



Detection of Pulmonary Tuberculosis in Chest X-Ray Images Using CNN-VGG19: Evidence from a Local Indonesian Clinical Dataset

Emil Agusalm Habi Talib¹, Mar'atul Azizah^{2*}, Muhammad Faisal³, Abd Rakhim Nanda⁴, Nizirwan Anwar⁵,
Indriyanti Azis⁶, Rizki Yusliana Bakti⁷, M. Agusalm⁸, Desi Anggreani⁹

^{1,2,3,7,9}Department of Informatics, Universitas Muhammadiyah Makassar, Makassar, Indonesia

^{4,6,8}Department of Water Resources Engineering, Universitas Muhammadiyah Makassar, Makassar, Indonesia

⁵Department of Engineering Informatics, Universitas Esa Unggul, Jakarta, Indonesia

*105841105222@student.unismuh.ac.id

Abstract

Pulmonary tuberculosis (TB) remains a major public health challenge, especially in developing countries where early and accurate diagnosis is critical. Although deep learning models have shown promising performance in TB detection using chest X-ray (CXR) images, most studies rely on large-scale public datasets that may not adequately represent the variability and limitations of real-world clinical data. This study aimed to evaluate the robustness and generalization capabilities of the CNN-VGG19 architecture when applied to small-scale clinical datasets. The dataset consisted of 434 CXR images collected from Haji Makassar Hospital, Indonesia, and categorized into tuberculosis, pleural effusion, and normal classes. A transfer learning approach was applied, and model performance was evaluated using analyses of accuracy, precision, recall, F1 score, and confusion matrices, with comparisons to ResNet50 and InceptionV3 under identical experimental conditions. The results showed that CNN-VGG19 achieved an accuracy of 0.8807 and a macro F1 score of 0.8800, outperforming the comparative models. These findings suggest that CNN-VGG19 provides stable performance in a data-constrained environment and highlight the impact of class overlap on classification performance.

Keywords: chest X-ray (CXR); convolutional neural network (CNN); pulmonary tuberculosis; transfer learning; VGG19

How to Cite: E. A. H. . Talib, "Detection of Pulmonary Tuberculosis in Chest X-Ray Images Using CNN-VGG19: Evidence from a Local Indonesian Clinical Dataset", *J. RESTI (Rekayasa Sist. Teknol. Inf.)*, vol. 10, no. 5, pp. 1550–1558, Oct. 2026, doi: [10.29207/resti.v10i5.7607](https://doi.org/10.29207/resti.v10i5.7607).
Permalink/DOI: <https://doi.org/10.29207/resti.v10i5.7607>

Received: February 26, 2026

Accepted: July 16, 2026

Available Online: October 18, 2026

This is an open-access article under the [CC BY 4.0 License](https://creativecommons.org/licenses/by/4.0/)

Published by [Ikatan Ahli Informatika Indonesia](https://www.iaii.or.id/)

1. Introduction

Tuberculosis (TB), a highly infectious disease, remains a major public health challenge worldwide, particularly in low- and middle-income countries such as Indonesia [1]. Timely and accurate detection is essential to curb transmission and improve patient outcomes. For initial screening, chest X-ray (CXR) imaging is widely used because of its accessibility, rapid image acquisition, and relatively low cost compared with computed tomography (CT) and molecular testing. However, interpreting CXR images demands substantial radiological expertise. Differences in imaging equipment and exposure settings can affect image quality, while overlapping pulmonary abnormalities can further complicate interpretation [2].

Developments in deep learning have highlighted the considerable potential of convolutional neural networks (CNNs) for automated TB detection using CXR images.

Rather than relying solely on manually designed features, CNNs learn hierarchical representations of radiographic images through successive processing layers. Early layers extract basic visual features, including edges and textures, while deeper layers integrate these features into increasingly complex patterns relevant to classification. Previous studies have demonstrated that CNN-based methods can achieve better diagnostic performance than conventional approaches [3], [4]. These capabilities have encouraged the application of deep learning to CXR analysis and reflect the growing role of artificial intelligence in computer-aided diagnosis (CAD) systems for TB screening [5].

Despite these advances, most existing studies rely heavily on publicly available datasets such as Montgomery, Shenzhen, and Kaggle [6]. These datasets are generally curated under relatively controlled conditions, resulting in more consistent image quality

and limited variability. In contrast, real-world clinical datasets often exhibit substantial heterogeneity in terms of contrast, noise levels, imaging artefacts, acquisition protocols, and patient conditions. Such variability may significantly affect model generalization and robustness when deployed in clinical practice. This discrepancy raises concerns regarding the applicability of models trained on public datasets to real clinical environments, particularly in developing countries such as Indonesia, where variations in radiological data are more pronounced [1], [6].

In the Indonesian context, studies specifically investigating the performance of deep learning models on localized hospital-based datasets remain limited. Clinical datasets collected from single institutions can better reflect real diagnostic conditions, including variations in imaging equipment, patient demographics, and disease manifestations. However, these datasets are typically limited in size, which introduces challenges related to overfitting and model stability. To address this limitation, transfer learning has been widely adopted as an effective approach to leverage knowledge from large-scale datasets and improve performance on smaller medical datasets [7].

Although more recent architectures, such as Vision Transformers and attention-based CNN models, have shown promising results in medical image analysis, their effectiveness often depends on access to large-scale datasets and substantial computational resources. In scenarios where data are limited and computational efficiency is required, classical architectures such as VGG19 remain relevant due to their relatively simple structures and stable feature-extraction capabilities. Previous studies have reported that VGG19, when combined with transfer learning, can achieve competitive performance in various medical imaging tasks, including TB detection and multi-class lung disease classification [7], [8].

Based on these considerations, this study evaluates the performance of a CNN-VGG19 model on a localized clinical dataset collected at Haji Makassar Hospital, Indonesia. Unlike prior studies that primarily rely on public datasets, this research emphasizes real-world variability in radiological data and examines its influence on model performance. In addition, a comparative evaluation with ResNet50 and InceptionV3 is conducted under identical experimental conditions to ensure a fair and objective assessment. The contribution of this study lies in providing empirical evidence regarding the performance and stability of CNN-VGG19 on a small-scale, real-world clinical dataset, while supporting the development of practical and reliable computer-aided diagnosis (CAD) systems for TB screening in healthcare environments with limited resources.

2. Methods

2.1 Research Framework

This study employed a structured experimental framework to support reproducibility, methodological clarity, and clinical relevance in developing an automated system for pulmonary tuberculosis detection. The framework combines data-driven modelling with radiological diagnostic workflows through a sequence of stages comprising retrospective clinical data acquisition, preprocessing, model development, and performance evaluation.

This structured design enables a transparent assessment of model behaviour while minimizing potential bias during data preparation and training processes. Furthermore, the framework is intended to reflect real-world clinical scenarios, where models are trained on available historical data and evaluated on unseen cases, thereby supporting practical deployment considerations.

2.2 Data Collection

The dataset used in this study was obtained from the Department of Radiology at Haji Makassar Hospital, Indonesia. It consists of 434 chest X-ray (CXR) images categorized into three classes: tuberculosis (TB) ($n = 160$), pleural effusion (EP) ($n = 160$), and normal ($n = 114$). TB cases were clinically confirmed by bacteriological examinations, such as GeneXpert, and correlated with radiological findings, while normal cases were selected from patients without significant lung abnormalities based on radiological interpretation. The pleural effusion class encompasses non-TB conditions that may exhibit radiographic features similar to TB, thereby complicating classification.

All images were originally stored in DICOM format and exported in anonymized form to preserve patient confidentiality. The dataset reflects real-world clinical variability, including differences in image quality, acquisition protocols, and potential radiographic artefacts. Although the dataset size is relatively limited, this condition is representative of practical constraints in single-institution clinical settings. The class distribution shows a moderate imbalance, particularly in the normal category, which contains fewer samples. However, this distribution mirrors real clinical conditions and is therefore retained to maintain ecological validity in model evaluation [8], [9].

2.3 CNN-VGG19 Architecture

This study presents a CNN-VGG19 model that incorporates transfer learning to effectively extract characteristic features from chest X-ray (CXR) images. The backbone is VGG19 pretrained on ImageNet, selected for its ability to learn stable, hierarchical feature representations, particularly in medical imaging settings with limited data [10], [11].

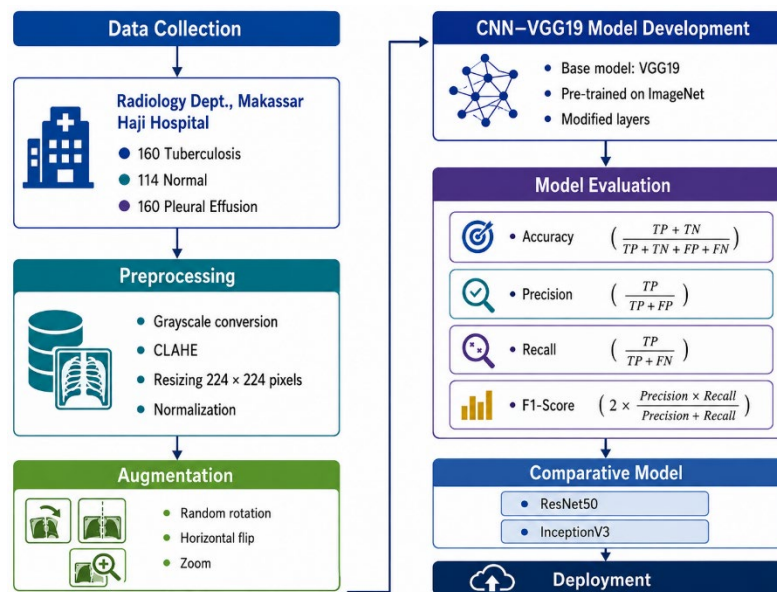


Figure 1. Research framework for tuberculosis detection system in the lungs based on CNN-VGG19

The workflow shown in Figure 1 outlines the stages of CXR classification using CNN-VGG19. A total of 434 images were obtained from Haji Makassar Hospital, comprising 160 tuberculosis, 114 normal, and 160 pleural effusion images. To standardize the inputs and improve the visibility of radiographic structures, preprocessing included conversion to grayscale, contrast enhancement with Contrast Limited Adaptive Histogram Equalisation (CLAHE), resizing to 224×224 pixels, and normalization. Random rotations, horizontal flips, and zooming were used for data

augmentation to introduce greater variation into the training data and reduce the risk of overfitting. Through transfer learning, the ImageNet-pretrained VGG19 network was adapted to classify images into the three categories. Performance was evaluated using accuracy, precision, recall, and F1-score, with comparisons against ResNet50 and InceptionV3 conducted under the same experimental conditions. The deployment stage refers to the model's planned future use in screening support; it does not indicate that the model has already been implemented in clinical practice.

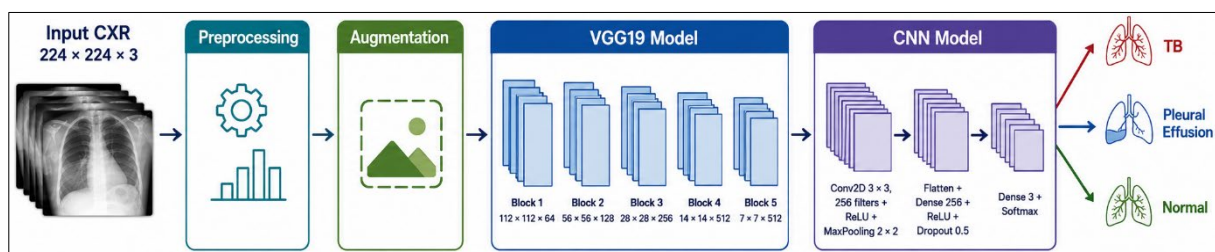


Figure 2. The CNN-VGG19 framework developed, aimed at the categorization of chest X-ray images (CXR), separates them into three different categories: tuberculosis (TB), normal, and pleural effusion

As depicted in Figure 2, a pre-processed chest X-ray (CXR) image, measuring $224 \times 224 \times 3$, is fed into the VGG19 model configured with `include_top = False`. This arrangement ensures that only five convolution blocks, each consisting of 3×3 convolutions followed by max-pooling 2×2 , serve as a feature extraction mechanism. To reduce the number of parameters and address the overfitting issues common in limited medical image datasets, the built-in fully connected layer of VGG19 was eliminated. Furthermore, the feature map generated by the backbone, measuring $7 \times 7 \times 512$, is compressed through Global Average Pooling 2D, resulting in a feature vector with 512 dimensions.

Furthermore, the feature vectors were analyzed through the classification head, which included a Dense 512 layer (ReLU) paired with Dropout 0.5, a Dense layer 256 (ReLU) combined with Dropout 0.3, an additional

Dropout layer at 0.2, and a SoftMax output layer consisting of three neurons to distinguish tuberculosis (TB), normal cases, and pleural effusion. The relatively sleek design of the classification head architecture seeks to balance the model's complexity and generalization capacity. This architecture has been used in a variety of studies related to the classification of convolutional neural networks (CNN)-based chest x-rays (CXRs) [8], [12].

At the training stage, the weight of most of the VGG19 convolutional layers is frozen first, while the classification head is fully trained. Next, some of the upper convolution blocks are unfrozen and fine-tuned with a small learning rate to match the features to the characteristics of the local CXR image [10], [13]. This gradual transfer learning strategy effectively improves

model generalization while reducing overfitting to limited medical data.

This training strategy is particularly suitable for small-scale clinical datasets, as it helps prevent overfitting while preserving the ability of the model to adapt to domain-specific characteristics.

2.4 Confusion Matrix

The dataset comprised 434 images, including 160 tuberculosis (TB), 160 pleural effusion (EP), and 114 normal images. Stratified partitioning was used to allocate the images to training, validation, and test subsets while preserving the class proportions. This stratified training-validation-test design was adopted to reflect conditions in which a model is developed using historical data and subsequently evaluated on unseen cases. Although k-fold cross-validation is widely recommended for small datasets to strengthen robustness and reduce variability in performance estimates, a fixed data partition was retained in this study to provide a consistent basis for comparing different model architectures. Future work will incorporate k-fold cross-validation to examine model stability and generalisation performance more thoroughly. Model performance was assessed using accuracy, precision, recall, and F1-score, which are established evaluation metrics in medical image classification [14], [15]. These metrics were calculated from the confusion matrix, which summarises correct and incorrect predictions through true-positive (TP), true-negative (TN), false-positive (FP), and false-negative (FN) counts. For the multiclass task, a one-versus-all approach was applied: each class was considered positive in turn, with the remaining classes grouped as negative. This procedure provided a consistent basis for evaluating all categories. The corresponding calculations are presented in Equations 1–4.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

Through the integration of these metrics, the performance of the system can be evaluated more comprehensively, not only in terms of overall accuracy but also in its ability to correctly identify tuberculosis cases and reduce misclassification among categories with similar radiographic characteristics [8], [16]. Additionally, when multiple performance metrics are considered simultaneously, a structured evaluation approach can support more systematic and objective decision-making.

2.5 Comparative Model

To provide an objective assessment of the efficacy of the main CNN-VGG19 model, the study combined two

recognized benchmark architectures: ResNet50 and InceptionV3 [17], [18]. These comparative models were developed, trained, and evaluated under identical experimental settings using the same chest X-ray (CXR) dataset. In addition, consistent pre-processing techniques, data partitioning methods, and evaluation metrics were applied to ensure fair and unbiased performance comparisons. Among the evaluated architectures, ResNet50 was utilized as a comparative model because of its deep residual learning architecture, which employs skip connections to maintain stable gradient propagation during the training process. Like other Convolutional Neural Network (CNN)-based approaches, ResNet50 can perform effective feature extraction and learn complex image representations for classification tasks. Previous studies have demonstrated that CNN architectures achieve strong performance in various image classification applications, including medical image analysis, making ResNet50 an appropriate benchmark model for comparison with VGG19 in terms of network depth and feature extraction capability [19–21].

The InceptionV3 architecture was selected due to its distinctive multi-path convolutional structure, which incorporates kernels of different sizes within a cohesive module. This design enables the concurrent extraction of both fine-grained local features and broader, more comprehensive features. Empirical evidence indicates that InceptionV3 achieves superior performance metrics while effectively balancing accuracy with computational efficiency [22]. Consequently, it serves as an appropriate benchmark for evaluating the comparative performance of CNN-VGG19 when applied to the same Chest X-Ray (CXR) dataset.

3. Results and Discussions

3.1 Image Pre-processing

The image pre-processing stage was applied to standardize the quality of chest X-ray (CXR) images before they were used as input data for the CNN-VGG19 model [23]. This standardization aims to reduce non-diagnostic differences arising from variations in imaging equipment, lighting settings, and image artefacts, thereby ensuring that extracted features more accurately reflect the pathological attributes of lung tissue.

As depicted in Figure 3, the CXR image undergoes an initial transformation into a grayscale representation, then undergoes local contrast augmentation using Contrast-Limited Adaptive Histogram Equalisation (CLAHE), with the aim of highlighting lung structure and anatomical boundaries. Next, the image is scaled to 224×224 pixels, transformed into a three-channel configuration, and normalized in intensity before the training process. Overall, this pre-processing improves the visibility of lung structures without over-amplifying artefacts, encouraging stronger feature extraction, more uniform feature depictions, and improved model adaptability to clinical evaluation datasets.

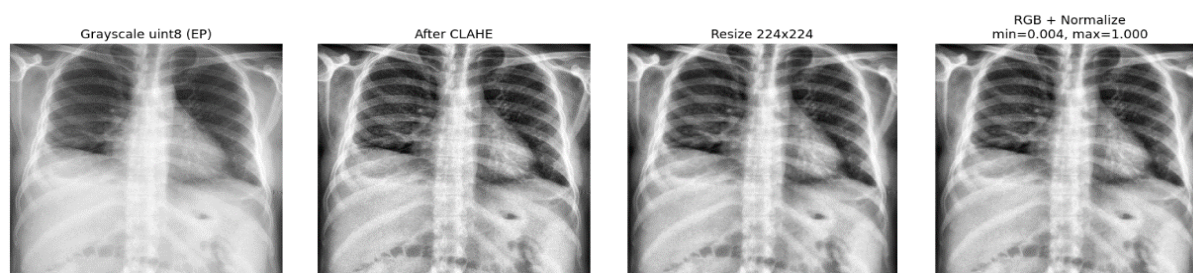


Figure 3. Chest X-ray (CXR) image pre-processing pipeline that includes grayscale conversion

3.2 Model Performance

The CNN-VGG19 model was evaluated on a test set comprising 109 chest X-ray (CXR) images: 40 pleural effusion (EP), 40 tuberculosis (TB), and 29 normal cases. Its multiclass classification performance was quantified using accuracy, precision, recall, and F1-score. Detailed evaluation results are provided in Table 1.

Table 1. Performance Of The CNN-VGG19 Model

Classes	Precision	Recall	F1-Score
EP	0.9459	0.8750	0.9091
Normal	0.8387	0.8966	0.8667
TB	0.8537	0.8750	0.8642
Accuracy: 0.8807			
Macro avg	0.8794	0.8822	0.8800
Weighted avg	0.8835	0.8807	0.8813

Table 1 shows that the proposed CNN-VGG19 model achieved an overall accuracy of 0.8807 and a macro-average F1-score of 0.8800, indicating balanced classification performance across all categories. The weighted-average scores were also comparable to the macro-average values, suggesting that the moderate

class imbalance did not significantly affect the overall model performance [11], [24].

Among the evaluated classes, pleural effusion achieved the highest precision (0.9459) and F1-score (0.9091), while the normal class obtained the highest recall value (0.8966). For tuberculosis detection, the model achieved a precision of 0.8537 and a recall of 0.8750, demonstrating stable performance despite the radiographic overlap between TB and other thoracic abnormalities. Overall, the results indicate that CNN-VGG19 provides reliable performance for multi-class pulmonary disease classification using a localized clinical dataset [15], [25], [26].

3.3 Evaluation of Result

The confusion matrix enables a thorough evaluation of the CNN-VGG19 model's efficacy by revealing the distribution of accurate and incorrect predictions across distinct categories. This examination proves instrumental in identifying classification errors that might be overlooked when relying exclusively on aggregate metrics. Figure 4 presents the confusion matrix, depicted in raw counts, for the standard configurations of the EP, normal, and TB classes.

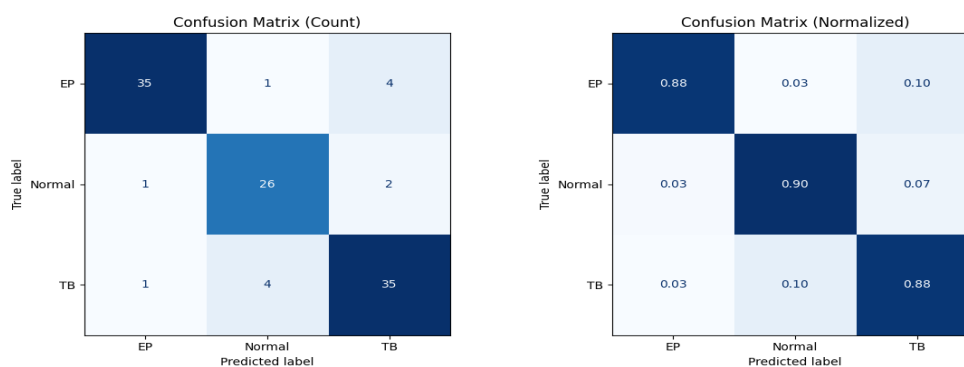


Figure 4. Confusion matrix and normalized confusion matrix model CNN-VGG19

The model shows an overall accuracy of 0.8807, as illustrated in Figure 4. In the context of pleural effusion (EP), 35 of the 40 images were accurately categorized (recall = 0.875), while the rest were erroneously classified as TB ($n = 4$) or normal ($n = 1$). The normal class also showed stable performance, with 26 of 29 images correctly identified (recall = 0.8966); misclassifications were limited to EP ($n = 1$) and TB ($n = 2$). In the TB class, 35 of 40 images were correctly classified (recall = 0.875), whereas a small number were predicted as normal ($n = 4$) or EP ($n = 1$). Overall,

the dominant errors occur between classes with overlapping radiographic characteristics, particularly TB versus EP and TB versus normal, which is consistent with challenges in CXR-based TB recognition, especially for subtle manifestations or lower-contrast findings [9].

3.4 Loss and Accuracy Curve

Figure 5 illustrates the learning dynamics of the CNN-VGG19 model by presenting the accuracy and loss metric trajectories for the training and validation dataset

in two different phases of the training process: (i) the preliminary transfer learning stage in which the VGG19 backbone remains frozen, and (ii) the enhancement stage that focuses on the top layer. Vertical dotted lines mark the start of the refinement process.

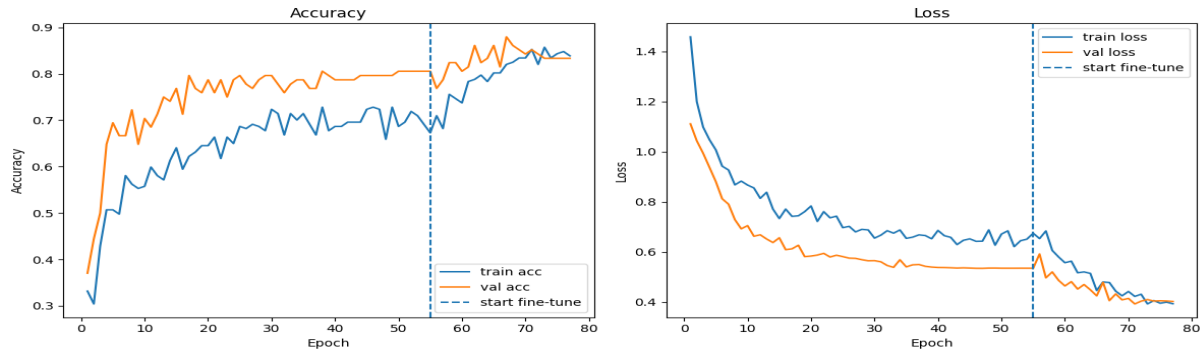


Figure 5. An illustration of the accuracy and loss trajectories demonstrated by the CNN-VGG19 architecture across the training and validation datasets

Concurrently, the loss metric exhibits a consistent downward trajectory across both the training and validation datasets throughout the entirety of the training period. Post-finetuning, the divergence between these two curves is negligible, signifying an absence of substantial overfitting in the model's performance. Collectively, these observed patterns suggest that the synergistic application of transfer learning, gradual fine-tuning, and regularization techniques facilitates the attainment of robust training stability, even when working with comparatively limited datasets [27].

The accuracy curves show a progressive improvement from the early epochs, followed by a more stable increase after fine-tuning begins. Validation accuracy shows a similar pattern, stable in a consistent range as training approaches completion, thus demonstrating increased generalization capabilities and negligible differences between training and validation metrics.

3.5 Comparison of Model

In pursuit of an impartial appraisal of the CNN-VGG19 model's efficacy, a direct empirical comparative analysis was undertaken against two prevalent convolutional neural network designs: ResNet50 and InceptionV3. All models were subjected to equivalent training and evaluation protocols utilizing the same CXR datasets. This was facilitated by a standardized experimental paradigm that incorporated uniform pre-processing techniques, data segmentation strategies, and performance evaluation metrics, thereby ensuring a rigorous and equitable examination.

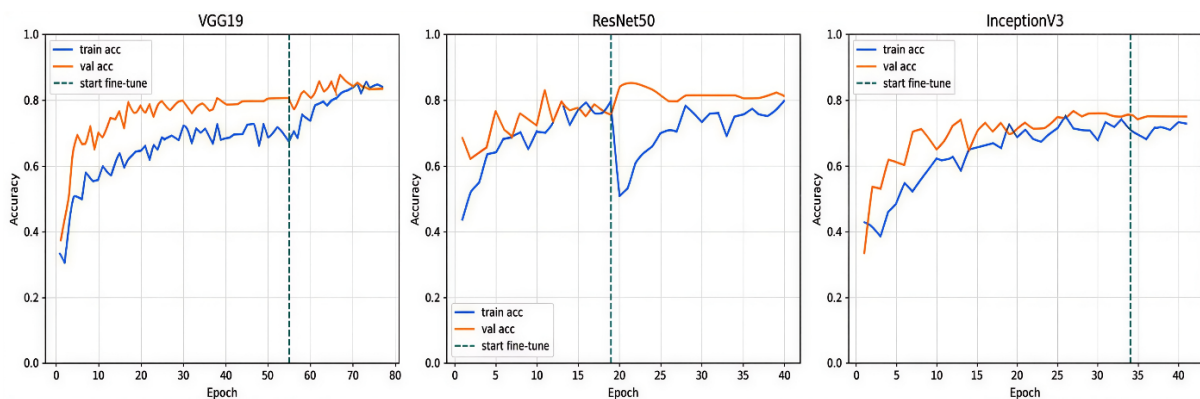


Figure 6. Comparison of the accuracy of CNN-VGG19, ResNet50, and InceptionV3

Figure 6 presents the training and validation accuracy curves for CNN-VGG19, ResNet50, and InceptionV3. Evaluation on the held-out test set from Haji Makassar Hospital showed that CNN-VGG19 achieved the highest accuracy (0.8807), followed by ResNet50 (0.8624) and InceptionV3 (0.6972).

The observed advantages of CNN-VGG19 can be associated with its relatively straightforward and stable architecture, which may be beneficial when learning radiological representations from limited clinical

datasets [28]. By contrast, deeper and more complex architectures such as ResNet50 and InceptionV3 often require larger and more diverse training data to reach optimal generalization. Overall, these findings suggest that VGG19-based transfer learning may be a promising approach for pulmonary TB classification in data-constrained settings, although further validation is needed before clinical implementation [29].

To further validate the findings, a comparison with previous studies is presented in Table 2.

Table 2. Comparison with Previous Studies on TB Detection Using CXR

Study	Dataset	Model	Classes	Accuracy
Mohan et al. [6]	Public dataset	VGG19 + Optimization	TB vs Normal	0.91
Sharma et al. [12]	Public dataset	CNN-based	Multi-class	0.89
Wajgi et al. [29]	Public dataset	Transfer Learning CNN	TB vs Normal	0.93
This Study	Local clinical dataset	CNN-VGG19	3 classes	0.8807

As shown in Table 2, most previous studies report higher accuracy values; however, these studies are predominantly conducted using large-scale and well-curated public datasets. In contrast, the dataset used in this study reflects real-world clinical conditions, which inherently involve greater variability, noise, and class overlap.

The slightly lower accuracy achieved in this study (0.8807) can therefore be attributed to the increased complexity of the dataset, rather than limitations of the model itself. This finding highlights an important trade-off between experimental performance and real-world applicability.

In addition, many earlier studies have focused on binary classification, such as distinguishing TB from normal cases. By contrast, the present study examines a more demanding three-class task involving tuberculosis, pleural effusion, and normal cases. Overlap in the radiographic features of these classes makes classification more difficult, providing a more realistic basis for assessing model performance in relation to the challenges encountered in clinical practice.

3.6 Limitations

Some of the inherent constraints in this study require careful consideration during the interpretation of the results. First, although the dataset was obtained from a real clinical setting, the total number of CXR images remains relatively limited ($n = 434$), and the class distribution, which is moderately imbalanced (TB = 160, EP = 160, normal = 114), may still restrict the model's ability to represent a broader spectrum of radiographic appearances. This constraint is particularly relevant for TB cases with subtle manifestations and for images affected by low quality or acquisition artefacts, which may reduce generalizability when the model is applied to more heterogeneous patient populations [24]. In addition, a moderate class imbalance remains, particularly in the normal class, and was not explicitly addressed using re-sampling or cost-sensitive learning techniques, which may introduce bias toward majority classes. Furthermore, the evaluation strategy relied on a fixed train-validation-test split without incorporating k-fold cross-validation, which may limit the robustness of performance estimation, especially for small-scale datasets. Moreover, explainability techniques such as Gradient-weighted Class Activation Mapping (Grad-CAM) were not incorporated, limiting the ability to verify whether the model focuses on clinically relevant lung regions. Future work will address these limitations by expanding the dataset, incorporating cross-validation strategies, handling class imbalance more

effectively, and integrating explainable artificial intelligence methods to enhance model transparency and clinical trust.

4. Conclusions

This study evaluated the performance of a CNN-VGG19 model for the detection of pulmonary tuberculosis using chest X-ray (CXR) images obtained from a localized clinical dataset at Haji Makassar Hospital, Indonesia. The proposed approach, which integrates transfer learning with a structured pre-processing pipeline, achieved an overall accuracy of 0.8807 and a macro F1-score of 0.8800 in a three-class classification task involving tuberculosis, pleural effusion, and normal cases. These results indicate that the model can deliver stable and balanced performance across classes, despite the inherent variability and limited size of the dataset.

Under the fixed data partition and experimental configuration employed in this study, CNN-VGG19 achieved higher test accuracy than ResNet50 and InceptionV3. This result provides preliminary evidence that the relatively straightforward VGG19 architecture may be suitable for small-scale and heterogeneous clinical datasets. However, the observed performance difference should not be interpreted as definitive evidence of model superiority because the evaluation was based on a single train-validation-test split. Repeated stratified k-fold cross-validation, statistical comparison across multiple runs, and external validation using data from other healthcare institutions are required to confirm the stability and generalizability of these findings. From a practical perspective, this highlights the importance of selecting model architectures that align with data availability and real-world clinical constraints, rather than relying solely on more recent or computationally intensive approaches.

The use of a real-world clinical dataset represents an important contribution of this study, as it reflects the variability in image quality, acquisition protocols, and patient conditions commonly encountered in healthcare settings. Therefore, the findings provide relevant insights into the feasibility of implementing deep learning-based computer-aided diagnosis (CAD) systems for tuberculosis screening, particularly in resource-limited environments where data and computational resources are restricted.

Several limitations should nevertheless be considered when interpreting these findings. The dataset was relatively small and obtained from a single institution, which may restrict the generalizability of the results. The evaluation did not include k-fold cross-validation,

and explainability methods such as Grad-CAM were not used to support clinical interpretation of the model's predictions. Future work should therefore expand the dataset through collaboration across multiple centres, adopt more robust validation procedures, and incorporate explainable AI methods to improve transparency and strengthen clinical trust. Comparisons with newer architectures could also offer further insight into ways to optimize performance under similar conditions.

Declarations

The following declarations are required for every submission. They support transparency, enable proper editorial decision-making, and are a mandatory component of the published article.

AI Usage Statement

Authors use ChatGPT only to help with English editing, grammar improvements, structuring texts, and improving the clarity of the manuscript during the writing process. All scientific content, data analysis, and conclusions are entirely the result of the author's thoughts and responsibilities.

Author Contributions

Conceptualization, E.A.H.T., M.A., and M.F.; methodology, E.A.H.T., M.F., and A.R.N.; software, N.A. and I.A.; Validation, M.F., A.R.N., and R.Y.B.; formal analysis, E.A.H.T., M.A., and M.A. (M. Agusalim); investigation, E.A.H.T., I.A., and D.A.; resources, A.R.N., N.A., and M.A. (M. Agusalim); data curation, E.A.H.T. and D.A.; drafting of original drafts, E.A.H.T. and M.A.; editing and manuscript review, M.F., A.R.N., and R.Y.B.; visualization, N.A. and I.A.; Supervision, M.F., A.R.N., and N.A.; project administration, R.Y.B., M.A. (M. Agusalim), and D.A. All authors have read and approved this published version of the manuscript.

Funding Statement

This research does not receive funding from external parties.

Conflict of Interest Statement

The authors state that there is no conflict of interest in this study.

Ethical Clearance

Ethical approval is not required for this study because it only relies on secondary data in the form of retrospective chest X-ray images that have been anonymized from the Department of Radiology of Haji Makassar Hospital. This research was conducted in accordance with the applicable ethical standards regarding the use of anonymized medical data.

Data and Code Availability

The dataset used in this study was obtained from the Department of Radiology of Haji Makassar Hospital

and cannot be accessed publicly due to institutional confidentiality and patient privacy. The dataset of processed results and program codes produced in this study can be obtained from the authors of correspondence based on reasonable requests.

References

- [1] M. Oloko-Oba and S. Viriri, "A Systematic Review of Deep Learning Techniques for Tuberculosis Detection From Chest Radiograph," *Front. Med.*, vol. 9, no. March, pp. 1–11, 2022, doi: [10.3389/fmed.2022.830515](https://doi.org/10.3389/fmed.2022.830515).
- [2] A. Mirugwe, L. Tamale, and J. Nyirenda, "Improving Tuberculosis Detection in Chest X-Ray Images Through Transfer Learning and Deep Learning: Comparative Study of Convolutional Neural Network Architectures," *JMIRx Med*, vol. 6, pp. e66029–e66029, 2025, doi: [10.2196/66029](https://doi.org/10.2196/66029).
- [3] E. Showkati, M. Salehi, H. Ghaffari, R. Reiazi, and N. Sadighi, "Deep learning-based automatic detection of tuberculosis disease in chest X-ray images," *Polish J. Radiol.*, vol. 87, no. 1, pp. 118–124, 2022, doi: [10.5114/pjr.2022.113435](https://doi.org/10.5114/pjr.2022.113435).
- [4] L. Mangeri, O. S. Gnana Prakasi, P. Kanmani, and N. Puppala, "Chest Diseases Prediction from X-ray Images using CNN Models: A Study," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 10, pp. 236–243, 2021, doi: [10.14569/IJACSA.2021.0121026](https://doi.org/10.14569/IJACSA.2021.0121026).
- [5] C. Ieracitano et al., "A fuzzy-enhanced deep learning approach for early detection of Covid-19 pneumonia from portable chest X-ray images," *Neurocomputing*, vol. 481, pp. 202–215, 2022, doi: [10.1016/j.neucom.2022.01.055](https://doi.org/10.1016/j.neucom.2022.01.055).
- [6] R. Mohan, S. Kadry, V. Rajinikanth, A. Majumdar, and O. Thinnukool, "Automatic Detection of Tuberculosis Using VGG19 with Seagull-Algorithm," *Life*, vol. 12, no. 11, 2022, doi: [10.3390/life12111848](https://doi.org/10.3390/life12111848).
- [7] S. Matarnah, F. Elghaish, F. Pour Rahimian, E. Abdellatif, and S. Abrishami, "Evaluation and optimization of pre-trained CNN models for asphalt pavement crack detection and classification," *Autom. Constr.*, vol. 160, no. December 2023, p. 105297, 2024, doi: [10.1016/j.autcon.2024.105297](https://doi.org/10.1016/j.autcon.2024.105297).
- [8] M. Venkateswarlu and V. R. R. Chirra, "CNN-ViT: A multi-feature learning based approach for driver drowsiness detection," *Array*, vol. 27, no. June, p. 100425, 2025, doi: [10.1016/j.array.2025.100425](https://doi.org/10.1016/j.array.2025.100425).
- [9] V. Sharma, Nillmani, S. K. Gupta, and K. K. Shukla, "Deep learning models for tuberculosis detection and infected region visualization in chest X-ray images," *Intell. Med.*, vol. 4, no. 2, pp. 104–113, 2024, doi: [10.1016/j.imed.2023.06.001](https://doi.org/10.1016/j.imed.2023.06.001).
- [10] M. A. Sayed et al., "Automated lumbar intervertebral disc identification and herniation detection in MR images using cascade CNN architecture," *Informatics Med. Unlocked*, vol. 55, no. February, p. 101648, 2025, doi: [10.1016/j.imu.2025.101648](https://doi.org/10.1016/j.imu.2025.101648).
- [11] I. A. Kandhro et al., "Performance evaluation of E-VGG19 model: Enhancing real-time skin cancer detection and classification," *Heliyon*, vol. 10, no. 10, p. e31488, 2024, doi: [10.1016/j.heliyon.2024.e31488](https://doi.org/10.1016/j.heliyon.2024.e31488).
- [12] A. Sharma, K. Singh, and D. Koundal, "A novel fusion based convolutional neural network approach for classification of COVID-19 from chest X-ray images," *Biomed. Signal Process. Control*, vol. 77, no. April, p. 103778, 2022, doi: [10.1016/j.bspc.2022.103778](https://doi.org/10.1016/j.bspc.2022.103778).
- [13] F. I. AlZobi et al., "Optimized soft-voting CNN ensemble using particle swarm optimization for endometrial cancer histopathology classification," *Comput. Methods Programs Biomed. Updat.*, vol. 8, no. August, p. 100217, 2025, doi: [10.1016/j.cmpbup.2025.100217](https://doi.org/10.1016/j.cmpbup.2025.100217).
- [14] M. Faisal et al., "A Machine Learning-Driven Framework for Evaluating and Clustering City-Level Carbon Trading Strategies," *Int. J. Math. Eng. Manag. Sci.*, vol. 10, no. 6, pp. 2171–2200, 2025, doi: [10.33889/IJMEMS.2025.10.6.101](https://doi.org/10.33889/IJMEMS.2025.10.6.101).
- [15] M. Gour and S. Jain, "Uncertainty-aware convolutional neural network for COVID-19 X-ray images classification," *Comput. Biol. Med.*, vol. 140, no. July 2021, p. 105047, 2022, doi: [10.1016/j.compbiomed.2021.105047](https://doi.org/10.1016/j.compbiomed.2021.105047).
- [16] M. Faisal et al., "A Hybrid MOO, MCGDM, and Sentiment

- Analysis Methodologies for Enhancing Regional Expansion Planning: A Case Study Luwu - Indonesia,” *Int. J. Math. Eng. Manag. Sci.*, vol. 10, no. 1, pp. 163–188, 2025, doi: [10.33889/IJMEMS.2025.10.1.010](https://doi.org/10.33889/IJMEMS.2025.10.1.010).
- [17] P. Arafin and A. H. M. M. Billah, “Deep learning-based concrete defects classification and detection using semantic segmentation,” 2024, doi: [10.1177/14759217231168212](https://doi.org/10.1177/14759217231168212).
- [18] M. A. Rahman et al., “Classification of fungal genera from microscopic images using artificial intelligence,” *J. Pathol. Inform.*, vol. 14, 2023, doi: [10.1016/j.jpi.2023.100314](https://doi.org/10.1016/j.jpi.2023.100314).
- [19] H. Nurrani, A. K. Nugroho, and S. Heranurweni, “Image Classification of Vegetable Quality using Support Vector Machine based on Convolutional Neural Network,” *J. RESTI*, vol. 7, no. 1, pp. 168–178, 2023, doi: [10.29207/resti.v7i1.4715](https://doi.org/10.29207/resti.v7i1.4715).
- [20] R. D. Ramadhani, A. N. A. Thohari, C. Kartiko, A. Junaidi, T. G. Laksana, and N. A. S. Nugraha, “Optimization of the Accuracy of Convolutional Neural Network Methods for Waste Type Identification,” *J. RESTI*, vol. 5, no. 2, pp. 312–318, 2021, doi: [10.29207/resti.v5i2.2754](https://doi.org/10.29207/resti.v5i2.2754).
- [21] M. N. Ichsan, N. Armita, A. E. Minarno, F. D. S. Sumadi, and Hariyady, “Increased Accuracy on Image Classification of Game Rock Paper Scissors using CNN,” *J. RESTI*, vol. 6, no. 4, pp. 606–611, 2022, doi: [10.29207/resti.v6i4.4222](https://doi.org/10.29207/resti.v6i4.4222).
- [22] R. K. Bhukya, D. D. Kayande, A. Jadhav, R. Dawane, and D. Athawale, “Automated Classification of Skin Diseases Using CNNs and Outbreak Implications,” *Procedia Comput. Sci.*, vol. 260, pp. 201–208, 2025, doi: [10.1016/j.procs.2025.03.194](https://doi.org/10.1016/j.procs.2025.03.194).
- [23] M. Assaduzzaman, M. Dutta, A. Saha, and S. G. Paul, “ALSA-3: Customized CNN model through ablation study for Alzheimer’s disease classification,” *Informatics Med. Unlocked*, vol. 50, no. September, p. 101584, 2024, doi: [10.1016/j.imu.2024.101584](https://doi.org/10.1016/j.imu.2024.101584).
- [24] N. L. Yadav, S. Singh, R. Kumar, and D. K. Nishad, *Transfer learning with fuzzy decision support for multi-class lung disease classification: performance analysis of pre-trained CNN models*. 2025. doi: [10.1038/s41598-025-19114-3](https://doi.org/10.1038/s41598-025-19114-3).
- [25] P. Aggarwal, N. K. Mishra, B. Fatimah, P. Singh, A. Gupta, and S. D. Joshi, “COVID-19 image classification using deep learning: Advances, challenges and opportunities,” *Comput. Biol. Med.*, vol. 144, no. March, p. 105350, 2022, doi: [10.1016/j.combiomed.2022.105350](https://doi.org/10.1016/j.combiomed.2022.105350).
- [26] Y. Wang and C. A. Hargreaves, “A Review Study of the Deep Learning Techniques used for the Classification of Chest Radiological Images for COVID-19 Diagnosis,” *Int. J. Inf. Manag. Data Insights*, vol. 2, no. 2, p. 100100, 2022, doi: [10.1016/j.jjime.2022.100100](https://doi.org/10.1016/j.jjime.2022.100100).
- [27] S. Kumar and H. Kumar, “Classification of COVID-19 X-ray images using transfer learning with visual geometrical groups and novel sequential convolutional neural networks,” *MethodsX*, vol. 11, no. July, p. 102295, 2023, doi: [10.1016/j.mex.2023.102295](https://doi.org/10.1016/j.mex.2023.102295).
- [28] T. Le, F. Shi, M. Ge, R. Dong, and D. Shan, “Tuberculosis detection based on VGG19 transfer learning and Zebra Optimization Algorithm,” *EAI Endorsed Trans. Pervasive Heal. Technol.*, vol. 10, pp. 1–7, 2024, doi: [10.4108/eetpht.10.5981](https://doi.org/10.4108/eetpht.10.5981).
- [29] R. Wajgi et al., “Optimized tuberculosis classification system for chest X-ray images: Fusing hyperparameter tuning with transfer learning approaches,” *Eng. Reports*, vol. 6, no. 11, pp. 1–18, 2024, doi: [10.1002/eng2.12906](https://doi.org/10.1002/eng2.12906).