JCBI
Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
Volume 11 Issue 02
2026
Research Article
https://doi.org/10.56979/1102/2026/1578

Automated Detection and Segmentation of Pneumonia from Chest X-Ray Images Using a Convolutional Neural Network and U-Net

Mithoo Kailash[1], Umer Iqbal[2], Muhammad Kashif Siddhu[3], Muhammad Asif Saleem[2*], and Muhammad Saeed[2]

[1]Department of Artificial Intelligence, Islamia University Bahawalpur, Pakistan.
[2]Department of Artificial Intelligence, Faculty of Computer Sciences, Lahore Garrison University, Lahore, Pakistan.
[3]Department of Computer Science, Faculty of Computer Sciences, Lahore Garrison University, Lahore, Pakistan.
[*]Corresponding Author: Muhammad Asif Saleem. Email: asif.saleem@lgu.edu.pk

Received: March 19, 2026 Accepted: July 14, 2026

Abstract: Pneumonia is a severe disease of the lungs with inflammation of the air sacs which can be a cause of severe sickness and end of life, particularly in children and elderly people. The most frequently used method to diagnose pneumonia is to use chest X-ray imaging, however manual diagnosis by radiologists may be timely and prone to errors. This paper presents a deep learning-based automated approach for lung segmentation and pneumonia diagnosis in chest X-ray (CXR) pictures. A Convolutional Neural Network (CNN) model was used to categorize the CXR pictures as either normal or pneumonia, and a U-Net model was used to segment the lung area. The models were trained and tested using a publicly available library of CXR pictures of both normal and pneumonia patients. Before being input into the model during training, each picture was shrunk to 224 × 224 pixels, normalized, and enhanced for optimal performance. It was compared with ResNet-50, MobileNet, Inception-V3, Vision Transformer (ViT), and the Swin Transformer models for the proposed CNN, and with a VGG U-Net model for the proposed U-Net. Compared to all the other comparison models, which varied from 90.43% to 93.34% accuracy and 0.82 to 0.88 recall value for pneumonia cases, the proposed CNN obtained an accuracy of 94.00% using a test set of 1150 CXR pictures and a 0.95 recall value for the pneumonia cases.

Keywords: Pneumonia Detection; Chest X-ray; Convolutional Neural Network; U-Net Segmentation; Deep Learning; Medical Image Analysis

1. Introduction

Pneumonia is one of the leading causes of mortality and morbidity worldwide. It causes inflammatory condition of oxygen sacs in the lungs and decreases oxygen exchange, and is most concerning in older adults, weak immune systems and children [1]. Early and proper diagnosis helps in therapy and reduces the probability of severe complications. CXR (CXR) is the greatest frequently used tool for diagnosis of pneumonia as it is easily available, cost-effective and rapid. Manual reading is, however, slow and varies with the expertise of the reader [2]. This burden can be alleviated using automated analysis, which can aid in quicker and more uniform decision making [3] of medical images has changed by the application of Deep Learning (DL). CNNs independently identify the characteristics from the images and are capable of evaluating pneumonia in CXR images with high accuracy [4]. Careful architecture design and hyper parameter tuning further boosts robustness on the standard CXR dataset [5] and when there is a limited amount of labeled data, transfer learning with pre-trained backbones like DenseNet remains a potent baseline [6].

Recently transformers have been making headway in this endeavor. ViTs employ self-attention to gain the global context and space information, and have outperformed CNN baselines on public CXR data [7]. Local and global features can be combined by hybrid models containing both CNN and transformer

components, such as CNN encoders with modified Swin Transformer blocks, which achieve a higher accuracy with a reduced number of misclassifications [8] and CNN–transformer models with early multi-channel fusion that provides additional explain ability and uncertainty estimation, necessary for clinical use [9].

Attention-guided methods combining both spatial and temporal and biologically inspired processing achieve higher than 99% accuracy with high accuracy in clinically relevant regions [10] and multi-scale attention fusion of ViT and Efficient Net provides good multi-class pneumonia classification support [11]. It is as crucial to mark the affected area as a diagnosis since pneumonia can be seen as hazy areas of varying shape and size and cannot be differentiated from the normal tissue. The U-Net is the standard network used for this type of task and recent studies have bolstered the network for CXR. A variant of the Convolutional based block focus element is used to enhance lung border accuracy in a U-Net [12] and a lightweight residual U-Net with attention and Atrous spatial pyramid pooling can achieve strong results with just a few million parameters [13]. Better segmentation for diseased lungs using a dual attention network with boundary-aware loss [14] and comparative studies demonstrate that encoder–decoder models are still the most reliable option for the screening of infectious-diseases on CXR [15].

While these advances have been made, there are still challenges to be addressed. The number of studies published which test their model with a small or single-source data set makes it difficult to judge the generalizability of the models to other clinical settings. Moreover, pneumonia detection and lung segmentation are normally studied as independent problems, although they are both needed in clinical practice. Another drawback is that there is no consistent evaluation protocol, with varying datasets, preprocessing and evaluation metrics used across the different studies, which makes it hard to compare each model. They limit the usefulness of the current methods and indicate the need for a more thorough and systematically assessed system.

Automated system should not only be able to indicate the presence of pneumonia but also specify the region of the lung that is affected by pneumonia from a diagnostic perspective. The diagnosis and segmented region can help to improve medical professional's comprehension of the model's prediction, while making their decision making confident with the help of computer-aided decision making. This integrated approach could increase the efficiency of workflow by helping radiologists with initial screening, while ensuring diagnostic accuracy.

This study is mainly contributed as follows:

1. An integrated deep learning framework comprised of a CNN for pneumonia identification and a U-Net architecture for lung segmentation was created to give accurate disease categorization and localization of the damaged lung regions from chest X-ray images.
2. Using a conventional experimental technique, a lightweight CNN model is created and compared with five well-known DL architectures, including ResNet-50, MobileNet, Inception-V3, Vision Transformer (ViT), and Swin Transformer.
3. The U-net architecture-based lung segmentation model is put into practice and assessed by contrasting its segmentation performance with that of the VGG U-net model utilizing dice coefficient, intersection over union (IoU), and pixel accuracy.

The structure of the paper is as follows. Section 2 presents previous work and the research gap is established. In section 3, materials and procedures are described. Section 4 discusses the proposed framework. Section 5 discusses the findings and outcomes, while Section 6 presents the study's conclusions.

2. Literature Review

Previous studies have demonstrated how well DL approaches work for automatically identifying pneumonia in CXR images. In these methods, ensemble learning and feature fusion have demonstrated the best results by leveraging the advantage of multiple deep neural networks.

In this work [16], an ensemble learning design was introduced that removes the irrelevant regions of images prior to the classification process, and sets the weights of multiple CNNs by employing a genetic algorithm. The studies demonstrated an F1-score of 97.45% and a performance of 97.23% on the public pneumonia chest X-ray dataset. Using the Pediatric CXR dataset, the study [17] , developed a feature fusion-based ensemble CNN for the detection of infection in children. They achieve 98.94% accuracy and

99.12% F1 score by employing many refined CNN models with two distinct feature selection techniques, namely Chi-Square and minimal Redundancy Maximum Relevance (mRMR), prior to feature fusion.

The classification of pneumonia by adding DenseNet169, MobileNetV2 and ViT to the Kaggle CXR dataset, is presented in this work [18]. A representation of the ensemble obtained an accuracy of 93.91% and F1 score of 93.88% which shows that the classification accuracy of ensemble model is enhanced by using the complementary feature representations. Recent studies have paid more attention to lightweight CNN architectures integrated with attention mechanisms due to their computational efficiency and suitability to resource-constrained clinical environments. In this study [19] a CBAM-augmented EfficientNetB2 framework that combines a lightweight EfficientNetB2 backbone with Convolutional Block Attention Module (CBAM), Squeeze-and-Excitation (SE) blocks, and multiscale feature fusion to enhance the feature representation of chest X-ray images. The model was tested on a dataset of COVID-19, Viral pneumonia and Normal chest X-ray images, and achieved an accuracy of 99.3% on the unseen data. The results show the capability of lightweight attention-based CNNs for highly accurate and computationally efficient automated lung disease diagnosis. A lightweight weighted-average ensemble of MobileNetV2 and NASNetMobile is described in study [20], were created to decrease the model's computational complexities without compromising the high performance of the classification. The accuracy obtained on CXR images was 98.63%.

In this study [21] proposed an ensemble modified transfer learning approach with deep CNNs for chest radiograph classification, and showed that ensemble learning models gave more robust predictions than individually fine-tuned models. To boost medical image classification accuracy, [22] introduced the Dynamic Multi-scale CNN (DM-CNN) by integrating multi-scale feature extraction with uncertainty quantification to quantify the prediction uncertainty and enhance the reliability of automated medical image classification. XAI has also been highlighted as a means to enhance the clinical validity of pneumonia detection systems. An Inception-V3 based CNN with the use of the Integrated Gradients technique [23] was able to get an accuracy of 97.23% and provide visual explanations for the image regions that are important for the classification. In this study [24], an enhanced image by using a modified CNN with Contrast Limited Adaptive Histogram equalization (CLAHE) preprocessing was used for the image classification and the accuracy, precision and recall were measured as 98.70%, 99.30% and 98.60%, respectively. This work [25], compared the performance of custom designed CNNs with transfer learning models on pediatric CXR images and reported that the max results achieved by fine-tuned model with the backbone of ImageNet was 99.48% compared to the max accuracy of 97.04% by the custom CNN architectures.

The features are represented using EfficientNetB0 and DenseNet121 along with the advanced attention mechanisms for improved detection of pneumonia from chest X-ray images. To enhance the detection accuracy of the object [26] proposed a deep convolutional neural network with multi-head self-attention, channel attention-based feature fusion, residual learning, attention-augmented feature enhancement, and dynamic attention pooling. The model was tested on chest X-ray dataset containing both normal and pneumonia images and the performance of the model is evaluated by the accuracy of 95.19%, precision of 98.38%, recall of 93.84%, F1 score of 96.06%, specificity of 97.43%, and AUC of 0.9564. Multiple attention mechanisms fused with light pre-trained CNN backbones can extract more features and improve the diagnostic performance of automated pneumonia detection with high computational efficiency, which were demonstrated in the results. Using a MobileNetV2 based Pneumonia Detection Framework (MPDF) to automate pneumonia detection from chest x-ray images [27]. It features the MobileNetV2, EfficientNetB0 and an attention-guided U-Net inspired model, which can boost feature extraction and classification. The public chest X-ray dataset available on Kaggle was used as the data and underwent data augmentation and normalization before being used for training and testing the model. Experimental results showed that the proposed method outperforms the baseline models and it has shown a good generalization ability. The study showed that the transfer learning with attention guided feature learning is an effective and computation-efficient solution for automated pneumonia diagnosis. In study [28] MobileNetV2 combined with Grad-CAM, showed that it was able to identify pneumonia efficiently and showed that it could increase the interpretation of predictive models by revealing areas that affect classification decisions. The comparative analysis of the different architectures VGG16, ResNet50 and Inception-V3 presented in [29] showed that ResNet50 architecture outperformed the other two for overall classification performance on

CXR images. Although the classification accuracy of the DL model has improved remarkably, as demonstrated in [30], the interpretation of the model must be analyzed along with the model's prediction results so that physicians can have more confidence in the diagnosis process using AI. DL algorithms exhibit potential for lung detection; there are still a number of challenges to overcome. To the best of our knowledge, very few studies have derived a single framework for combining both tasks, and most of the current research has focused on either lung segmentation or the classification of illnesses. However, in addition to the differences in datasets, preparation methods, and evaluation techniques, there are other factors that make it difficult to make a fair comparison. Hence, an effective framework that performs both tasks of lung segmentation and pneumonia diagnosis with accurate results and assesses the framework performance with a consistent experimental method is required.

3. Materials and Methods

The research design, dataset, pre-processing and baseline architectures, experimental set up and evaluation parameters have been discussed in this section.

3.1. Research Design

This study used an experimental design. It formulates pneumonia detection as a binary image classification problem. It formulates the problem of marking the lung area as a segmentation problem. The proposed CNN and U-Net are the primary models. We compare with five detection baselines and one segmentation baseline.

3.2. Dataset

In this study, a public CXR dataset with paired lung masks was used. The dataset has two classes: normal and pneumonia. The masks help the segmentation task. Class sizes are nearly balanced after the split. Figure 1 illustrates the class distribution of training and test sets.

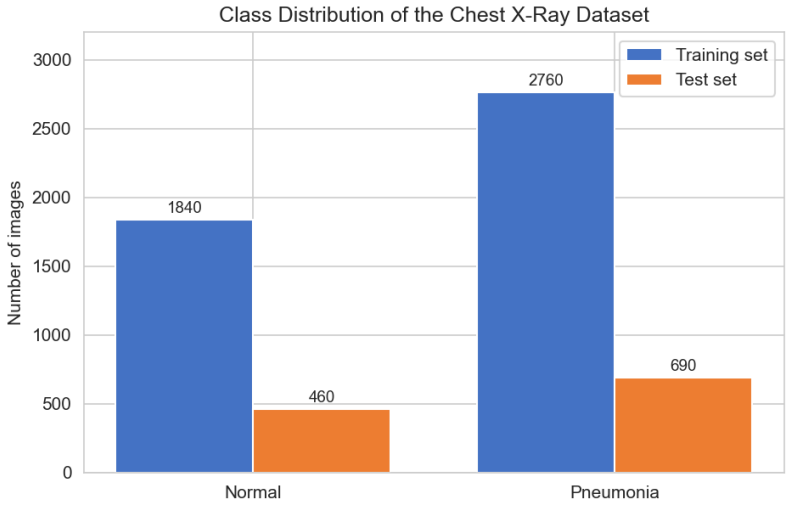

Figure 1. Class-wise distribution of the CXR Dataset

3.3. Data Pre-processing

Images are pre-processed for stable training. The source images raw scans weren't all the same size. To ensure consistency, all input photos were scaled to 224x224. Pixel values were divided by 255 to scale them into the interval [0, 1]. This step is to keep the input images scale in the same size. The learning set was enhanced with light. Augmentation included small rotations, shifts, zoom and horizontal flips. This increases the training set and reduces over fitting. Then the data are separated into 20% for testing and 80% for validation. The split resulted in 4,600 training images and 1,150 testing images.

3.4. Baseline Architectures

We selected five detection baselines that span common design families. ResNet-50 adds deep residual connections. MobileNet offers a light and efficient design. Inception-V3 uses multi-scale convolution. The Swin Transformer is a hierarchical attention model. The Vision Transformer is a plain self-attention model. The convolutional baselines used ImageNet weights as a starting point. For segmentation, the VGG U-Net is the baseline. It uses a VGG encoder inside the U-Net design. Table 1 lists the main training settings. All

models used a batch size of 32 and the same data split. The proposed CNN trained for 40 epochs with the Adam optimizer.

Table 1. Training settings of the baseline and proposed models

Model	Task	Optimizer	Epochs	Batch	Learning rate
ResNet-50	Detection	Adam	30	32	0.0001
MobileNet	Detection	Adam	30	32	0.0001
Inception-V3	Detection	Adam	30	32	0.0001
Swin Transformer	Detection	AdamW	50	32	0.0003
Vision Transformer	Detection	AdamW	50	32	0.0003
Proposed CNN	Detection	Adam	40	32	0.001
VGG U-Net	Segmentation	Adam	40	16	0.0001
Proposed U-Net	Segmentation	Adam	40	16	0.0001

3.5. Experimental Setup and Tools

The models were built with TensorFlow and Keras in Python. Experiments ran on a GPU-enabled environment. The Keras image generator loaded and labeled the X-ray images. The scikit-learn library handled the split and the evaluation reports. Matplotlib and Seaborn produced the training curves, confusion matrices, and ROC curves.

3.6. Evaluation metrics

The accuracy is the percentage of correct identification of photos. Precision is the percentage of true positives in the projected positives. The ratio of true positives to real positives is known as recall. The accuracy and recall harmonic mean is represented by the F1-score. True positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) are used in the metrics.

The overall classification correctness described in equation (1) as:

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

Equation (2) states that accuracy is the ratio of real positive cases for every instance that the classifier determined to be positive. By splitting the quantity of real achievements by the number of outcomes calculated by the classifier, precision may be calculated.

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

According to equation (3), recall is the proportion of true instances to all positive occurrences in the test set.

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

The F1 score, which provides a single evaluation of memory and accuracy, is defined by equation (4) as the balanced average of recall and precision.

$$F1\ score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (4)$$

We evaluated segmentation with two overlap metrics. The Dice coefficient calculates the gap between the estimated and the true mask. The shared area over the connected area is measured by the intersection over union, or IoU. Both have a range of 0 to 1, with greater being preferable. In this case, G is the reality on the field mask and P is the forecast mask, as defined in equation (5) and (6) as:

$$Dice = \frac{2|P \cap G|}{|P|+|G|} \quad (5)$$

$$IoU = \frac{|P \cap G|}{|P \cup G|} \quad (6)$$

4. Proposed Framework

The framework has two stages. The first stage detects pneumonia with a CNN. The second stage segments the lung region with a U-Net. Both stages share the same preprocessing. Figure 2 shows the proposed CNN used for detection.

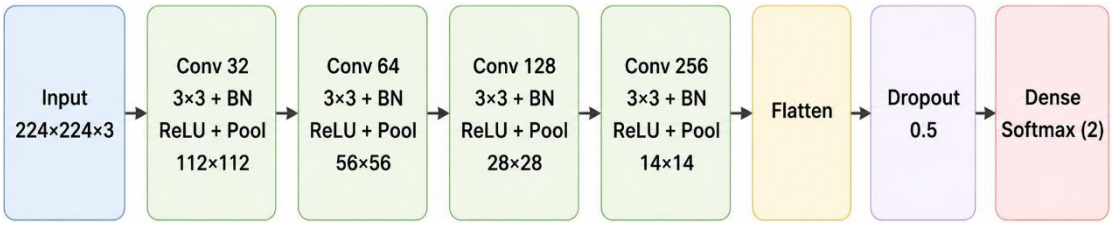

Figure 2. Architecture of the proposed CNN for pneumonia detection

4.1. Proposed CNN for Detection

The CNN captures an image of size 224×224×3. It uses four convolutional blocks for feature extraction. Each block has a convolution, equalization of batches max pooling, and a ReLU activation. As in Equation (7), the convolution applies learned filters throughout the picture. In this case, w is the filter, b is the bias, and x is the input.

$$z = (x \times w) + b \tag{7}$$

The ReLU activation adds non-linearity. It replaces negative values with zero, as in Equation (8). Max pooling then reduces the spatial size and keeps the strongest features. The four blocks use 32, 64, 128, and 256 filters, so the network learns features from simple to complex.

$$ReLU(z) = max(0, z) \tag{8}$$

After the blocks, a flatten layer forms one feature vector. A layer of dropouts with rate 0.5 reduces overfitting. A last layer of density with softmax gives the two class probabilities, as in Equation (9). Here g is the feature vector, and class c is normal or pneumonia.

$$p_c = \frac{e^{(Wg+b)_c}}{\sum(j=1)^C e^{(Wg+b)_j}} \tag{9}$$

The Adam Optimizer with Qualitative Cross-Entropy-Loss is used to train the network. Equation (10) defines the loss. In this case, p_(i,c) is the predicted frequency of image i in the class, and y_(i,c) is the real label. The detecting training loop is summed up in Algorithm 1.

$$L_{det} = -(1/N) \sum(i = 1)^N \sum(c = 1)^C y_{(i,c)log}\left(p_{(i,c)}\right) \tag{10}$$

The configuration of proposed CNN is given in table 2.

Algorithm 1. Training of the proposed CNN for detection.

Input : training set (X_train, y_train); epochs E = 40; batch size B = 32
Output: trained parameters theta

```
Resize images to 224x224 and scale pixels to [0, 1]
Apply augmentation (rotation, shift, zoom, flip)
Initialize theta; set Adam optimizer
for epoch = 1 to E do
    for each mini-batch (x, y) of size B do
        for block = 1 to 4 do
            x <- MaxPool( ReLU( BN( Conv(x) ) ) )
        g <- Dropout( Flatten(x), 0.5 )
        p <- softmax( Dense(g) )
        Update theta with Adam on cross-entropy loss
return theta
```

Table 2. Configuration of the proposed CNN

Parameter	Value
Input shape	224 × 224 × 3
Convolutional blocks	4 (Conv + BN + ReLU + MaxPool)
Filters per block	32, 64, 128, 256
Kernel size	3 × 3
Pooling	Max pooling 2 × 2
Dropout	0.5
Output layer	Dense, softmax (2 classes)

Optimizer	Adam
Loss	Categorical cross-entropy
Learning rate	0.001
Batch size	32
Epochs	40

4.2. Proposed U-Net for Segmentation

The U-Net marks the lung region on a scan. It has a coding machine path and a decoder path. The encoder uses convolution and down sampling to capture context. The decoder uses up sampling to restore the spatial size. Remove link matching encoder and decoder levels, which keeps fine detail. Figure 3 shows the suggested U-Net.

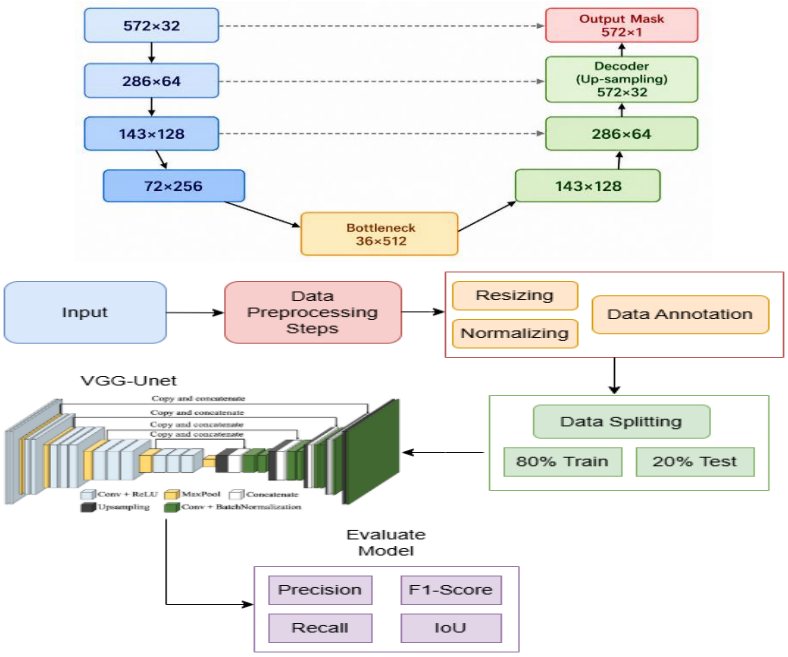

Figure 3. The proposed U-Net's architecture for lung segmentation

U-Net trains with a combined loss. The loss adds binary cross-entropy and a Dice loss, as in Equation (11). The Dice loss directly improves the overlap between masks. This combination handles class imbalance between lung and background pixels.

$$L\_seg = BCE(P, G) + (1 - Dice(P, G)) \quad (11)$$

4.3. Design Decisions

Each design choice has a clear reason. Four convolutional blocks give enough depth for CXR without heavy cost. Batch normalization speeds training and stabilizes the gradients. A dropout of 0.5 limits over-fitting on the limited data. The U-Net skip connections keep the lung borders sharp in the output mask.

5. Results and Discussion

This section depicts the experimental results with discussion of proposed algorithm.

5.1. Training and Testing protocols

All detection models training on a divide of the data 80% in training and 20% in test. They were tested on an unseen test set of 1150 images. There were total of 460 normal and 690 pneumonia images in the test set. We report the result, precision, recall, and F1-score. Additionally, we examined ROC curves, confusion matrices, and training curves.

5.2. Training Behavior

The suggested evaluation and training curves CNN is illustrated in figures 4 and 5. The training and validation curve of the suggested CNN is presented in figures 4 and 5, respectively. The accuracy improves rapidly, and then gradually approaches the level of 0.94. The loss looks like it is dropping and leveling off, indicating that it is not significantly over-fitting. The validation curves are closely following the training curves, suggesting good learning stability.

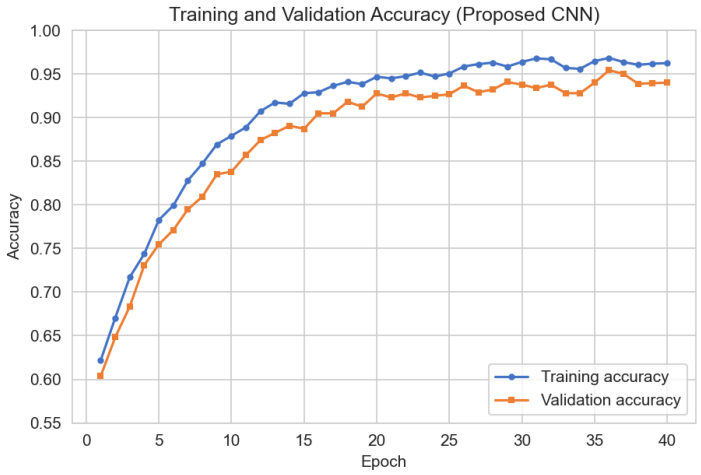

Figure 4. Accuracy graph of the Proposed CNN

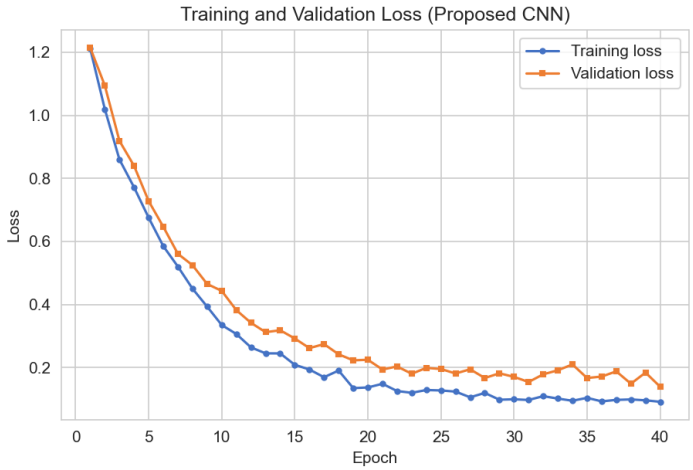

Figure 5. Loss graph of the suggested CNN

5.3. Detection performance of the Proposed Model

The results of the categorization by class, achieved by the proposed CNN model, are presented in Table 3. The model achieved accuracy of 0.926, a recall of 0.924, and an F1 score of 0.925 on the 460 test images, successfully identifying Normal class images that are healthy. The model achieved a remarkable performance with a high precision of 0.949, high recall of 0.951, and high F1 score of 0.950 when it comes to the detection of Pneumonia cases out of 690 test pictures for the Pneumonia class. The proposed model had an overall accuracy score of 94.00% for more than 1,150 test photos. While the weighted-average precision, recall, and F1-score were all 0.940, the macro-average precision, recall, and F1-score were 0.938, 0.937, and 0.937, respectively. Overall, the high level of agreement between the macro and weighted averages indicates that the proposed CNN model demonstrates promising classification accuracy for both classes, even with the low/high distribution difference, and is robust and has a balanced classification capability.

Table 3. Per-class detection performance of the proposed CNN

Class	Precision	Recall	F1-score	Support
Normal	0.926	0.924	0.925	460
Pneumonia	0.949	0.951	0.950	690
Accuracy			0.9400	1150
Macro average	0.938	0.937	0.937	1150
Weighted average	0.940	0.940	0.940	1150

The confusion matrix for the proposed CNN is given in Figure 6. For the 425 normal and 656 pneumonia images, the model was correct in its classifications. The total errors made were 69. It was used to scan 35 normal samples as pneumonia and failed to detect 34 pneumonia samples. The errors were kept equal between the two classes.

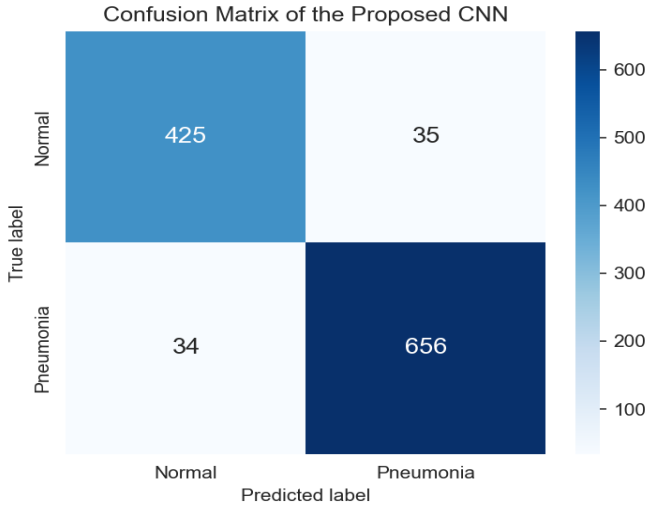

Figure 6. The proposed CNN confusion matrix

Figure 7 displays the proposed CNN's ROC curve. An area under the curve of 0.97 indicated good ranking quality. Figure 8 plots the precision, recall, and F1-score for each class. In every metric, the pneumonia class outperformed the normal class.

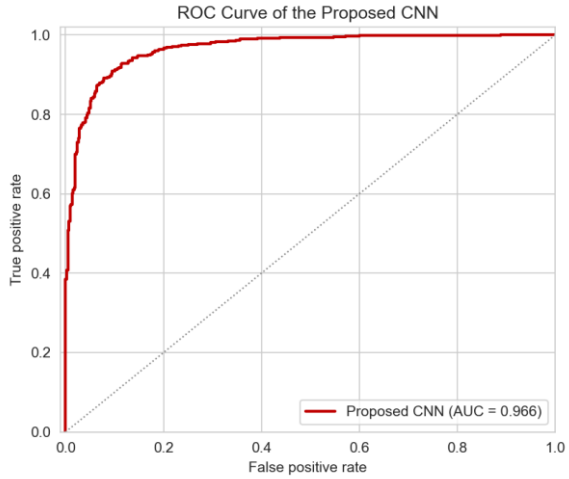

Figure 7. ROC curve of the proposed CNN

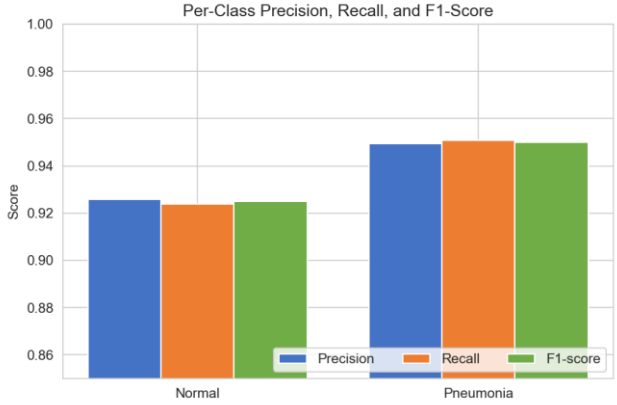

Figure 8. The proposed CNN's per-class accuracy, recall, and F1-score

5.4. Segmentation Performance

The validation Dice of the suggested U-Net throughout the training process is also illustrated in figure 9. The Dice starts to climb high and end up around 0.95. Table 5 displays the difference between proposed U-Net and VGG U-Net. The proposed U-Net obtained a Dice of 0.951 and an IoU of 0.907. It outperforms VGG U-Net on all metrics.

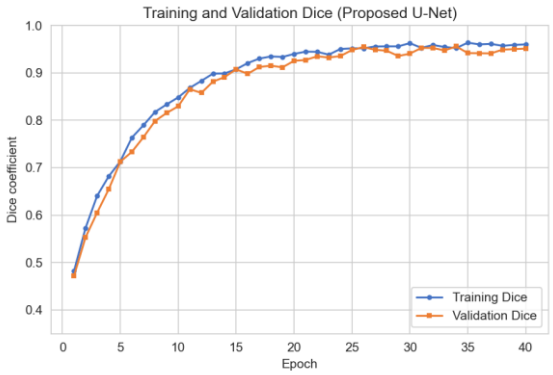

Figure 9. Training and validation Dice of the suggested U-Net

The segmentation results of the assessed models utilizing the Dice coefficient, Intersection over Union (IoU), and pixel accuracy is shown in Table 4. Strong segmentation performance was demonstrated by the VGG U-Net's Dice coefficient of 0.928, IoU of 0.867, and pixel accuracy of 0.961. Nevertheless, the suggested U-Net attained a Dice coefficient of 0.951, an IoU of 0.907, and a pixel accuracy of 0.974, outperforming the baseline model in every evaluation parameter, as displayed in figure 10.

Table 4. U-Net models segmentation result

Model	Dice	IoU	Pixel accuracy
VGG U-Net	0.928	0.867	0.961
Proposed U-Net	0.951	0.907	0.974

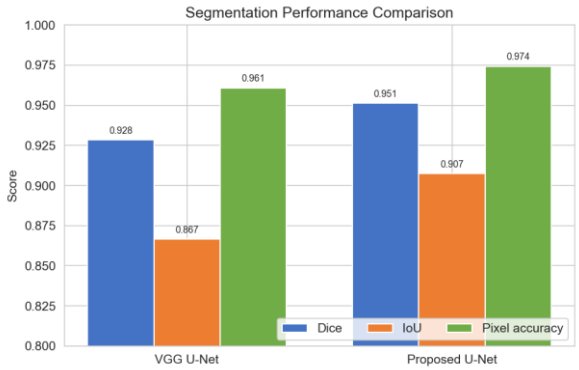

Figure 10. Segmentation metric comparison between the VGG U-Net and the suggested U-Net

A sample segmentation by the proposed U-Net is depicted in figure 11. The predicted mask shape closely matches the ground-truth mask of the lung region. There is a considerable overlap; only marginal differences at the borders. The high Dice and IoU scores are in line with this visual result.

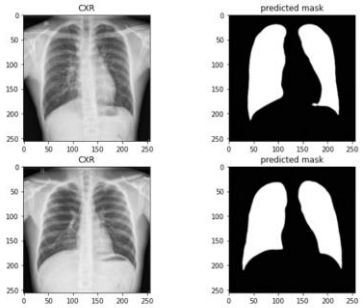

Figure 11. Illustrative lung segmentation by the proposed U-Net

5.5. Discussions

The results in a categorization of the assessed DL models for pneumonia identification is shown in Table 5 With results of 90.43%, accuracy of 0.905, recall of 0.904, and F1-score of 0.904, Inception-V3 outperformed all other baseline models. ResNet-50 improved the classification results with 91.20% accuracy, 0.912 precision, 0.911 recall, and 0.911 F1-score. MobileNet considerably enhanced performance, demonstrating the effectiveness of lightweight CNN architectures for pneumonia classification with an accuracy of 92.05%, precision of 0.921, recall of 0.920, and F1-score of 0.920. With a precision of 93.34%, precision of 0.934, recall of 0.933, and F1-score of 0.933, the Vision Transformer (ViT) produced slightly better results than the Swin Transformer, which attained an accuracy of 92.88%, precision of 0.929, recall of 0.928, and F1-score of 0.928. These results demonstrate that transformer topologies effectively capture global contextual information from CXR images, outperforming conventional CNN models in classification.

Figure 12 illustrates how the proposed CNN beat all evaluated models with an accuracy of 94.00%, precision of 0.940, recall of 0.940, and F1-score of 0.940. The consistent improvement across all assessment parameters shows that the proposed architecture is successful in learning discriminative characteristics for pneumonia diagnosis.

Table 5. Evaluation of the proposed and baseline models comparative performances

Model	Accuracy	Precision	Recall	F1-score
Inception-V3	90.43%	0.905	0.904	0.904
ResNet-50	91.20%	0.912	0.911	0.911
MobileNet	92.05%	0.921	0.920	0.920
Swin Transformer	92.88%	0.929	0.928	0.928
Vision Transformer	93.34%	0.934	0.933	0.933
Proposed CNN	94.00%	0.940	0.940	0.940

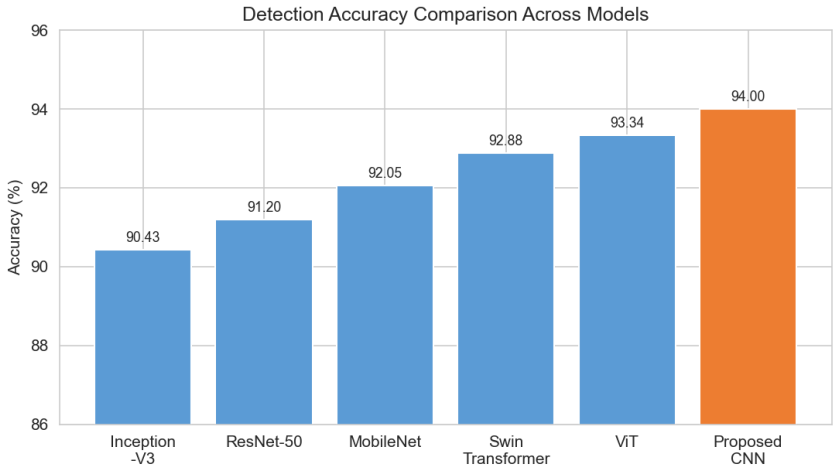

Figure 12. Detection accuracy comparison across all evaluated models

5.6. Limitations

There are serious problems with the study. The recommended method's applicability to other clinical settings and imaging systems may be limited because it was assessed using a single CXR dataset. The methodology was validated offline rather than in real-time clinical deployment, and the segmentation model only identifies lung areas without identifying lesion-specific boundaries.

6. Conclusion

In this study, authors proposed a pipeline for the recognition of pneumonia and lung segmentation in CXR. The compact CNN was used for the verification and U-Net for the division. Out of the six detection algorithms, the CNN model had the best accuracy 94.00%. U-Net achieved a Dice of 0.951, which is better than the VGG U-Net baseline. The key contribution is a single-protocol pipeline that integrates the detection and segmentation. The work provides per class metrics, confusion matrix, ROC analysis and

overlap scores. The findings indicate that a small CNN is able to perform as well as larger models on a small CXR dataset. The pipeline should be tested on multiple external datasets and cross-validated in future. It should validate the source of the data-set and look for label noise. Incorporating explain-ability techniques would help to make the model decisions more understandable to clinicians. More extensive and expansive data sets would enhance robustness of imaging sources. Future research will concentrate on verifying the suggested framework on larger, multi-center chest X-ray datasets to increase its resilience and applicability. Explainable AI approaches and lightweight deployment strategies will be investigated to improve model interpretability and enable real-time clinical implementation in computer-aided diagnosis systems.

References

1. A. A. Gelagay, T. Azale, L. D. Gezie, Z. Tigabu, and K. Alemu, "High proportion of pneumonia morbidity and risk factors in sick under-five children, northwest Ethiopia: a health facility based cross-sectional study," *Scientific Reports*, 2026.
2. C. Jayawardena, N. Wedasingha, N. Kolambage, and S. Perera, "Pneumonia Detection and Lung Disease Assessment from Chest X-rays: Developing A Diagnostic Support System," 2025.
3. H. Slimi, A. Balti, S. Abid, and M. Sayadi, "Trustworthy pneumonia detection in chest X-ray imaging through attention-guided deep learning," *Scientific Reports*, vol. 15, no. 1, p. 40029, 2025.
4. V. Tanwar, "Transforming Pneumonia Classification through Deep Learning: A CNN-Based Approach to Enhance Accuracy and Efficiency," in *2024 3rd International Conference for Advancement in Technology (ICONAT)*, 2024: IEEE, pp. 1-5.
5. V. Yadav, S. Shrivastava, and A. P. Singh, "Enhanced pneumonia detection using deep learning techniques on chest X-Rays," in *2024 2nd International Conference on Advances in Computation, Communication and Information Technology (ICAICCIT)*, 2024, vol. 1: IEEE, pp. 923-929.
6. O. Iparraguirre-Villanueva, W. Robles-Espiritu, M. Suxe-Ramírez, and R. O. Flores-Castañeda, "Use of CNN with Transfer Learning to Improve the Accuracy of Pneumonia Diagnosis," *International Journal of Engineering Trends and Technology*, vol. 73, no. 6, pp. 382-395, 2025.
7. S. Singh, M. Kumar, A. Kumar, B. K. Verma, K. Abhishek, and S. Selvarajan, "Efficient pneumonia detection using Vision Transformers on chest X-rays," *Scientific reports*, vol. 14, no. 1, p. 2487, 2024.
8. B. Mustapha, Y. Zhou, C. Shan, and Z. Xiao, "Enhanced pneumonia detection in chest X-rays using hybrid convolutional and vision transformer networks," *Current Medical Imaging*, vol. 21, no. 1, p. e15734056326685, 2025.
9. M. A. Ather, K. Sardar, Z. Fatima, J. L. O. Rodríguez, and C. G. Sánchez-Mejorada, "HybridCATNet: Explainable and Uncertainty-Aware Pneumonia Detection from Chest X-Rays Using a CNN–Transformer with Early Multi-Channel Fusion," *Frontiers in Digital Health*, vol. 8, p. 1769444.
10. M. A. Al-Ebrahim, "Spike-based attention mechanisms for enhanced medical image segmentation," *Engineering, Technology & Applied Science Research*, vol. 15, no. 5, pp. 28273-28285, 2025.
11. S. Potharaju, S. N. Tambe, K. Dasari, N. Srikanth, R. Venkatarao, and S. Tambe, "Enhanced X-ray image Classification for Pneumonia Detection Using Deep Learning Based CBAM and SE Mechanisms," *Intelligence-Based Medicine*, p. 100299, 2025.
12. M. A. L. Khaniki and M. Manthouri, "A novel approach to chest X-ray lung segmentation using U-net and modified convolutional block attention module," *arXiv preprint arXiv:2404.14322*, 2024.
13. M. Jannat *et al.*, "Lung segmentation with lightweight convolutional attention residual u-net," *Diagnostics*, vol. 15, no. 7, p. 854, 2025.
14. A. K. Dubey *et al.*, "Comparative Evaluation of Deep Learning Models for Lung Segmentation in Chest X-rays: Applications in Infectious Disease Screening," 2026.
15. M. N. Alam *et al.*, "Comparative evaluation of deep learning models for lung segmentation in chest X-rays: applications in infectious disease screening," *BMC Infectious Diseases*, vol. 26, no. 1, p. 922, 2026.
16. M. Kaya and Y. Cetin-Kaya, "A novel ensemble learning framework based on a genetic algorithm for the classification of pneumonia," *Engineering Applications of Artificial Intelligence*, vol. 133, p. 108494, 2024.
17. M. Kaya, "Feature fusion-based ensemble CNN learning optimization for automated detection of pediatric pneumonia," *Biomedical Signal Processing and Control*, vol. 87, p. 105472, 2024.
18. A. Mabrouk, R. P. Diaz Redondo, A. Dahou, M. Abd Elaziz, and M. Kayed, "Pneumonia detection on chest X-ray images using ensemble of deep convolutional neural networks," *Applied Sciences*, vol. 12, no. 13, p. 6448, 2022.
19. A. B. Godbin and S. G. Jasmine, "Advanced Lung Disease Detection: CBAM-Augmented, Lightweight EfficientNetB2 with Visual Insights," *Current Medical Imaging*, vol. 20, no. 1, p. e15734056344651, 2024.
20. O. Irshad, U. G. Khan, "Formalization and Semantic Integration of Heterogeneous Omics Annotations for Exploratory Searches," *Current Bioinformatics* 2020 p.1-17, vol 15 DOI: 10.2174/1574893615666200127122818.
21. S. Rajeashwari and K. Arunesh, "Enhancing pneumonia diagnosis with ensemble-modified classifier and transfer learning in deep-CNN based classification of chest radiographs," *Biomedical Signal Processing and Control*, vol. 93, p. 106130, 2024.
22. Q. Han *et al.*, "DM-CNN: Dynamic Multi-scale Convolutional Neural Network with uncertainty quantification for medical image classification," *Computers in biology and medicine*, vol. 168, p. 107758, 2024.

23. J. Rabbah, M. Ridouani, and L. Hassouni, "Improving pneumonia diagnosis with high-accuracy CNN-Based chest X-ray image classification and integrated gradient," *Biomedical Signal Processing and Control,* vol. 101, p. 107239, 2025.
24. A. Shavkatovich Buriboev, A. Abduvaitov, and H. S. Jeon, "Binary classification of pneumonia in chest X-ray images using modified contrast-limited adaptive histogram equalization algorithm," *Sensors,* vol. 25, no. 13, p. 3976, 2025.
25. O. Irshad, U. Ghani. A Comparative Analysis of Biological Data Integration Systems Famous for Data Exploitation and Knowledge Discovery. Current Bioinformatics 2021; Vol 16(5): pp. 662 - 681. DOI: 10.2174/1574893615999210101125442.
26. Q. An, W. Chen, and W. Shao, "A deep convolutional neural network for pneumonia detection in x-ray images with attention ensemble," *Diagnostics,* vol. 14, no. 4, p. 390, 2024.
27. T. Geetha, M. Suganya, B. Jamalpur, S. S. Vidhya, and R. Nanmaran, "Pneumonia Segmentation in Chest X-Ray Images using MobileNetV2-based Pneumonia Detection Framework (MPDF)," in *2026 International Conference on Computer Networks and Inventive Communication Technologies (ICCNCT)*, 2026: IEEE, pp. 1071-1078.
28. K. T. Reddy, "Explainable deep learning for automated pneumonia detection from chest X-ray images using MobileNetV2 and Grad-CAM CNN," *Authorea Preprints,* 2025.
29. T. Dalvi, M. Patley, V. Bhatt, K. Kadam, S. Gite, and B. Pradhan, "Pneumonia Detection Using Chest X-Rays: A Comparative Study of CNN Architectures and Explainable AI Techniques," in *Advancements in IoT Sensors and Security: Harnessing AI/ML for Secured IoT Sensor Data*: Springer, 2026, pp. 219-245.
30. J. Colin and N. Surantha, "Interpretable deep learning for pneumonia detection using chest x-ray images," *Information,* vol. 16, no. 1, p. 53, 2025.