Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 04, ICTEM 2.0 (2026)

ChestXplain: An Explainable CNN Framework with Grad-CAM for Trustworthy Multi-Disease Classification in Chest X-Rays

C. Malathy
Networking and Communication
SRM Institute of Science and Technology
Kattankaluthur, India
malathyc@srmist.edu.in

CH. Vasanth Kumar
Networking and Communication
SRM Institute of Science and Technology
Kattankaluthur, India
vasanthc2@srmist.edu.in

V. Arvindh Kumar
Networking and Communication
SRM Institute of Science and Technology
Kattankaluthur, India
av0872@srmist.edu.in

J. Ajay
Networking and Communication
SRM Institute of Science and Technology
Kattankaluthur, India
aj8674@srmist.edu.in

P. Asuwin
Networking and Communication
SRM Institute of Science and Technology
Kattankaluthur, India
ap0719@srmist.edu.in

M.G. Ahil Raj
Networking and Communication
SRM Institute of Science and Technology
Kattankaluthur, India
mr9911@srmist.edu.in

***Abstract*- In spite of being a highly elementary clinical task, the diagnostic interpretation of chest X-rays (CXRs) remains vulnerable to mistakes made by the human practitioner. Although the diagnostic accuracy of Convolutional Neural Networks (CNNs) is well documented, inherent to their use is the so-called "blackbox" nature of these models that still poses a significant obstacle to the widespread implementation of AI technologies in the medical field and the establishment of trust in them as a reliable source of guidance and information for doctors. In addition to that, it can be expected that physicians will feel hesitant and reluctant to rely on predictions of future events that cannot be independently verified. To bridge the gap that exists between the artificial intelligence (AI) and the physician's confidence in the technology's prognostic ability, we present ChestXplain as a framework with established reliability for performing multi-disease classification of module diseases of thoracic pathology (e.g. pneumonia, COVID-19, tuberculosis) using a CNN based on ResNet50 architecture incorporated with Gradient-weighted Class Activation Mapping (Grad-CAM) for the production of user-friendly visual heatmaps to provide evidence of transparency and to assist in decision-making. By using the generated heatmaps that identify all CXR regions of interest that were identified by CNN-generated models, radiologists will be able to validate their own diagnostic conclusions against those from an AI-based model. Unlike prior studies that primarily have focused on achievement of superior classification accuracy, ChestXplain uniquely combines the production of visually-explainable results with high-performance predictive accuracy into one comprehensive, integrated system. This multi-factorial approach to AI technology will help support the responsible incorporation of AI technologies into medical practice by ensuring patient safety through the development of diagnostic transparency and reducing risk of harm or injury resulting from improper use of AI.**

Keywords—Explainable AI (XAI), Chest X-ray (CXR), ResNet50, Grad-CAM, Deep Learning, Multi-Disease Classification, Pneumonia, COVID-19, Tuberculosis.

I. INTRODUCTION

Diseases Diseases of the respiratory system like pneumonia, COVID-19, and tuberculosis (TB) are highly infectious and present similar symptoms, causing tremendous strain on healthcare systems across the world and resulting in the death of millions of people each year. Diseases of the respiratory system are also a continuing global healthcare issue. CXRs are the cheapest, most accessible and widely used diagnostic tool. However, in busy environments like emergency rooms, the work from the over-stretched staff is slow, subjective, and highly error-prone. The problem is made worse by the absence of an expert radiologist in many low resource areas of the world.

In recent years, deep learning, especially CNNs, has shown great possibilities for automating and assisting with CXR interpretation. These models can be trained on large datasets to achieve high levels of accuracy, often matching or even surpassing human performance on specific tasks. However, this high accuracy is overshadowed by a major issue which stands as the primary obstacle to understanding the system. Most cutting-edge CNNs work as "black boxes," which provide predictions of "pneumonia" without showing their reasoning process. The medical field faces a significant problem because of this. Clinicians need to understand AI-generated outputs before they can trust them yet they should also avoid trusting anything which they find difficult to comprehend. The "black-box" problem presents severe dangers because it allows models to incorrectly diagnose patients through their excessive confidence while it creates ethical and legal dilemmas which depend on Information that remains ambiguous and it prevents us from discovering and rectifying biases that exist within our algorithms. All medical artificial intelligence systems need to attain comprehension for their safe and ethical use within clinical environments.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 04, ICTEM 2.0 (2026)

The Explainable AI (XAI) framework ChestXplain was created to solve this particular problem. The system uses two fundamental elements for its diagnostic evaluation which requires trustworthy results to function accurately.

1. **Accurate Classification:** The high-performance diagnostic engine uses a CNN which is built on the ResNet50 architecture. The model develops its skills to identify multiple pneumonia, COVID-19 and tuberculosis conditions from a single CXR.

2. **Visual Explanation:** The system employs Grad-CAM technology to produce user-friendly visual heatmaps. The resulting heatmaps display the anatomical regions of the CXR which the model considered most essential for making its predictions.

ChestXplain uses a comprehensive Kaggle database of thoracic pathologies to provide doctors an effective tool which predicts outcomes with a visual heatmap that shows which factors affected the final choice. The "glass-box" method enables evaluation beyond standard accuracy testing because radiologists can validate their performance directly. The system establishes trust and ensures secure AI medical assessments while transforming AI from a concealed system into an open diagnostic tool which works together with medical professionals

II. RELATED WORKS

The literature review demonstrates multiple important research patterns which ChestXplain intends to address through its future research. The majority of existing research studies which exist from both past and present continue to prioritize accurate classification results while they disregard the essential requirement for system understanding. The research team from Varshni et. al., [14] conducted their pneumonia detection study by testing different hybrid CNN and SVM models which they developed to achieve better AUC results. The researchers Nag et al. [11] built a CNN system which detects pediatric pneumonia and they measured its accuracy plus recall performance. The systems provide value to users but they function as "black boxes," which restrict their usability in medical settings.

Current research studies show an increasing requirement for systems which provide clear understanding. The researchers Sharma et al. [7] developed a Tuberculosis detection system which successfully used GradCAM to show infected areas through visual representation. The visualizations showed good matching with radiological knowledge according to the researchers who conducted the study. The researchers Antunes et al. [12] developed a system called "PneumoNet" which used Grad-CAM and LIME XAI tools to create better system visibility. The studies demonstrate that XAI functions as a critical component which healthcare professionals need to develop trust within their medical practices.Researchers keep investigating the problem of classifying multiple diseases but they still ignore this research topic. Researchers most commonly investigate single diseases which include pneumonia and TB and COVID-19 as their primary focus. A clinical model that detects only one medical condition does not provide useful results because patients present with multiple conditions which show overlapping symptoms.

The ChestXplain framework provides a complete solution for this crucial research need. The existing research contains only a few studies which successfully combine advanced multi-disease classification capabilities for Pneumonia and COVID-19 and TB with visual explanation capabilities through Grad-CAM.

III. PROPOSED SYSTEM

The proposed ChestXplain framework is a complete system for reliable diagnostic support. The system design includes a robust classification backbone which combines with an explainability system to deliver precise predictions while showing how the model reached its results. The end-to-end operational flow of the system is visually represented in Fig. (1).

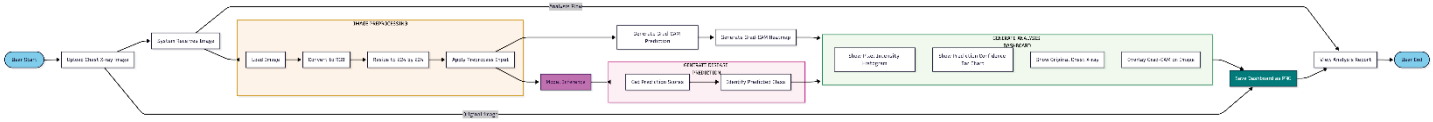

Fig. 1. ChestXplain Proposed System Flowchart.

A. Input and Preprocessing

The system uses a single CXR image as input. From a public Kaggle set which includes labeled examples of Pneumonia, COVID-19, Tuberculosis, and normal chest cavities, images are taken. Each image goes through a which is the standard pre processing pipeline. Each image is resized to 224x224 pixels to make it a fit for the ResNet50 model. Then, it undergoes pixel-value normalization, such as scaling to a range of [0, 1], to ensure the model trains reliably.

B. Classification Model: ResNet50

A ResNet50-based Convolutional Neural Network (CNN) sits at the center of the framework. ResNet50 is one of the best architectures for medical imaging as it learns deep features effectively. It implements residual connections to solve the problem of disappearing gradients in deep networks. The framework performs transfer learning and refines model that was pretrained on a large dataset called ImageNet to efficiently utilizes the already learnt low-level features, like edges, textures and customize them for recognition of thoracic problems, such as opacities, effusions, and consolidations. The last layers of the network are specifically modified for multi-label classification. A Global Average Pooling (GAP) layer and a Dense layer with a sigmoid activation function take the place of the original classifier to produce a probability for each target disease (TB, COVID-19, and pneumonia).

C. Explainability Module: Grad-CAM

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 04, ICTEM 2.0 (2026)

The framework incorporates Grad-CAM to deal with the 'black box' phenomenon. It utilizes Grad-CAM from the top techniques in post-hoc explainable AI that provides visual explanations for CNN predictions. Grad-CAM uses the weighted gradients of the output return to the last convolutional layer to create class specific saliency map or heatmap to visualize the regions in the input that the model focused in making the predictions

D. System Output

For each input CXR, the ChestXplain framework provides a two-part output:

- **The Disease Prediction:** A set of probability scores and labels that show whether each target disease is present or absent (for example, "Pneumonia: Detected").

The Visual Explanation: The original CXR overlaid with the color-graded Grad-CAM heatmap. In an overlay like this, a radiologist can quickly see why the model has made this prediction and whether this reasoning matches that of set clinical knowledge.

IV. METHODOLOGY

The ChestXplain framework was created as a comprehensive solution that included three components which processed data and trained models in two stages and provided users with a web application for explainable inference.

A. Environment and Tools

Framework was implemented using the Python. We constructed, taught, and assessed the deep learning model by using the TensorFlow framework together with its high-level Keras API. The primary ResNet50 network design was obtained from tensorflow.keras.applications. The OpenCV library (cv2) functioned as the tool for executing image input and output operations together with image enhancement tasks which included image size adjustment and color space transformation.

The user-facing diagnostic application was built using Flask, a lightweight web micro-framework. All static visualizations were created through Matplotlib which produced both the training history plots and the complete 2x2 analysis dashboard.

B. Dataset and Preprocessing

The model was trained by using public dataset for its train and validation processes. The dataset supports multiclass classification which are COVID19 NORMAL PNEUMONIA and TURBERCULOSIS.

Created tf.data.Dataset pipelines through the use of tf.keras.utils.image_dataset_from_directory function.

- **Image Resizing:** resized all images to match an input dimension of 224x224 pixels
- **Label Mode:** The label_mode setting established categorical which converted labels into 4-element one-hot vectors.
- **Batching:** a BATCH_SIZE of 32 for their training and validation processes.
- **Performance:** .cache() and .prefetch() methods to create a performance chain that improved I/O performance while reducing GPU idle time for their datasets.

C. Model Architecture and Augmentation

The classifier's architecture is based on ResNet50 with a custom head. A data augmentation pipeline was integrated directly into the model using keras.Sequential to apply RandomFlip, RandomRotation, and RandomZoom on the fly.

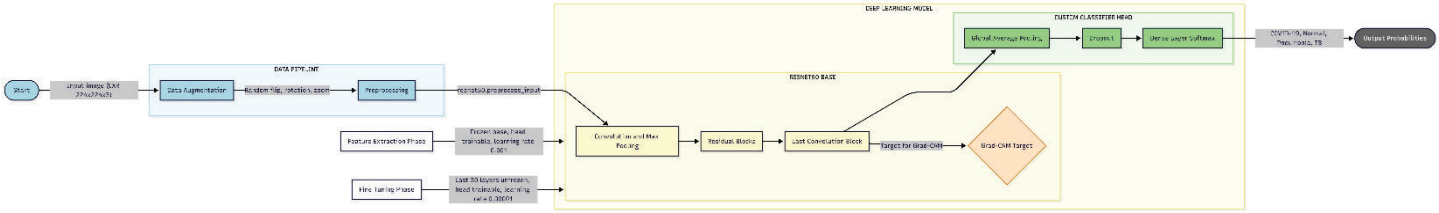

Fig. 2. ChestXplain Proposed System Flowchart.

Fig. (2). Detailed Deep Learning Architecture and Training Flow of ChestXplain. The model's data flow is as follows:

1. **Input:** A 224x224x3 tensor.
2. **Augmentation:** The input passes through the data_augmentation layer.
3. **Preprocessing:** The resnet50.preprocess_input function is applied to normalize pixel values as expected by the ResNet architecture.
4. **Base Model:** The data is fed into the ResNet50 base model (with include_top=False), pre-trained on ImageNet.
5. **Classification Head:** The output feature maps from the base model are passed through:
    - To reduce spatial dimensions, GlobalAveragePooling2D layer is used.
    - To prevent overfitting, Dropout(0.5) layer is controlled.
    - To create a probability distribution for each of the four output classes, a final Dense layer with softmax activation is used.

D. Training and Fine-Tuning Strategy

A A two-phase transfer learning strategy was implemented as detailed in train.py:

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 04, ICTEM 2.0 (2026)

1. **Phase 1 (Feature Extraction):** The entire ResNet50 base model was stopped (trainable = False). The new classification head's weights were ony trained for 10 epochs. The Adam optimiser with a learning rate of 0.001 and categorical_crossentropy as the loss function are used in this phase.
2. **Phase 2 (Fine-Tuning):** The base model was unfrozen, but all layers except for the top 30 were frozen again. This allows the model to make small adjustments to its high-level feature detectors. The model was compiled again with a very low learning rate of 1e-5 and trained for an additional 10 epochs, continuing from the end of Phase 1.

The final trained model was saved as chest_xray_model.h5 for use in the application. The learning curves for this training process are shown in Fig. (3).

Fig. 3. Training and Fine-Tuning History of the Proposed ResNet50 Classification Model

E. Explainable AI Web Application

For practical interface, a Flask web application app.py was built. The application operates in full capacity by loading its chest_xray_model.h5 model file when the system starts. The application conducts complete image analysis after the user submits a CXR image.

1. **Inference:** The image undergoes loading and preprocessing before its model execution which produces a prediction vector outputting the results as [0.05 0.10 0.80 0.05]. The final prediction is the class with the highest chance of being correct..
2. **Grad-CAM Generation:** The make_gradcam_heatmap function is called. The function creates a heatmap which shows the image areas that contributed to model class prediction through its targeting of ResNet50's final convolutional layer (conv5_block3_out).
3. **Dashboard Visualization**: A comprehensive 2x2 analysis dashboard is dynamically generated using Matplotlib. The dashboard saves as a distinct PNG file which delivers complete insights through its display of The original uploaded X-ray.
    - The Grad-CAM heatmap overlaid on the original image.
    - The model's confidence scores for four classes are displayed in a bar chart.
    - A pixel intensity histogram of the original image.

This dashboard Fig. (4) is then presented to the end-user in their web browser, fulfilling the "glass-box" objective of the ChestXplain framework.

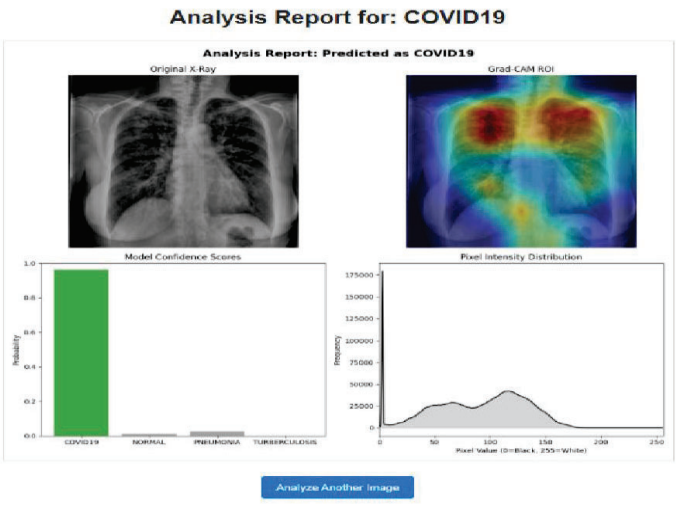

Fig. 4. Diagnostic Dashboard showing Grad-CAM Visualization and Confidence Scores for COVID-19

V. VALIDATION AND RESULTS

The thorough validation procedures developed to ensure the ChestXplain framework's reliability and usability in a clinical setting examined the system from qualitative (visual interpretive) and quantitative (statistical performance) perspectives.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 04, ICTEM 2.0 (2026)

A. Validation Method:

A Hold-Out Validation methodology was employed for the sake of objective accuracy, and to avoid any data leakage resulting from the data's dimensionality, the chest x-ray images represented high dimensionality, and the amount of resources required to fine-tune the deep ResNet50 architecture would be immense. Thus, in order to achieve these goals, the dataset was partitioned into three independent (mutually exclusive) sets:

1. **Training Set (80%):** The data was used to learn the weights of the model and to optimize the model's parameters.
2. **Validation Set (10%):** This data was used to monitor the accuracy and loss curves during training in order to modify any hyperparameters and implement early stopping.
3. **Test Set (10%):** This data was used solely for the final performance evaluation of the study.

B. Quantitative Performance Analysis

The model completed a 20-epoch learning phase to maintain its performance at stable levels without experiencing any overfitting issues. The training process develops through two separate stages, which Fig. (5) demonstrates. The custom classifier head experienced a rapid loss reduction during Phase 1 because it learned to utilize general features that ResNet50 base had collected. The pre-trained weights for thoracic disease showed successful adaptation to the specific medical field which resulted in validation loss decrease after Phase 2 started at Epoch 10.

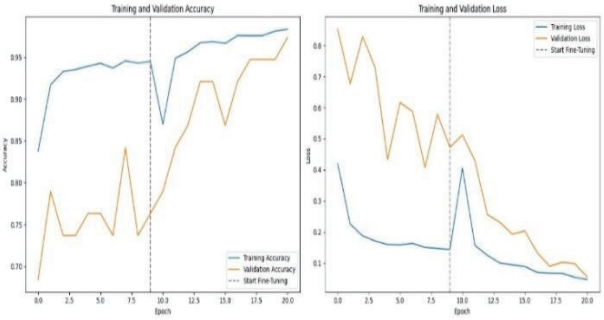

Fig. 5. Training and Validation Performance

The Precision, Recall (Sensitivity), and F1-Score are three metrics used to evaluate the final classification outcomes from the proposed framework. The proposed framework is capable of providing good diagnostics from all four categories of classifier output, thus the proposed framework demonstrates excellent performance. The results of the proposed framework have been compiled into Table I.

CLASSIFICATION PERFORMANCE METRICS			
CLASS	PRECISION	RECALL	F1-SCORE
COVID-19	0.99	0.93	0.97
NORMAL	0.91	0.70	0.79
PNEUMONIA	0.84	0.97	0.90
TUBERCULOSIS	0.98	0.98	0.97
WEIGHTED AVG	0.89	0.88	0.88

Table. 1. Diagnostic Dashboard showing Grad-CAM Visualization and Confidence Scores for COVID-19

A confusion matrix is used to give a more thorough picture of class performance and confusion between classes (Fig 6). Based on this alone, COVID-19 was correctly detected in nearly all of the tested instances (99 out of 106), while Tuberculosis was found with perfect accuracy (41 out of 41). These results indicate that both are highly accurately differentiated from each other, as these two are serious infections.

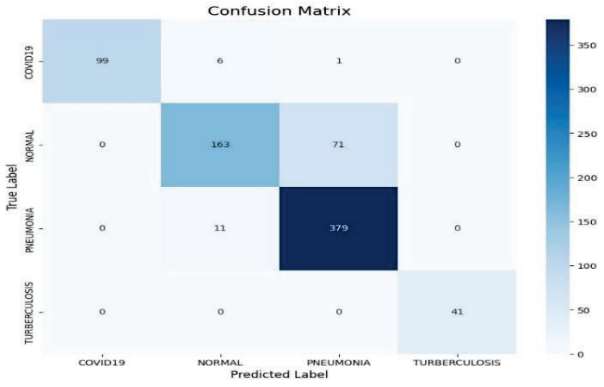

Fig. 6. Confusion Matrix of the Test Set results.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 04, ICTEM 2.0 (2026)

However, the confusion matrix revealed a major issue for the Normal class. The normal class has 71 instances of patients without pneumonia that were misdiagnosed as having pneumonia. The lower recall (0.70) for normals is likely due to this misdiagnosis, since on some lower-resolution X-rays of patients with modest pneumonic abnormalities, the milky or cloudy appearance of the lung tissue would appear similar to normal healthy lung tissue. Regardless of the fact that the algorithm also has a higher sensitivity (recall) than would be expected because the number of true sick patients was found by the algorithm, a higher proportion of false positive (normal patients incorrectly flagged by the algorithm) than false negative (sick patients not found by the algorithm) is often preferred when screening for the chance of flagging a patient for further follow-up evaluation by a physician.

C. Qualitative Validation via Explainability

Qualitative validation of this system was accomplished, in addition to statistical methods, in order to demonstrate that "right reasoning" resulted in "right predictions." With respect to models using high accuracy, those based on artifacts (e.g., hospital tags or anatomical structures) will have limited clinical utility as compared to models which are based upon pathology.

The Grad-CAM interpretability module was used to audit whether or not the focus of the model was valid. The resulting heat maps (see Analysis Report, Fig. (4)) consistently indicated relevant areas of interest within the lung fields (e.g., pneumonia has opacities; COVID has consolidation), but would fail to indicate any ambient noise that did not meet the aforementioned criteria as relevant; thus, confirming that the model's ability to be used as a diagnostic support tool correlates with that of clinically relevant anatomical structures in terms of their Region of Interest (ROI).

VI. CONCLUSION

The ChestXplain framework was developed to enhance the accuracy of deep learning models used for diagnosing chest X-rays. The framework resolved the critical "black-box" problem which prevents accurate AI models from being implemented in medical practice.

The contribution is a unified, two-pronged system that incorporates:

- A precise ResNet50-based classifier for COVID-19, tuberculosis, and pneumonia multi-label detection
- A Grad-CAM explainability module that produces visual heatmaps to support the model's forecasts.

ChestXplain enhances diagnostic safety together with system transparency which builds provider trust because it enables clinicians to confirm AI results through access to the model's decision-making process instead of its output. The work offers a useful, comprehensible, and multi-disease solution, in contrast to many studies that only report on accuracy or single disease analysis. The authors plan to expand their model by adding different thoracic pathologies and they will conduct clinical validation through the system's integration with hospital Picture Archiving and Communication Systems (PACS).

Link: https://colab.research.google.com/drive/19GAlZgonvZlQbe98StbkuK0kMMzEtt9U?usp=sharing

REFERENCE

[1] Chibuzo J. Ejiyi, Zhen Qin, Andrew O. Nnani, Fei Deng, Tochukwu U. Ejiyi, Michael B. Ejiyi, Victor K. Agbesi, and Olugbenga Bamisile, "ResfEANet: ResNet-fused external attention network for tuberculosis diagnosis using chest X-ray images," *IEEE Access*, vol. 11, pp. 1–14, 2023.

[2] Jianwei Zhang, Huan Chao, Girish Dasegowda, Ge Wang, Mannudeep K. Kalra, and Pingkun Yan, "Revisiting the trustworthiness of saliency methods in radiology AI," *IEEE Transactions on Medical Imaging*, vol. 42, no. 9, pp. 2566–2578, 2023.

[3] Vishal Sharma, Nillmani, Sanjay Kumar Gupta, and K. K. Shukla, "Deep learning models for tuberculosis detection and infected region visualization in chest X-ray images," *Biomedical Signal Processing and Control*, vol. 86, pp. 104897, 2023.

[4] Parvez Rahman and Golam Mahbub Islam, "Interpretable deep learning for pediatric pneumonia diagnosis through multi-phase feature learning," *Electronics*, vol. 14, no. 2, pp. 1–19, 2025.

[5] Yong Xie, Bin Zhu, Yifan Jiang, Bo Zhao, and Hong Yu, "Diagnosis of pneumonia from chest X-ray images using YOLO deep learning," *IEEE Access*, vol. 13, pp. 1–12, 2025.

[6] Tushar Nag, Shubham S. Rajawat, and Ankit Rana, "Detection of pneumonia using chest X-ray images and convolutional neural network," *Biomedical Engineering Letters*, vol. 15, no. 1, pp. 89–101, 2025.

[7] Catarina Antunes, José Manuel F. Rodrigues, and Adriano Cunha, "PneumoNet: Artificial intelligence assistance for pneumonia detection on X-rays," *Artificial Intelligence in Medicine*, vol. 151, pp. 102876, 2025.

[8] Shubham Sharma and Kiran Guleria, "A systematic literature review on deep learning approaches for pneumonia detection using chest X-ray images," *Multimedia Tools and Applications*, vol. 83, pp. 24101–24151, 2023.

[9] Rida Siddiqi and Shahrukh Javaid, "Deep learning for pneumonia detection in chest X-ray images: A comprehensive survey," *Journal of Imaging*, vol. 10, no. 7, pp. 176, 2024.

[10] Waleed Mohammed Eido and Huda Mohammed Yasin, "Pneumonia and COVID-19 classification and detection based on convolutional neural network: A review," *Asian Journal of Research in Computer Science*, vol. 18, pp. 174–183, 2025.

[11] Mohammed A. M. Abueed, Dayang Norwati Nor, Noraini Ibrahim, and Jean-Marc Ogier, "Pneumonia detection using transfer learning: A systematic literature review," *International Journal of Advanced Computer Science and Applications*, vol. 16, no. 2, pp. 1–15, 2025.

[12] Dnyaneshwar G. Bhalke and Anjum S. Shaikh, "Classification of pneumonia subtypes in chest X-rays using a custom CNN," in *Proceedings of the 1st International Conference on AIML-Applications for Engineering & Technology*, Pune, India, IEEE, 2025, pp. 1–6.