IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, July 2025
Impact Factor: 7.67

Hybrid Deep Learning Framework for Pneumonia Diagnosis: A CNN and YOLO-Based Approach

Ms. Shilpa Bhandare
Bharati Vidyapeeth College of Engineering, Navi Mumbai

Abstract: *Pneumonia remains a leading global cause of hospitalization, especially in resource-constrained areas lacking expert radiologists. Chest X-rays (CXRs) are the standard diagnostic tool, yet manual interpretation is slow and skill-intensive. This study investigates deep learning as a scalable solution for automated pneumonia detection and localization. A CNN was applied for classification, and YOLO for region-of-interest (ROI) localization. Four models were trained using preprocessed, augmented datasets. The best model achieved 96.8% training accuracy, 83% validation accuracy, and F1-scores of 0.799 (normal) and 0.819 (pneumonia). YOLO-based localization yielded F1-scores of 0.82 (normal) and 0.54 (pneumonia). These AI-driven methods offer cost-effective, high-speed diagnostics, alleviating radiologist burden and improving healthcare access in underserved regions.*

Keywords: Pneumonia

I. INTRODUCTION

Pneumonia is a critical respiratory condition that affects individuals of all ages and can result in severe complications or death if not identified early. Accurate and timely diagnosis is essential for effective treatment. Traditionally, chest X-rays (CXRs) interpreted by radiologists have been the primary method for diagnosing pneumonia. However, this manual process is both labor-intensive and prone to subjective variability. While various computer-assisted techniques have been developed to support radiologists, their diagnostic accuracy has often been limited.

With the rapid evolution of machine learning, particularly deep learning, significant strides have been made in medical image analysis. Convolutional Neural Networks (CNNs) have emerged as powerful tools for feature extraction and classification in medical imaging, including the detection of pneumonia in CXRsThe paper by Rathnakannan and Balasubramanian introduces a deep learning approach for pneumonia detection, combining Convolutional Neural Networks (CNN) and YOLO. Their model demonstrates promising results in accurately identifying pneumonia from medical images. This method offers a significant improvement in automated medical imaging systems [1].

Salido and Ruiz [2] explored grayscale transformation combined with CNN-based classification to enhance pneumonia detection accuracy. Furthermore, the You Only Look Once (YOLO) framework—a high-speed object detection model—has shown promising results in various computer vision applications [3], [4]. However, conventional pre-trained YOLO models, built on large-scale datasets, often come with high computational demands and a tendency to overfit when applied to smaller, specialized datasets [5], [6].

To overcome these challenges, this study develops a custom YOLO architecture from scratch, eliminating the reliance on pre-trained weights. The model is evaluated using publicly available datasets from RSNA and Kaggle [7]. Data augmentation techniques are applied to improve performance, particularly under data-constrained conditions. Recent advancements, such as the use of Generative Adversarial Networks (GANs) for synthetic image generation, have also demonstrated effectiveness in enhancing dataset diversity [8].

Manual threat detection in medical X-rays, especially when the targets are partially obscured or rotated, remains a significant challenge [9]–[11]. To address this, a YOLO-based detector is proposed to automate the process. The use of dropout, introduced by Srivastava et al. [12], is incorporated to reduce overfitting and enhance model generalization. Notably, Ozturk et al. [13] achieved high classification accuracy for chest X-rays using a YOLO-DarkNet hybrid, effectively handling binary (e.g., COVID-19 vs. Normal) and multi-class problems (e.g., COVID-19, Pneumonia, Normal).

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, July 2025
Impact Factor: 7.67

Yao et al. [14] proposed Pneumonia-YOLO (PYolo), which integrates dilated convolutions and attention mechanisms to improve lesion detection in CXRs, achieving a mean Average Precision (mAP) of 46.84 on the RSNA dataset. In another approach, Nithya et al. [15] employed a multisampling image filtering method alongside an enhanced CNN to further improve classification outcomes.

These studies underscore the transformative potential of deep learning in automating and improving pneumonia detection from CXRs. Leveraging such advancements can significantly reduce diagnostic time, improve accuracy, and alleviate the burden on medical professionals.

The primary goal of this research is to design a **low-cost, efficient machine learning model** for automatic pneumonia detection from chest X-rays. The proposed model addresses two main challenges:

Efficiency: Radiologist-based CXR analysis is time-consuming. Automating this task with a YOLO-based detector enhances speed and consistency in diagnosis.

Cost-effectiveness: Many advanced diagnostic solutions are expensive and impractical in low-resource settings. Developing a lightweight and accessible model ensures broader adoption in underprivileged regions.

By addressing both these aspects, the proposed model aims to improve healthcare accessibility and diagnostic accuracy on a global scale.

II. PROPOSED SYSTEM

The proposed system's workflow for pneumonia prediction is illustrated in Fig. 1.

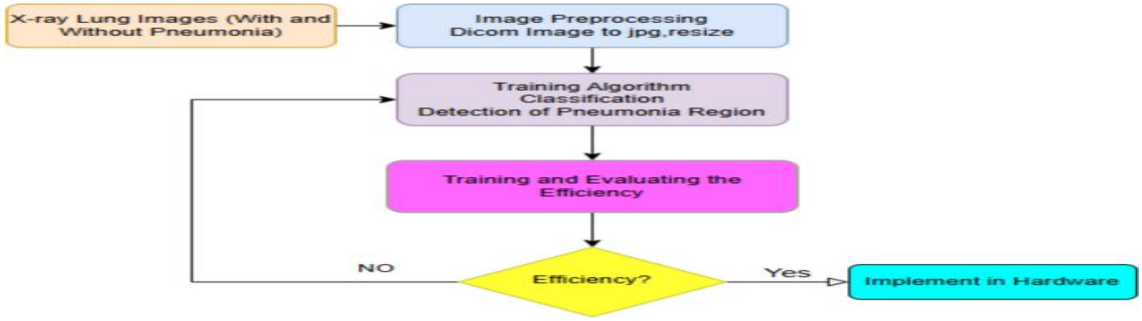


Fig. 1 Block diagram of the proposed system

The pneumonia detection pipeline begins with the preprocessing of chest X-ray (CXR) images. Since medical images are commonly stored in the DICOM format [15], they are initially converted to the widely supported JPG format to ensure compatibility with the deep learning framework. All images are then resized to a standardized resolution of $200 \times 200 \times 3$ pixels to maintain uniformity in input dimensions. To further streamline processing, the images are converted from RGB to grayscale, effectively reducing the channel dimensionality from three to one. This reduction not only decreases computational complexity but also improves training efficiency, as grayscale intensity values are sufficient for accurate feature extraction in pneumonia diagnosis.

The preprocessed grayscale images are then input into the deep learning model. Within the model, multiple convolutional layers work together to identify and extract key features associated with pneumonia. These features are subsequently flattened into a one-dimensional vector and passed through a fully connected (dense) layer. This layer leverages the extracted features to predict the presence of pneumonia in previously unseen X-ray images. To maintain high performance, the model's accuracy and efficiency are continually evaluated and optimized throughout the training process.

2.1 Tools and Algorithm Used for Pneumonia Detection

This research utilizes Google Colab with GPU/TPU support, along with Python, OpenCV, and Keras/PyTorch libraries, for detecting pneumonia from chest X-ray images. Two CNN architectures—one with dropout and one without—were developed using ReLU and Sigmoid activations [16, 17]. Models were trained on datasets with and without augmentation to enhance feature learning and diagnostic accuracy.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, July 2025
Impact Factor: 7.67

The CNN with dropout includes 23 layers, comprising three convolutional layers (ReLU activated), each followed by pooling and dropout layers. The final Sigmoid layer handles binary classification. Input images of size $200 \times 200 \times 3$ are processed using 3×3 filters starting with 16 channels. Features from initial convolutional layers are flattened and passed through three dense layers (512, 256, 64 neurons). The training uses binary cross-entropy loss, the Adam optimizer, and accuracy as the evaluation metric [17, 18].

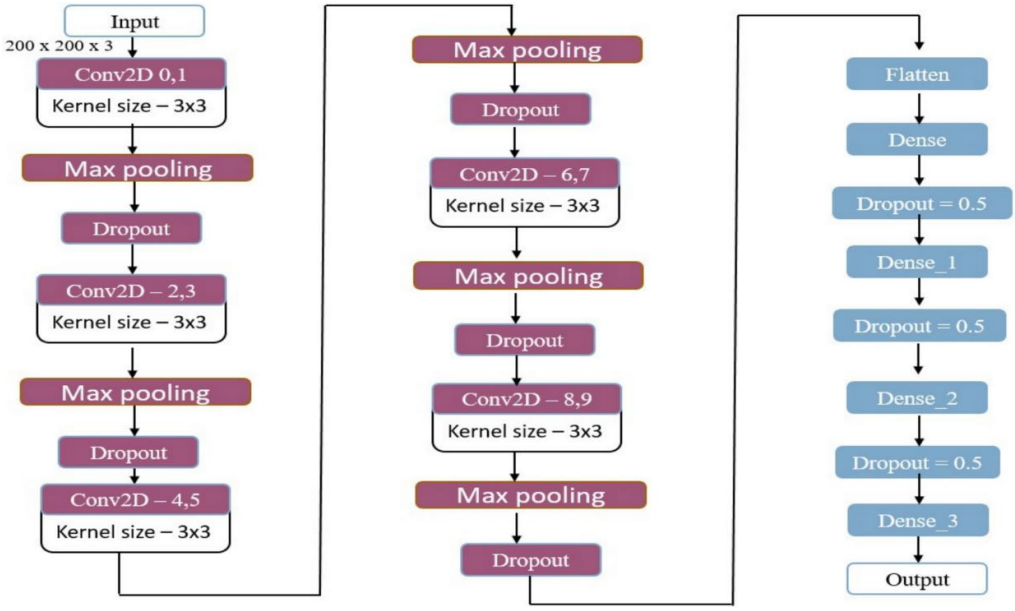


Fig. 2 Architecture of the CNN using dropout

2.2 YOLO

YOLO (You Only Look Once) is a high-performance, real-time object detection model that balances speed and accuracy by processing entire images in a single pass. Unlike traditional two-stage methods, YOLO uses a unified architecture for rapid detection.

The CSPDarknet53 backbone enhances feature extraction by splitting layers into processing and bypass branches, reducing parameters while retaining critical information. The SPP module improves detection across various object scales, and PANet enhances feature flow for precise localization. The YOLO head outputs bounding boxes, confidence levels, and class probabilities.

2.3 Chest X-Ray for Pneumonia and Dataset

Chest X-rays offer a fast, non-invasive way to diagnose pneumonia, revealing signs of lung infection and helping guide treatment. They are also useful for tracking recovery through follow-up imaging.

For this study, a dataset of 6,289 chest X-ray images was compiled from Google Images, Kaggle, and the RSNA Pneumonia Detection Challenge. This diverse, high-quality dataset supports robust model training and evaluation across various patient demographics and pneumonia cases.

This division provides a robust foundation for training and evaluating the proposed models. The training set, comprising 3875 normal and 1341 pneumonia images, serves as the primary resource for the model to learn distinguishing features. The validation set, with 390 normal and 234 pneumonia images, is used to fine-tune the hyperparameters and prevent overfitting. Finally, the testing set, which includes 170 normal and 279 pneumonia images, ensures a comprehensive and unbiased evaluation of the model's performance on unseen data.arechest X-ray without pneumonia, while X-rays with pneumonia manifesting in various positions. This visual diversity further enriches the dataset, helping the model to generalize better across a variety of real-world clinical scenarios. The

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, July 2025
Impact Factor: 7.67

inclusion of such a representative dataset underpins the study's goal of developing a reliable, efficient, and scalable solution for automated pneumonia detection and localization in clinical practice.

The collected DICOM-format pneumonia images were first converted to JPG and resized to $200 \times 200 \times 3$ for uniformity.Images were then converted to grayscale to educedimensionality and improve training efficiency.Feature extraction was performed through multiple CNN layers, followed by flattening and a fully connected layer for classification.Training used the Adam optimizer (learning rate 0.001), with 80 epochs and a batch size of 32.Data augmentation (rotations, flips, zoom) and dropout (0.5) with early stopping helped prevent overfitting and improve model robustness.

III. ANALYSIS OF PNEUMONIA DETECTION USING CNN AND YOLO ALGORITHMS

This section presents the training workflow and outcomes of both CNN with dropout and YOLO algorithms applied to pneumonia detection. The models were trained on a dataset comprising 4295 normal and 1652 pneumonia-labeled chest X-rays. CNN was trained for 80 epochs, while YOLO underwent 90 epochs of training. Both models used the same input dataset, with YOLO also performing object localization to identify affected lung areas. Prior to training, all images were formatted uniformly to optimize processing efficiency.

4Prediction Result

Following the training of the CNN model using both validation and training datasets, predictions were made on a separate testing dataset. Notably, Model demonstrated the most promising performance among the evaluated models.

Table 1.presents the precision, recall, and F1-score metrics achieved by Model 2 for both normal and pneumonia-affected chest X-rays (CXRs) in the testing dataset.

Metrics		Model 2
Precision	0	0.78
	1	0.83
Recall	0	0.80
	1	0.81
F1-score	0	0.799
	1	0.819

The testing dataset is analyzed using the YOLO algorithm and the result is investigated. Figure 3.displays the printed message of actual and predicted counts for both pneumonia and normal X-rays separately. Based on this data, a classification report is generated, as shown in Fig. 4. Examining this report, we see that the prediction and recall for pneumonia are 0.99 and 0.37, respectively, while for normal X-rays, these values are 0.70 and 1.0. This indicates that the model exhibits high precision for identifying pneumonia but lower recall, meaning it may miss some actual pneumonia cases.

```
print("Actual and Predicted count of Pneumonia cases - "+str(pneumonia_act),str(pneumonia_pred))

print("Actual and Predicted count of Normal cases - "+str(Normal_act),str(Normal_pred))
```

```
Actual and Predicted count of Pneumonia cases - 1202 450
Actual and Predicted count of Normal cases - 1770 2522
```

Fig. 3 Actual and predicted count cases

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, July 2025
Impact Factor: 7.67

```
print(classification_report(target_act,target_pred))
```

	precision	recall	f1-score	support
0	0.70	1.00	0.82	1770
1	0.99	0.37	0.54	1202
accuracy			0.74	2972
macro avg	0.84	0.68	0.68	2972
weighted avg	0.82	0.74	0.71	2972

Fig. 4 Classification report of YOLO model

$$\text{F1 score} = 2 \times \frac{\text{PrcisionX Recall}}{\text{Prcision} + \text{Recall}} \quad (2)$$

While accuracy is useful, the F1-score (Eq. 2) provides a more balanced view by incorporating both precision and recall—especially important in medical datasets with class imbalance. Model 2 demonstrated the highest F1-score, making it more effective than others. However, to reduce missed diagnoses, further improvements are needed to boost recall for pneumonia cases.

Figure 29 highlights the typical trade-off: as precision increases, recall tends to drop. Figure 30 presents the ROC curve, showing high sensitivity and a low false positive rate. The curve's proximity to the top-left corner and an AUC of ~0.93 indicate strong classification performance. Figure 31 shows the confusion matrix, affirming the model's reliability in identifying pneumonia in chest X-rays.

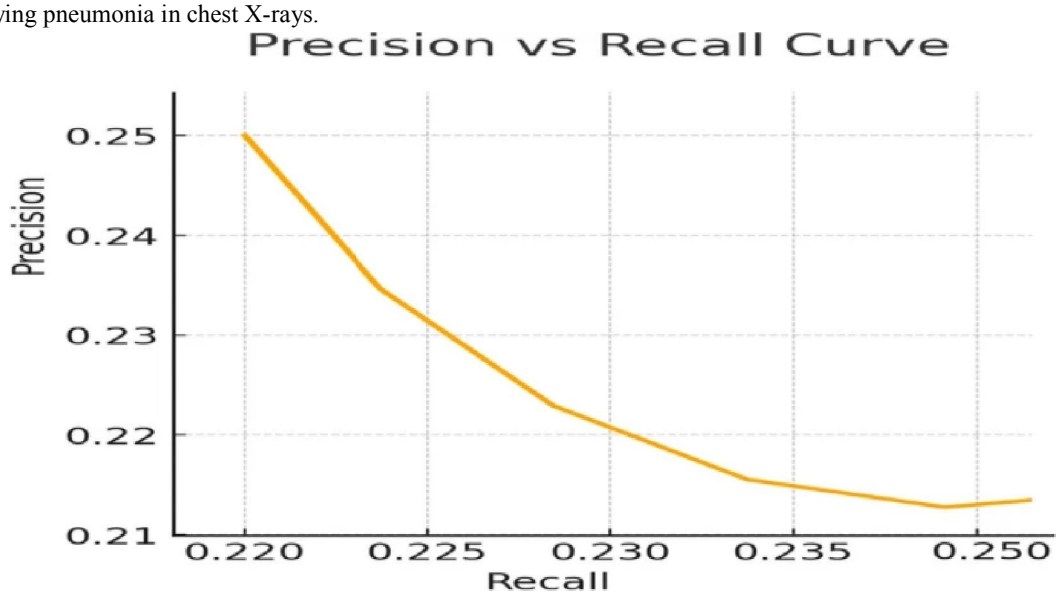


Fig 5.Precision vs recall curve

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, July 2025
Impact Factor: 7.67

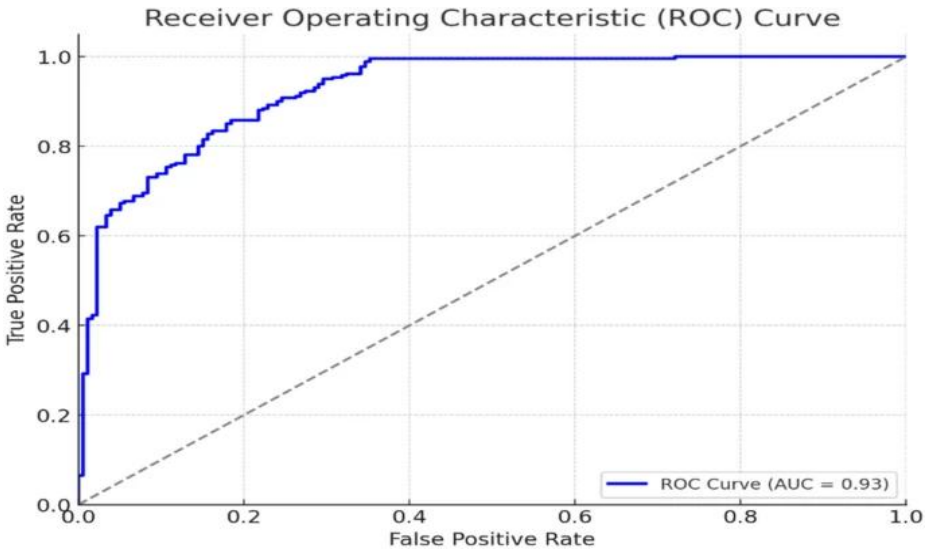


Fig 6.ROC curve

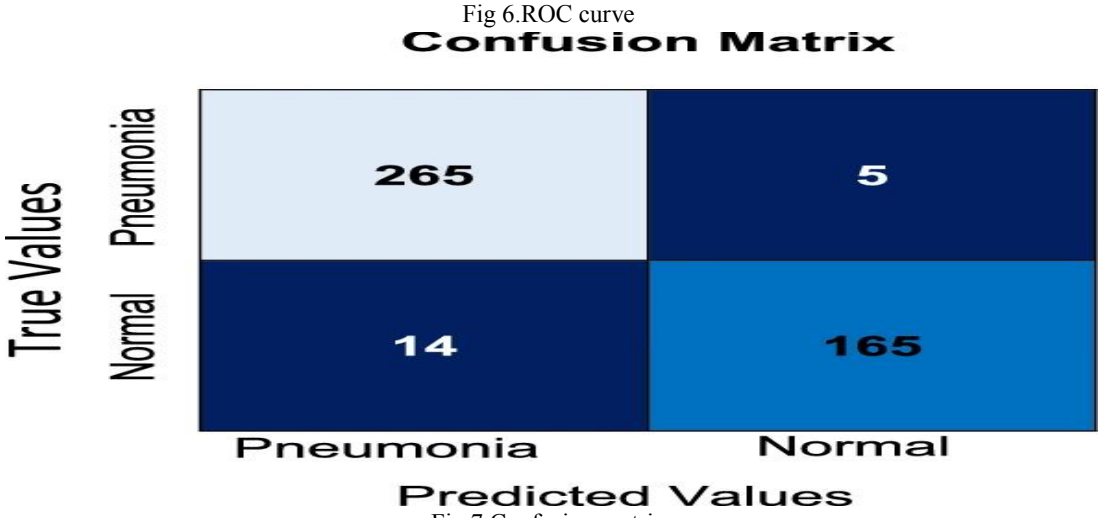


Fig 7.Confusion matrix

IV. QUALITATIVE RESULTS

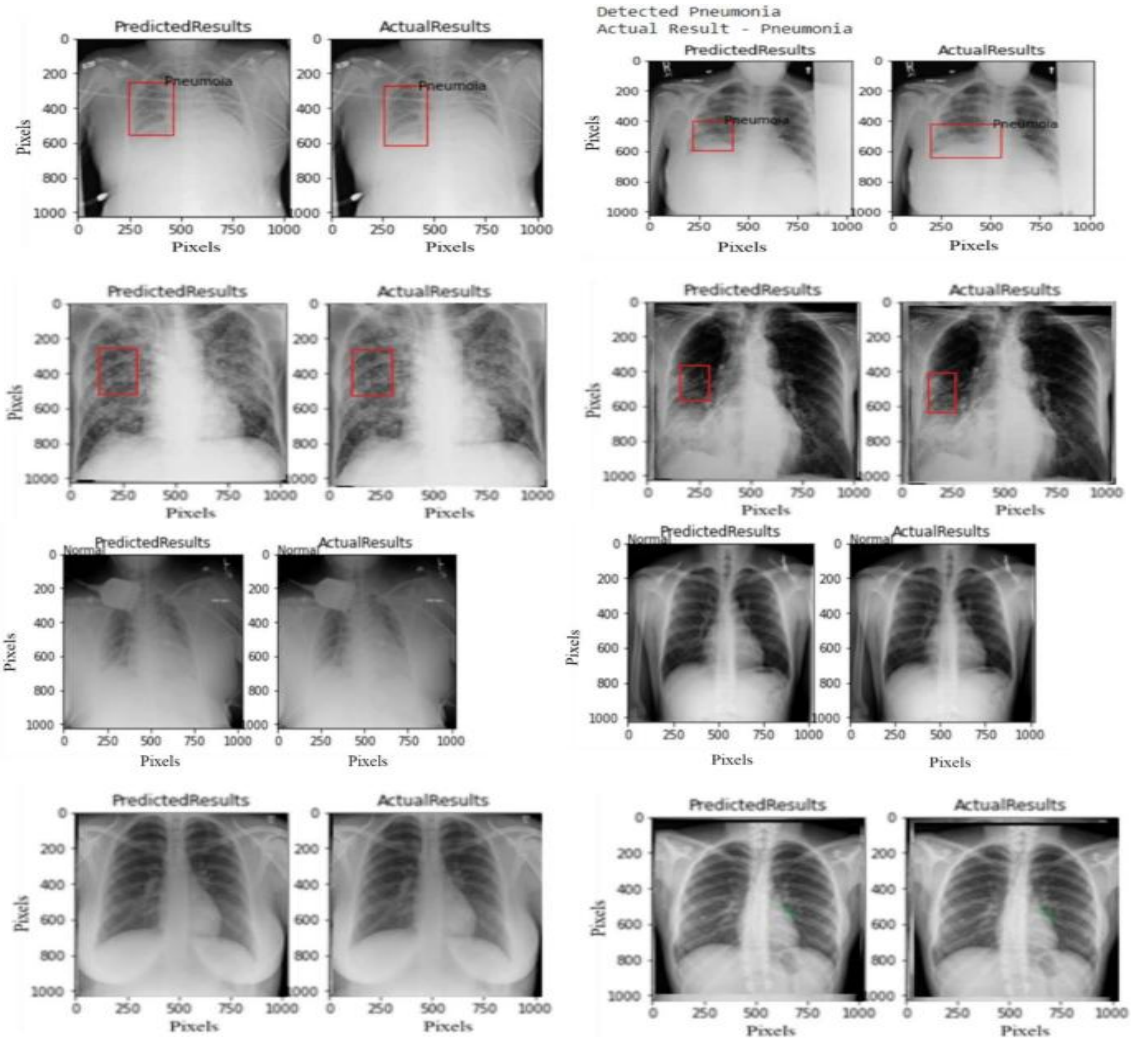
4.1 Prediction Visualization

After training and validation, the YOLO model was tested on unseen data. Figure 32 presents examples where pneumonia was correctly identified. Each instance includes the original chest X-ray, the model's predicted output, and a red bounding box highlighting the predicted pneumonia region. The model classifies each image as either pneumonia or normal, with classification and ROI accuracy shown in Figure 33.

Figures 33 depict normal chest X-rays where the model correctly identified the absence of pneumonia, validating its classification ability. In addition to successful predictions, a few misclassified cases were observed. These errors included incorrect class predictions, inaccurate ROI positioning, or discrepancies in the ROI size—pointing to areas for further model refinement.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, July 2025
Impact Factor: 7.67

Correctly detected pneumonia image



Correctly detected normal image

Comparative studies in recent literature have primarily focused on classification, often overlooking localization. A baseline CNN model [19] using a similar CXR dataset achieved an F1-score of 0.77 and a validation accuracy of 0.80. In contrast, our CNN approach, enhanced with dropout and data augmentation, reached a validation accuracy of 0.83 and an F1-score of 0.819.

Another CNN study [13], while achieving an accuracy of 0.81, did not include localization. Our model, integrating YOLO, provided both classification and localization, achieving a training accuracy of 0.968, validation accuracy of 0.83, and F1-scores of 0.82 for normal and 0.54 for pneumonia—resulting in a weighted average F1-score of 0.71. This localization capability offers clinicians more actionable diagnostic information compared to traditional CNNs.

Together, the results from CNN and YOLO confirm the effectiveness of our hybrid approach in both detecting and localizing pneumonia-affected regions in chest X-rays. This dual capability enhances diagnostic utility and clinical decision-making.

However, limitations remain. Differentiating pneumonia from other lung diseases (e.g., tuberculosis, cancer) continues to pose a challenge. Variations in image quality and inconsistencies in annotations can also impact performance.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 2, July 2025 Impact Factor: 7.67

Addressing these issues may involve advanced techniques such as attention mechanisms, ensemble learning, and curating larger, high-quality annotated datasets.

Future integration of this model into embedded edge computing devices could further streamline pneumonia detection, enabling fast and reliable on-site diagnosis. This would be especially valuable in remote or resource-limited healthcare environments, improving access to timely treatment and clinical support.

V. CONCLUSION

This work demonstrates an effective deep learning approach for automated pneumonia detection in chest X-rays. After preprocessing varied CXR datasets, a baseline CNN model was trained, followed by three enhanced versions using data augmentation and dropout. The best model achieved 96.8% training accuracy and 83% validation accuracy, with F1-scores of 0.799 for normal and 0.819 for pneumonia cases.

Additionally, a YOLO-based detector was employed for both classification and localization. It achieved F1-scores of 0.82 for normal and 0.54 for pneumonia, with a weighted average of 0.71 and macro average of 0.68. The combined CNN and YOLO framework shows strong potential for accurate, efficient, and interpretable pneumonia detection, especially in clinical or resource-limited environments.

REFERENCES

[1] K. Rathnakannan and S. Balasubramanian, "Deep learning for pneumonia detection: A combined CNN and YOLO approach," *Human-Centric Intelligent Systems*, vol. 5, no. 1, pp. 44–62, Mar. 2025, doi: 10.1007/s44230-025-00091-9.

[2] J. A. A. Salido and C. Ruiz Jr., "Using deep learning to detect melanoma in dermoscopy images," *Int. J. Mach. Learn. Comput.*, vol. 8, no. 1, pp. 18–23, 2018, doi: 10.18178/ijmlc.2018.8.1.664.

[3] Y. Lu, L. Zhang, and W. Xie, "YOLO-compact: An efficient YOLO network for single-category real-time object detection," in *Proc. Chinese Control and Decision Conf. (CCDC)*, 2020, pp. 1931–1936, doi: 10.1109/CCDC49329.2020.9164580.

[4] R. Pathaka, M. Pandeya, and S. Rautaraya, "Application of deep learning for object detection," *Procedia Comput. Sci.*, vol. 132, pp. 1706–1717, 2018, doi: 10.1016/j.procs.2018.05.147.

[5] S. Rajaraman, I. Kim, and S. K. Antani, "Detection and visualization of abnormality in chest radiographs using modality-specific convolutional neural network ensembles," *PeerJ*, vol. 8, p. e8693, 2020, doi: 10.7717/peerj.8693.

[6] A. K. Jaiswal *et al.*, "Identifying pneumonia in chest X-rays: A deep learning approach," *Measurement*, vol. 145, pp. 511–518, 2019, doi: 10.1016/j.measurement.2019.05.076.

[7] Kaggle, "Chest X-ray pneumonia dataset." [Online]. Available: https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia

[8] R. Joby *et al.*, "PneumoGAN: A GAN-based model for pneumonia detection," in *Proc. 2nd Int. Conf. Advances in Computing, Communication, Embedded and Secure Systems (ACCESS)*, 2021, pp. 102–106, doi: 10.1109/ACCESS51619.2021.9563341.

[9] P. Lakhani and B. Sundaram, "Deep learning at chest radiography: Automated classification of pulmonary tuberculosis by using convolutional neural networks," *Radiology*, vol. 284, no. 2, pp. 574–582, 2017, doi: 10.1148/radiol.2017162326.

[10] S. A. Patil and V. R. Udpi, "Chest X-ray features extraction for lung cancer classification," *J. Sci. Ind. Res.*, vol. 69, no. 4, pp. 271–277, 2010.

[11] H. Mohsen, E. A. El-Dahshan, E. S. M. El-Horbaty, and A. M. Salem, "Classification using deep learning neural networks for brain tumors," *Future Comput. Inform. J.*, vol. 3, no. 1, pp. 68–71, 2018, doi: 10.1016/j.fcij.2017.12.001.

[12] N. Srivastava *et al.*, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

[13] T. Ozturk *et al.*, "Automated detection of COVID-19 cases using deep neural networks with X-ray images," *Comput. Biol. Med.*, vol. 121, p. 103792, 2020, doi: 10.1016/j.compbiomed.2020.103792.

[14] S. Yao, Y. Chen, X. Tian, R. Jiang, and S. Ma, "An improved algorithm for detecting pneumonia based on YOLOv3," *Appl. Sci.*, vol. 10, no. 5, p. 1818, 2020, doi: 10.3390/app10051818.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, July 2025
Impact Factor: 7.67

[15] T. M. Nithya, P. R. Kanna, S. Vanithamani, and P. Santhi, "An efficient PM—multisampling image filtering with enhanced CNN architecture for pneumonia classification," *Biomed. Signal Process. Control*, vol. 86, p. 105296, 2023, doi: 10.1016/j.bspc.2023.105296.

[16] T. B. Chandra *et al.*, "Automatic detection of tuberculosis-related abnormalities in chest X-ray images using hierarchical feature extraction scheme," *Expert Syst. Appl.*, vol. 158, p. 113514, 2020, doi: 10.1016/j.eswa.2020.113514.

[17] L. T. Duong *et al.*, "Detection of tuberculosis from chest X-ray images: Boosting the performance with vision transformer and transfer learning," *Expert Syst. Appl.*, vol. 184, p. 115519, 2021, doi: 10.1016/j.eswa.2021.115519.

[18] B. P. Rajan, A. J. Shree, and K. Rathnakannan, "Binarized neural network with depth imaging techniques for stock return direction prediction," *J. Ambient Intell. Human Comput.*, vol. 14, pp. 3899–3912, 2023, doi: 10.1007/s12652-022-04460-1.

[19] H. Liz, J. Huertas-Tato, M. Sánchez-Montañés, J. Del Ser, and D. Camacho, "Deep learning for understanding multilabel imbalanced Chest X-ray datasets," *Future Gener. Comput. Syst.*, vol. 144, pp. 291–306, 2023, doi: 10.1016/j.future.2023.03.005