Journal of Health Informatics in Africa

Improving Tuberculosis Detection with Deep Learning in X-Ray Imaging

Asia Turra [a], Gilbert M. Gilbert [a, *], Tabu Kondo [b,]

[a] Department of Information Systems and Technology, The University of Dodoma, Tanzania
[b] Department of Computer Science and Engineering, The University of Dodoma, Tanzania

Background and Purpose: Tuberculosis (TB) continues to pose a major global health burden, particularly in low-resource settings, where timely and accurate diagnosis remains a challenge. Conventional diagnostic methods such as sputum smear microscopy and bacterial culture are often slow, labour-intensive, and susceptible to human error. Although several studies have explored alternative diagnostic tools, there remains a critical need for efficient, accurate, and scalable solutions. This study aimed to address this gap by developing a deep learning (DL)-based diagnostic model using chest X-ray images to detect TB with high precision and speed in a resource-limited setting.
Methods: A Convolutional Neural Network (CNN) model based on the VGG19 architecture was developed and trained on a dataset of labelled chest X-ray images. The model was optimized to identify the radiographic features indicative of TB. Performance metrics, including precision, recall, F1-score, and overall accuracy, were used to evaluate the diagnostic effectiveness of the model.
Results: The proposed DL model demonstrated strong diagnostic performance, achieving a precision of 96%, recall of 96%, F1-score of 96% and an overall accuracy of 96%. These results indicate the robustness of the model in identifying TB cases from chest X-ray images with minimal false positives and false negatives.
Conclusions: The findings underscore the potential of AI-driven diagnostic tools to significantly enhance TB detection, particularly in settings with limited access to laboratory facilities. The VGG19-based model offers a promising pathway toward faster, more reliable, and scalable TB diagnosis, contributing to improved disease management and global control efforts.

Keywords: Medical imaging, Tuberculosis (TB) detection, Deep learning, Computer-aided diagnosis, Chest X-ray analysis, VGG19 architecture.

1 Introduction

Tuberculosis (TB) is a life-threatening infectious disease caused by *Mycobacterium tuberculosis*, primarily affecting the lungs and transmitted through airborne droplets released when infected individuals cough, sneeze or speak [1]. Despite decades of global control efforts, TB remains one of the leading causes of death from infectious diseases, with an estimated 10 million new cases and approximately 1.5 million deaths reported annually[2]. The burden of TB is disproportionately high in low- and middle-income countries, where healthcare systems often face significant diagnostic and infrastructural constraints.

Early and accurate diagnosis remains the most effective strategy for controlling and preventing TB [3]. Conventional TB diagnostic approaches include chest radiography, sputum smear microscopy, and bacterial culture testing [4]. Chest X-ray examinations rely heavily on the expertise of trained radiologists, making the process time-consuming and prone to inter-observer variability, particularly in resource-limited settings. Bacterial culture, although considered the gold standard, is slow, typically requiring three to four weeks to yield results [5]. Sputum smear microscopy is the most widely used diagnostic method because of its low cost and simplicity; however, its sensitivity decreases significantly in patients with low bacterial loads or TB–HIV co-infection [2], [6], [7]. These limitations often delay the initiation of treatment and increase the risk of disease transmission.

*Corresponding author address: The University of Dodoma, Dodoma, Tanzania. Email: gilbert.gilbert@udom.ac.tz
 J Health Inform Afr. 2026;13(1):138-152. DOI: 10.12856/JHIA-2026-v13-i1-627

Advances in digital health technologies have introduced automated approaches to TB diagnosis, aimed at supporting clinical decision-making and reducing diagnostic delays. Computer-Aided Diagnosis (CAD) systems leverage image processing, machine learning (ML), and, more recently, deep learning (DL) techniques to analyze medical images and identify patterns associated with TB [6], [8], [9]. CAD systems have demonstrated promising results in the analysis of chest radiography and computed tomography (CT) images, offering rapid and scalable screening solutions. In parallel, radiomics-based approaches have emerged as an alternative automated strategy for extracting quantitative features from medical images to support TB classification and diagnosis [10]. However, radiomics pipelines are often complex, requiring specialized expertise for implementation and interpretation, which limits their practical adoption in routine clinical settings.

Despite their technical promise, the adoption of AI-powered TB diagnostic systems remains uneven across different regions. While several high-income countries have successfully integrated CAD solutions into clinical workflows, their implementation in Africa is still at an early stage [11]. Pilot deployments have been reported in countries such as South Africa, Kenya, and Tanzania; however, widespread uptake is constrained by insufficient clinical infrastructure, restricted access to advanced technologies, high licensing costs associated with proprietary software, unreliable Internet connectivity, and a shortage of skilled personnel [12]. In Tanzania, TB screening at peripheral healthcare facilities often relies on TB score charts to guide referrals to specialized centers. Although useful, these tools are subjective and prone to misclassification, necessitating confirmation through laboratory-based testing [13], [14].

Simultaneously, many hospitals in Tanzania have transitioned to digital medical record systems and routinely acquire digital chest X-ray images, creating an opportunity for the integration of AI-driven diagnostic support tools. For TB diagnosis to be clinically effective in such environments, diagnostic systems must be fast, accurate, affordable, and usable with minimal equipment and expertise [5]. Ideally, these systems should operate with limited computational resources, deliver results within hours, and provide outputs that frontline clinicians can interpret. However, existing diagnostic techniques which include microscopy, culture, and rapid diagnostic tests remain limited by low sensitivity, high costs, and delayed turnaround times, particularly in rural and resource-constrained facilities [7].

Against this background, there is a clear need for deployment-aware, deep learning–based diagnostic tools that not only achieve high diagnostic accuracy but are also tailored to the infrastructural, technological, and human resource realities of low-resource settings such as Tanzania. This study responds to that need by proposing and evaluating a deep learning model for TB detection from chest X-ray images, designed with contextual feasibility and clinical usability in mind.

2 Related Work in Deep Learning–Based Medical Image Analysis

Deep learning, particularly through convolutional neural networks (CNNs), has become the dominant paradigm in medical image analysis owing to its ability to automatically learn complex hierarchical features from raw imaging data. CNN-based models have demonstrated state-of-the-art performance in disease detection, classification, and prognosis in diverse medical imaging domains. Among the widely adopted CNN architectures, the Visual Geometry Group (VGG) family, which includes VGG16 and VGG19, remains influential owing to its simple and modular design, strong feature extraction capability, and reproducibility across datasets and applications.

2.1 VGG Architectures in General Medical Image Diagnosis

Globally, VGG-based architectures have been successfully applied to a wide range of medical-image analysis tasks. The foundational work by Simonyan and Zisserman demonstrated that increasing the network depth using small, uniform convolutional kernels significantly improves the image classification performance. Building on this architecture, [15] applied CNNs, including VGG variants, to dermatological image classification and achieved a performance comparable to that of expert dermatologists in skin cancer detection. Similarly, [16] employed VGG-inspired deep learning models for diabetic retinopathy screening and reported high sensitivity and specificity suitable for large-scale clinical deployment.

During the COVID-19 pandemic, VGG-based transfer learning models have been extensively explored for thoracic image analysis. [17] demonstrated that VGG architectures achieved competitive accuracy in detecting COVID-19 from chest radiographs, reinforcing their suitability for lung disease classification

 J Health Inform Afr. 2026;13(1):138-152. DOI: 10.12856/JHIA-2026-v13-i1-627

tasks. The continued relevance of VGG networks in medical imaging is largely attributed to their sequential architecture and use of 3×3 convolutional filters, which enable stable training and effective extraction of fine-grained pathological features from images. These characteristics are particularly valuable in medical contexts, where disease manifestations may be subtle and spatially localized.

2.2 Global Studies on VGG-Based Tuberculosis Detection

Several studies, in the global context, have specifically investigated the use of VGG architectures for TB detection from chest X-ray images. [18] conducted one of the earliest large-scale evaluations of deep CNNs for pulmonary TB detection, reporting classification accuracies exceeding 90% and demonstrating the feasibility of automated screening for TB. [5] proposed a VGG-based framework that combined lung segmentation, feature extraction, and visualization techniques, achieving robust diagnostic performance while improving model interpretability through heatmap analysis. Similarly, scholars compared multiple CNN architectures, including VGG, for TB screening and found that transfer learning significantly enhanced performance, particularly when the annotated training data were limited [19].

These studies consistently show that VGG-based models remain competitive with deeper and more complex architectures, such as ResNet and Inception. Their effectiveness is particularly evident in scenarios characterized by moderate dataset sizes, limited computational resources, and a need for transparent and reproducible models, which are also the conditions that closely align with TB screening requirements in low- and middle-income countries.

2.3 African Studies on AI and Deep Learning for TB Diagnosis

In the African context, the application of AI and deep learning to TB diagnosis has gained increasing attention due to the high disease prevalence and chronic shortage of radiological expertise. [4] evaluated computer-aided TB screening systems in Zambia and demonstrated that automated chest X-ray analysis could achieve a sensitivity comparable to that of expert radiologists, supporting the feasibility of CAD tools in sub-Saharan Africa. The other study examined the integration of AI in medical imaging practice across multiple African countries and highlighted the potential of CNN-based systems to alleviate workforce shortages, particularly in rural healthcare settings [8].

More recently, a study conducted a systematic review of deep learning approaches for TB detection from chest radiographs and reported consistently high diagnostic accuracies across African datasets [20]. However, they emphasized that many existing models fail to account for contextual challenges, such as infrastructural limitations, data heterogeneity, and limited technical expertise. These findings underscore the need for context-aware AI solutions that are not only accurate but also deployable in real-world African healthcare settings.

2.4 Studies Related to Tanzania and Comparable Settings

Research on AI-assisted TB diagnosis in Tanzania is still emerging but demonstrates a growing potential. Pilot deployments of AI-powered TB screening tools in East African countries, including Tanzania, noted improvements in screening efficiency and case detection rates, particularly in peripheral facilities [12]. Other studies conducted in Tanzanian healthcare settings have highlighted the persistent challenges associated with conventional TB diagnostic workflows, including reliance on sputum microscopy, delayed laboratory results, and limited access to expert radiologists [21], [22], [23], [24].

Although most Tanzanian studies focus on feasibility assessments and pilot implementations rather than algorithm development, they provide a strong justification for locally adaptable deep learning artifacts. By leveraging a VGG19-based architecture trained on chest X-ray images and designed with infrastructural and human resource constraints in mind, this study contributes to bridging the gap between technical innovation and practical deployment in Tanzania.

Overall, the existing literature demonstrates that VGG-based deep learning models are effective for medical image analysis and TB detection across global, African, and regional contexts. However, many prior studies emphasize algorithmic performance without explicitly addressing the feasibility of deployment in low-resource environments. Few studies have integrated considerations such as offline operation, low-cost hardware compatibility, and reduced dependence on specialized personnel into model

 J Health Inform Afr. 2026;13(1):138-152. DOI: 10.12856/JHIA-2026-v13-i1-627

design. This study addresses these gaps by combining a proven VGG19 architecture with a deployment-aware experimental setup, offering a technically robust and contextually appropriate solution for TB diagnosis in Tanzania and in similar settings.

3 Materials and methods

3.1 Study Design

This study used an experimental research design to develop a deep learning model for TB detection. In this design, the dataset was divided into two distinct groups: the training and testing datasets. The input chest X-ray images constituted the independent (cause) variables, whereas the model's TB classification outcomes represented the dependent (effect) variables. The Design Science Research (DSR) methodology was employed to create a practical artifact in the form of a deep learning model for TB detection. The study followed established DSR guidelines, progressing through problem identification, objective definition, artifact design and development, demonstration, evaluation, and communication, as detailed in Table 1.

Table 1 Summary of Design Science Research (DSR) Process

DSR Process	Techniques / Methods Used	Outputs / Outcomes
Problem Identification	Systematic literature review using Scopus, Google Scholar, ScienceDirect, and WHO reports; analysis of TB diagnostic workflows in Tanzania; review of existing CAD and AI-based TB detection systems. In this study, the independent variable is the set of features extracted from chest X-ray images, while the dependent variable is the predicted tuberculosis classification outcome (TB-positive or TB-negative)	Clearly defined research problem highlighting limitations of existing TB diagnostic methods and CAD systems in low-resource settings (infrastructure constraints, limited accessibility, shortage of skilled personnel)
Definition of Objectives and Requirements	The requirements analysis phase was conducted through a literature study, which involved reviewing existing research on computer-aided diagnosis (CAD) systems for tuberculosis detection, deep learning applications in medical imaging, and challenges associated with TB diagnosis in low-resource settings. Key sources included peer-reviewed articles, WHO reports, and prior studies such as Rahman et al.[5] .	Defined artefact objectives and requirements: use of standard chest X-ray images, low computational cost, offline operability, high diagnostic accuracy, and interpretability for clinicians [5].
Artefact Design and Development	Dataset acquisition from Benjamin Mkapa Hospital and public datasets; image preprocessing (resizing, normalization, noise filtering); data augmentation; CNN architecture selection (VGG16, VGG19, ResNet50); transfer learning; hyperparameter tuning using TensorFlow and Keras	Developed and optimized VGG19-based TB detection model capable of extracting discriminative radiological features from chest X-ray images
Demonstration	Application of trained model to unseen chest X-ray images; classification into TB-positive and TB-negative cases; visualization of prediction probabilities	Functional demonstration of the artefact's ability to classify TB cases using standard digital chest X-ray inputs
Evaluation	Quantitative model evaluation using accuracy, sensitivity, specificity, precision, recall, and F1-score. Usability evaluation using structured Likert-scale questionnaires administered to 21 healthcare professionals (medical doctors, clinical officers, and radiologists) from Benjamin Mkapa Hospital, selected through purposive sampling based on their experience in interpreting chest X-rays.	Empirical evidence of model performance and usability. Questionnaire results indicated high user acceptance, with 90.5% agreement on model performance, 85.7% on clinical relevance, and 95.2% on system reliability, confirming the system's clinical applicability.

 J Health Inform Afr. 2026;13(1):138-152. DOI: 10.12856/JHIA-2026-v13-i1-627

Communication	Scientific documentation and dissemination through peer-reviewed publication	Contribution of a validated, deployment-aware AI artefact and methodological knowledge to the health informatics community

3.2 Data Collection

To develop a successful deep learning model, the training data must possess several key qualities. The data should be representative, properly labeled, resistant to noise, and contain specific patterns [8]. Three primary methods for obtaining data are described in the literature: data discovery, data augmentation, and data generation. The first method, data discovery, involves searching for new datasets. Data augmentation enhances the data discovery phase by providing new data from external sources when the discovered datasets are insufficient [5]. Data synthesis is necessary when external datasets are limited or unavailable; it involves generating artificial data that mimic real-world data. The sample dataset of the chest x-ray images used in this study is shown in Figure 1.

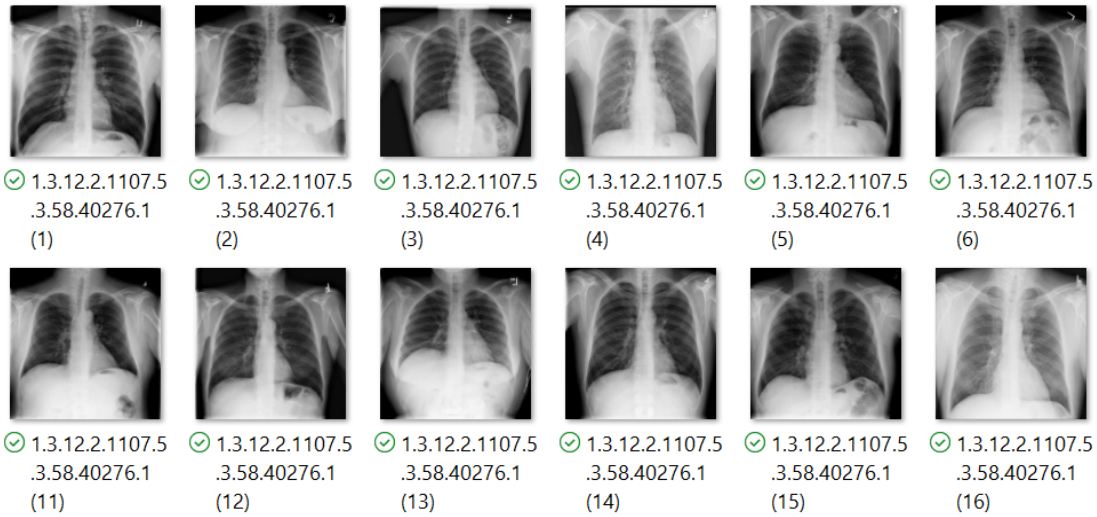

Figure 1 Collected chest X-ray images

To meet the study's objectives, the dataset included 401 TB-negative and 531 TB-positive images collected from Benjamin Mkapa Hospital, along with 6,012 TB-negative and 6,851 TB-positive images from the Kaggle repository, resulting in a total of 6,413 TB-negative and 7,382 TB-positive images. Before addressing the class imbalance, initial data preprocessing was conducted, which involved removing images with artifacts, such as medical devices and text overlays, to ensure data quality. After preprocessing, random undersampling was applied a technique that reduces the size of the majority class to match the minority class. A final sample size of 10,674 images was obtained through a structured preprocessing and class-balancing procedure. Initially, 13,795 chest X-ray images were collected from two sources. Images containing medical devices, annotations, or severe artifacts were excluded to enhance the data quality during preprocessing.

To address the class imbalance and reduce the model bias, random undersampling was applied to the majority classes in both datasets. As a result, 337 TB-positive and 337 TB-negative images were retained from the Benjamin Mkapa Hospital dataset, whereas 5,000 TB-positive and 5,000 TB-negative images were selected from the Kaggle dataset, yielding a total of 10,674 images used for model development.

The two datasets were combined through a structured integration process. First, class labels were harmonized to ensure consistent binary categorization of tuberculosis-positive and tuberculosis-negative cases across sources. All images were resized to a uniform spatial resolution and normalized to standardize pixel intensity distributions. Duplicate patient records and overlapping images were removed to prevent data leakage. Finally, stratified sampling was applied to preserve class balance after merging. This approach enabled the construction of a unified and representative dataset suitable for robust deep learning model development.

 J Health Inform Afr. 2026;13(1):138-152. DOI: 10.12856/JHIA-2026-v13-i1-627

The study included presumptive and confirmed cases of TB aged 2 years or older, regardless of sex, with symptoms such as cough, fever, weight loss, loss of appetite, night sweats, or a history of TB, as well as drug or alcohol use. Pregnant women were not included as they were not eligible for X-ray image examination.

3.3 Experimental Setup

To ensure the practical deployment of the proposed system in low-resource healthcare environments, the model was designed to operate on low-cost hardware commonly available in Tanzanian healthcare facilities in which they operate with limited ICT infrastructure and shared computing resources [25]. In this context, a standard laptop refers to a device with minimum specifications in the range of:

- Intel Core i5 (or equivalent) processor
- 8 GB RAM
- At least 256 GB storage
- Optional entry-level GPU such as NVIDIA GTX 1050 or equivalent

During model development and training, Google Colab with NVIDIA Tesla T4 GPU (12 GB VRAM), 12.7 GB RAM, and approximately 33 GB temporary storage was used to accelerate training processes. However, once trained, the model can run on standard laptops without requiring high-performance computing infrastructure.

This configuration reflects the typical computing infrastructure available in many district hospitals and diagnostic centers in Tanzania [26].

The software stack consisted of TensorFlow and Keras, which offer high-level application programming interfaces (APIs) that support the efficient construction, training, and optimization of deep learning models.

Beyond computational considerations, the experimental setup was intentionally designed to address the contextual challenges associated with deploying computer-aided diagnostic (CAD) systems in low-resource settings, such as Tanzania. The VGG19-based model was trained and validated using standard two-dimensional chest X-ray images, which are widely available even in district and peripheral health facilities, thereby eliminating the need for specialized imaging modalities, such as computed tomography. To ensure feasibility within constrained environments, the model was developed to operate on low-cost hardware, including standard laptops and entry-level GPUs, which require minimal computational infrastructure. Furthermore, unlike many existing CAD solutions that depend on continuous Internet connectivity, the proposed system can operate offline once trained, making it suitable for healthcare facilities with intermittent or unreliable network access.

The artifact was designed to be deployable either as a standalone application or through integration with existing Picture Archiving and Communication Systems (PACS), enhancing accessibility without reliance on cloud-based services. In addition, the system minimizes dependence on specialized personnel by employing automated image preprocessing and classification pipelines, supported by intuitive outputs, such as probability scores and heatmap visualizations, that enable frontline clinicians to interpret results without advanced expertise in artificial intelligence. Finally, the model and accompanying documentation are intended for open-source release, supporting scalability, local customization, and capacity building through knowledge transfer and training initiatives aimed at strengthening human resources in medical imaging and artificial intelligence.

3.4 Image Preprocessing and Labelling

The acquired digital X-rays from Benjamin Mkapa Hospital were labeled as TB positive or negative to facilitate proper model training, and also the Kaggle dataset was downloaded already labelled. Labeling was based on clinical symptoms and radiological features, with the guidance of a radiologist. Factors such as malnutrition, HIV, alcoholism, smoking, and diabetes contribute to TB susceptibility [27]. The radiological features that were considered included the presence of cavities in lung tissues, which are commonly seen in advanced TB, the presence of infiltrates, which are areas with increased density suggesting the presence of an acute infection or inflammation caused by TB, and the presence of nodules on X-ray images indicating TB, which are small rounded opacities scattered throughout the lungs [18]. Images were labeled as positive if the patient's history and X-ray image indicated susceptibility to TB, and

 J Health Inform Afr. 2026;13(1):138-152. DOI: 10.12856/JHIA-2026-v13-i1-627

negative otherwise. A range of pre-processing approaches were performed to improve the image brightness and contrast, including noise reduction, smoothing irregularities, and geometric transformation.

Data normalization was essential in ensuring the viability of the deep learning model for TB detection [28]. This technique involves normalizing pixel values in X-ray images, resulting in a faster training process and improved stability during learning. In addition, data augmentation techniques were incorporated into the workflow. Techniques like rotation, flipping, and zooming were used to enhance the dataset's diversity and improve the model's ability to generalize effectively.

3.5 Architecture Selection

Using CNN architecture, this study developed a model to correctly classify TB-positive and TB-negative images. A CNN is a network architecture designed for pixel data processing and image recognition [29]. CNNs have the ability to analyze digital images, assign importance to different elements and objects within the image, and distinguish between them. When it comes to classification techniques, a CNN requires minimal pre-processing [30]. This study employed VGG19 (Visual Geometry Group), a CNN architecture, to develop a model for TB detection. Further, compared to deeper architectures such as ResNet and Inception, VGG19 offers a favorable balance between performance, interpretability, and computational efficiency, which is critical for low-resource clinical environments.

The VGG architecture was chosen for its strong generalization to unseen data, thanks to its deep convolutional layers that progressively capture hierarchical features. Its use of smaller 3x3 filters reduces parameters while retaining the ability to extract fine details, making it effective for image recognition [31]. VGG balances good performance with moderate computational demands, making it more accessible than deeper models like ResNet or Inception. Its straightforward, sequential design is easier to understand and interpret, supporting transparency, with tools like Grad-CAM helping to visualize and explain predictions.

3.6 Model Development with VGG19

The VGG19 architecture was first presented in 2014 by Simonyan and Zisserman in their influential publication titled "Very Deep Learning CNN for Large-Scale Image Recognition" [32]. The researchers successfully showed that a deep neural network could be trained effectively by utilizing 3x3 filters. which allowed the network to reach a considerable depth of 19 layers, this approach enabled the network to achieve a significant depth of 19 layers, resulting in outstanding performance on the complex ImageNet dataset [33]. VGG19, unlike prior architectures, consistently used 3x3 kernels in all of its layers, instead of using a combination of different filter sizes [34]. This architecture of VGG19 is fundamental for its ability to effectively classify tasks that go beyond its original training scope.

To develop a better predictive model using VGG19 architecture, the dataset comprising 674 chest X-ray images, was prepared to aid in the model's development. The finalized dataset was divided into training (80%), validation (15%), and testing (15%) subsets using stratified sampling to maintain class distribution consistency and to ensure that no patient-level overlap occurred across the subsets. The training data played a crucial role in enabling the model to learn and distinguish between patterns associated with TB-positive and TB-negative cases. The remaining 536 images were set aside for validation purposes equating to 5% of the entire dataset, serving as a continuous monitoring tool that provided an unbiased evaluation of the model's performance.

The following formula was used for each convolutional layer to perform a discrete convolution operation in feature extraction. This shows how the model extracts hierarchical features (edges, textures, patterns) from the images.

$$F_{i,j}^{(l)} = \sum_{m=1}^{M} \sum_{n=1}^{N} X_{i+m,j+n}^{(l-1)} \cdot K_{m,n}^{(l)} + b^{(l)}$$

where:

- $X^{(l-1)}$ is the input feature map from the previous layer
- $K^{(l)}$ is the convolution kernel
- $b^{(l)}$ is the bias term

 J Health Inform Afr. 2026;13(1):138-152. DOI: 10.12856/JHIA-2026-v13-i1-627

- $F^{(l)}$ is the resulting feature map at layer l

This operation allows the network to learn spatial patterns associated with TB-related abnormalities.

A modified VGG19-like architecture was adopted for model development. The network is characterized by a series of convolutional layers (Conv2d), each followed by batch normalization (BatchNorm2d) and ReLU activation. Max pooling layers (MaxPool2d) are used to reduce spatial dimensions at several points in the network.

The architecture consists of:

- initial Conv2d-BatchNorm2d-ReLU blocks with 64 filters
- Conv2d-BatchNorm2d-ReLU blocks with 128 filters
- Conv2d-BatchNorm2d-ReLU blocks with 256 filters
- 8 Conv2d-BatchNorm2d-ReLU blocks with 512 filters

After the convolutional layers, the network uses both adaptive average pooling and adaptive max pooling, followed by flattening. The classifier part of the network consists of a series of fully connected layers (Linear), batch normalization, ReLU activation, and dropout, culminating in a final output layer with 2 units.

All convolutional layers use 3x3 filters, and all 20,563,776 parameters in the network are trainable. The network maintains spatial dim3ensions through the convolutional layers and reduces dimensions through max pooling. This architecture has no non-trainable parameters, allowing for full control of the entire network during training.

Table 1 indicates a comprehensive breakdown of the network architecture with the initial input shape being 32 x 3 x 256 x 256. Also, Table 2 provides the VGG19 architecture, which includes output volume sizes for each layer and relevant information about convolutional filter sizes and pooling sizes.

Table 2: Overview of the VGG19 architecture

Layer Type	Output Size	Filter Size
INPUT IMAGE	256×256×3	
Conv2d +BatchNorm2d+ReLU	256×256×64	3×3, K = 64
Conv2d +BatchNorm2d+ReLU	256×256×64	3×3, K = 64
MaxPool2d	128×128×64	2×2
Conv2d +BatchNorm2d+ReLU	128×128×128	3×3, K = 128
Conv2d +BatchNorm2d+ReLU	128×128×128	3×3, K = 128
MaxPool2d	64×64×128	2×2
Conv2d +BatchNorm2d+ReLU	64×64×256	3×3, K = 256
Conv2d +BatchNorm2d+ReLU	64×64×256	3×3, K = 256
Conv2d +BatchNorm2d+ReLU	64×64×256	3×3, K = 256
Conv2d +BatchNorm2d+ReLU	64×64×256	3×3, K = 256
MaxPool2d	32×32×256	2×2
Conv2d +BatchNorm2d+ReLU	32×32×512	3×3, K = 512
Conv2d +BatchNorm2d+ReLU	32×32×512	3×3, K = 512
Conv2d +BatchNorm2d+ReLU	32×32×512	3×3, K = 512
Conv2d +BatchNorm2d+ReLU	32×32×512	3×3, K = 512
MaxPool2d	16×16×512	2×2
Conv2d +BatchNorm2d+ReLU	16×16×512	3×3, K = 512
Conv2d +BatchNorm2d+ReLU	16×16×512	3×3, K = 512
Conv2d +BatchNorm2d+ReLU	16×16×512	3×3, K = 512
Conv2d +BatchNorm2d+ReLU	16×16×512	3×3, K = 512
MaxPool2d	8×8×512	2×2
AdaptiveAvgPool2d+AdaptiveMaxPool2d	1×1×512	
Flatten	1024	
BatchNorm1d + Dropout	1024	
Linear + ReLU	512	
BatchNorm1d + Dropout	512	
Linear	2	

 J Health Inform Afr. 2026;13(1):138-152. DOI: 10.12856/JHIA-2026-v13-i1-627

The development of TB detection model combined VGG19 architectural strengths with the essential components of a CNN This integration leveraged VGG19's efficient use of 3×3 kernels throughout its architecture while incorporating the core functions of CNN layers. This blending empowered the model to analyze chest X-ray images efficiently, transforming raw pixel data into a detailed understanding of TB-related patterns.

3.7 Model Optimization

To optimize the model's performance hyperparameter tuning was employed with key parameters including learning rate, batch size, dropout rate, number of epochs, and the choice of optimization algorithm. The systematic approach involved iterative experimentation using a grid-based search strategy, where one hyperparameter was varied at a time while others were held constant. Model performance was evaluated on the validation set using accuracy and loss metrics. Configurations that minimized validation loss and improved generalization were selected, leading to the final hyperparameter values reported in this study. The learning rate of range of $0.0001(1^{e-4})$ to $0.01(1^{e-2})$ was employed to minimize validation loss while optimizing accuracy, as presented in Figure 2.

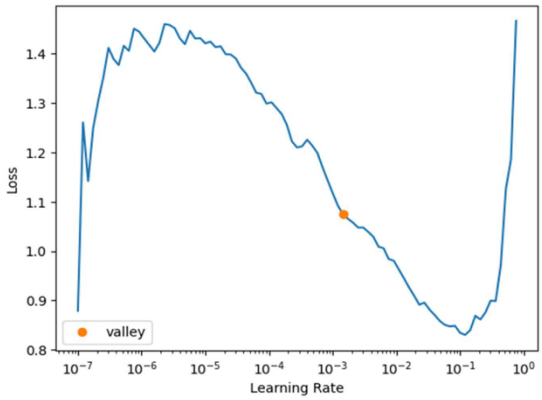

Figure 2 Finding of the Learning Rate

Various batch sizes were tested, with 32 chosen to balance computational efficiency and model accuracy, given hardware constraints. Although larger sizes like 64 have been optimal in other studies [6], the VGG19 architecture required a smaller batch size for stable performance. A 20% dropout rate was used to reduce overfitting by deactivating neurons randomly, maintaining model performance by balancing dropout and avoiding reliance on specific neurons.

To further enhance the model's performance, 50 training epochs were used, influencing the trade-off between better performance on training data and the potential risks of overfitting. The Adam optimizer [35] was chosen for its adaptive learning rate and ability to handle sparse gradients, which were critical for achieving accelerated convergence and improved accuracy in TB detection.

3.8 Model Evaluation

The process of evaluation began at the initial stage of model development, where data was divided into three distinct groups: the training set, the validation set, and the test set. To ensure a thorough evaluation, the validation set played a crucial role in the development of the DL model. The validation set acted as an ongoing monitoring tool, enabling continuous evaluation of the model's performance. The dataset was split into training (80%), validation (5%), and testing (15%) sets, corresponding to 8,538, 536, and 1,600 images respectively.

The effectiveness of the model was assessed to determine its performance in detecting TB by using various metrics and techniques to evaluate the ability of the model in detecting TB through the study of X-ray images. The performance metrics used were Accuracy, Precision, Recall, and F1-Score. Which were computed based on the following mathematical formulars as shown in Table 3:

 J Health Inform Afr. 2026;13(1):138-152. DOI: 10.12856/JHIA-2026-v13-i1-627

- TP (True Positive): Indicates the instances where TB images were correctly detected as TB cases.
- TN (True Negative): Indicates the instances where normal images were correctly detected as non-TB cases.
- FP (False Positive): Indicates the instances where normal images were falsely classified as TB cases.
- FN (False Negative): Indicates the instances where TB images were falsely classified as normal cases.

Further, the proposed artifact was subjected to usability evaluation which involved 21 healthcare professionals from Benjamin Mkapa Hospital, including medical doctors, clinical officers, and radiologists who are routinely involved in TB diagnosis using chest X-ray images. Participants were selected using purposive sampling, ensuring that only professionals with relevant expertise in interpreting radiological images were included.

Table 3: Performance metrics and their corresponding formula

Performance Metric	Formula
Accuracy is the ratio of correctly classified predictions out of the total number of predictions generated from the test set.	$Accuracy\ (A) = \frac{(TP + TN)}{(TP + FN) + (FP + TN)}$
A model's recall is the number of positive cases it correctly detected,	$Recall\ (R) = \frac{TP}{(TP + FN)}$
precision is the accuracy of the model in positive predictions.	$Precision\ (P) = \frac{TP}{(TP + FP)}$
The F1 score is a performance evaluation metric that measures the accuracy of a model by taking into account both its precision and recall values	$F1\ Score = \frac{(2 * TP)}{(2 * TP + FN + FP)}$

The validation of the developed TB detection system was conducted in real clinical settings at Benjamin Mkapa Hospital (BMH) in Dodoma, Tanzania. During this phase, healthcare professionals interacted with the system by uploading chest X-ray images, reviewing the model's predictions, and assessing the usability of the web-based interface. The evaluation followed a naturalistic validation approach, allowing the system to be tested within the real clinical workflow of the hospital environment. This approach ensured that the results reflect the practical applicability of the model in real healthcare settings.

The evaluation was conducted using structured questionnaires based on a Likert scale, allowing respondents to rate different aspects of the system, including model performance, clinical relevance, reliability, security, and overall usability.

The results indicated strong positive feedback regarding the system's usability and effectiveness. Specifically, 90.5% of respondents agreed that the model performs well in distinguishing TB-positive and TB-negative cases, 85.7% agreed on its clinical relevance, and 95.2% confirmed the reliability of the system. These findings suggest that the proposed system has strong potential to support healthcare professionals in TB diagnosis.

3.9 Ethical Consideration

The ethical considerations of research must be considered before, during, and after the research process, encompassing all stages of the research, such as data collection, data storage, processing, and sharing (Ishtiaq, 2019). The researcher obtained research permit and ethical clearance from the University of Dodoma Review Board to ensure compliance with all University ethical and legal regulations. The study has also obtained a research permit from Benjamin Mkapa Hospital to ensure compliance with relevant national regulations and ethical standards. The privacy and confidentiality of participants have been protected by storing data securely. The study is inclusive, and no gender-based discrimination was tolerated. On the other hand, participants were not offered any incentives that may unduly influence their decision to participate.

 J Health Inform Afr. 2026;13(1):138-152. DOI: 10.12856/JHIA-2026-v13-i1-627

4 Results

The DL model for TB detection achieved a remarkable accuracy of 96%, with a precision of 95% and a recall of 98% for the TB-negative class, leading to an F1-Score of 96%. This indicates the model's strong ability to correctly identify all TB-negative instances. For the TB-positive class, the model recorded a precision of 98% and a recall of 95%, resulting in an F1-Score of 96%. Table 4 illustrates more of the performance of the model. Overall, the performance of the model highlights the high effectiveness in distinguishing between TB and non-TB cases, potentially enhancing clinical diagnostic accuracy.

Table 4 Results of Model Development Using VGG19

	Precision	Recall	F1-score	Support
Negative class	95%	98%	96%	800
Positive class	98%	95%	96%	800
Macro average	96%	96%	96%	1600
Weighted average	96%	96%	96%	1600
Accuracy	96%			1600

The confusion matrix, in Table 5, for the TB detection model shows that it correctly identified 783 instances as negative (TN = 783), and misclassify 41 instances as positive (FP-17) indicating a strong accuracy in recognizing negative cases. However, the model did misclassify 7 positive instances as negative (FN = 41), reflecting a minor shortfall in detecting some positive cases. Despite this, the model successfully identified 43 positive instances as positive (TP = 759), demonstrating a solid ability to detect positive cases accurately.

Table 5 Confusion Matrix for VGG19 Model

		Predicted Values	
		Negative	**Positive**
Actual Values	**Negative**	783	17
	Positive	41	759

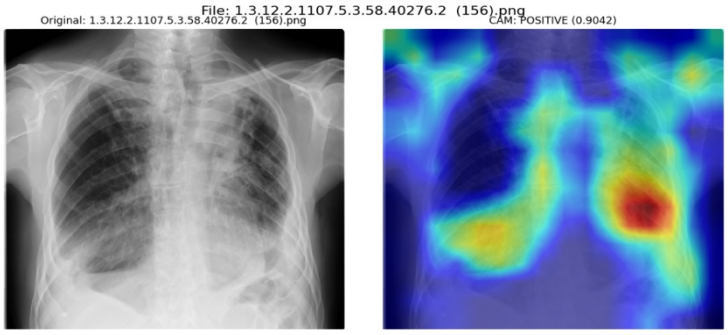

Figure 2 Showing heat map of an X-ray image detected by the model as TB positive.

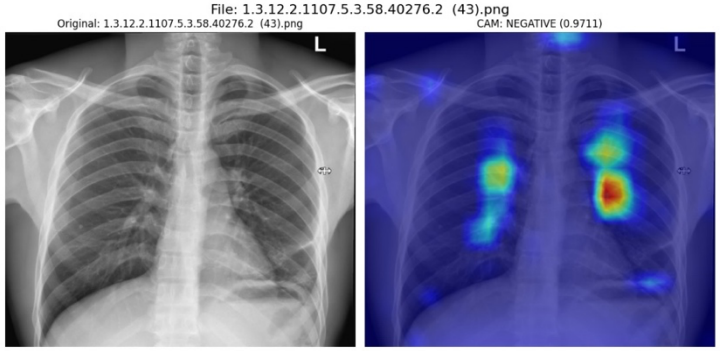

Figure 3 Showing heat map of an X-ray image detected by the model as TB negative.

 J Health Inform Afr. 2026;13(1):138-152. DOI: 10.12856/JHIA-2026-v13-i1-627

The output of the X ray images after being analyzed by deep learning model for TB detection has been indicated in Figures 2 and 3 using heatmaps to visually represent how the deep learning model focuses on specific areas of the X-ray images to make its predictions for TB detection.

In Figure 2, the heat map highlights regions with higher color concentration, indicating the areas the model prioritized when classifying the X-ray image as TB-positive. In contrast, Figure 3 shows the heat map for a TB-negative classification, where the model focused on different regions of the X-ray image and the color concentration shows the area where the model prioritized when making such decision. These visualizations provide insight into the decision-making process of the model, helping to explain how it distinguishes between TB-positive and TB-negative cases based on image features.

5 Discussion

The performance of the VGG19-based deep learning (DL) model in this study in terms of 96% accuracy, 96% F1-score, and balanced precision-recall across TB-positive and TB-negative classesaligns with and in some cases exceeds benchmarks reported in recent literature. For instance, Hansun et al. [27] noted that DL models using chest X-rays often achieve accuracy rates between 85% and 95%, but performance varies significantly depending on dataset quality and model architecture. The high recall for TB-negative cases (98%) and high precision for TB-positive cases (98%) in this study suggest that the model is particularly effective at minimizing both false negatives and false positives, which is critical in TB screening contexts where misdiagnosis can lead to delayed treatment or unnecessary isolation.

A study by Lakhani and Sundaram evaluated several deep CNN architectures, including VGG variants, for TB classification from chest X-rays [18]. Using transfer learning, they reported classification accuracies exceeding 90% on public TB datasets, affirming the feasibility of CNN-based TB detection. The current model exhibits similar performance levels while being tailored for deployment under constrained infrastructure, which was not the primary focus in Lakhani and Sundaram's work.

Further, a study by Rahman et al**.** extended this line of research by incorporating lung segmentation and visualization methods into a VGG-based TB detection framework [5]. Their approach achieved robust classification performance, particularly in scenarios where unauthorized image regions could confound predictions. In contrast, this study emphasizes the importance of infrastructural feasibility and clinician interpretability under real-world constraints (e.g., offline operation and low-cost hardware). While Rahman et al. focused on segmentation-assisted improvements, the present model achieves strong diagnostic results with a simpler pipeline, making it more practical for environments with limited computational resources.

In a comparative evaluation, Hwang et al. analyzed multiple CNN architectures, including VGG19, ResNet50, and Inception networks, for TB screening tasks [19]. They found that models using transfer learning achieved better performance than those trained from scratch, particularly when dataset sizes were moderate. Consistent with their findings, our model benefits from transfer learning, but further distinguishes itself by optimizing for deployment feasibility and interpretability which are key considerations for clinical adoption in low-resource regions.

Moreover, the use of class activation maps (CAMs) to visualize model attention enhances interpretability which is an essential factor for clinical adoption. This aligns with findings by Rahman et al. [28], who emphasized that heatmap visualizations can increase clinician trust in AI systems by revealing the decision-making process. The model's ability to focus on diagnostically relevant regions in TB-positive and TB-negative cases, as shown in Figures 2 and Figure 3, supports its potential as a decision-support tool.

The current study's results not only align with established performance benchmarks but also address practical limitations that have hindered the adoption of AI-based TB diagnostics in low-resource settings. Unlike many prior studies that prioritize maximizing accuracy within controlled experimental environments, this work explicitly integrates deployment-aware design principles such as offline capability, low-cost hardware compatibility, and outputs that clinicians can interpret without advanced technical training.

For example, heatmap-based visualization of model predictions provides clinicians with intuitive insight into model reasoning, thereby addressing a major barrier to trust and adoption identified in the broader literature.

The proposed artefact addresses key challenges in Tanzania by leveraging low-cost digital X-ray systems, requiring minimal additional infrastructure, and operating with limited computational resources.

 J Health Inform Afr. 2026;13(1):138-152. DOI: 10.12856/JHIA-2026-v13-i1-627

Its automated nature reduces reliance on highly specialized personnel, making it suitable for deployment in peripheral healthcare facilities.

However, the study's reliance on a limited dataset introduces concerns about generalizability. As highlighted by Oloko-Oba and Viriri, DL models trained on homogeneous datasets may underperform when exposed to images from different populations or imaging protocols [20]. Future work should incorporate multi-institutional datasets and explore domain adaptation techniques to enhance robustness. The model's performance should be prospectively validated in operational clinical settings to assess external generalizability. Additionally, integration with existing hospital information systems and workflow evaluation remains an important step toward translation into practice.

6 Conclusion

The study has successfully developed and validated a deep learning model using a VGG19 architecture to detect tuberculosis (TB) from chest X-ray images. The model demonstrated exceptional accuracy in distinguishing TB-positive from TB-negative cases, The findings indicate that deep learning approaches, particularly those based on CNN, can greatly enhance TB detection, offering a rapid, cost-effective, and accessible diagnostic tool. This technology holds promise for improving TB management, especially in resource-limited settings where traditional diagnostic methods may be inadequate.

Despite achieving high accuracy, the study has several limitations. First, the dataset primarily consisted of X-ray images from a limited number of sources, which may not represent all potential variations seen in different populations or healthcare settings. Additionally, the model's performance is influenced by the quality and characteristics of the input images, meaning that poor-quality or heavily artifact-laden images could degrade the model's effectiveness.

The research paves the way for broader applications in other areas of medical imaging. Future research should focus on integrating other imaging modalities, such as CT scans or MRI to further enhance the model's accuracy. Lastly developing mobile applications that can incorporate this TB detection model would increase accessibility in low-resource areas, while ensuring compatibility with existing healthcare systems.

Acknowledgements

This work was carried out with the aid of a grant from the Artificial Intelligence for Development in Africa Program, a program funded by Canada's the International Development Research Centre, Ottawa, Canada and the Swedish International Development Cooperation Agency as part of capacity building under AI4D Lab, Grant no. 110470-001.

Statement on conflicts of interest

Not Applicable.

References

[1] V. Acharya and others, "AI-Assisted Tuberculosis Detection and Classification from Chest X-Rays Using a Deep Learning Normalization-Free Network Model," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–19, 2022, doi: 10.1155/2022/2399428.

[2] J. Chakaya and others, "Global Tuberculosis Report 2020 – Reflections on the Global TB burden, treatment and prevention efforts," *Int J Infect Dis*, vol. 113, pp. S7–S12, 2021, doi: 10.1016/j.ijid.2021.02.107.

[3] T. Lane and others, "Supplemental data Comparing and Validating Machine Learning Models for Mycobacterium tuberculosis Drug Discovery." 2022.

[4] M. Muyoyeta and others, "The sensitivity and specificity of using a computer aided diagnosis program for automatically scoring chest X-rays of presumptive TB patients compared with Xpert MTB/RIF in Lusaka Zambia," *PLoS One*, vol. 9, no. 4, 2014, doi: 10.1371/journal.pone.0093757.

[5] T. Rahman and others, “Reliable tuberculosis detection using chest X-ray with deep learning, segmentation and visualization,” *IEEE Access*, vol. 8, pp. 191586–191601, 2020, doi: 10.1109/ACCESS.2020.3031384.
[6] E. N. Cui and others, “Radiomics model for distinguishing tuberculosis and lung cancer on computed tomography scans,” *World J. Clin. Cases*, vol. 8, no. 21, pp. 5203–5212, 2020, doi: 10.12998/wjcc.v8.i21.5203.
[7] S. Kadry, G. Srivastava, V. Rajinikanth, S. Rho, and Y. Kim, “Tuberculosis Detection in Chest Radiographs Using Spotted Hyena Algorithm Optimized Deep and Handcrafted Features,” *Comput Intell Neurosci*, vol. 2022, pp. 1–12, 2022, doi: 10.1155/2022/9263379.
[8] B. O. Botwe and others, “The integration of artificial intelligence in medical imaging practice: Perspectives of African radiographers,” *Radiography*, vol. 27, no. 3, pp. 861–866, 2021, doi: 10.1016/J.RADI.2021.01.008.
[9] Z.-L. Han *et al.*, “A systematic review and meta-analysis of artificial intelligence software for tuberculosis diagnosis using chest X-ray imaging.,” *J. Thorac. Dis.*, vol. 17, no. 5, pp. 3223–3237, May 2025, doi: 10.21037/jtd-2025-604.
[10] F. Pérez-García, R. Sparks, and S. Ourselin, “TorchIO: A Python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning,” *Comput Methods Programs Biomed*, vol. 208, 2021, doi: 10.1016/j.cmpb.2021.106236.
[11] D. A. Winkler, “The impact of machine learning on future tuberculosis drug discovery,” *Expert Opin. Drug Discov.*, vol. 17, no. 9, pp. 925–927, 2022, doi: 10.1080/17460441.2022.2108785.
[12] A. Baddeley and others, “Acknowledgements 197 countries and territories that reported data >500 people who contributed to reporting and review of data,” World Health Organization, 2021. [Online]. Available: https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2021
[13] “TB Treatment Outcome | National Tuberculosis & Leprosy Programme.” Accessed: Feb. 07, 2026. [Online]. Available: https://ntlp.go.tz/tuberculosis/treatment-outcome/
[14] M. Krafft, “Unlocking insights and driving action in Tanzania with DHIS2 scorecards,” DHIS2. Accessed: Feb. 07, 2026. [Online]. Available: https://dhis2.org/tanzania-scorecard-app/
[15] A. Esteva *et al.*, “Dermatologist-level classification of skin cancer with deep neural networks,” *Nature*, vol. 542, no. 7639, pp. 115–118, 2017, doi: 10.1038/nature21056.
[16] V. Gulshan *et al.*, “Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs,” *JAMA*, vol. 316, no. 22, pp. 2402–2410, 2016, doi: 10.1001/jama.2016.17216.
[17] I. D. Apostolopoulos and T. A. Mpesiana, “COVID-19: Automatic detection from X-ray images utilizing transfer learning with convolutional neural networks,” *Phys. Eng. Sci. Med.*, vol. 43, no. 2, pp. 635–640, 2020, doi: 10.1007/s13246-020-00865-4.
[18] P. Lakhani and B. Sundaram, “Deep learning at chest radiography: Automated classification of pulmonary tuberculosis by using convolutional neural networks,” *Radiology*, vol. 284, no. 2, pp. 574–582, 2017, doi: 10.1148/radiol.2017162326.
[19] S. Hwang, H.-E. Kim, J. Jeong, and H. J. Kim, “A novel approach for tuberculosis screening based on deep convolutional neural networks,” *Radiology*, vol. 290, no. 3, pp. 697–706, 2019, doi: 10.1148/radiol.2018180932.
[20] M. Oloko-Oba and S. Viriri, “Deep learning for tuberculosis detection: A systematic review,” *Diagnostics*, vol. 12, no. 3, p. 748, 2022, doi: 10.3390/diagnostics12030748.
[21] S. Mugusi *et al.*, “Diagnostic accuracy of sputum smear microscopy compared with culture in patients with tuberculosis confirmed by Xpert Mycobacterium tuberculosis/rifampicin assay in Tanzania: A cross-sectional study,” *J. Int. Med. Res.*, vol. 53, no. 10, p. 3000605251390703, 2025, doi: 10.1177/03000605251390703.
[22] D. Pamba and others, “Tuberculosis service delivery challenges and their mitigations during the COVID-19 pandemic in Tanzania: a qualitative study,” *BMJ Open*, 2025.
[23] S. Mfinanga and others, “Increased tuberculosis case detection in Tanzanian children and adults using African giant pouched rats,” *BMC Infect. Dis.*, vol. 24, p. 9313, 2024, doi: 10.1186/s12879-024-09313-0.
[24] M. of H. United Republic of Tanzania, “National Tuberculosis and Leprosy Programme Epidemiological Review January 2023,” National Tuberculosis and Leprosy Programme, 2023.

 J Health Inform Afr. 2026;13(1):138-152. DOI: 10.12856/JHIA-2026-v13-i1-627

[25] United Republic of Tanzania, "Data Systems, Data Generation and Data Use in the Health and Social Welfare Sector," Tanzania Development Partners Group, 2022. [Online]. Available: https://tzdpg.or.tz/wp-content/uploads/2022/06/Data-Systems-Data-Generation-and-Data-Use.pdf
[26] A. Mwogosi, "Factors influencing electronic health record system use in public health facilities in Tanzania," *Discov. Public Health*, vol. 22, no. 1, p. 681, Nov. 2025, doi: 10.1186/s12982-025-01094-4.
[27] G. B. Migliori *et al.*, "Tuberculosis and COVID-19 co-infection: description of the global cohort," *Eur. Respir. J.*, vol. 59, no. 3, Mar. 2022, doi: 10.1183/13993003.02538-2021.
[28] C. Garbin, X. Zhu, and O. Marques, "Dropout vs. batch normalization: an empirical study of their impact to deep learning," *Multimed. Tools Appl.*, vol. 79, no. 19–20, pp. 12777–12815, 2020, doi: 10.1007/s11042-019-08453-9.
[29] Y. Sun, B. Xue, M. Zhang, G. G. Yen, and J. Lv, "Automatically Designing CNN Architectures Using the Genetic Algorithm for Image Classification," *IEEE Trans. Cybern.*, vol. 50, no. 9, pp. 3840–3854, Sep. 2020, doi: 10.1109/TCYB.2020.2983860.
[30] J. J. Yeh, H. C. Lin, Y. C. Yang, C. Y. Hsu, and C. H. Kao, "Asthma Therapies on Pulmonary Tuberculosis Pneumonia in Predominant Bronchiectasis–Asthma Combination," *Front. Pharmacol.*, vol. 13, Mar. 2022, doi: 10.3389/fphar.2022.790031.
[31] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," Sep. 2014.
[32] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition." 2014.
[33] A. Mahmoud *et al.*, "Advanced Deep Learning Approaches for Accurate Brain Tumor Classification in Medical Imaging," *Symmetry*, vol. 15, no. 3, 2023, doi: 10.3390/sym15030571.
[34] U. Muhammad, W. Wang, S. P. Chattha, and S. Ali, "Pre-trained VGGNet Architecture for Remote-Sensing Image Scene Classification," *Proc. - Int. Conf. Pattern Recognit.*, vol. 2018-Augus, pp. 1622–1627, 2018, doi: 10.1109/ICPR.2018.8545591.
[35] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," Jan. 30, 2017, *arXiv*: arXiv:1412.6980. doi: 10.48550/arXiv.1412.6980.

 J Health Inform Afr. 2026;13(1):138-152. DOI: 10.12856/JHIA-2026-v13-i1-627