International Journal of Advanced Engineering and Management Research

Vol. 11, No. 04; 2026

ISSN: 2456-3676

Development of AI Model for Tuberculosis Detection in Chest CT Images

Sri Widodo

[1]Health Science Faculty, Universitas Duta Bangsa Surakarta, Central Java, Indonesia

Jalan K.H. Samanhudi 93, Surakarta, 57147, Central Java, Indonesia

doi.org/10.51505/ijaemr.2026.11409 URL: http://dx.doi.org/10.51505/ijaemr.2026.11409

Received: Aug 11, 2026 Accepted: Aug 18, 2026 Online Published: Aug 27, 2026

Abstract

Tuberculosis (TB) remains a major public health issue in Indonesia, characterized by challenges such as delayed early detection and low patient adherence to long-term treatment. Current TB prevention and control efforts still rely heavily on healthcare facilities, while support for technology-based patient self-management has not yet been optimally integrated. Advances in Artificial Intelligence (AI), specifically deep learning, offer opportunities to develop intelligent models that support sustainable patient self-empowerment. The objective of this study is to develop an AI-based application for the early detection of tuberculosis using CT scan images. The research employs an AI model development approach for the early detection of tuberculosis (TB) based on Vision Transformers and CNN-Transformer hybrids, as well as multimodal deep learning that integrates medical imaging and clinical data. A key aspect of this study is the integration of Explainable AI to enhance model transparency and interpretability, thereby fostering trust within the digital health context. The study's novelty lies in the development of an AI-based TB detection model designed to support patient self-management. The findings are expected to serve as a scientific foundation for the future development of digital health technologies for TB control.

Keywords: Artificial Intelligence, CNN, Digital Health, Patient Self-Management, Tuberculosis.

1. Introduction

Tuberculosis (TB) represents a significant global health burden, particularly in countries with a high disease burden such as Indonesia (Heroine et al., 2025). Caused by mycrobacterium tuberculosis, the disease can affect the lungs as well as other organs (*Identification of Tuberculosis : A Review of the Guidelines*, n.d.) and carries a high risk of mortality if not detected and treated in a timely manner. Data indicate that Indonesia has the third-highest number of TB cases globally (Ramadhan & Wulandari, 2023), with hundreds of thousands of new cases occurring annually. This highlights the persistent gap between the actual number of cases and the number of cases formally diagnosed (Sastramandala et al., 2025). Conventional TB diagnostic methods, such as sputum examination and manual X-ray image interpretation, have limitations regarding accuracy, time requirements, and service accessibility, particularly in areas

Vol. 11, No. 04; 2026

ISSN: 2456-3676

with limited healthcare resources (Shrivas & Singh, 2025). This highlights a compelling need to implement advanced technologies capable of enhancing TB screening and early detection (Widodo et al., 2019). With the advancement of artificial intelligence (AI) technology, specifically deep learning, various studies have demonstrated that AI models can assist in disease detection from medical images with high levels of accuracy (Hampiholi & Engineer, 2023)(Ge et al., 2022)(Khosravi et al., 2024)(Vazquez et al., 2025)(Hajiheydari et al., 2025)(Pinto-coelho, 2023)(Santamato et al., 2024). For instance, the development of Convolutional Neural Network (CNN) models for classifying TB from X-ray images has yielded promising performance, achieving high accuracy and F1-scores (Rahman et al., 2025)(Shahreki et al., 2024)(Garud et al., 2023)(Yulvina et al., 2024). Furthermore, digital-based interventions, including the use of reminder systems and chatbots, have been shown to improve patient engagement and treatment adherence, offering potential for adaptation in the care of TB patients (Lee et al., 2023).

The use of Artificial Intelligence (AI) for Tuberculosis (TB) detection has advanced rapidly in recent years. Initially, most research utilized Convolutional Neural Networks (CNNs) as the primary method for classifying medical images, particularly chest X-rays (CXRs). Research conducted by Imron, demonstrated that CNN models could distinguish between images of normal lungs and those of TB patients with an accuracy of approximately 98%. However, these studies relied on two-dimensional X-ray images, resulting in relatively limited anatomical information regarding the lungs, particularly in early-stage TB cases. Subsequent developments have been marked by the use of various modern deep learning architectures, such as AlexNet, ResNet-50, VGG-16, and GoogLeNet (Imron et al., 2020). Based on a systematic review conducted by Hansun, deep learning methods consistently yield higher sensitivity and accuracy compared to conventional machine learning methods. However, the majority of the studies analyzed focused on chest X-ray images, meaning the utilization of chest CT scans as a source of diagnostic information remains underexplored (Hansun et al., 2023).

A more comprehensive study was conducted by Han, through a meta-analysis of various AI software tools for CXR-based TB diagnosis. The findings indicate that AI-based Computer-Aided Diagnosis (CAD) systems can enhance the effectiveness of clinical screening processes. However, the study also revealed heterogeneity in datasets, algorithmic variations, and the need for threshold adjustments, making it difficult to generalize the results across diverse patient populations (Han et al., 2025).

In addition to improved accuracy, recent research has begun to focus on the aspect of model interpretability. Patel developed a self-supervised deep learning approach integrated with Explainable Artificial Intelligence (XAI) techniques. This approach not only yields high classification performance but also provides visualizations of the image regions that form the basis of the model's decision-making. However, the study relied on chest X-ray images and thus did not explore the three-dimensional volumetric information inherent in chest CT scans (Patel & Wong, 2022). The most recent development is marked by the emergence of the Vision Transformer (ViT) architecture. Capell et al. developed a Vision Transformer method combined with Self-Supervised Learning (SSL) to detect tuberculosis in pediatric patients. Their findings indicate that the self-attention mechanism within the Vision Transformer is capable of capturing global spatial relationships that are difficult for conventional CNNs to learn. Nevertheless, Vision Transformers face limitations regarding the need for large amounts of training data and high computational resources; consequently, their performance may degrade when applied to relatively small medical datasets (Capell et al., n.d.).

On the other hand, Agus compared several CNN models, specifically AlexNet and VGG-19, combined with the CLAHE preprocessing technique. Their study demonstrated that VGG-19 yielded superior classification performance compared to AlexNet. However, the research was limited to conventional CNN models and did not incorporate Transformer architectures, which currently represent one of the most promising approaches in computer vision (Agus et al., 2025). Another study by Lumamba developed a Patch-based CNN model integrated with Logistic Regression to classify microscopic TB specimen images. Although this approach produced good results with histopathology images, it was not designed to detect tuberculosis from pulmonary radiological images, making it less relevant for chest CT-based diagnostic applications (Lumamba et al., 2024).

Based on these studies, it can be concluded that most research still focuses on chest X-rays, while studies utilizing chest CT scans remain relatively limited. Furthermore, while CNN models excel at extracting local features, they are less capable of capturing global spatial relationships between lung regions. Conversely, Vision Transformers possess excellent global representation capabilities but require large datasets to achieve optimal performance. To date, research integrating CNNs and Vision Transformers specifically for tuberculosis detection in chest CT images remains very limited.

Based on this research gap, this study proposes a hybrid CNN–Transformer model for tuberculosis detection in chest CT images. The proposed model combines the CNN's ability to extract local features with the Vision Transformer's self-attention mechanism, which is capable of capturing global information. This approach is expected to yield more comprehensive feature representations, thereby enhancing accuracy, sensitivity, and generalization capabilities compared to standalone CNN or Vision Transformer models.

2. Method

2.1. Materials

This study utilizes chest Computed Tomography (CT) images to develop an artificial intelligence model for detecting tuberculosis (TB). The dataset was obtained from two sources: an open/institutional research dataset from the ICMR–National Institute for Research in Tuberculosis (NIRT), India, and clinical data from Ir. Soekarno Regional General Hospital (RSUD) in Sukoharjo, Indonesia. Data from both sources were combined to create a multi-source dataset aimed at enhancing the diversity of image characteristics and evaluating the

ISSN: 2456-3676

model's generalization capability. Using datasets from different institutions also allows the model to be tested against variations in acquisition protocols, patient characteristics, and CT image characteristics. CT data are stored in DICOM format or converted into a numerical format suitable for deep learning processes. Each case is labeled based on available clinical or radiological diagnoses—specifically, TB and non-TB. To ensure evaluation independence, the dataset is split at the patient level rather than the slice level, preventing slices from the same patient from being distributed across both the training and testing sets.

2.2. Research Methodology

The research methodology flowchart is shown in Figure 1.

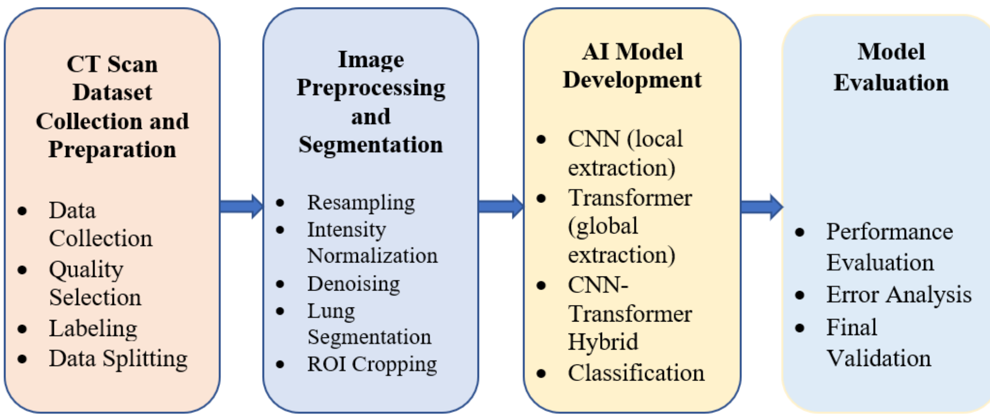

Figure 1. Research Methodology Flowchart

2.2.1. CT-Scan Dataset Collection and Preparation

The first stage involves the collection and preparation of the CT-scan dataset. In this stage, chest CT images are obtained from two sources: the ICMR–National Institute for Research in Tuberculosis (NIRT) and Ir. Soekarno Sukoharjo Regional General Hospital (RSUD). The acquired data subsequently undergo a quality selection process to ensure the images are suitable for model development. Each case is labeled based on available diagnostic information—specifically, TB and Non-TB groups. Finally, the dataset is partitioned into training, validation, and testing sets using a patient-level split approach to prevent data leakage between the groups.

2.2.2. Image Preprocessing and Segmentation.

The second stage involves image preprocessing and segmentation. This stage aims to standardize CT image characteristics and focus the analysis on the lung region. The process begins with resampling to standardize voxel resolution, followed by intensity normalization based on Hounsfield Unit (HU) value characteristics. Subsequently, a denoising process is applied to reduce image noise. Once the image is standardized, lung segmentation is performed to generate a lung mask. The segmentation result is then used for Region of Interest (ROI) cropping, thereby eliminating irrelevant areas outside the lungs and allowing the model to focus more effectively on the structures and patterns of abnormalities within the lungs. The preprocessing stage aims to standardize image quality and improve the signal-to-noise ratio, while segmentation is used to extract the lung area (region of interest).

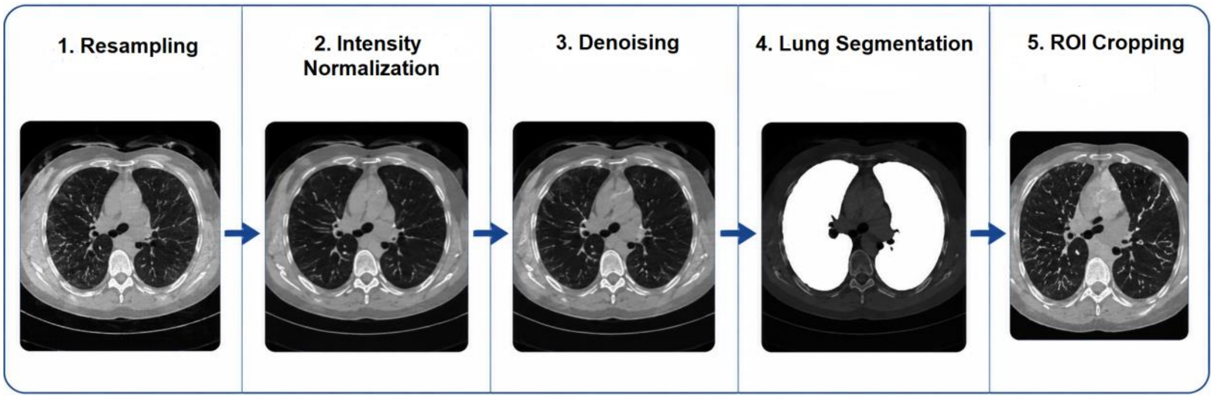

Figure 2. Image Preprocessing and Segmentation Stages

The preprocessing stage is performed to enhance the consistency and quality of CT scan images before they are used as input for the deep learning model. This process comprises five main stages: resampling, intensity normalization, denoising, lung segmentation, and ROI cropping, as illustrated in Figure 2.

a. Resampling.
   Resampling aims to standardize the spatial resolution of CT images obtained from different data sources. Differences in slice thickness, voxel size, and spatial resolution can lead to variations in image characteristics across patients and datasets. Therefore, CT images are resampled to a predetermined resolution to ensure a more consistent spatial representation for the model. In the figure, this process is illustrated using a chest CT image as the initial input.
b. Intensity Normalization.
   CT images have intensity values expressed in Hounsfield Units (HU). While the range of these values can be very broad, the information relevant to lung tissue analysis lies within a specific range. Therefore, intensity range adjustment and normalization are performed to produce a more uniform distribution of pixel values. This process aims to ensure that the model is not unduly influenced by variations in intensity characteristics across different CT scans.
c. Denoising.
   Denoising is used to reduce noise—or unwanted intensity variations—in CT images. Noise reduction is performed while preserving critical lung tissue structures, ensuring that pathological patterns relevant to TB detection are not lost. The output of this stage is a cleaner CT image that is better prepared for segmentation and feature extraction.
d. Lung Segmentation.
   After the images are normalized and noise is reduced, segmentation is performed to separate the lung regions from surrounding structures. This segmentation process generates a lung mask that delineates the lung areas within the CT images. With this mask, lung regions can be

Vol. 11, No. 04; 2026

ISSN: 2456-3676

distinguished from extra-pulmonary tissue, allowing subsequent analysis to focus on the relevant anatomical areas. Unlike X-ray imaging, segmentation in this study is performed directly on chest CT images, ensuring that the lung mask preserves the spatial information contained within the CT slices.

e. ROI Cropping.
The lung mask obtained during the segmentation stage is used to define the boundaries of the lung region to be analyzed. Areas outside the lung ROI are then removed, resulting in an image focused on lung tissue. The final preprocessing output is a standardized CT image focused on the lung area, which subsequently serves as input for the CNN feature extraction and Transformer learning stages.

2.2.3. Development of Artificial Intelligence Models

The third stage involves the development of an Artificial Intelligence model. In this stage, a CNN-Transformer hybrid approach is developed by combining the CNN's capability to extract local features with the Transformer's ability to learn global relationships within the image. CNNs are used to obtain local feature representations through convolution operations, while these features are subsequently processed using the Transformer's self-attention mechanism to capture broader relationships between image regions. Local and global features are then integrated via a feature fusion mechanism to produce a hybrid feature representation. This representation is subsequently fed into a classification head to determine the TB or Non-TB class. Figure 3 illustrates the architecture of the developed model.

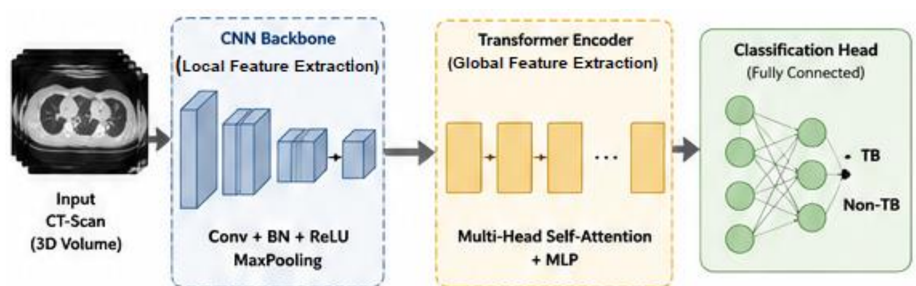

Figure 3. Architecture of the developed model

This study develops a CNN–Transformer hybrid model to detect tuberculosis (TB) using chest CT images. The model architecture is designed by combining the capability of Convolutional Neural Networks (CNNs) to extract local image features with the ability of Transformers to learn global relationships and context across image regions. This hybrid approach is employed because TB-related lung abnormalities exhibit variations in shape, size, texture, and location, necessitating a model capable of understanding both local-level information and broader spatial relationships. The developed architecture comprises three main components: a CNN backbone, a Transformer encoder, and a classification head. These components operate sequentially, beginning with local feature extraction from the CT images, followed by global representation

ISSN: 2456-3676

learning using the Transformer, and concluding with classification into TB and non-TB categories. The details are described as follows:

a. CT Scan Image Input

The model input consists of chest CT images that have undergone quality control, preprocessing, lung segmentation, and ROI extraction. These steps are performed prior to feeding the images into the model to reduce irrelevant information and enhance the consistency of data characteristics. During preprocessing, the CT images are standardized through resampling, intensity normalization, and noise reduction. Subsequently, lung segmentation is employed to isolate the lung regions from surrounding tissues or structures. The segmented areas serve as the Region of Interest (ROI), allowing the model to focus more specifically on lung tissue characteristics associated with the presence of TB. Utilizing the lung ROI is expected to minimize the likelihood of the model learning patterns unrelated to the disease—such as image backgrounds, examination tables, or anatomical structures outside the lung area. Consequently, the input received by the model represents an image format that is more relevant to the classification objective.

b. CNN Backbone for Local Feature Extraction

The first component of the architecture is the CNN backbone. A CNN is employed as the initial feature extractor due to its proficiency in learning local spatial characteristics from medical images. At this stage, CT images are processed through multiple convolutional layers to hierarchically generate feature maps. The initial CNN layers capture fundamental characteristics such as edges, textures, intensities, and local patterns. Subsequently, deeper layers produce increasingly complex and abstract feature representations. In the architecture illustrated in Figure 3, the CNN backbone comprises several key operations: convolution, batch normalization, ReLU, and max pooling. The convolution operation is used to extract local patterns via a filter that moves across the image area. In the context of lung CT images, this process enables the model to learn textural characteristics and local patterns associated with pathological changes caused by TB. Following the convolution process, Batch Normalization is employed to stabilize learning and help maintain the activation distribution during training. Subsequently, the ReLU activation function introduces non-linearity to the network, allowing the model to learn more complex patterns. Max Pooling is used to reduce the spatial dimensions of the feature map while preserving dominant feature information; this dimensionality reduction also helps alleviate the computational load for subsequent stages. The final output of the CNN backbone is a feature map containing representations of local characteristics from the CT image. This feature map then serves as input for the Transformer component.

c. Feature Map Transformation for Transformers

Feature maps generated by a CNN are not used directly for classification. Instead, they are passed to the Transformer stage so that the local information captured by the CNN can be analyzed within a broader context. At this stage, the CNN feature maps are converted into a representation suitable for Transformer processing. This representation is structured into a set of feature tokens or patch-based representations, with each token representing a specific portion of the feature map. Consequently, information originally represented as spatial feature maps can be processed by the Transformer as a sequence of representations. This process is a crucial component of the hybrid architecture, as it enables features learned by the CNN to be further analyzed using the attention mechanism.

d. Transformer Encoder for Global Feature Extraction

The second component is the Transformer Encoder. Unlike CNNs, which predominantly learn local relationships through convolution operations, Transformers are used to learn broader relationships between parts of the image representation. The Transformer Encoder employs a Multi-Head Self-Attention mechanism. This mechanism allows each token to interact with others, enabling the model to determine which parts of the image representation hold significant relevance for the classification process. In the context of TB detection, this capability is crucial because the characteristics of lung abnormalities are not always confined to a single local area; information from various parts of the lung may be interconnected in determining whether an image exhibits TB or non-TB characteristics. The use of multiple attention heads enables the model to learn distinct relationships simultaneously. Each attention head can generate a different relational representation from the feature tokens; this information is then combined to form a more comprehensive representation. Following the Multi-Head Self-Attention process, the feature representations are processed through a Multi-Layer Perceptron (MLP) within the Transformer Encoder. The combination of self-attention and the MLP allows the Transformer to perform deeper transformations of the feature representations. Consequently, the Transformer's output encompasses not only information regarding local characteristics derived from the CNN but also information concerning global relationships between image regions.

e. Integration of CNN and Transformer

The primary advantage of the developed approach lies in the integration of CNNs and Transformers within a single architecture. CNNs are employed to capture local feature representations, while Transformers are used to obtain global contextual representations. These two representations complement each other. CNNs possess a strong capability for recognizing local patterns in images, yet explicitly representing relationships between distant image regions can be challenging. Conversely, Transformers have the ability to learn global relationships through self-attention but rely heavily on the quality of input representations and may require relatively large amounts of data. Therefore, this research utilizes a CNN for the initial feature extraction stage and a Transformer for the global context learning stage. Through this approach, the model is expected to derive richer image representations than would be achieved by using either architecture in isolation.

f. Classification Head

The final component of the model is the Classification Head. This section is responsible for transforming the feature representations obtained from the Transformer into a classification decision. The final representation from the Transformer is passed to a fully connected layer. This layer maps the feature representations to the class space relevant to the research objective. Since the study employs binary classification, the model's final output consists of two classes: TB, indicating that the image is predicted to exhibit characteristics of Tuberculosis; and Non-TB, indicating that the image is predicted not to belong to the Tuberculosis group. The Classification Head generates a probability value for each class, and the class with the highest probability serves as the model's final prediction. Thus, the classification process is not performed directly based on image pixel characteristics, but rather based on feature representations that have undergone two stages of learning: local feature extraction by the CNN and global context learning by the Transformer.

2.2.4. Evaluation

The fourth stage is model evaluation. Model performance is evaluated using several metrics, including accuracy, sensitivity, specificity, precision, F1-score, and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC). In addition to performance evaluation, an error analysis is conducted to identify the characteristics of misclassified cases. Subsequently, hyperparameter optimization is performed to obtain the optimal model configuration. The best model is determined based on its performance on the validation data, followed by a final validation using test data that was not utilized during the training and optimization processes. Evaluation is carried out using data not used during the training phase. A confusion matrix is used to determine True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). Finally, accuracy, sensitivity, specificity, precision, and F1-score values are calculated.

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

$$Sensitivity = \frac{TP}{TP + FN} \quad (2)$$

$$pecificity = \frac{TN}{TN + FP} \quad (3)$$

$$Precision \frac{TP}{TP + FP} \quad (5)$$

$$F1 - Score = 2\frac{Precision\ x\ Sensitivity}{Precision + Sensitivity} \quad (6)$$

3. Result

3.1. Model Training Results

Vol. 11, No. 04; 2026

ISSN: 2456-3676

The CNN–Transformer hybrid model was trained for 10 epochs using the training dataset and evaluated periodically using the validation dataset. Based on the training results, the model demonstrated a rapid increase in accuracy starting from the first epoch. In the first epoch, the model yielded a training loss of 0.3518, a training accuracy of 86.01%, and a validation accuracy of 91.67%. In the second epoch, the training loss decreased to 0.2555, while the training accuracy rose to 89.21% and the validation accuracy reached 97.22%. Subsequently, in the third epoch, the model showed even more significant improvement, achieving a training accuracy of 98.54% and a validation accuracy of 94.44%. Training performance steadily improved, reaching a training accuracy of 99.13% by the fourth epoch. In the fifth epoch, the model achieved its highest validation accuracy of 100%, with a training accuracy of 96.21%. A validation accuracy of 100% was achieved again in the seventh epoch, alongside a training accuracy of 99.13%. In the subsequent epoch, validation accuracy dropped to 94.44%, although training accuracy remained high. By the tenth epoch, the model yielded a training loss of 0.0495, a training accuracy of 99.42%, and a validation accuracy of 94.44%.

Overall, the results indicate that the model is capable of effectively learning patterns from the training data. However, the discrepancy between the training accuracy, which approaches 100%, and the validation accuracy, which fluctuates between 91.67% and 100%, suggests that the model is beginning to overfit, consequently, evaluation using truly independent data is essential.

3.2. Model Testing Results

The best-performing model was then used to test 174 samples. The confusion matrix resulting from the test is shown in Figure 4.

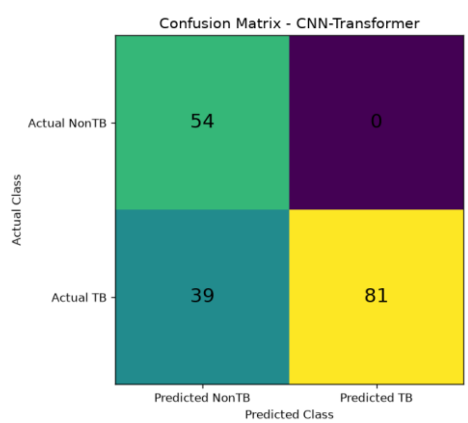

Figure 4 Confusion Matrix

The confusion matrix indicates that 54 Non-TB cases were correctly classified. However, out of 120 TB cases, 81 were successfully identified as TB, while 39 were misclassified as Non-TB. This demonstrates that the model excels at avoiding false positives but still faces challenges in detecting all TB cases.

ISSN: 2456-3676

3.2. Accuracy

Based on the confusion matrix, the model accuracy was obtained as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$= \frac{81+54}{81+54+0+39}$$

= 0.7759

Accuracy = 77,59%

This value indicates that the model is able to correctly classify 135 out of 174 samples.

3.3. Sensitivity

Sensitivity is a crucial metric in tuberculosis detection research, as it indicates the model's ability to identify TB cases.

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

$$= \frac{81}{81 + 39}$$

= 0.675

Sensitivity = 67,50%

These results indicate that the model successfully detected 81 out of 120 TB cases, while 39 cases went undetected. From the perspective of TB screening applications, this is an aspect requiring attention, as false negatives can result in TB cases being missed by the system.

3.4. Specificity

$$\text{Specificity} = \frac{TN}{TN+FP}$$

$$= \frac{54}{54+0}$$

= 1.0

Specificity = 100%

The results indicate that all 54 non-TB cases were correctly identified, and no non-TB cases were classified as TB. In other words, the model produced no false positives on the test dataset.

3.5. Precision

$$\text{Precision} = \frac{TP}{TP+FP}$$

ISSN: 2456-3676

$= \frac{81}{81+0}$

$= 1.0$

Precision = 100%

A precision value of 100% indicates that all images predicted as TB by the model indeed belong to the TB group in the test dataset.

3.6. F1-Score

F1-score = 80,60%

This value represents a balance between precision and recall (sensitivity). Although the model's precision reached 100%, the F1-score did not reach 100% because sensitivity was only 67.50%. This indicates that the model's primary issue lies not in incorrectly labeling non-TB cases as TB, but rather in failing to detect some TB cases.

3.7. AUC-ROC

The evaluation results show:
AUC-ROC = 1,0000
or:
AUC = 100%

The resulting ROC curve shows an AUC value of 1.0000, which theoretically indicates perfect discrimination capability based on the predicted probability rankings generated by the model.
The proposed CNN–Transformer hybrid model was trained for 10 epochs to classify chest CT images into TB and Non-TB categories. The training results showed a rapid improvement in model performance, with training accuracy increasing from 86.01% in the first epoch to 99.42% in the tenth epoch. The highest validation accuracy of 100% was obtained at epochs 5 and 7, while the validation accuracy at the final epoch was 94.44%. The decreasing training loss and increasing training accuracy indicate that the model was able to learn discriminative representations from the training data. However, the fluctuation between training and validation performance suggests the possibility of overfitting, which requires further evaluation using independent validation strategies.

On the test dataset consisting of 174 samples, the model achieved an accuracy of 77.59%. The confusion matrix showed 54 true negatives, 81 true positives, 39 false negatives, and no false positives. The model therefore achieved a sensitivity of 67.50% and a specificity of 100.00%. The sensitivity indicates that 81 of 120 TB cases were correctly identified, while 39 TB cases were incorrectly classified as Non-TB. In contrast, all 54 Non-TB cases were correctly classified, resulting in a specificity of 100.00%. The precision for the TB class reached 100.00%, indicating that all samples predicted as TB were actually TB cases in the test dataset. The resulting F1-score for TB was 80.60%.

Vol. 11, No. 04; 2026

ISSN: 2456-3676

The ROC analysis produced an AUC of 1.0000, indicating excellent discriminatory ability based on the predicted probability scores. However, the difference between the perfect AUC and the sensitivity of 67.50% at the operating classification threshold indicates that the probability threshold used for final classification may not yet be optimal for TB screening. Therefore, threshold optimization should be investigated in subsequent experiments to improve sensitivity while maintaining an acceptable level of specificity.

4. Discussion

This study develops a CNN–Transformer hybrid model based on ResNet18 and a Transformer Encoder to detect tuberculosis (TB) in chest CT images. Test results indicate that the model achieved an accuracy of 77.59%, sensitivity of 67.50%, specificity of 100%, precision of 100%, and an F1-score of 80.60%. Furthermore, ROC analysis yielded an AUC value of 1.0000. These results demonstrate the model's excellent capability in distinguishing between TB and non-TB groups based on the generated probability scores, although its performance at the applied classification threshold still reveals some undetected TB cases.

The model's ability to achieve 100% specificity indicates that all 54 non-TB cases in the test dataset were correctly identified. There were no false positives; consequently, every image predicted as TB in the test dataset was indeed a TB case. This characteristic is also reflected in the precision value of 100%. From a decision support system perspective, these results demonstrate the model's excellent ability to maintain accuracy when making TB predictions.
However, the model's sensitivity of 67.50% indicates that 39 out of 120 TB cases were still classified as Non-TB. Thus, the model's primary weakness in this experiment lies not in a tendency to over-diagnose TB, but in its ability to detect all TB cases. This is a crucial aspect to consider, as a system used for TB screening ideally requires high sensitivity to minimize the number of missed positive cases. These findings align with previous research indicating that while the development of AI systems for TB detection in medical imaging can yield high diagnostic performance, this performance may differ substantially between development studies and clinical evaluations. A systematic review and meta-analysis of 61 studies on AI for TB diagnosis reported pooled sensitivity and specificity of 91% and 65%, respectively, in clinical evaluations, whereas model development studies yielded a sensitivity of 94% and a specificity of 95%. This discrepancy demonstrates that model performance on development datasets does not necessarily reflect performance in external populations. In the context of CT imaging, previous studies have also demonstrated the potential of deep learning to detect and characterize TB. The MAResNet model 3D network-based architecture featuring a multi-scale attention mechanism developed for CT images, reportedly achieves an accuracy of approximately 94% in TB classification. Other studies employing 3D architectures such as VGG-16, EfficientNet, and ResNet-50 for CT analysis of active and inactive TB demonstrate that deep learning can effectively capture complex radiological characteristics from CT scans. Thus, these findings reinforce the evidence that CT serves as a promising source of information for developing deep learning-based AI systems for TB, while the hybrid approach utilized in this study provides a framework for integrating local and global image characteristics.

Vol. 11, No. 04; 2026

ISSN: 2456-3676

This study has several limitations. First, the amount of test data used is relatively limited, comprising 174 samples, so the performance results need to be confirmed using a larger and more diverse dataset. Second, there is a class imbalance in the test dataset, consisting of 120 TB cases and 54 non-TB cases; this factor must be taken into account when interpreting accuracy and the weighted average. Third, the discrepancy between training accuracy and validation accuracy suggests potential overfitting. Fourth, an AUC of 1.0000 combined with a sensitivity of 67.50% indicates that the classification threshold requires further analysis. Fifth, the model's generalizability to data from other institutions cannot be determined based solely on internal testing. Future research will focus on patient-level cross-validation, threshold optimization, external validation using data from different institutions, and transitioning CT representation from 2D slices to 3D CT volumes. Furthermore, explainable AI methods, such as Grad-CAM or attention visualization can be employed to ensure that the model makes decisions based on clinically relevant lung regions. Such interpretability approaches have been utilized in previous studies on CT-based TB analysis to visualize the lesion locations contributing to the model's predictions.

Conclusion

This study developed a CNN–Transformer Hybrid model based on ResNet18 and Transformer Encoder for tuberculosis detection from chest CT images. The proposed model achieved an accuracy of 77.59%, sensitivity of 67.50%, specificity of 100%, precision of 100%, and F1-score of 80.60% on the test set. The model also achieved an AUC-ROC of 1.0000, indicating strong discriminatory capability based on prediction scores. However, the presence of 39 false-negative cases indicates that sensitivity remains the main limitation of the proposed approach. Therefore, further optimization of the classification threshold, patient-level validation, external validation, and three-dimensional CT representation are required to improve model robustness and generalizability. Overall, the findings demonstrate the potential of CNN–Transformer hybrid learning as a computer-aided approach for tuberculosis detection from chest CT images.

Acknowledgment

We would like to thank Universitas Duta Bangsa Surakarta for funding our research through an Applied Research Grant in 2026.

References

Agus, I. P., Darma, E., Karang, I. M., Prawira, S., & Arya, I. G. B. (2025). *Comparison of Artificial Intelligence Methods for Tuberculosis Detection Using X-Ray Images*. *19*(1), 49–60.

Capell, D., Parida, A., Juan, J. G., Sanchez-jacob, R., Roshanitabrizi, P., Linguraru, M. G., Ledesma-carbayo, J., Anwar, S. M., Technologies, B. I., & Telecomunicaci, E. (n.d.). *3 , 2 2,4*.

Garud, P., Patel, G., Kadlag, N., Bhandari, S., & Patil, R. Y. (2023). *Classification of Chest X-ray Images using CNN for Medical Decision Support System*. *April*.

Ge, F., Zhang, Y., Xu, J., Muhammad, A., Song, J., & Yu, D. J. (2022). Prediction of disease-associated nsSNPs by integrating multi-scale ResNet models with deep feature fusion. *Briefings in Bioinformatics*, *23*(1). https://doi.org/10.1093/bib/bbab530

Hajiheydari, N., Soltani, M., & Saheb, T. (2025). Social Science & Medicine AI in medical diagnosis : A contextualised study of patient motivations and concerns. *Social Science & Medicine*, *371*(May 2024), 117850. https://doi.org/10.1016/j.socscimed.2025.117850

Hampiholi, N., & Engineer, D. (2023). *MEDICAL IMAGING ENHANCEMENT WITH AI MODELS FOR AUTOMATIC DISEASE DETECTION AND CLASSIFICATION BASED ON*. *8*(05), 31–37.

Han, Z., Zhang, Y., Li, J., Gao, S., Liu, W., & Yang, W. (2025). *A systematic review and meta-analysis of artificial intelligence software for tuberculosis diagnosis using chest X-ray imaging*. *17*(5), 3223–3237. https://doi.org/10.21037/jtd-2025-604

Hansun, S., Argha, A., Liaw, S., Celler, B. G., & Marks, G. B. (2023). *Machine and Deep Learning for Tuberculosis Detection on Chest X-Rays : Systematic Literature Review Corresponding Author : 25*, 1–22. https://doi.org/10.2196/43154

Heroine, N., Murni, N. S., Suryani, L., Utami, D., Ariani, S., Tinggi, S., Kesehatan, I., & Husada, B. (2025). *Kejadian Tuberkulosis Paru Pada Masyarakat di Wilayah Kerja Puskesmas Megang Sakti Kabupaten Musi Rawas*. *5*(2), 13809–13818.

Identification of Tuberculosis : A Review of the Guidelines. (n.d.). 1–54.

Imron, M., Widyawati, M. N., & Masrochah, S. (2020). *I nternational J ournal of A llied M edical S ciences and C linical R esearch (IJAMSCR)*. *8*(3).

Khosravi, M., Zare, Z., Mojtabaeian, S. M., & Izadi, R. (2024). *Arti fi cial Intelligence and Decision-Making in Healthcare : A Thematic Analysis of a Systematic Review of Reviews*. https://doi.org/10.1177/23333928241234863

Lee, S., Rajaguru, V., Baek, J. S., & Shin, J. (2023). *Digital Health Interventions to Enhance Tuberculosis Treatment Adherence : Scoping Review*. *11*, 1–12. https://doi.org/10.2196/49741

Lumamba, K. D., Wells, G., Naicker, D., Naidoo, T., Steyn, A. J. C., & Gwetu, M. (2024). Computer vision applications for the detection or analysis of tuberculosis using digitised human lung tissue images - a systematic review. *BMC Medical Imaging*. https://doi.org/10.1186/s12880-024-01443-w

Patel, N., & Wong, A. (2022). *E MPOWERING T UBERCULOSIS S CREENING WITH E XPLAINABLE S ELF -S UPERVISED D EEP N EURAL N ETWORKS*. 1–9.

Pinto-coelho, L. (2023). *How Artificial Intelligence Is Shaping Medical Imaging Technology : A Survey of Innovations and Applications*.

Rahman, K. K. M., Zulaikha, S., Dhafer, B., & Ahmed, R. (2025). Intelligence-Based Medicine Advancing tuberculosis screening : A tailored CNN approach for accurate chest X-ray analysis and practical clinical integration. *Intelligence-Based Medicine*, *11*(September 2024), 100196. https://doi.org/10.1016/j.ibmed.2024.100196

Ramadhan, M. D., & Wulandari, L. (2023). *Profile of Pulmonary Tuberculosis Patients in Dr . Soetomo General Academic Hospital*. *29*(0672), 272–276.

ISSN: 2456-3676

Santamato, V., Tricase, C., Faccilongo, N., & Iacoviello, M. (2024). *Exploring the Impact of Artificial Intelligence on Healthcare Management : A Combined Systematic Review and Machine-Learning Approach*.

Sastramandala, M. B., Sari, R. N., Informatika, P., Industri, F. T., & Gunadarma, U. (2025). *Pemanfaatan Deep Learning menggunakan Convolutional Neural Network Terhadap Penyakit Tuberkulosis melalui Citra Rontgen*. *14*, 631–642.

Shahreki, A. P., Hosseini-baharanchi, F. S., & Roudbari, M. (2024). Diagnosis of Tuberculosis Using Medical Images by Convolutional Neural Networks. *Kerman University of Medical Sciences*, *31*(4), 180–186. https://doi.org/10.34172/jkmu.2024.29

Shrivas, A., & Singh, S. (2025). *Tuberculosis Diagnosis and Management : Recent Advances*. *June 2024*, 3–9. https://doi.org/10.4103/jgid.jgid

Vazquez, B., Rojas-garcía, M., Rodríguez-esquivel, J. I., Marquez-acosta, J., Aranda-flores, C. E., Cetina-pérez, L. C., Soto-lópez, S., Estévez-garcía, J. A., Bahena-román, M., Madrid-marina, V., & Torres-poveda, K. (2025). *Machine and Deep Learning for the Diagnosis , Prognosis , and Treatment of Cervical Cancer : A Scoping Review*. *Cc*, 1–42.

Widodo, S., Rohmah, R. N., Handaga, B., & Arini, L. D. D. (2019). Lung diseases detection caused by smoking using support vector machine. *Telkomnika (Telecommunication Computing Electronics and Control)*, *17*(3), 1256–1266. https://doi.org/10.12928/TELKOMNIKA.V17I3.9799

Yulvina, R., Putra, S. A., Rizkinia, M., Pujitresnani, A., Tenda, E. D., Yunus, R. E., Djumaryo, D. H., Yusuf, P. A., & Valindria, V. (2024). *Hybrid Vision Transformer and Convolutional Neural Network for Multi-Class and Multi-Label Classification of Tuberculosis Anomalies on Chest X-Ray*. 1–29.