Journal of
Image Processing & Pattern Recognition Progress

ISSN: 2394-1995
Volume 12, Issue 3, 2025
September–December
DOI (Journal): 10.37591/JoIPPRP

https://journals.stmjournals.com/joipprp

Review JoIPPRP

Stacked Generalization-Based Deep Learning Approach for Pneumonia Detection

Jyoti Dabass[1,*], Bhupender Singh Dabass[2]

Abstract

The proposed work focuses on a stacked generalization-based approach for diagnosing pneumonia from chest X-ray images. It utilizes regularization, early stopping, and data augmentation to deal with overfitting. It uses safe level SMOTE to deal with class imbalance and attention-based feature fusion to adaptively weigh features based on their importance. It uses two publicly available datasets (RSNA and Kermany) with ground truth provided by expert radiologists. The proposed work used ChexNet, SqueezeNet, and EfficientNet-B0, as ChexNet has already been fine-tuned on chest X-ray images, SqueezeNet uses its parameters efficiently, and EfficientNet-B0 balances accuracy and efficiency with limited parameters. We experimented with different regularization methods and optimizers to identify the most effective combinations of hyperparameters for our model training. Ensembling these basic classifiers adds diversity and makes it more generalizable for adaptation to unseen data with limited training parameters. Experimental results show that the proposed work is better compared to other state-of-the-art techniques, providing 99.4% accuracy on the Kermany dataset and 98.5% accuracy on the RSNA dataset. It shall be useful to the radiologists in improving the efficiency of pneumonia detection and thus will be beneficial in saving lives.

Keywords: Pneumonia detection, stacked generalization-based approach, chest X-ray images, RSNA and Kermany, SMOTE

INTRODUCTION

Pneumonia affects over four million people worldwide and can be dangerous if left untreated [1]. Early diagnosis helps in timely treatment, and a Chest X-ray is mostly preferred for early diagnosis due to its accessibility and low cost [2]. But detecting pneumonia from chest X-rays can be time-consuming and error-prone in resource-constrained scenarios due to the unavailability of expert doctors [3].

This problem can be mitigated to some extent with the help of a stacked generalization-based approach, which inhibits the strength of multiple deep learning models and provides better accuracy [4]. Motivated by this, the proposed work combines the strengths of SqueezeNet, EfficientNet-B0, and ChexNet by fusing their features using an attention mechanism to detect pneumonia with high accuracy. It provides reliability, diversity, and generalizability by integrating the strength of fine-tuned pretrained models and shall be beneficial for radiologists by providing them a second opinion in pneumonia diagnosis and thus saving lives in limited resource areas [5].

***Author for Correspondence**
Jyoti Dabass
E-mail: jyotidabas91@gmail.com

[1]Postdoctoral Research Associate, Department of Electronics & Electrical Communication, Yardi School of Artificial Intelligence, Indian Institute of Technology Delhi, Hauz Khas, New Delhi, India
[2]Student, Department of Law, Institute of Law and Research, Faridabad, Haryana, India

Received Date: June 25, 2025
Accepted Date: July 16, 2025
Published Date: September 19, 2025

Citation: Jyoti Dabass, Bhupender Singh Dabass. Stacked Generalization-Based Deep Learning Approach for Pneumonia Detection. Journal of Image Processing & Pattern Recognition Progress. 2025; 12(3): 20–31p.

The major highlights of the proposed methodology are as follows: (1) Using a stacked generalization-based approach and ensembling finetuned models having limited training parameters by fusing their features using an attention mechanism to provide reliability and diversity in pneumonia detection. (2) Hyperparameter selection, early stopping, regularization, data augmentation, and safe level SMOTE (Synthetic Minority Over-sampling Technique) techniques to deal with overfitting and class imbalance issues in publicly available datasets (Kermany and RSNA) with ground truth provided by expert radiologists. (3) It provides better results, i.e., 99.4% on Kermany and 98.5% accuracy on the RSNA dataset with a high degree of specificity and sensitivity. It shall be useful to radiologists in timely pneumonia detection by automating the diagnosis process with limited training time and parameters (i.e., 12.2 M) [6].

LITERATURE REVIEW

Recent studies have made significant progress in developing CNN-based algorithms for detecting pneumonia from chest X-ray images. A deep learning-based framework proposed by Asnake *et al*. utilized YOLOv3 detection, threshold segmentation, and SVM to achieve an F1-score of 99% [7]. Das *et al*. presented a deep Convolutional Neural Network-based solution that achieved a classification accuracy of 91.62% [8]. COVIDXrayNet, a CNN model proposed by Monshi *et al*. diagnosed COVID-19 with 95.82% accuracy, outperforming others through data augmentation and hyperparameter tuning [9]. Usman *et al*. trained a deep learning-based algorithm on 26,684 chest X-ray images, which led to a recall of 0.73, a precision of 0.76, an accuracy of 0.79, and an F1-score of 0.74, thus providing prompt diagnosis [10]. Aljawarneh *et al*. compared the performance of deep learning models in diagnosing pneumonia, where the enhanced CNN model performed better by providing the highest accuracy of 92.4% [11]. Multiple CNN models were ensembled by fusing their features optimally by Kaya *et al*. [12]. It addressed the challenges of image noise and limited labeled data, resulting in an accuracy of 98.94% and an F1-score of 99.12%. Farhan *et al*. combined a convolutional Neural Network and machine learning classifiers to build a framework named Multimodal Classification of Lung Disease and Severity Grading (MCLSG) [13]. It improved pneumonia diagnosis by 3.5% and severity level analysis by 6.8%. A computer-aided diagnosis framework based on a weighted average ensemble proposed by Kundu *et al*. achieved 86.85 and 98.91% accuracy on two datasets and showed its robustness in detecting pneumonia [14]. Bhatt *et al*. presented a Convolutional Neural Network (CNN) with a weighted ensemble method that provided an F1-score of 88.56% and a high recall of 99.23% [15]. A Dilated Convolution and residual structure-based deep learning framework proposed by Liang *et al*. achieved a recall rate of 96.7% in detecting childhood pneumonia [16]. Similarly, transfer learning-based models helped in improving pneumonia detection. The VGG-16 model-based system proposed by Soares *et al*. attained an accuracy of 97.3%, with 100% accuracy in categorizing COVID-19 cases and 88.2% accuracy in classifying normal lungs with other infections [17]. The simplified VGG-based method proposed by Zhang *et al*. enhanced the contrast of chest X-ray images to attain an accuracy of 96.07% in diagnosing pneumonia [18]. A customized VGG19-based Deep-learning system proposed by Dey *et al*. provided a significantly high accuracy of 97.94% in examining chest radiographs [19]. A neural network and a VGG-16-based network proposed by Sharma *et al*. achieved an accuracy of 95.4 and 92.15% for the two datasets [20]. Chen *et al*. presented a deep convolutional neural network-based method that combined an attention mechanism-based DenseNet 121 and EfficientNetB0 to achieve an accuracy of 95.19% [21]. Nettur *et al*. proposed a combination of NASNet and MobileNetV2, which provided an accuracy of 98.63% [22]. COVIDX-Net presented by Hemdan *et al*. achieved good performance with VGG19 and DenseNet models [23]. An unsupervised learning approach in a federated setting proposed by Rana *et al*. performed better than transfer learning, achieving an accuracy of 94% without compromising patient privacy [24].

In contrast, transformer-based models have also shown better performance than CNN models in detecting pneumonia. A cutting-edge pneumonia detection method using Vision Transformer (ViT) architecture proposed by Singh *et al*. achieved high accuracy, with 97.61% accuracy, 95% sensitivity, and 98% specificity on chest X-rays [25]. A BERT and ResNet-50-based framework proposed by Xu *et al*. built a multimodal transformer, which provided an F1 score of 93%, accuracy of 94%, and recall of 95% [26]. Also, CNNs with modified Swin Transformer blocks proposed by Mustapha *et al*. attained an accuracy of 98.72% in detecting pneumonia [27]. From the literature review, it can be deduced that most of the proposed work is based on CNN models, transfer learning, and ensemble methods using models having high training parameters, and very little work has been done on utilizing transformer-based models in detecting pneumonia. CNN-based models may lack generalizability if not trained properly, and ensembling models with large training parameters increase computational complexity. Although these models achieve good results, there remains a need for further research to develop more robust, lightweight, reliable, accurate, transparent, and deployable diagnostic-aid solutions that can improve generalizability and can be clinically integrated.

DATASETS USED

For the proposed work, we have used two publicly available datasets:

1. *RSNA Dataset:* This dataset comprises of 30,227 DICOM-format X-ray images, each with a resolution of 1024×1024 pixels, courtesy of the National Institutes of Health (NIH) and annotated by the Radiological Society of North America in association with the Society for Thoracic Radiology and MD.ai [28, 29].
2. *Kermany Dataset:* This dataset contains 5,856 pediatric chest X-ray images, primarily taken from children aged 1 to 5 years at Guangzhou Women's and Children's Medical Centre. After quality control, the dataset was narrowed down to include 1,583 images labeled as normal and 4,273 images annotated as pneumonia [30].

For each of these datasets, we have used the ratio for the training, validation, and testing process as 60, 20, and 20%, respectively. To address the challenges of overfitting that normally occur due to ensemble depth or dataset biases, the proposed work utilizes regularization, early stopping, and data augmentation techniques, including rotation, horizontal flipping, shearing, and scaling. It is observed that Safe Level SMOTE (Safe Level Synthetic Minority Oversampling Technique) is better than Borderline SMOTE, SMOTE, and ADASYN in dealing with class imbalances [31]. The datasets used in the work are imbalanced, so we used Safe-level SMOTE to deal with the class imbalance issue. It is a variant of the SMOTE technique used for oversampling the minority class by sampling the minority class in line with different weight degrees called safe levels. Traditional methods synthesize the minority instances by choosing their k-nearest neighbor, whereas Safe-level SMOTE achieves better performance by synthesizing more samples around the safe level.

METHODOLOGY

We have used two publicly available annotated datasets, namely RSNA and Kermany, for our proposed work. We have shown the sample normal and pneumonia images taken from the Kermany dataset in Figure 1.

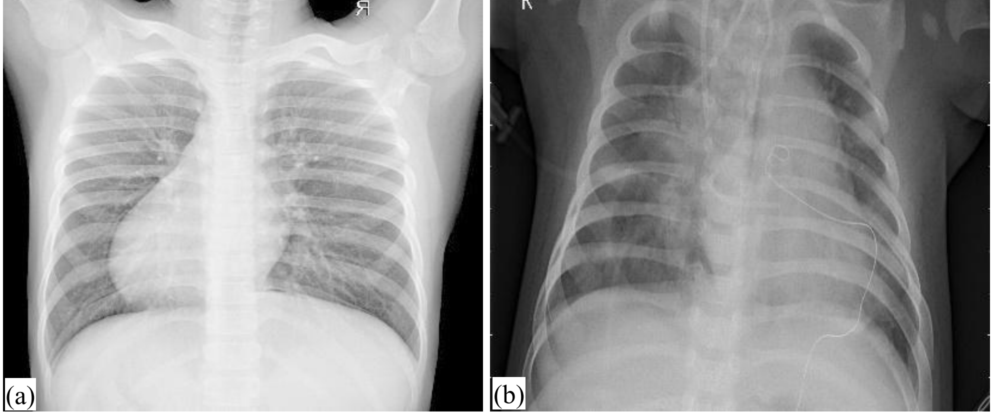


Figure 1. (a) Normal Image. (b) Pneumonia image.

The methodology of the proposed work uses the following steps.

1. *Preprocessing Data:* The chest X-ray images of the Kermany and RSNA datasets are pre-processed to remove noise and to improve the quality, followed by resizing them to the fixed size of 224×224×3 pixels for reducing memory requirements and optimizing training efficiency.
2. *Splitting Dataset:* The pre-processed images of both datasets are split using a random splitting strategy into training, testing, and validation ratios of 60, 20, and 20%, respectively.
3. *Stacked Generalization Approach:* A stacked generalization approach helps in improving accuracy. Motivated by this, we used an Attention-Based Feature Fusion layer to adaptively weigh the features of finetuned models, including ChexNet, SqueezeNet, and EfficientNet-B0, in their feature fusion.
4. *Transfer Learning:* All three pretrained models are fine-tuned on training datasets to adapt to the task of pneumonia detection.
5. *Model Training and Evaluation:* The proposed model, based on stacked generalization and an ABEF layer to fuse features of CNN models, is trained on both datasets. For evaluating the performance on the testing datasets, we have used metrics like AUC, recall, F1-score, and accuracy as discussed by Usman *et al*. [10].

Stacked Generalization Architecture

The stacked generalization architecture comprises the following mechanisms:

Base Models

Three fine-tuned pre-trained CNNs (e.g., ChexNet, SqueezeNet, EfficientNet-B0) are applied as base models.

ChexNet

ChexNet is a 121-layer DenseNet model fine-tuned over more than 100,000 chest X-ray images for predicting pneumonia [32]. Using this model will reduce the training cost as the network will converge earlier on Chest X-ray images, for it has been previously trained on large X-ray images. Figure 2 demonstrates the parameters for the ChexNet model.

SqueezeNet

It is a CNN architecture that utilizes fire modules comprising two layers, namely an expanding layer and a squeeze layer. It provides comparable accuracy to AlexNet on the ImageNet dataset with 50x fewer parameters. The expanding layer utilizes a mix of 3×3 and 1×1 convolution filters, while the squeeze layer employs 1×1 convolution filters. The squeeze layer lessens the number of parameters by a factor of 9. SqueezeNet provides the highest accuracy density, which shows its efficient use of parameters [33]. It is an appropriate choice for our model as it contains comprehensive and well-structured data that supports accurate training and evaluation.

EfficientNet-B0

It is a CNN method that uses the compound scaling method to scale dimensions, resolution, and width of the network with a fixed ratio.

```
global_average_pooling2d (Glob  (None, 1024)      0          ['relu[0][0]']
alAveragePooling2D)

dropout (Dropout)              (None, 1024)      0          ['global_average_pooling2d[0][0]'
                                                            ]

dense_1 (Dense)                (None, 3)         3075       ['dropout[0][0]']

==========================================================================================
Total params: 7,040,579
Trainable params: 6,956,931
Non-trainable params: 83,648
```

Figure 2. ChexNet model parameters.


```
 global_average_pooling2d (Glob  (None, 1280)        0           ['top_activation[0][0]']
 alAveragePooling2D)

 dropout (Dropout)              (None, 1280)        0           ['global_average_pooling2d[0][0]'
                                                                 ]

 dense (Dense)                  (None, 3)           3843        ['dropout[0][0]']

==================================================================================================
Total params: 4,053,414
Trainable params: 3,969,375
Non-trainable params: 84,039
```

Figure 3. EfficientNet-B0 parameters.

It has achieved a remarkable 93.3% accuracy with 0.39 billion floating-point operations (FLOPs) and 5.3 million parameters in the ILSVRC challenge, outperforming DenseNet-169 and ResNet-50, which require 14 M and 26M parameters, respectively [34].

As the input resolution surges, it is important to scale up the feature channels to deepen the network to expand the receptive fields, and capture finer details. The compound scaling method attains this by proportionately scaling all dimensions of the network. For our proposed work, we imported the EfficientNet-B0 model from TensorFlow Keras, which comprises around 4 million trainable parameters. Figure 3 displays the parameters of the EfficientNet-B0 model.

We selected ChexNet as it has already been fine-tuned over more than 100,000 chest X-ray images [32]. SqueezeNet uses its parameters proficiently as equated to other CNN models [33]. EfficientNet-B0 provides both efficiency and accuracy in image recognition tasks with limited training parameters. As compared to individual models, ensemble models provide better generalized predictions on unseen datasets [35]. The proposed work, having 12.2 M network parameters, integrates these models to add diversity and reliability.

Attention-based feature Fusion (ABFF) Layer and Output Layer

The Softmax function helps in computing the importance of each feature. The proposed work weighs features based on their importance using an attention mechanism. The outputs of the base models are passed through an ABEF layer, which gives combined features, which are then passed through the output layer having a softmax activation function to provide the probability distribution over two categories, i.e., normal and pneumonia, thus giving a single output.

The proposed work uses a stochastic descent algorithm for optimization and a binary cross-entropy loss function along with libraries with OpenCV, TensorFlow, Pandas, and Numpy on the Jarvis RTX5000 device, having a 16 GB GPU, 7 vCPUs, 32 GB RAM, and 20 GB SSD. It is trained end-to-end, and its performance is evaluated on testing sets of both datasets and compared with current state-of-the-art methods.

Hyperparameter Selection to Optimize Model Performance

Hyperparameter tuning indicates the process of getting the optimal values for hyperparameters that include batch size, regularization method, input size, number of training epochs, output size, loss function, learning rate, activation function, and optimizer choice.

Hyperparameter tuning helps in improving the performance of CNN models [36]. We tried different optimizers and regularization methods to choose the best combinations for our model training, as shown in Table 1. Using the results of Table 1, we selected the hyperparameters for our proposed work as shown in Table 2.

Table 1. Initial model evaluation to select hyperparameters.

Classifier	Regularization method	Optimizer	Accuracy (validation)	Validation loss
EfficientNet B0	Batch Normalization	Adam	96.36%	0.1161
EfficientNet B0	Dropout (0.4)	Adam	96.65%	0.1075
EfficientNet B0	Batch Normalization	SGD	92.32%	0.2107
SqueezeNet	Batch Normalization	Adam	94.05%	0.2069
SqueezeNet	Dropout (0.3)	Adam	92.18%	0.2013
SqueezeNet	Batch Normalization	SGD	88.58%	0.3050
ChexNet	Batch Normalization	Adam	89.07%	0.3253
ChexNet	Dropout (0.4)	Adam	96.56%	0.0982
ChexNet	Batch Normalization	SGD	87.11%	0.4393

Table 2. Hyperparameters for training the model.

Parameter	Value
Optimizer	Adam
Batch Size	32
Final Layer Activation	SoftMax Activation
Epochs	60–100 (based on model learning rate)
Patience	10/15
Learning Rate	1e-3 to 1e-5
Regularization Method	Dropout

RESULTS

The results obtained with the proposed methodology are compared with state-of-the-art methods in Tables 3 and 4.

Kermany Dataset

The proposed work categorizes most of the pneumonia cases correctly with low number of false positives (15) and false negatives (12) and achieves a highest accuracy of 99.4%, precision of 99.1%, recall of 99.3%, and F1-score of 99.2%, which is better compared to Ensemble ResNet18, GoogleNet, and DenseNet121 [14] with 98.81% accuracy, 0.983 AUC, 98.80% recall, 98.79% F1-score and 98.82% precision, and KNN-wavelet-GLCM with 99.0% accuracy, 0.990 AUC, 98.4% F1-score, 99.0% recall and 98.3% precision on the Kermany dataset, as shown in Table 3 [37].

RSNA Dataset

On the RSNA dataset, the proposed work categorizes normal and pneumonia cases correctly with a low number of false positives (181) and false negatives (241) and attains an accuracy of 98.50%, a recall of 99.5%, a precision of 99.6% and an F1-score of 99.6% which is better compared to KNN-wavelet-GLCM with 97.0% accuracy, 0.970 AUC, 97.6% recall, 97.5% precision, F1-score of 97.5% and InceptionNetV3, ResNet50, VGG19, and VGG-16 with highest accuracy of 88%, AUC of 0.918, recall of 88.0%, F1-score of 88% and precision of 91% achieved with VGG-19 [37]. Also, our proposed model is lightweight with 12.2 M training parameters compared to VGG-19 with 144 million parameters [38]. Overall, the proposed model provides excellent performance on both datasets, with high accuracy, recall, precision, and F1-score, demonstrating its ability to diagnose pneumonia from chest X-ray images as shown in Table 4.

Figure 4 demonstrates the confusion matrix for the Kermany dataset. As per the confusion matrix, we have 618 true negatives (TN), 1697 true positives (TP), 15 false positives (FP), and 12 false negatives (FN). We get Recall: TP/(TP+FN)=1697/(1697+12)=0.993, Precision: TP/(TP+FP)=1697/(1697+15)=0.991, F1-score: 2× (Precision × Recall)/(Precision + Recall) =0.992 and Accuracy: (TN+TP)/(TN+TP+FP+FN)=(618+1697)/(618+1697+15+12)=0.994. Also, we get a Low false negative rate (12/1709=0.7%) and a Low false positive rate (FPR=FP/FP+TN=15/633=15/15+618=2.4%), which shows that the model is accurate enough to detect pneumonia.

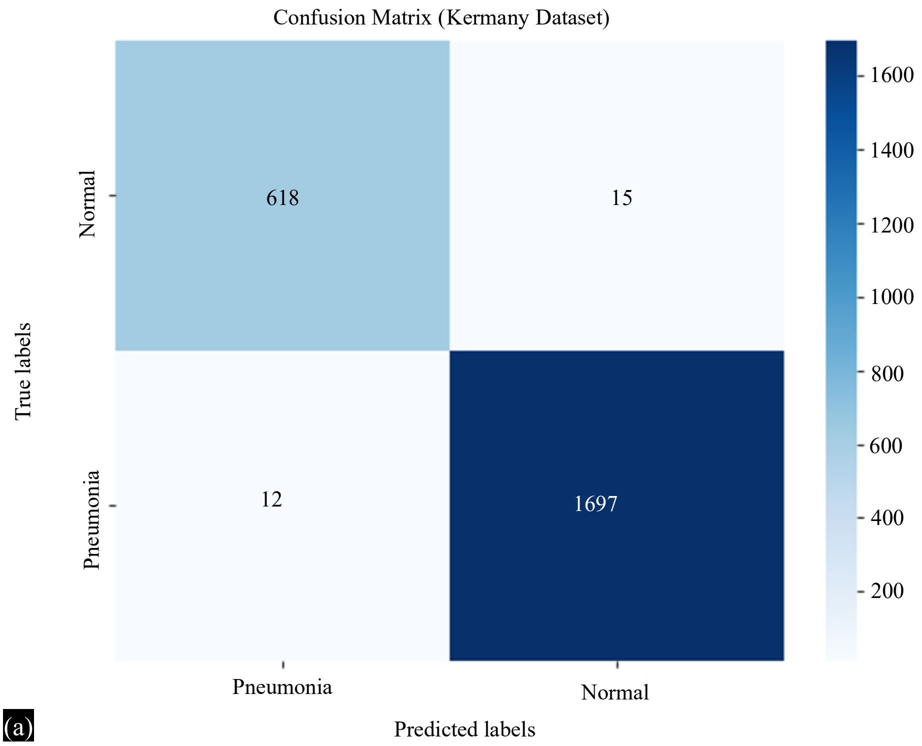
The ROC curve indicates that the model is correctly able to predict pneumonia as the true positive rate surges and the false positive rate declines when the threshold is increased. Also, a higher AUC implies better performance. Also, the proposed model achieves a **high specificity** (TN/(TN+FP) of 0.976 (i.e. 618/(618+15)=0.976) and **high sensitivity** (TP/(TP+FN)) or **recall** of 0.993 (1697/(1697+12)=0.993), which indicates that the model can predict most pneumonia images correctly while minimizing the number of false positives. **Figure 5** displays the confusion matrix and ROC curve for the RSNA dataset.

Table 3. Assessment of results on the Kermany dataset.

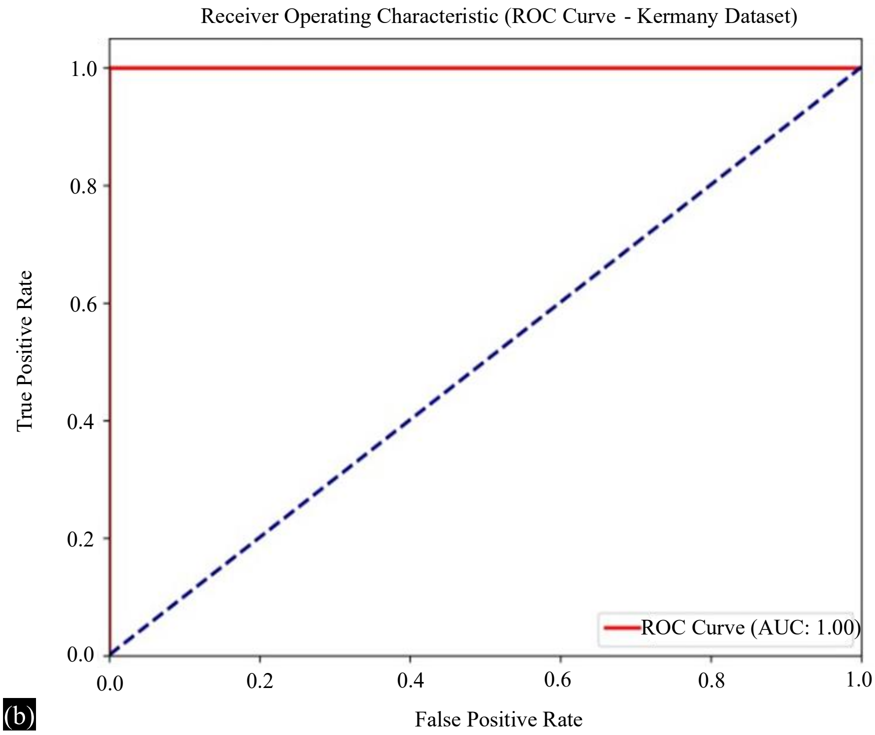
Authors and Year	Model	Precision	F1-score	Accuracy	AUC	Recall
Shati *et al*., 2025 [37]	KNN-wavelet-GLCM	98.3%	98.4%	99.0%	0.990	99.0%
Kundu *et al*.., 2021 [14]	Ensemble ResNet18, GoogleNet, DenseNet121	98.82%	98.79%	98.81%	0.983	98.80%
Proposed work	**Stack generalization-based approach**	**99.1%**	**99.2%**	**99.4%**	**-**	**99.3%**

Table 4. Assessment of results on the RSNA dataset.

Study	Model	Precision	F1-score	Accuracy	AUC	Recall
Shati *et al*., 2025 [37]	KNN-wavelet-GLCM	97.5%	97.5%	97.0%	0.970	97.6%
Study	Model	Precision	F1-score	Accuracy	AUC	Recall
Chiwariro *et al*., 2023 [38]	InceptionNetV3 ResNet50 VGG19 VGG16	76.0% 78.0% 86.0% 91.0%	75.0% 78.0% 82.0% 88.0%	79.0% 73.0% 80.0% 88.0%	0.870 0.809 .825 0.918	74.0% 78.0% 80.0% 88.0%
Proposed work	**Stack generalization-based approach**	**99.6%%**	**99.6%**	**98.50%**	-	**99.5%**



(a)



(b)

Figure 4. (a and b) Confusion matrix and ROC curve (Kermany dataset).

As shown in Figure 5, we have 49860 true positives (TP), 241 false negatives (FN), 181 false positives, and 184 true negatives (TN). Also we get Recall: TP/(TP+FN)=49860/(49860+241)=0.995, Accuracy: (TP+TN)/(TP+TN+FP+FN)=(49860+184)/(49860+184+181+241)=0.985, Precision: TP/(TP+FP)=49860/(49860+181)=0.996 and F1-score: 2×(Precision × Recall)/(Precision + Recall)=2×(0.996×0.995)/(0.996+0.995)=0.996. From the confusion matrix, it can be deduced that the proposed model is working well with a low number of false negatives (241) and false positives (181). From the ROC curve, it can be deduced that the model can attain a high true positive rate (TPR: TP/(TP+FN)=49860/(49860+241)=0.995) while keeping a low false positive rate (FPR: FP/(FP+TN)=181/(181+184)=0.496). A higher AUC implies better performance on the RSNA dataset. The proposed work achieves these results with 12.2 M training parameters and focuses on binary classification on publicly available datasets.

DISCUSSION

To achieve our goal of developing a robust, accurate, lightweight, reliable model with high generalizability, we have carefully analyzed the performance of various deep learning models and selected the ones that give comparatively better results without increasing computational complexity. We have utilized the best combination of optimizers and regularization methods for our training after comparing their results, which makes our method more reliable compared to existing literature. We have focused on improving the accuracy of pneumonia detection with high generalizability and diversity. Normally, algorithms work well on one dataset, but when applied to a different dataset, they perform badly.

Contrary to the existing works, we have tested our results on multiple datasets to check their robustness and reliability. With the proposed methodology, we can get comparatively better results than existing literature with 12.2 M training parameters, which makes our model suitable to be deployed in real real-life scenario. We have ensembled three lightweight pretrained models compared to existing literature that have utilized two models having large training parameters to give our model diversity. In the future, one can focus on proposing more lightweight models with attention map visualizations for better interpretability.

ISSN: 2394-1995

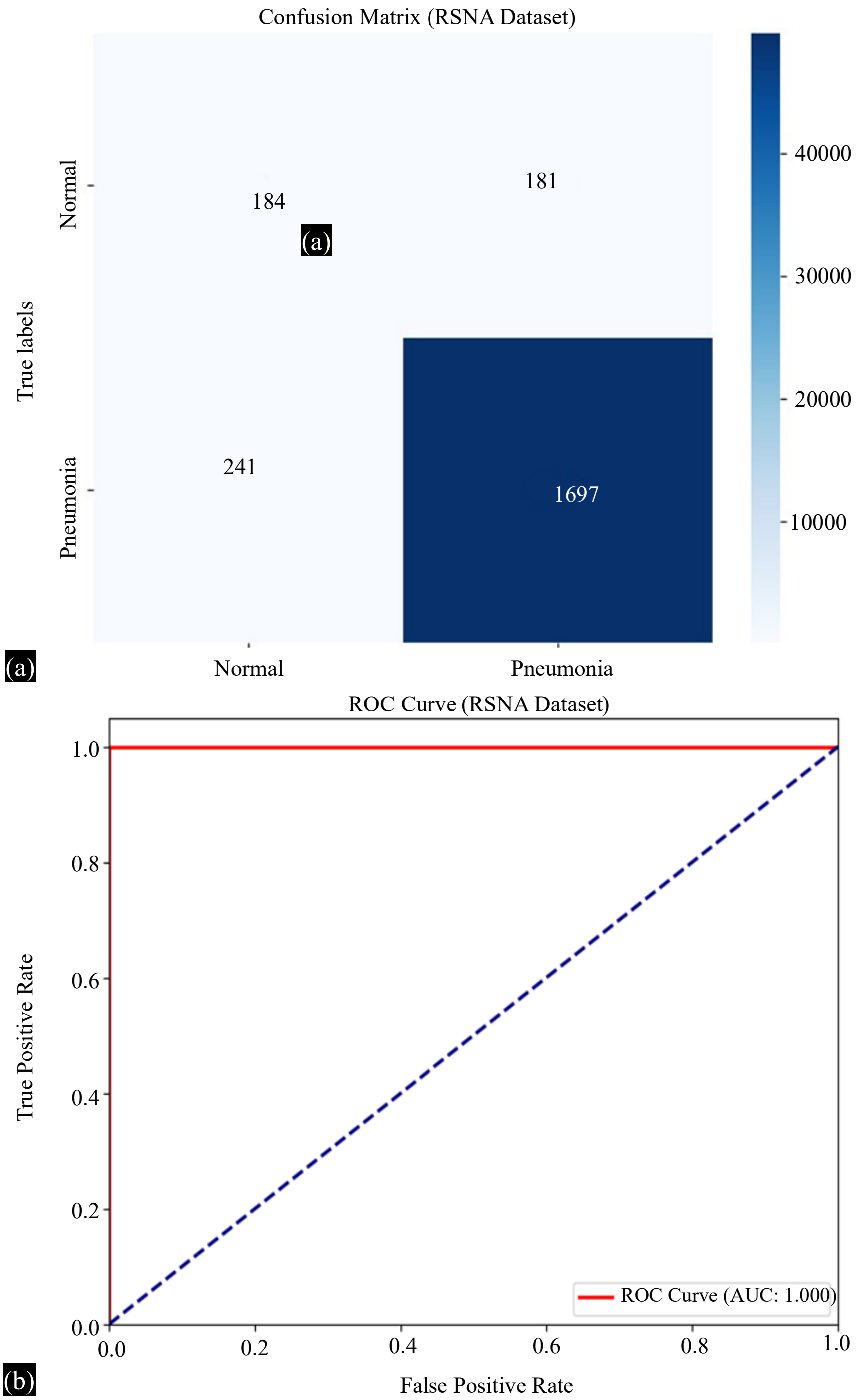


Figure 5. Confusion matrix and ROC curve for the RSNA dataset.

Our proposed work gave the best results on two-class classification with a focus on better sensitivity and specificity. In the future, this work can be extended to multi-class classification of images and classifying private datasets of pneumonia images.

CONCLUSION

The proposed work used a stack generalization approach with an ABFF layer to weigh features based on their importance to diagnose pneumonia from Chest X-ray images. It has 12.2 M training parameters that are fewer than other state-of-the-art methods, including VGG-16 (134.7M) and AlexNet (61M). It is less error-prone, for it avoids the segmentation of the region of interest followed by classification.

Ensembling of fine-tuned ChexNet, SqueezeNet, and EfficientNet aids in making it more generalizable for adaptation to testing data. Early stopping, data augmentation, safe level SMOTE, and regularization help in achieving high accuracy, avoiding overfitting and class imbalance issues. It achieves 99.4% accuracy on Kermany and 98.5% accuracy on the RSNA dataset, which is better compared to other state-of-the-art techniques. The proposed work has focused more on detecting pneumonia images correctly while reducing the number of false positives, which can be helpful to radiologists to detect pneumonia in resource-constrained settings. In the future, one can explore the applications of the proposed work in other medical imaging domains, including colored images. Also, the proposed work can be extended for multi-class classification on private datasets acquired from the hospitals, focusing on optimizing the model for deployment in resource-constrained settings. Currently, the model achieves high accuracy with 12.2 M network parameters. In the future, one can improve it further by reducing computational cost, resource requirements, and training time.

Conflict of Interest

The authors of this research work have no conflicts of interest to disclose.

Acknowledgment

The authors are grateful to the team of radiologists from RSNA, the Society of Thoracic Radiology, and the University of California, San Diego, who have annotated the datasets. Without their support, we could not have checked the feasibility of the proposed work. We have followed the guidelines provided by them to use the dataset and have ensured that the images are de-identified and anonymized.

REFERENCES

1. Rehman MU, Shafique A, Khan KH, Khalid S, Alotaibi AA, Althobaiti T, Ramzan N, Ahmad J, Shah SA, Abbasi QH. Novel privacy preserving non-invasive sensing-based diagnoses of pneumonia disease leveraging deep network model. Sensors. 2022 Jan 8;22(2):461.
2. Kareem A, Liu H, Sant P. Review on pneumonia image detection: A machine learning approach. Hum-Centric Intell Syst. 2022; 2(1): 31–43.
3. Harsono IW, Liawatimena S, Cenggoro TW. Lung nodule detection and classification from Thorax CT-scan using RetinaNet with transfer learning. J King Saud Univ-Comput Inf Sci. 2022; 34(3): 567–577.
4. Aydogdu M, Ozyilmaz E, Aksoy H, Gursel G, Ekim N. Mortality prediction in community-acquired pneumonia requiring mechanical ventilation; values of pneumonia and intensive care unit severity scores. Tuberk Toraks. 2010 Jan 1; 58(1): 25–34.
5. QH. Novel privacy-preserving non-invasive sensing-based diagnoses of pneumonia disease leveraging a deep network model. Sensors. 2022; 22(2): 461.
6. Colin J, Surantha N. Interpretable Deep Learning for Pneumonia Detection Using Chest X-Ray Images. Information. 2025; 16(1): 53.
7. Asnake NW, Salau AO, Ayalew AM. X-ray image-based pneumonia detection and classification using deep learning. Multimed Tools Appl. 2024 Jun; 83(21): 60789–807.
8. Das AK, Ghosh S, Thunder S, Dutta R, Agarwal S, Chakrabarti A. Automatic COVID-19 detection from X-ray images using ensemble learning with a convolutional neural network. Pattern Anal Appl. 2021; 24: 1111–1124.
9. Monshi MMA, Poon J, Chung V, Monshi FM. CovidXrayNet: Optimizing data augmentation and CNN hyperparameters for improved COVID-19 detection from CXR. Comput Biol Med. 2021; 133: 104375.
10. Usman C, Rehman SU, Ali A, Khan AM, Ahmad B. Pneumonia Disease Detection Using Chest X-Rays and Machine Learning. Algorithms. 2025; 18(2): 82.
11. Aljawarneh SA, Al-Quraan R. Pneumonia detection using enhanced convolutional neural network model on chest x-ray images. Big Data. 2025 Feb 1; 13(1): 16–29.
12. Kaya M. Feature fusion-based ensemble CNN learning optimization for automated detection of pediatric pneumonia. Biomed Signal Process Control. 2024; 87: 105472.

13. Farhan AM, Yang S, Al-Malahi AQ, Al-antari MA. MCLSG: Multi-modal classification of lung disease and severity grading framework using consolidated feature engineering mechanisms. Biomed Signal Process Control. 2023; 85: 104916.
14. Kundu R, Das R, Geem ZW, Han GT, Sarkar R. Pneumonia detection in chest X-ray images using an ensemble of deep learning models. PloS one. 2021; 16(9): e0256630.
15. Bhatt H, Shah M. A Convolutional Neural Network ensemble model for Pneumonia Detection using chest X-ray images. Healthcare Analytics. 2023 Nov 1; 3: 100176.
16. Liang G, Zheng L. A transfer learning method with the deep residual network for pediatric pneumonia diagnosis. Comput Methods Programs Biomed. 2020; 187: 104964.
17. Soares LP, Soares CP. Automatic detection of COVID-19 cases on X-ray images using convolutional neural networks. arXiv preprint arXiv:2007.05494. 2020.
18. Zhang D, Ren F, Li Y, Na L, Ma Y. Pneumonia detection from chest X-ray images based on a convolutional neural network. Electronics. 2021; 10(13): 1512.
19. Dey N, Zhang YD, Rajinikanth V, Pugalenthi R, Raja NSM. Customized VGG19 architecture for pneumonia detection in chest X-rays. Pattern Recognit Lett. 2021; 143: 67–74.
20. Sharma S, Guleria K. A deep learning-based model for the detection of pneumonia from chest X-ray images using VGG-16 and neural networks. Procedia Comput Sci. 2023; 218: 357–366.
21. An Q, Chen W, Shao W. A deep convolutional neural network for pneumonia detection in X-ray images with attention ensemble. Diagnostics. 2024 Feb 11; 14(4): 390.
22. Nettur SB, Karpurapu S, Nettur U, Gajja LS, Myneni S, Dusi A, Posham L. Lightweight Weighted Average Ensemble Model for Pneumonia Detection in Chest X-Ray Images. arXiv preprint arXiv:2501.16249. 2025.
23. Hemdan EED, Shouman MA, Karar ME. Covidx-net: A framework of deep learning classifiers to diagnose COVID- 19 in X-ray images. arXiv preprint arXiv:2003.11055. 2020.
24. Rana N, Marwaha H. Pneumonia detection from X-ray images using federated learning–An unsupervised learning approach. Meas: Sens. 2025; 37: 101410.
25. Singh S, Kumar M, Kumar A, Verma BK, Abhishek K, Selvarajan S. Efficient pneumonia detection using Vision Transformers on chest X-rays. Sci Rep. 2024; 14(1): 2487.
26. Xu J, Wang Y. FMT: A Multimodal Pneumonia Detection Model Based on Stacking MOE Framework. arXiv preprint arXiv:2503.05626. 2025.
27. Mustapha B, Zhou Y, Shan C, Xiao Z. Enhanced Pneumonia Detection in Chest X-Rays Using Hybrid Convolutional and Vision Transformer Networks. Curr Med Imaging. 2025; 21: e15734056326685.
28. Shih G, Wu CC, Halabi SS, Kohli MD, Prevedello LM, Cook TS, Sharma A, Amorosa JK, Arteaga V, Galperin-Aizenberg M, Gill RR. Augmenting the National Institutes of Health chest radiograph dataset with expert annotations of possible pneumonia. Radiol: Artif Intell. 2019; 1(1): e180041.
29. Stein A, Wu C, Carr C, Shih G, Dulkowski J, Kalpathy-Cramer J. RSNA pneumonia detection challenge. Mountain View: Kaggle. 2018.
30. Kermany D. Labeled optical coherence tomography (OCT) and chest x-ray images for classification. Mendeley data. 2018.
31. Gosain A, Sardana S. Handling class imbalance problem using oversampling techniques: A review. In: 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI). 2017; 79–85.
32. Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T, Ding D, Bagul A, Langlotz C, Shpanskaya K, Lungren MP. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. arXiv preprint arXiv:1711.05225. 2017.
33. Bianco S, Cadene R, Celona L, Napoletano P. Benchmark Analysis of Representative Deep Neural Network Architectures. *IEEE Access*. 2018; 6: 64270–64277.
34. Tan M, Le Q. EfficientNet: Rethinking model scaling for convolutional neural networks. In International Conference on machine learning; PMLR. 2019 May; 6105–6114.
35. Polikar R. Ensemble-based systems in decision making. IEEE Circuits Syst Mag. 2006; 6(3): 21–45.

36. Weerts HJ, Mueller AC, Vanschoren J. Importance of tuning hyperparameters of machine learning algorithms. arXiv preprint arXiv:2007.07588. 2020.
37. Shati A, Hassan GM, Datta A. A comprehensive fusion model for improved pneumonia prediction based on KNN- wavelet-GLCM and a residual network. Intell Syst Appl. 2025; 26: 200492.
38. Chiwariro R, Wosowei JB. Comparative analysis of deep learning convolutional neural networks based on transfer learning for pneumonia detection. Int J Res Appl Sci Eng Technol. 2023; 11(1): 1161–1170.

