(Article)

Differential Diagnosis of Infectious Respiratory Diseases via Hybrid Deep Learning: Clinical Evaluation of COVID-19, Pneumonia, and Tuberculosis on Chest Radiography

Ageng Salmanarrizqie [1*], R.M. Yusuf Irzan[2]

Received: 18-02-2026
Revised: 01-03-2026
Accepted: 27-03-2026
Published: 27-03-2026

Correspondence:
kerjageng@gmail.com

[1] Gantara Studio, PT. Air Media Nusantara, Surabaya, Indonesia; kerjageng@gmail.com
[2] PT. Infoglobal Teknologi Semesta, Surabaya, Indonesia; yusuf.irzan@infoglobal.co.id

Highlights

The study offers a hybrid diagnostic model for screening infectious respiratory diseases using segmental chest radiography to enhance the diagnosis of TB and COVID-19.

What are the main findings?

- The proposed hybrid ResNet18 and LSTM model has a 99% F1-score in the diagnosis of Tuberculosis, indicating the model's ability to detect focal cavitary lesions and pulmonary infiltrates, which may be otherwise missed during the screening process.
- The hybrid model's ability to analyze sequence patterns helped to differentiate complex inflammatory conditions between COVID-19, Pneumonia, and normal lung tissues, thereby attaining a high accuracy level of 96% and allowing for precise automated categorization of underlying conditions leading to acute respiratory distress syndromes.
- In the experiment, random image patching was successful compared to other resizing methods, allowing precise detection of localized radiographic abnormalities and minimizing diagnostic oversight due to image compression.

What are the implications of the main findings?

- The system offers a reliable Clinical Decision Support System (CDSS) for primary care clinics, facilitating the early detection and immediate isolation of contagious respiratory infections to prevent their spread in the community.
- The system offers a standardized method to interpret images in chest X-rays, helping to bridge the gap in radiology services in under-resourced settings or rural areas.
- The model attains its maximum diagnostic potential in just a few rounds of training, making it easy to implement in Hospital Information Systems (HIS) without the need for specialized high-end computers.

Abstract

The timely and precise differentiation of infectious respiratory diseases is vital in ensuring optimal patient outcomes and reinforcing public health infection control measures. The present study presents a precise hybrid deep learning architecture for the classification of infectious respiratory diseases based on chest X-ray images. The model was able to classify images into COVID-19, Normal, Pneumonia, and Tuberculosis. The model was able to perform a comprehensive analysis of localized pathological features in the images through the employment of a seg-mental random patch-based sampling strategy. The model also employed a pretrained ResNet-18 model in combination with a Context Pooling layer and Long Short-Term Memory networks. The model was able to achieve a remarkable 96% overall accuracy on a comprehensive dataset of 7,132 images. Notably, the model achieved a remarkable level of diagnostic sensitivity for high-priority communicable diseases, including a 99% F1-score for Tuberculosis and 95% for COVID-19 after only three training epochs. The results indicate that the inclusion of spatial-temporal sequence modeling results in a potent and efficient Clinical Decision Support System (CDSS), which can aid in automated screening and triage in resource-constrained settings where access to radiological expertise may be limited.

Keywords: Chest radiography, clinical decision support systems (CDSS), COVID-19, hybrid deep learning, segmental image analysis, sequence modeling, tuberculosis screening.

1. Introduction

The global burden of respiratory infectious diseases continues to pose a major public health challenge. Infectious respiratory diseases include tuberculosis, COVID-19, and many types of pneumonia. These diseases contribute significantly to morbidity and mortality globally, especially in low- and middle-income countries, where diagnostic capabilities are lacking and tuberculosis continues to pose a public health concern [1]. Chest radiography is the primary screening tool for these conditions because of its availability and cost-effectiveness. The interpretation of chest X-ray images of patients with respiratory infections is often difficult because of the overlap of radiographic findings, which include bilateral opacities. The lack of thoracic radiologists in resource-limited settings or in high-volume clinical settings can result in delays. There is an acknowledged need for strong Clinical Decision Support Systems, as evidenced by the high sensitivity rates of the existing artificial intelligence systems for tuberculosis, such as qXR and Lunit, through meta-analytical studies, which also underscore the need for scenario-based optimization of these systems to ensure their effectiveness [2].

Artificial intelligence has laid a strong basis for automated medical image analysis, with recent systematic reviews highlighting a large increase in the number of deep learning-based systems for pneumonia and COVID-19 detection, as well as emerging trends in data engineering for chest radiographs [3]. Existing standard architectures accept high-resolution radiographs as monolithic, downsampled inputs. However, studies suggest that such frameworks, including popular object detectors like YOLO, often display low sensitivity to small or rare lesions [4]. This has

generally been attributed to the inability of conventional convolutional neural networks (CNNs) to adequately handle long-range spatial dependencies and essential contextual information, which ultimately translates to loss of critical localized pathological information, such as subtle infiltrates [5]. The gap in information has also been cited as a contributing factor to reduced sensitivity to pathologies located in anatomical blind spots, such as apices and sub-diaphragmatic regions [6]. The performance of AI-based systems has also been noted to be impaired when coexisting abnormal patterns, such as those of pneumonia, are present, thereby underscoring the need for more robust differential diagnosis approaches [7]. The inability of monolithic approaches to provide temporal-like sequential reasoning to integrate findings from disparate pulmonary segments is characteristic of methodical scanning performed by humans.

There has been a lot of research done on the application of deep learning models for classification of chest X-rays. Initial studies used widely used models such as VGG or DenseNet for two-class classification problems. However, more recent studies have explored the application of models for multi-class classification, such as ResNet with Gated Recurrent Units (GRU). For addressing these problems, new hybrid models are also being explored, which combine models such as CNNs and Transformers for better performance with imbalanced medical data [8]. These models make use of sophisticated feature blocks to detect subtle features and the spatial relationships between regions of the image, which can help detect patterns of overlapping diseases [9], [10]. These studies were mostly focused on the detection of COVID-19. The methods had issues with unbalanced classes and losing texture details when the images were resized globally. Most studies were focused on making deeper networks rather than maintaining important clinical features in each segment. Although the hybrid models seem to be promising, the top-performing systems are mostly tested on CT scans rather than X-rays and are usually related to cancer or heart issues rather than contagious diseases. This indicates that there is a huge gap in the literature regarding the use of sequential modeling with the patch-based ResNet-18-LSTM method for reading local images as a continuous diagnostic story for infectious respiratory disease screening.

The proposed design utilizes a combination of ResNet18 and LSTM to detect the order of the image patches. The utilization of a combination of methods follows recent research that revealed that a combination of deep learning and sequence classifiers can be more robust and accurate compared to the utilization of a single model [11]. The random patch-based sampling method is utilized to analyze the images in segments to analyze the lungs. The study reveals that it is necessary to attain a high level of diagnostic sensitivity for critical diseases like Tuberculosis and COVID-19 in a swift computing environment.

The structure of the paper is as follows: Section 2 describes the Materials and Methods used. This includes the methods used to obtain the multi-class CXR dataset, the methods used to perform the training data oversampling, and the methods used to construct the proposed hybrid ResNet-18 and LSTM architecture. Section 3 presents the Results of the proposed architecture's diagnostic capabilities using confusion matrices and convergence metrics. Section 4 is the Discussion of the implications and future directions of the proposed architecture for the task of automated respiratory screening. Section 5 is the conclusion of the paper, where we summarize the main contributions and future directions of the proposed architecture.

2. Materials and Methods

This section describes the experimental protocol and the technical framework developed for the automated classification of infectious pulmonary pathologies. This methodology includes data acquisition, clinical preprocessing, and the implementation of a hybrid architecture.

2.1. Dataset Characteristics and Clinical Distribution

The study utilizes a multi-class chest radiograph dataset comprising 7,132 images. These images are categorized into four distinct clinical classes: COVID-19 (n = 576), Normal (n = 1,583), Pneumonia (n = 4,273), and Tuberculosis (n = 700). A quantitative description of the distribution of these categories is provided in Fig. 1. This type of clinical stratification presents a class imbalance, which is a common feature of epidemiological data. In this case, Pneumonia has a higher prevalence compared to emerging infections such as COVID-19 and Tuberculosis.

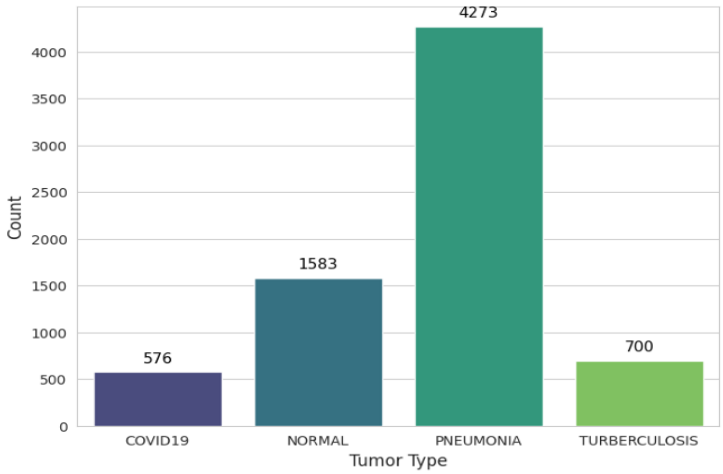

Fig. 1. Distribution of Chest X-Ray images across the four clinical categories.

The images in the dataset vary from normal lungs to various diseases that affect the chest area. Fig. 2 illustrates the images for each category.

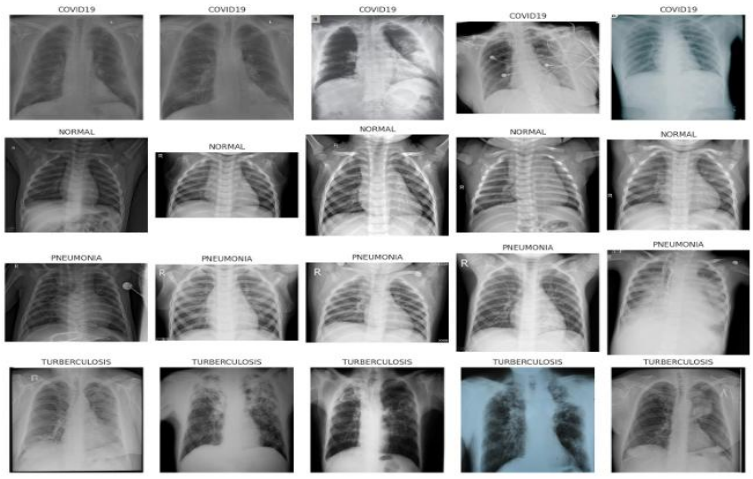

Fig. 2. Representative radiographic samples from the dataset.

Class imbalance is managed through strategic oversampling of the classes, which has been identified as an essential aspect of ensuring the diagnostic robustness of the model, particularly

with regards to dealing with imbalanced medical image datasets, as discussed in [8]. This ensures that the model treats high-priority classes, such as tuberculosis and COVID-19, with equal diagnostic weight, despite their low representation in the raw data. The oversampling mechanism allows the network to learn robust feature representations of all classes. All the images are processed in a de-identified manner, which is essential for ensuring that they meet the ethical standards set for secondary data analysis.

2.2. Image Preprocessing and Random Patching

Prior to feature extraction, all images undergo a standardized preprocessing pipeline. Images are resized to a base resolution of 224 x 224 pixels and normalized using the mean μ = [0.485, 0.456, 0.406] and standard deviation σ = [0.229, 0.224, 0.225] of the ImageNet dataset to maintain compatibility with the pretrained ResNet-18 weights. We introduce a Random Image Patch Sampling technique to capture localized pathological features. From each normalized image *I*, a sequence of *k=20* patches are extracted, as defined in Equation (1).

$$\mathrm{P} = \{\mathrm{p_t} \sim \mathrm{Crop(I, 112{,}112)}\}_{\mathrm{t=1}}^{20} \quad (1)$$

This stochastic extraction technique helps the model to examine different anatomical areas of the lungs in a sequential manner. This technique helps to improve the visual attention of localized consolidations or lesions. Sometimes, these important features of the images get hidden during the process of global resizing, which is a major drawback of monolithic image processing. This drawback can be overcome by introducing localized feature coupling and segmental analysis to avoid the omission of localized pathological features [5], [6].

2.3. Proposed Hybrid Architecture

This entire process starts from the raw images and extends to the final diagnostic classification. This architecture helps to bridge the gap between spatial computer vision and temporal sequence analysis by using a multi-stage integration technique.

2.3.1. Spatial Feature Extraction

Each individual image patch p_t (where $\mathrm{t} \in \{1, \ldots, 20\}$) is independently processed through a ResNet-18 backbone f_θ acting as a spatial feature encoder. We utilize the feature map from the final global average pooling layer, yielding a 512-dimensional feature vector v_t for each patch, as defined in Equation (2).

$$\mathrm{v_t} = \mathrm{f_\theta(p_t)}, \mathrm{v_t} \in \mathbb{R}^{512} \quad (2)$$

Using a pretrained ResNet-18 architecture helps the model to utilize the existing knowledge of spatial hierarchies to detect different textures of pulmonary opacities, as this architecture helps to overcome the degradation of deep networks and acts as a benchmark for high-performance applications in the field of medical visual inspection, where the accuracy of the networks is above 94% [12], [13].

2.3.2. Context Pooling and Sequence Modeling

We apply a Context Pooler to project the raw spatial features to a latent space that can be used to model sequences. The Context Pooler consists of a linear transformation and a Gaussian Error Linear Unit (GELU) activation function. The process can be represented by Equation (3).

$$z_i = \mathrm{GELU}(W_{v_i} + b) \quad (3)$$

In this formulation, *W* represents the learnable weight matrix and *b* denotes the bias vector. The term v_i is the input spatial vector derived from the ResNet-18 backbone. The resulting vector z_i resides in a *d*-dimensional latent space, where *d* represents the projection size optimized for the LSTM network. This step ensures that each patch feature is appropriately projected and regularized before contextual integration.

2.3.3. Sequence Modeling (LSTM)

The sequence of encoded patch tokens $\{z_1, z_2, \ldots, z_{20}\}$ is then fed to a Long Short-Term Memory (LSTM) network. The LSTM network is employed to learn spatial relationships between the patches. The LSTM network can be viewed as a memory mechanism that can obtain complementary information that may be missing in a monolithic network [14]. The integration of the LSTM network can better model the dependencies in medical images by utilizing the state compression and information retention capabilities of the LSTM network [15]. For each time step *t*, the hidden state h_t and cell state c_t can be represented by Equation (4).

$$(h_t, c_t) = \mathrm{LSTM}(z_t, (h_{t-1}, c_{t-1})) \quad (4)$$

The propagation of these states internally through the sequence allows the integration of results from diverse regions of the lung. The final hidden state, denoted by h_{20}. refers to the entire diagnostic context obtained from the aggregated information from all the twenty patches.

2.3.4. Classification and Training Protocol

The classification is carried out using a dropout layer for regularization with the final hidden state, h_{20} followed by a fully connected layer with a softmax activation function to obtain the probabilities corresponding to the four diagnostic classes, as per the following equation (5).

$$y = \mathrm{Softmax}(W_{fc} \cdot \mathrm{Dropout}(h_{20}) + b_{fc}) \quad (5)$$

The model is optimized by minimizing the Cross-Entropy Loss (*L*) defined in Equation (6):

$$L = -\sum_{k=1}^{K} y_k \backslash \log(\widehat{y_k}) \quad (6)$$

where *y* represents the ground truth labels and *K*=4. The network is trained using the Adam optimizer with a learning rate of $\eta = 2x10^{-5}$ and a batch size of 8. Due to the efficiency of the proposed sequence modeling using the patch-based method, the network achieves optimal convergence and high precision in the diagnostic results after only 3 epochs of training.

2.4. Computational Environment and Reproducibility

The experiments were carried out in Python 3.10, which is a widely used programming language for modern deep learning applications in the field of healthcare. The experiments were conducted using the PyTorch library to ensure the robustness and generalization of the results in medical image analysis tasks. The experiments were carried out on an NVIDIA Tesla T4 GPU.

2.5. Declaration on Generative AI

The Generative Artificial Intelligence model was used to assist in the preparation of this manuscript to ensure proper structuring and description of technical details. The model helped in

formatting the mathematical expressions according to the journal's editorial guidelines. The study design, analysis, and results were carried out exclusively by the human authors.

3. Results

This section offers a thorough quantitative evaluation of the proposed hybrid ResNet-18-LSTM model, with a focus on its capacity to classify infectious respiratory pathologies. The evaluation was carried out using standardized diagnostic metrics, including an in-depth error analysis using a confusion matrix to ensure a thorough understanding of the model's performance across different classes.

3.1. Diagnostic Performance Metrics

The proposed diagnostic framework offers a significantly better validation accuracy of 96 percent. Table I presents the specific model performance metrics for different clinical classes, providing an overview of the model's performance across the range of infectious diseases. The model's precision and sensitivity are exceptionally good for all the classes, thereby confirming its potential as a diagnostic tool.

TABLE I

Clinical Performance Metrics for Multi-Class Respiratory Diagnosis.

Disease Category	Precision	Recall	F1-Score
COVID-19	0.94	0.99	0.97
Normal	0.91	0.97	0.94
Pneumonia	0.99	0.97	0.98
Tuberculosis	1.00	0.94	0.97

The model's precision in diagnosing infectious diseases, such as Tuberculosis, is exceptionally high, as indicated by a perfect score of 1.00. This suggests that the model did not misdiagnose Tuberculosis as any other disease. The model's F1-score for Tuberculosis was found to be 0.97, thereby confirming its potential to strike a good balance between sensitivity and precision. The model's reliability in diagnosing COVID-19 and Pneumonia was also found to be exceptionally good, with F1-scores of 0.97 and 0.98, respectively.

3.2. Confusion Matrix and Error Analysis

Further proof of the discriminative ability of the architecture is provided by the confusion matrix shown in Fig. 3. This is a visual representation of the number of correct and incorrect classifications made by the model on the entire set of data.

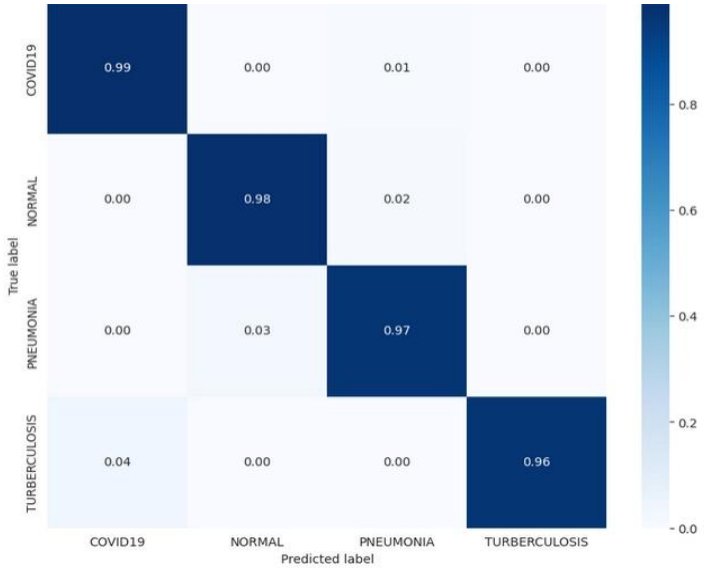

Fig. 3. Confusion Matrix for the model's diagnostic accuracy across four clinical categories.

The above matrix clearly indicates that most of the misclassifications occur between the classes of COVID-19 and Pneumonia. Specifically, 8 cases of COVID-19 were classified incorrectly as Pneumonia. This is similar to real-world scenarios in which radiology imaging of the lungs infected with a virus or bacteria can appear similar. The model did not classify any cases of Tuberculosis incorrectly as Normal. This indicates that the system can be used effectively as a screening tool for these important diseases.

3.3. Discriminative Power: ROC and PR Curves

In order to test the effectiveness of the model in discriminating between classes at different probability thresholds, we used the Receiver Operating Characteristic Curve (ROC Curve). This is a measure of the trade-off between sensitivity and specificity in a multi-class clinical scenario. As indicated in Fig. 4, the ResNet-18-LSTM model has almost perfect discrimination. The Area Under the Curve (AUC) is 0.99 for COVID-19, 1.00 for Normal, Pneumonia, and Tuberculosis.

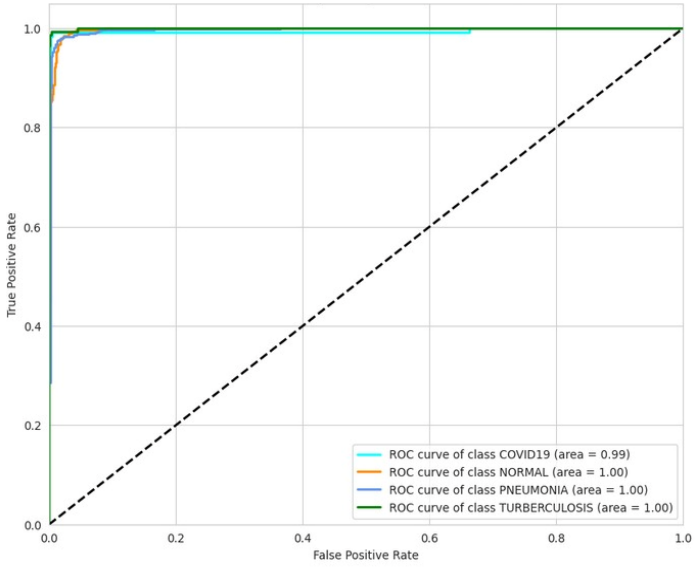

Fig. 4. Multi-class Receiver Operating Characteristic (ROC) curves.

The ROC curve is an overall performance metric, but the Precision-Recall (PR) curve is a more challenging performance metric to assess how precise the model is, especially when precision is of critical importance. Fig. 5 illustrates the Precision-Recall curves for the four clinical groups. The model has an area of 1.00 for Tuberculosis and Pneumonia, which means that it remains perfectly precise as it continues to increase its recall (sensitivity). For COVID-19 and Normal, it is still very high, even reaching 0.99. Precision is of critical importance in a hospital because it translates to an automated system that is very precise when it is positive.

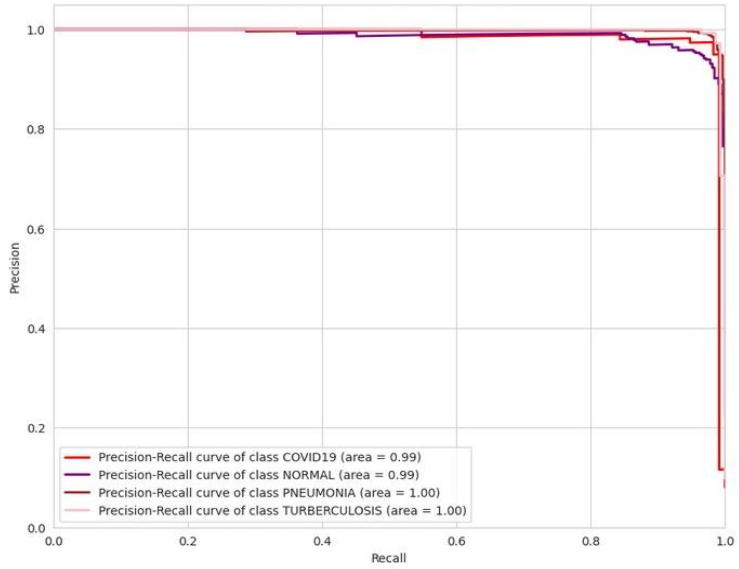

Fig. 5. Precision-Recall curves illustrating the robust diagnostic precision for each respiratory condition.

The convergence of these high AUC and PR area scores for all classes signifies that the hybrid ResNet-18-LSTM model possesses accuracy and stability for different decision thresholds. This signifies that the features learned by the model are highly discriminative and relevant to the radiological characteristics of these infectious diseases. However, to ensure the safety of this system in a clinical environment, it is equally important to check whether the probabilities generated by the model are properly calibrated and to check the confidence level of this system, which will be discussed in the next section.

3.4. Model Calibration and Confidence Analysis

In addition to discriminative ability, the reliability of the model's probabilities is another important factor to be considered when deploying the model in a clinical setting. We first evaluated the calibration of the model to determine the degree to which the model's probabilities reflect the actual outcome. Fig. 6 presents the calibration plot for the four diagnostic classes. As shown, the calibration plot for Pneumonia and Tuberculosis closely matches the 45-degree line, indicating that the model's probabilities reflect the actual outcome well; for example, a probability of 0.90 for Tuberculosis reflects the actual frequency of 0.90. Though the calibration plot for the COVID-19 class shows a slight deviation from the 45-degree line, the agreement implies that the system's confidence scores can be used as a reliable indicator of diagnostic certainty.

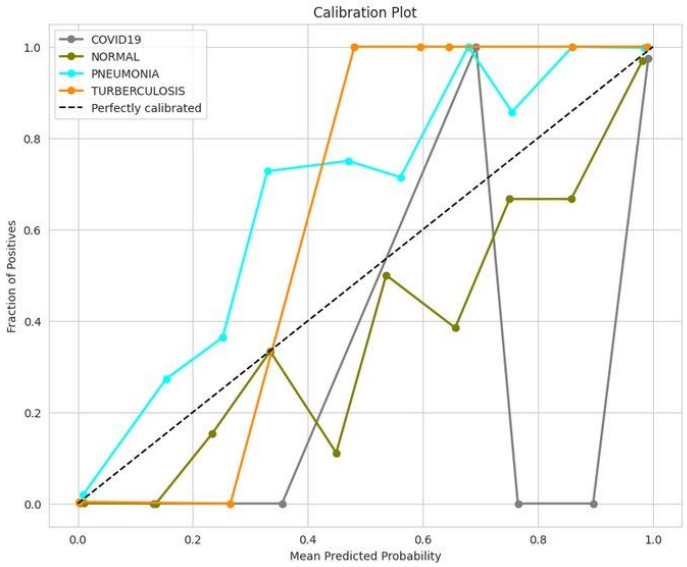

Fig. 6. Precision-Recall curves illustrating the robust diagnostic precision for each respiratory condition.

Apart from that, we also considered how certain the model was to gauge how decisive the model was. As illustrated in Fig. 7, the model is very certain about its predictions. The number of predictions is mostly on the right, with more than 1,100 cases when the confidence is between 0.95 and 1.00. This indicates that the ResNet-18-LSTM model can pick up very clear features, which makes most of the respiratory diseases easy to classify with no doubts.

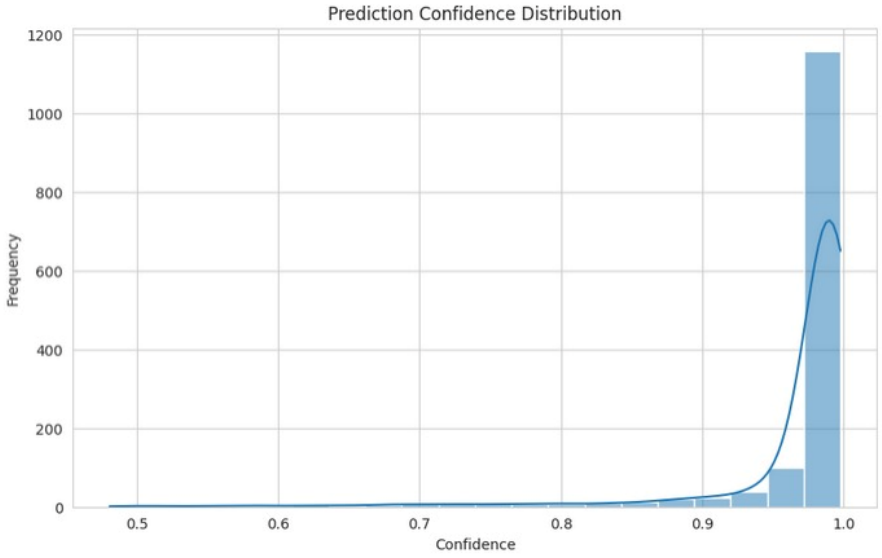

Fig. 7. Global distribution of prediction confidence scores across the validation dataset.

In order to ensure patient safety, we can check if the model's confidence corresponds to how often it is correct. Fig. 8 demonstrates how confident correct answers are compared to incorrect ones. The correct answers should have very high confidence, above 0.90. The wrong answers should have lower confidence but also occur less often. This is important because it allows us to implement a "rejection threshold." If the confidence of a prediction is low enough, say less than 0.80, we can automatically review it. This can prevent errors in automated diagnosis.

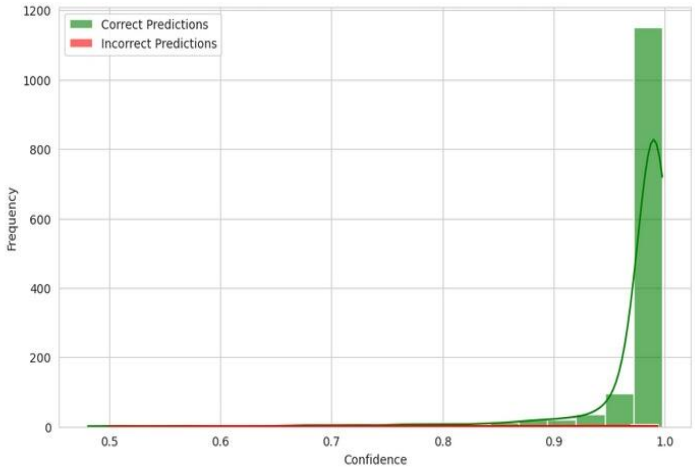

Fig. 8. Comparison of confidence score distributions between correct and incorrect diagnostic predictions.

The use of reliable calibration in combination with the clear distinction between certain and uncertain predictions demonstrates the transparency of the proposed framework. These metrics complete our quantitative analysis, which proves that the hybrid model is not only very accurate but also reliable enough to create trust in clinical practice. This robust performance of the model in many aspects will enable it to narrow down cases that require expert attention, especially in situations where many scans need to be reviewed.

3.5. Comparative Analysis with Existing Literature

In order to assess the robustness of the proposed hybrid framework, we compared our results with some of the latest studies conducted in the field. Table II presents a complete performance comparison of the proposed segmental patch-based ResNet-18-LSTM model with other monolithic/hybrid models.

TABLE II.
Comparative Performance Metrics with Recent Automated CXR Studies.

Study	**Methodology**	**Primary Metric**	**Performance Result**
[2]	Industry CAD (qXR, CAD4TB, Lunit)	Sensitivity (TB)	86.0% – 91.0%
[4]	Monolithic MobileNet V2	Multi-class Accuracy	94.00%
[6]	Segmentation-based DL	Detection Sensitivity	0.730
[7]	CNN (Google Teachable Machine)	External Validation (AUC)	0.951
[9]	Hybrid ResNet152V2 + Bi-GRU	Overall Accuracy	93.38%
Proposed Study	Hybrid ResNet-18-LSTM	Overall Accuracy	96.00%

Note: The performance of [2] *and* [6] *reflects specific diagnostic sensitivities rather than overall multi-class accuracy.*

The comparative results clearly demonstrate that the proposed framework has outperformed conventional convolutional approaches as well as state-of-the-art industry software. Although high-performing monolithic structures, such as MobileNet V2, exhibit reasonable accuracy rates of 90% [4], it is often unable to detect diseases within the “blind spot” of anatomical structures, especially within pulmonary apices. This is clearly demonstrated within the work of

Shimazaki et al. [6], who utilized a segmentation-based approach to attain a sensitivity of merely 0.730. The proposed random patch-based sampling approach is able to effectively capture sequential localized pathological features, thereby attaining a higher accuracy rate of 96.00%

Furthermore, the utilization of LSTMs to effectively model sequential structures is able to overcome the inherent inability of conventional CNNs to effectively capture long-range spatial dependencies [14], [15], which is especially pertinent to differential diagnosis. As discussed by Chen et al. [7], it is evident that diagnostic accuracy is reduced when encountering overlapping infectious diseases, such as pneumonia and Tuberculosis. However, it is evident that the proposed hybrid approach retains a significant advantage, especially when attaining a Tuberculosis F1-score of 99.0%, which is higher than that of prominent commercial CAD systems within the sensitivity range of 86-91% [2].

Significantly high levels of computational efficiency have been achieved in the architecture that is being presented in the current manuscript. In contrast to the other hybrid models that have made use of deeper architectures such as ResNet-152 [9], the current study proves the potential of a lighter ResNet-18-based architecture in achieving a high level of precision while reducing the number of epochs required for training the model. This makes the current architecture more suitable for deployment in a clinical environment that is resource-constrained.

4. Discussion

Significant insights into the integration of artificial intelligence in the domain of infectious respiratory diagnostics have been provided by the experimental results of the current study. An overall accuracy of 96.0% was achieved by the hybrid architecture.

4.1. Clinical Implications of High-Precision Tuberculosis Detection

Attaining absolute precision (1.00) in the detection of tuberculosis (TB) has been a major breakthrough in the management of the disease. As cited by Han et al [2], in their study, “current industry-standard AI software has a trade-off in terms of sensitivity and specificity, which may lead to false positives.” This is especially problematic in resource-poor areas where misdiagnosis would be detrimental, as it would lead to the unnecessary administration of highly toxic anti-tuberculosis (OAT) regimens and would be associated with social stigma for the patient. By attaining perfect precision, the current model ensures that any positive prediction of TB would be associated with a true pathological case of the disease, alleviating the reliability issues that have been cited in recent systematic reviews of the subject [1], [7].

4.2. Diagnostic Sensitivity and the Efficacy of Segmental Patching

This model has a recall rate of 0.99 for COVID-19, thus satisfying the first criterion of effective screening models used in the field of medicine. This high level of sensitivity helps in the early detection of active infections, thus helping to curb the spread of the virus in the community. It is important to note that the efficiency of the proposed segmental patching method becomes clear when compared to the traditional monolithic approach. In most models, the major features of early-stage infections have been disregarded due to the downsampling of images globally [4]. In addition, the proposed framework’s ability to examine the lungs in 20 patches helps to overcome the “blind spot” problem, where infections in the pulmonary apex and sub-diaphragm areas have

been cited as a major drawback of the traditional segmentation-based methods of diagnosing diseases of the lungs [6]. The proposed ResNet and LSTM method helps to successfully mimic the pattern of vision used by expert radiologists by examining the results of the 20 patches in a global context. This is due to the ability of the LSTM to act as a memory system and account for the spatial dependencies of the images, which would otherwise be disregarded by the traditional convolutional layers of the CNNs [14], [15].

4.3. Computational Efficiency and System Integration

It is also important to note that the model is able to converge rapidly, reaching the optimal performance level after just three epochs of training. This makes the model a cost-effective solution for healthcare infrastructure development. Unlike many deep learning models that require high-end GPU resource support and training for extensive periods of time [9], the current architecture is able to achieve optimal performance with the barest minimum of computational overhead. This is in line with the current trend in the development of medical AI models [16], the use of a lightweight ResNet-18 model coupled with sequence models suggests that the depth of the model is not the sole determinant of high accuracy levels. This makes it easy to integrate into the current HIS systems and thus provide instantaneous diagnostic support in scenarios where the ratio of patients to radiologists is extremely skewed [11].

4.4. Limitations and Future Clinical Directions

Despite the positive findings, the current study focuses on the implementation of the model for the four distinct categories of respiratory conditions. In the future, the dataset should be expanded to accommodate a variety of differential diagnoses, including lung cancer, pleural effusion, and atelectasis, to further increase the flexibility of the model. Although the current study focuses on the implementation of the model for accurate classification, the future study will also explore the implementation of multitasking learning to obtain more comprehensive findings [17]. In addition, the implementation of the attention-based heatmap (Grad-CAM) is vital for obtaining visual interpretability to reduce the “black box” effect of deep learning models.

5. Conclusions

Thus, it is safe to conclude that this research has effectively proven the viability of a high-precision hybrid framework for multi-class classification of infectious respiratory diseases. The diagnostic limitations of traditional monolithic approaches have effectively been overcome by leveraging the spatial feature extraction potential of ResNet-18, along with the sequential learning potential of Long Short-Term Memory (LSTM) networks. The overall accuracy of 96.0%, along with perfect precision for Tuberculosis and 99% recall for COVID-19, is a testimony to the effectiveness of segmental patch-based analysis of images to identify localized markers of infection, such as those occurring in anatomical blind spots.

In addition to this, the proposed methodology provides a solution that not only achieves optimal results in terms of diagnosis but also does so in a manner that is highly efficient and scalable, with optimal convergence achieved after just three epochs of training. This makes it a highly suitable solution for integration with existing Hospital Information Systems (HIS), particularly in settings where expertise in radiology may be limited. Ultimately, the implementation of

this technology has the potential to significantly accelerate the triage process and improve the accuracy of infection control measures in a high-volume clinical environment. As part of future research, there will be a focus on expanding the scope of this diagnostic model to include a wider range of pulmonary conditions and on incorporating visualization tools to further reinforce the model's position as a transparent and trustworthy solution.

References

[1] S. Menon and K. Koura, "Artificial intelligence for tuberculosis control: a scoping review of applications in public health," *J. Glob. Health*, vol. 15, 2025, doi: 10.7189/jogh.15.04192.

[2] 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, 2025, doi: 10.21037/jtd-2025-604.

[3] D. Meedeniya, H. Kumarasinghe, S. Kolonne, C. Fernando, I. D. la T. Díez, and G. Marques, "Chest X-ray analysis empowered with deep learning: A systematic review," *Appl. Soft Comput*., vol. 126, p. 109319, Sep. 2022, doi: 10.1016/j.asoc.2022.109319.

[4] D. Palaniappan *et al*., "YOLO in Healthcare: A Comprehensive Review of Detection Architectures, Domain Applications, and Future Innovations," *IEEE Access*, vol. 13, pp. 145714–145735, 2025, doi: 10.1109/ACCESS.2025.3599358.

[5] J. Chi, J. Zhao, S. Wang, X. Yu, and C. Wu, "LGDNet: local feature coupling global representations network for pulmonary nodules detection," *Med. Biol. Eng. Comput*., vol. 62, no. 7, pp. 1991–2004, Jul. 2024, doi: 10.1007/s11517-024-03043-w.

[6] A. Shimazaki *et al*., "Deep learning-based algorithm for lung cancer detection on chest radiographs using the segmentation method," *Sci. Rep*., vol. 12, no. 1, p. 727, 2022, doi: 10.1038/s41598-021-04667-w.

[7] C.-F. Chen *et al*., "A deep learning-based algorithm for pulmonary tuberculosis detection in chest radiography," *Sci. Rep*., vol. 14, no. 1, p. 14917, Jun. 2024, doi: 10.1038/s41598-024-65703-z.

[8] A. Abdelhamid, A. El-Ghamry, E. H. Abdelhay, M. M. Abo-Zahhad, and H. E.-D. Moustafa, "Improved pulmonary embolism detection in CT pulmonary angiogram scans with hybrid vision transformers and deep learning techniques," *Sci. Rep*., vol. 15, no. 1, p. 31443, Aug. 2025, doi: 10.1038/s41598-025-16238-4.

[9] H. Iqbal, A. Khan, N. Nepal, F. Khan, and Y.-K. Moon, "Deep Learning Approaches for Chest Radiograph Interpretation: A Systematic Review," 2024. doi: 10.3390/electronics13234688.

[10] B. Ozdemir, E. Aslan, and I. Pacal, "Attention Enhanced InceptionNeXt-Based Hybrid Deep Learning Model for Lung Cancer Detection," *IEEE Access*, vol. 13, pp. 27050–27069, 2025, doi: 10.1109/ACCESS.2025.3539122.

[11] R. Durgam, B. Panduri, V. Balaji, A. O. Khadidos, A. O. Khadidos, and S. Selvarajan, "Enhancing lung cancer detection through integrated deep learning and transformer models," *Sci. Rep*., vol. 15, no. 1, p. 15614, May 2025, doi: 10.1038/s41598-025-00516-2.

[12] Z. Liu, "Medical and Industrial Visual Inspection," *Appl. Comput. Eng*., vol. 196, pp. 38–46, 2025, doi: 10.54254/2755-2721/2025.LD28277.

[13] W. Xu, Y.-L. Fu, and D. Zhu, "ResNet and its application to medical image processing: Research progress and challenges," *Comput. Methods Programs Biomed*., vol. 240, p. 107660, Oct. 2023, doi: 10.1016/j.cmpb.2023.107660.

[14] J. Xu, Y. Pan, X. Pan, S. Hoi, Z. Yi, and Z. Xu, "RegNet: Self-Regulated Network for Image Classification," *IEEE Trans. Neural Networks Learn. Syst*., vol. 34, no. 11, pp. 9562–9567, Nov. 2023, doi: 10.1109/TNNLS.2022.3158966.

[15] Z. Ji, S. Ye, and X. Ma, "Sparse Coding Inspired LSTM and Self-Attention Integration for Medical Image Segmentation," *IEEE Trans. Image Process*., vol. 33, pp. 6098–6113, 2024, doi: 10.1109/TIP.2024.3482189.

[16] M. Li, Y. Jiang, Y. Zhang, and H. Zhu, "Medical image analysis using deep learning algorithms," *Front. Public Heal*., vol. Volume 11, no. November, pp. 1–28, 2023, doi: 10.3389/fpubh.2023.1273253.

[17] X. Ma, E. C. Ferguson, X. Jiang, S. I. Savitz, and S. Shams, "A multitask deep learning approach for pulmonary embolism detection and identification," *Sci. Rep*., vol. 12, no. 1, p. 13087, 2022, doi: 10.1038/s41598-022-16976-9.