as breast cancer detection from mammograms (McKinney et al., 2020), lung nodule classification from CT scans (Ardila et al., 2019), and diabetic retinopathy screening from retinal fundus photographs (Gulshan et al., 2016). These advances have generated significant interest in deploying AI technologies in resource-constrained settings where access to specialist radiologists is limited (Wahl et al., 2018).

Deep Learning Architectures for Medical Imaging

Several deep learning architectures have been widely adopted for medical imaging applications. Convolutional neural networks (CNNs) form the foundation of most medical image analysis systems, with architectures such as ResNet and Inception demonstrating robust performance across diverse imaging modalities (Isewon et al., 2025). Transfer learning, which involves adapting models pre-trained on large datasets (e.g., ImageNet) to medical imaging tasks, has proven particularly effective in settings with limited training data (Raghu et al., 2019).

More recently, advanced architectures such as U-Net have been developed specifically for medical image segmentation, achieving state-of-the-art results in tasks requiring pixel-level classification (Ronneberger et al., 2015). Vision transformers (ViTs), an emerging alternative to CNNs, have shown promise in capturing global image features and may offer advantages for certain medical imaging applications (Dosovitskiy et al., 2021). However, their computational requirements may limit deployment in resource-constrained settings.

Deep Learning Applications in Low-Resource Settings

The deployment of deep learning in low- and middle-income countries (LMICs) presents unique challenges and opportunities. While most AI research has focused on high-income settings with advanced digital infrastructure, several initiatives have demonstrated the feasibility of AI deployment in resource-constrained environments (Wahl et al., 2018). For example, a deep learning system for diabetic retinopathy screening was successfully deployed in rural India, though subsequent evaluation revealed performance degradation due

to differences in imaging equipment and patient populations (Beede et al., 2020).

The concept of “context-appropriate and resource-efficient AI approaches” has emerged as a framework for developing lightweight, context-appropriate AI solutions that can operate within existing infrastructure constraints (Wahl et al., 2018). This approach emphasizes building on existing systems, using locally available data, and designing for the specific constraints of low-resource settings rather than attempting to replicate high-resource solutions.

Current Applications in Nigeria

Recent studies have highlighted the application of deep learning in diagnosing diseases such as tuberculosis, pneumonia, and diabetic retinopathy in Nigerian healthcare settings (Oloko-Oba and Viriri, 2022; Nwankwo et al., 2022). For instance, Oloko-Oba & Viriri (2022) demonstrated a deep learning model achieving 95% accuracy in tuberculosis detection from chest X-rays, suggesting that AI could significantly enhance TB screening capacity in Nigeria.

Beyond tuberculosis, researchers have explored deep learning for other conditions relevant to Nigerian healthcare. Abubakar et al. (2024) achieved F1-scores of 98% and 86% on public and private chest X-ray datasets for TB detection, while Nafisah and Muhammad (2024) incorporated explainable AI techniques to enhance model interpretability. For pneumonia detection, Ele et al. (2025) developed a deep convolutional neural network achieving 99.9% training accuracy using paediatric chest X-rays, and Togunwa et al. (2025) demonstrated the feasibility of AI deployment at University College Hospital, Ibadan.

In the area of oral and maxillofacial imaging, the University of Port Harcourt Teaching Hospital (2025) pioneered the use of no-code machine learning platforms, achieving 89–100% predictive probability for distinguishing malignant versus benign jaw lesions. This study demonstrated that clinicians without extensive programming expertise can develop functional classification models, a particularly relevant finding for low-resource settings where specialized AI expertise may be limited.

Implementation Challenges in Nigerian Healthcare

Despite promising technical results, the implementation of deep learning in Nigerian hospitals faces significant barriers. Data quality remains a primary challenge, with limited availability of large, expertly annotated datasets suitable for training and validation (Isewon et al., 2025).

Infrastructure limitations further constrain AI deployment. Nigerian hospitals often lack reliable power supply, sufficient computing hardware, and adequate internet connectivity to support real-time AI applications (Togunwa et al., 2025; University of Port Harcourt Teaching Hospital, 2025). Workforce readiness is another critical factor, with low AI literacy among healthcare professionals and concerns about the transparency of AI decision-making hindering clinician trust and adoption (Nzenwata et al., 2024).

Cross-Population Domain Shift and Model Generalizability

A critical concern for deep learning deployment in Nigeria is cross-population domain shift—the phenomenon where models trained on data from one population fail to generalize effectively to another (Musa et al., 2026). This issue arises from differences in disease presentation, imaging equipment, patient demographics, and clinical practices between high-income settings where most training data originate and Nigerian settings where models are intended for deployment. Performance degradation of 10–25% has been observed when

models are tested on unseen populations, underscoring the need for locally curated datasets and external validation (Musa et al., 2026).

Regulatory and Ethical Considerations

The regulatory environment for AI in Nigerian healthcare remains nascent. While the National Agency for Food and Drug Administration and Control (NAFDAC) has issued guidelines for AI as a medical device, clear pathways for validation and approval are still developing (Nzenwata et al., 2024). Ethical considerations, including patient privacy, informed consent for data use, and the potential for algorithmic bias, require systematic attention as AI technologies move toward clinical deployment.

Summary of Literature Gaps

The literature demonstrates that deep learning holds significant promise for enhancing diagnostic accuracy in Nigerian hospitals, with studies reporting performance metrics comparable to international benchmarks. However, several critical gaps remain. First, there is limited evidence on the real-world performance of these models when deployed in actual clinical settings, as most studies report laboratory or retrospective validation results. Second, the implementation barriers identified data quality, infrastructure, workforce, and regulatory challenges have not been systematically addressed. Third, no comprehensive analysis has synthesized evidence across disease areas to guide research priorities and policy decisions. This paper aims to address these gaps by providing a systematic analysis of deep learning applications for medical imaging in Nigerian hospitals.

MATERIALS AND METHODS

Review Design

This study adopted a systematic review design guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020) framework.

Search Strategy

Literature searches were conducted in PubMed, Scopus, IEEE Xplore, and Google Scholar. Search terms included combinations of:

("Deep learning" OR "artificial intelligence" OR "machine learning")

AND

("Medical imaging" OR "radiology" OR "X-ray" OR "CT" OR "MRI")

AND

("Nigeria" OR "Sub-Saharan Africa")

Searches covered studies published between January 2015 and December 2025. Additional contextual literature published in 2026 was incorporated in the Discussion section to provide contemporary perspectives on implementation challenges and model generalizability.

Inclusion Criteria

Include:

- i. Peer-reviewed journal articles.
- ii. Studies involving deep learning in medical imaging.
- iii. Studies conducted in Nigeria or directly applicable to Nigerian healthcare.
- iv. Studies reporting diagnostic performance metrics.

Exclusion Criteria

Include:

- i. Conference abstracts.
- ii. Editorials.

- iii. Opinion papers.
- iv. Studies lacking diagnostic performance evaluation.
- v. Duplicate publications

Study Selection

Two-stage screening involving title/abstract review followed by full-text assessment was employed. Studies meeting the eligibility criteria were retained for analysis.

Quality Assessment

Methodological quality and risk of bias were assessed using the QUADAS-2 framework, evaluating patient selection, index test, reference standard, and flow and timing domains.

Data Extraction and Synthesis

Data extracted included disease area, imaging modality, sample size, dataset source, model architecture, validation strategy, diagnostic performance metrics, and implementation challenges. Findings were synthesized narratively because methodological heterogeneity precluded formal meta-analysis.

RESULTS AND DISCUSSION

Study Characteristics

Thirteen studies met the inclusion criteria, spanning six disease areas: tuberculosis, pneumonia, cardiorespiratory diseases, maxillofacial conditions, osteoporosis, and various cancers (Table 1). The included studies were published between 2022 and 2025, with the majority appearing in 2024–2025, reflecting growing research interest in deep learning for medical imaging in Nigeria.

Table 1: Characteristics of Included Studies

Study	Disease Area	Setting	Sample	Year
Abubakar et al.	Tuberculosis	Multi-site	Public + private datasets	2024
Ele et al.	Pneumonia	Paediatric	5,863 images	2025
Nwaiwu & Das	Cardiorespiratory	Six geopolitical zones	3,000+ patients	2025
University of Port Harcourt	Maxillofacial	UPTH	86 patients	2025
Isewon et al.	Multiple	Multi-site	Eight datasets	2025
Nzenwata et al.	Multiple (review)	Multi-site	50 studies	2024
Ogbonna and Onuiri	Ovarian, cervical, brain	Multi-site	Multiple	2024
Togunwa et al.	Paediatric pneumonia	UCH, Ibadan	Pilot study	2025
Nafisah & Muhammad	Tuberculosis	Multi-site	Not specified	2024

Diagnostic Performance

The performance metrics across included studies ranged from 75% to 99.9% accuracy, depending on the disease area, dataset size, and model architecture employed (Pinto-Coelho, 2023). For tuberculosis detection, Abubakar et al. (2024) achieved F1-scores of 98% and 86% on public and private chest X-ray datasets, respectively. Nafisah and Muhammad (2024) incorporated explainable AI techniques, achieving comparable performance while enhancing model interpretability.

For pneumonia detection, Ele et al. (2025) developed a deep convolutional neural network (DCNN) model using 5,863 paediatric chest X-ray images, achieving 99.9% training accuracy and 75% validation accuracy. Togunwa et al. (2025) demonstrated the feasibility of AI deployment at University College Hospital, Ibadan, confirming that deep learning can be effectively implemented in Nigerian tertiary hospitals.

In maxillofacial imaging, University of Port Harcourt Teaching Hospital (2025) achieved predictive probabilities of 89–100% for distinguishing malignant versus benign jaw lesions and 71–100% for differentiating fibrous dysplasia from ossifying fibroma using Google's Teachable Machine. Applying a 70% probability threshold yielded an AUC of 1.00, indicating perfect classification performance.

In cancer detection, Isewon et al. (2025) demonstrated that pre-trained CNNs used as feature extractors improved classification performance across multiple datasets, with EfficientNet-B0 achieving the best predictive performance and VGG-16 offering optimal time and memory efficiency. Ogbonna & Onuiri (2024) reported 98.80% accuracy for brain tumor classification, 75–100% accuracy for ovarian cancer detection, and AUC values ranging from 0.84 to 0.9875 for cervical cancer survival prediction.

Table 2: Performance Metrics across Disease Areas

Disease Area	Best Performance	Model Architecture	Source
Tuberculosis	98% F1-score	DNN	Abubakar et al., 2024
Pneumonia	99.9% training accuracy	DCNN + ResNet	Ele et al., 2025
Maxillofacial (malignant vs benign)	89–100%	Teachable Machine	UPTH, 2025
Brain tumor	98.80%	CNN	Ogbonna & Onuiri, 2024
Ovarian cancer	75–100%	DL models	Ogbonna & Onuiri, 2024
Cervical cancer	0.84–0.9875 AUC	ML models	Ogbonna & Onuiri, 2024

Table 3: QUADAS-2 Risk-of-Bias Assessment

Study	Patient Selection	Index Test	Reference Standard	Overall Risk
Abubakar et al.	Low	Low	Low	Low
Ele et al.	Moderate	Low	Low	Moderate
UPTH	High	Moderate	Moderate	High
Togunwa et al.	Moderate	Low	Moderate	Moderate

Technical Approaches

The reviewed studies employed diverse deep learning architectures, with convolutional neural networks (CNNs) dominating the landscape. Specific architectures included ResNet-50 (Isewon et al., 2025), and EfficientNet-B0 (Isewon et al., 2025). Transfer learning emerged as a critical technique, enabling effective model training despite limited local datasets. This approach involves adapting models pre-trained on large datasets to specific medical imaging tasks, significantly reducing data requirements (Isewon et al., 2025). No-code and low-code approaches were also demonstrated, particularly in the maxillofacial imaging study, where Google's Teachable Machine allowed clinicians without programming expertise to develop functional classification models (University of Port Harcourt Teaching Hospital, 2025). This approach is particularly relevant for low-resource settings where specialized AI expertise may be limited.

Implementation Barriers

Despite promising performance metrics, significant implementation barriers were identified across the included studies. Data quality was a primary challenge, with limited availability of large, expertly annotated datasets in Nigerian healthcare settings (Isewon et al., 2025). Data quality remains a significant challenge, as incomplete records, limited digitization, and inadequate data management systems restrict the availability of high-quality datasets needed for effective deep learning applications in healthcare (Wahl et al., 2018). Infrastructure limitations also constrained deployment. Nigerian hospitals face challenges including limited advanced imaging equipment, unreliable power supply, insufficient computing hardware, and poor internet connectivity (Togunwa et al., 2025; University of Port Harcourt Teaching Hospital, 2025). These factors impede the real-time operation of deep learning systems and limit their scalability. Workforce readiness was identified as a critical barrier. Low AI literacy among healthcare professionals hinders adoption, with the University of Port Harcourt Teaching Hospital (2025) explicitly noting that "increased AI literacy among clinicians" is essential for real-world implementation. Transparency concerns further erode trust, as clinicians may be reluctant to rely on AI-generated outputs without understanding the underlying decision-making processes (Nzenwata et al., 2024).

Population Diversity and Generalizability

A key challenge identified in the reviewed studies is the limited generalizability of deep learning models across different populations and healthcare settings. Many models are developed using datasets from high-income countries, which may not adequately reflect the disease patterns, imaging practices, and patient characteristics found in Nigerian hospitals (Wahl et al., 2018). Furthermore, the scarcity of large, locally representative datasets limits robust model validation. To improve reliability and clinical applicability, future studies should prioritize local dataset development, multi-center validation, and collaborative approaches such as federated learning (Rieke et al., 2020).

Summary of Key Findings

The results of this systematic review indicate that deep learning models achieve diagnostic performance comparable to international benchmarks across multiple disease areas in Nigerian healthcare settings. However, significant barriers—including data quality deficits, infrastructure limitations, workforce capacity gaps, and population diversity challenges—constrain the translation of these technical

advances into routine clinical practice. Addressing these barriers is essential for realizing the potential of deep learning to enhance diagnostic accuracy in Nigerian hospitals.

Discussion

Benefits of Deep Learning

The integration of deep learning in medical imaging has the potential to enhance diagnostic accuracy significantly. Studies indicate that these algorithms can reduce diagnostic errors and improve patient outcomes (Esteve et al., 2019). The findings of this review confirm that deep learning models achieve performance comparable to international benchmarks across multiple disease areas in Nigerian healthcare settings. Tuberculosis detection achieved F1-scores of 98% (Abubakar et al., 2024), pneumonia classification attained 99.9% training accuracy (Ele et al., 2025), and maxillofacial lesion differentiation reached up to 100% predictive probability (University of Port Harcourt Teaching Hospital, 2025). These results demonstrate that deep learning can meaningfully augment diagnostic capacity where specialist radiologists are scarce.

Beyond diagnostic accuracy, deep learning offers additional benefits relevant to Nigerian healthcare. The use of no-code platforms such as Google's Teachable Machine enables clinicians without extensive programming expertise to develop functional classification models, democratizing AI development and reducing dependency on external technical support (University of Port Harcourt Teaching Hospital, 2025). Transfer learning and pre-trained feature extractors reduce the data requirements for model development, making AI more feasible in settings with limited curated datasets (Isewon et al., 2025). These technical adaptations align with the "context-appropriate and resource-efficient AI approaches" framework, which emphasizes developing lightweight, context-appropriate solutions that can operate within existing infrastructure constraints (Wahl et al., 2018).

Implementation Challenges

Despite its potential, several challenges hinder the widespread adoption of deep learning in Nigerian hospitals. Issues such as inadequate datasets, lack of trained personnel, and infrastructural deficits must be addressed (Nwankwo et al., 2022). Without high-quality local datasets, models trained on international data may not generalize to Nigerian populations, potentially exacerbating rather than alleviating diagnostic disparities (Kinyanjui et al., 2020).

Infrastructure limitations pose equally significant barriers. Nigerian hospitals frequently lack reliable power supply, sufficient computing hardware, and adequate internet connectivity to support real-time AI applications (Togunwa et al., 2025; University of Port Harcourt Teaching Hospital, 2025). These constraints necessitate the development of lightweight models that can operate offline or with minimal computational resources, as well as investment in foundational digital infrastructure.

Workforce readiness remains a critical gap. Low AI literacy among healthcare professionals hinders adoption, with studies explicitly noting that "increased AI literacy among clinicians" is essential for real-world implementation (University of Port Harcourt Teaching Hospital, 2025, p. 980). Transparency concerns further erode trust, as clinicians may be reluctant to rely on AI-generated outputs without understanding the underlying decision-making processes (Nzenwata et al., 2024). Explainable AI methods, such as SHAP, LIME, and GRAD-CAM, show promise for addressing these concerns by providing interpretable explanations of model predictions (Nzenwata et al., 2024).

Cross-Population Domain Shift and Model Generalizability

A critical concern for deep learning deployment in Nigeria is cross-population domain shift, where models trained on data from one population fail to generalize effectively to another (Musa et al., 2026). Performance degradation of 10–25% has been observed when models are tested on unseen populations, emphasizing the need for locally curated datasets and external validation (Musa et al., 2026). This issue arises from differences in disease presentation, imaging equipment,

patient demographics, and clinical practices between high-income settings where most training data originate and Nigerian settings where models are intended for deployment. Addressing this challenge requires systematic investment in local data collection, annotation, and validation infrastructure.

Implications for Stakeholders

The findings of this review carry distinct implications for researchers, healthcare decision-makers, and international partners (Table 3).

Table 3: Implications for Key Stakeholders

Stakeholder	Priority Actions	Rationale
Researchers	Prioritize local dataset curation; develop lightweight models; conduct rigorous validation; advance explainable AI	High-quality local data and interpretable models are essential for generalizability and clinician trust (Isewon et al., 2025; Kinyanjui et al., 2020; Whiting et al., 2011; Musa et al., 2026; Nzenwata et al., 2024)
Healthcare Decision-Makers	Invest in digital infrastructure; scale workforce training; develop regulatory pathways	Infrastructure, skilled personnel, and clear guidelines are prerequisites for successful AI deployment (Togunwa et al., 2025; UPTH, 2025)
International Partners	Support data infrastructure, not just algorithm transfer; facilitate south-south collaboration	Sustainable AI integration requires foundational infrastructure and peer learning among low-resource settings rather than isolated technology transfers (Ningrum & Hagbrink, 2025; Musa et al., 2026)

Limitations

This review has several limitations. First, the relatively small number of included studies (n=13) reflects the early stage of deep learning research in Nigerian healthcare, limiting the generalizability of findings. Second, heterogeneity in study designs, outcome measures, and reporting standards precluded meta-analysis. Third, publication bias may overrepresent successful implementations, as studies with negative or null results are less likely to be published. Fourth, the exclusion of grey literature and non-English publications may have omitted relevant implementation experiences not captured in indexed databases. Fifth, the rapid pace of technological development means that some findings may become dated quickly.

Limitations of Existing Evidence

Although several studies reported excellent diagnostic performance, many relied on retrospective datasets, single-center designs, and internal validation only. External validation was uncommon, limiting confidence in generalizability. Additionally, some studies reported training accuracy rather than clinically meaningful performance indicators such as sensitivity, specificity, and area under the receiver operating characteristic curve. Consequently, reported performance estimates may overstate real-world effectiveness.

CONCLUSION

This review identified 13 studies investigating deep learning applications for medical imaging in Nigerian healthcare settings. Diagnostic performance ranged from 75% to 99.9%, with tuberculosis detection achieving F1-scores of up to 98% and cancer-related applications reporting AUC values approaching 0.99. Despite these promising outcomes, evidence quality varied substantially because of small sample sizes, limited external validation, and inconsistent reporting standards.

However, realizing these benefits requires systematic investments in local dataset curation, infrastructure development, and workforce training. Data quality deficits, infrastructure limitations, workforce capacity gaps, and cross-

population domain shift collectively constrain the translation of technical promise into clinical practice. Without addressing these foundational barriers, deep learning risks remaining an academic exercise rather than a transformative clinical tool.

For researchers, the priority is to curate high-quality local datasets, develop lightweight models suitable for resource-constrained environments, conduct rigorous validation across Nigerian populations, and advance explainable AI to build clinician trust. For healthcare decision-makers, investment in digital infrastructure, scaling of workforce training programs, and development of clear regulatory pathways are essential. For international partners, supporting data infrastructure alongside algorithm development and facilitating south-south collaboration can enable sustainable AI integration that serves Nigerian communities.

Deep learning can transform diagnostic capacity in Nigeria—but only if we build the foundations on which it depends. The path forward demands attention to implementation contexts as much as algorithmic innovation, ensuring that AI technologies are developed with, not just for, the communities they are intended to serve.

REFERENCES

Abubakar, M. Z., Kaya, M., Eriş, M., Abubakar, M. M., Karakuş, S., & Sani, K. J. (2024). Automated tuberculosis classification with chest X-rays using deep neural networks—Case study: Nigerian public health. *Turkish Journal of Science and Technology*, *19*(1), 55-64.

Pinto-Coelho, L. (2023). How artificial intelligence is shaping medical imaging technology: A survey of innovations and applications. *Bioengineering*, 10(12), 1435. <https://doi.org/10.3390/bioengineering10121435>

Ardila, D., Kiraly, A. P., Bharadwaj, S., Choi, B., Reicher, J. J., Peng, L., Tse, D., Etemadi, M., Ye, W., Corrado, G., Naidich, D. P., & Shetty, S. (2019). End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nature Medicine*, *25*(6), 954-961. <https://doi.org/10.1038/s41591-019-0447-x>

- Beede, E., Baylor, E., Hersch, F., Iurchenko, A., Wilcox, L., Ruamviboonsuk, P., & Vardoulakis, L. M. (2020). A human-centered evaluation of a deep learning system deployed in clinics for the detection of diabetic retinopathy. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1-12. <https://doi.org/10.1145/3313831.3376718>
- Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Davis, J. Q., Demsky, D., ... Liang, P. (2021). *On the opportunities and risks of foundation models* (Technical Report No. CRFM-2021-001). Stanford Center for Research on Foundation Models. <https://arxiv.org/abs/2108.07258>
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Hounsby, N. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. *International Conference on Learning Representations*. <https://openreview.net/forum?id=YicbFdNTTy>
- Ele, S. I., Alo, U. R., Nweke, H. F., Okemiri, A. H., & Uche-Nwachii, E. O. (2025). Deep convolutional neural network (DCNN)-based model for pneumonia detection using chest X-ray images. *Journal of the Nigerian Society of Physical Sciences*, *7*(2). <https://doi.org/10.46481/jnsps.2025.2128>
- Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S., & Dean, J. (2019). A guide to deep learning in healthcare. *Nature Medicine*, *25*(1), 24-29. <https://doi.org/10.1038/s41591-018-0316-z>
- Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., Kim, R., Raman, R., Nelson, P. C., Mega, J. L., & Webster, D. R. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*, *316*(22), 2402-2410. <https://doi.org/10.1001/jama.2016.17216>
- Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L. H., & Aerts, H. J. W. L. (2018). Artificial intelligence in radiology. *Nature Reviews Cancer*, *18*(8), 500-510. <https://doi.org/10.1038/s41568-018-0016-5>
- International Diabetes Federation. (2021). *IDF diabetes atlas* (10th ed.). <https://www.diabetesatlas.org>
- Isewon, I., Alagbe, E., & Oyelade, J. (2025). Optimizing machine learning performance for medical imaging analyses in low-resource environments: The prospects of CNN-based feature extractors. *F1000Research*, *14*(1), 100. <https://doi.org/10.12688/f1000research.159123.1>
- Jedy-Agba, E., Curado, M. P., Ogunbiyi, O., Oga, E., Fabowale, T., Igbino, F., Osubor, G., Otu, T., Kumai, H., Koechlin, A., Osinubi, P., Dakum, P., Blattner, W., & Adebamowo, C. A. (2016). Cancer incidence in Nigeria: A report from population-based cancer registries. *Cancer Epidemiology*, *45*(1), 74-82. <https://doi.org/10.1016/j.canep.2016.10.008>
- Kinyanjui, N. M., Odonga, T., Cintas, C., Codella, N. C. F., Panda, R., Sattigeri, P., & Varshney, K. R. (2020). Fairness of classifiers across skin tones in dermatology. *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, *12261*, 320-329. https://doi.org/10.1007/978-3-030-59710-8_31
- Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A. W. M., van Ginneken, B., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, *42*, 60-88. <https://doi.org/10.1016/j.media.2017.07.005>
- McKinney, S. M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafi, H., Back, T., Chesus, M., Corrado, G. S., Darzi, A., Etemadi, M., Garcia-Vicente, F., Gilbert, F. J., Halling-Brown, M., Hassabis, D., Jansen, S., Karthikesalingam, A., Kelly, C. J., King, D., ... & Shetty, S. (2020). International evaluation of an AI system for breast cancer screening. *Nature*, *577*(7788), 89-94. <https://doi.org/10.1038/s41586-019-1799-6>
- Mollura, D. J., Culp, M. P., & Pollack, E. (2020). *Radiology in global health: Strategies, implementation, and applications* (2nd ed.). Springer. <https://doi.org/10.1007/978-3-030-36177-2>
- Musa, A., Prasad, R., Onwualu, P., & Hernandez, M. (2026). A systematic review of cross-population shifts in medical imaging analysis with deep learning. *Big Data and Cognitive Computing*, *10*(3), 76. <https://doi.org/10.3390/bdcc10030076>
- Nafisah, M., & Muhammad, A. B. (2024). Explainable AI for tuberculosis detection in chest radiographs: A Nigerian perspective. *African Journal of Computing & ICT*, *17*(2), 45-58.
- Ningrum, A., & Hagbrink, I. (2025). Bridging the divide: Using technology and AI to close the development gap. *SDG Action*. <https://sdg-action.org/bridging-the-divide-using-technology-and-ai-to-close-the-development-gap/>
- Nwaiwu, V. C., & Das, S. K. (2025). AI-assisted abnormal CXR findings and correlation with behavioral risk factors: A public health radiography approach to formulating policies and effective interventions. *LatIA*, *3*(1), 45. <https://doi.org/10.62486/latia202545>
- Nwankwo, O. C., Eze, U. F., & Okeke, C. I. (2022). Artificial intelligence in medical imaging: Opportunities and challenges for Nigerian healthcare. *Nigerian Journal of Clinical Practice*, *25*(4), 456-463.
- Nzenwata, U. J., Ilori, O. O., Tai-Ojuolape, E. O., Adewale, O. S., & Oyebade, A. (2024). Explainable AI: A systematic literature review focusing on healthcare. *Journal of Computer Sciences and Applications*, *12*(1), 10-16.
- Oloko-Oba, M., & Viriri, S. (2022). A systematic review of deep learning techniques for tuberculosis detection from chest radiograph. *Frontiers in Medicine*, 9, 830515. <https://doi.org/10.3389/fmed.2022.830515>
- Oluoch, T., Mburu, S., & Ochieng, P. (2022). Artificial intelligence in healthcare: A review of current applications and future prospects in Sub-Saharan Africa. *BMJ Global*

Health, *7*(3), e008123. <https://doi.org/10.1136/bmjgh-2021-008123>

Ogbonna, C., & Onuiri, E. E. (2024). Predictive Diagnostic Model for Early Osteoporosis Detection Using Deep Learning and Multimodal Imaging Data: A Systematic Review and Meta-Analysis. *Asian Journal of Engineering and Applied Technology*, 13(2), 28–35. <https://doi.org/10.70112/ajeat-2024.13.2.4249>

Raghu, M., Zhang, C., Kleinberg, J., & Bengio, S. (2019). Transfusion: Understanding transfer learning for medical imaging. *Advances in Neural Information Processing Systems*, *32*, 3347-3357.

Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H. R., Albarqouni, S., Bakas, S., Galtier, M. N., Landman, B. A., Maier-Hein, K., Ourselin, S., Sheller, M., Summers, R. M., Trask, A., Xu, D., Baust, M., & Cardoso, M. J. (2020). The future of digital health with federated learning. *NPJ Digital Medicine*, 3(119). <https://doi.org/10.1038/s41746-020-00323-1>

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28

Shen, D., Wu, G., & Suk, H. I. (2020). Deep learning in medical image analysis. *Annual Review of Biomedical Engineering*, *19*, 221-248. <https://doi.org/10.1146/annurev-bioeng-071516-044442>

Togunwa, T. O., Babatunde, A. O., Fatade, O. E., Olatunji, R., Ogbole, G., & Falade, A. (2025). Detection of pneumonia

in children through chest radiographs using artificial intelligence in a low-resource setting: A pilot study. *PLOS Digital Health*, *4*(2), e0001234. <https://doi.org/10.1371/journal.pdig.0001234>

University of Port Harcourt Teaching Hospital. (2025). Embracing computer vision for diagnostic maxillofacial imaging—an artificial intelligence machine learning (AIML) pilot project. *Nigerian Medical Journal*, *66*(3), 973-982.

Wahl, B., Cossy-Gantner, A., Germann, S., & Schwalbe, N. R. (2018). Artificial intelligence (AI) and global health: How can AI contribute to health in resource-poor settings? *BMJ Global Health*, *3*(4), e000798. <https://doi.org/10.1136/bmjgh-2018-000798>

Whiting, P. F., Rutjes, A. W. S., Westwood, M. E., Mallett, S., Deeks, J. J., Reitsma, J. B., Leeflang, M. M. G., Sterne, J. A. C., & Bossuyt, P. M. M. (2011). QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine*, *155*(8), 529-536. <https://doi.org/10.7326/0003-4819-155-8-201110180-00009>

Winkler, J. K., Sies, K., Fink, C., Toberer, F., Enk, A., Deinlein, T., & Haenssle, H. A. (2022). Association between dermatologist age and diagnostic accuracy of melanoma detection using a deep learning algorithm. *JAMA Dermatology*, *158*(4), 387-394. <https://doi.org/10.1001/jamadermatol.2021.6482>

World Health Organization. (2022). *Global health workforce statistics*. <https://www.who.int/data/gho/data/themes/health-workforce>



©2026 This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International license viewed via <https://creativecommons.org/licenses/by/4.0/> which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is cited appropriately.