Enhancing Pneumonia Diagnosis: An Ensemble of Deep CNN Architectures for Accurate Chest X-Ray Image Analysis

Conference Paper · March 2024

DOI: 10.1007/978-981-99-8937-9_18

2 authors:

Md Rabiul Hasan
Khulna University of Engineering and Technology

Shah Muhammad Azmat Ullah
Khulna University of Engineering and Technology

Enhancing Pneumonia Diagnosis: An Ensemble of Deep CNN Architectures for Accurate Chest X-Ray Image Analysis

Md. Rabiul Hasan and Shah Muhammad Azmat Ullah

Abstract Various organisms, such as bacterial and viral infections, can cause a lung infection known as pneumonia. It is a significant health concern, particularly in developing and underdeveloped countries with high pollution rates, overcrowding, and limited healthcare infrastructure. In order to effectively treat pneumonia and improve survival rates, early detection is essential. The simplest technique for identifying pneumonia is a chest X-ray (CXR) study, but CXR analysis can be subjective and challenging. In this paper, we have developed a method for automatically detecting pneumonia from CXR images by combining transfer learning and an ensemble of three CNN network architectures (InceptionV3, MobileNetV2, and Xception) with the weighted average ensemble method. We evaluated our approach on chest X-ray datasets, achieving a maximum accuracy of 92% and F1-scores of 87% and 92% for normal and pneumonia, respectively. Our proposed method outperforms existing ensemble techniques and other cutting-edge approaches, demonstrating the potential for improving pneumonia diagnosis through deep learning-based approaches.

Keywords Deep learning · Transfer learning · Ensemble learning · Pneumonia · CNN · Sigmoid function · Max pooling

1 Introduction

Over four million untimely deaths worldwide from pollution in the air are now caused by pneumonia, making it one of the most hazardous diseases. Pneumonia is a primary cause of death in children under the age of 5 globally, with a death rate of approximately 95% in impoverished nations [1]. Of the total 119,000 children who passed away nationwide in 2015, 17,850 died in Bangladesh from pneumonia, or 15% of the total. They were kids, no more than 5 years old. About 24,300 children

Md. R. Hasan · S. M. A. Ullah (✉)
Department of Electronics and Communication Engineering, Khulna University of Engineering & Technology, Khulna, Bangladesh
e-mail: azmat@ece.kuet.ac.bd

M. S. Arefin et al. (eds.), *Proceedings of the 2nd International Conference on Big Data, IoT and Machine Learning*, Lecture Notes in Networks and Systems 867,
https://doi.org/10.1007/978-981-99-8937-9_18

per year in Bangladesh die from pneumonia. In 2016, pneumonia was the leading cause of mortality, ranking fourth among all other causes.

Identifying pneumonia cases in their early stages can be a challenge because the X-ray image produced is not clear. Despite the fact that chest X-rays are frequently helpful, they may not always indicate pneumonia as it may be misdiagnosed as other respiratory conditions. We were driven to develop an enhanced deep CNN network because of the drawbacks of several current machine learning models. With this ensemble of CNN-based methods, we have achieved precise pneumonia diagnosis, especially where conventional methods have fallen short.

Our suggested design involves combining Inception V3, MobileNet V2, and Xception models to create an ensemble architecture that will automatically diagnose CXR images and accurately classify them as either normal or indicating disease.

In this paper, we have structured the content as follows: Sect. 1 will present a concise introduction to pneumonia, offering an overview of the topic, while Sect. 2 reviews existing approaches to identifying it. The methodology we proposed is detailed in Sect. 3, and Sect. 4 details the dataset used. Image pre-processing, CNN models with transfer learning are covered in Sects. 5 and 6, respectively. Ensemble learning is discussed in Sect. 7, and Sect. 8 presents a visual representation of the performance of our proposed architecture. In the conclusion section of this paper, we provided a thorough review of our findings to tie up our discussion.

2 Literature Review

To diagnose pneumonia using a pediatric dataset, Ayan et al. [2] created the Xception and VGG16 network architectures. They used 80% data for training and 10% each for testing and validation. The accuracy results showed that VGG16 performed at an impressive rate of 87%, while Xception achieved a slightly lower accuracy of 82%. Rajpurkar et al. [3] suggested DenseNet 121 CNN architecture for pneumonia classification, achieving a F1-score of 76.8%. They identified the absence of patient history as a crucial factor that led to the suboptimal performance of their deep learning model. Omar et al. [4] presented a simplified CNN architecture of five convolutional layers (CL). They divided the dataset into 90% for training and 10% for testing, achieving an accuracy of 87.65%. The accuracy was unsatisfying and other metrics like sensitivity, precision, and specificity were not tested. Yao et al. [5] introduced a two-stage, end-to-end model. The model used a DenseNet as a DCNN encoder and an RNN (long short-term memory) as a decoder. Results demonstrated that the suggested model performed better in terms of classification, achieving an AUC of 0.71. Similarly, Hasan et al. [6] utilized a customized CNN for classifying pneumonia. They applied resizing, normalization, and target encoding in the image processing stage, but they did not employ image augmentation and got 96% accuracy. Varshni et al. [7] used CNN models with DenseNet169 and SVM and got an accuracy of 80.02%.

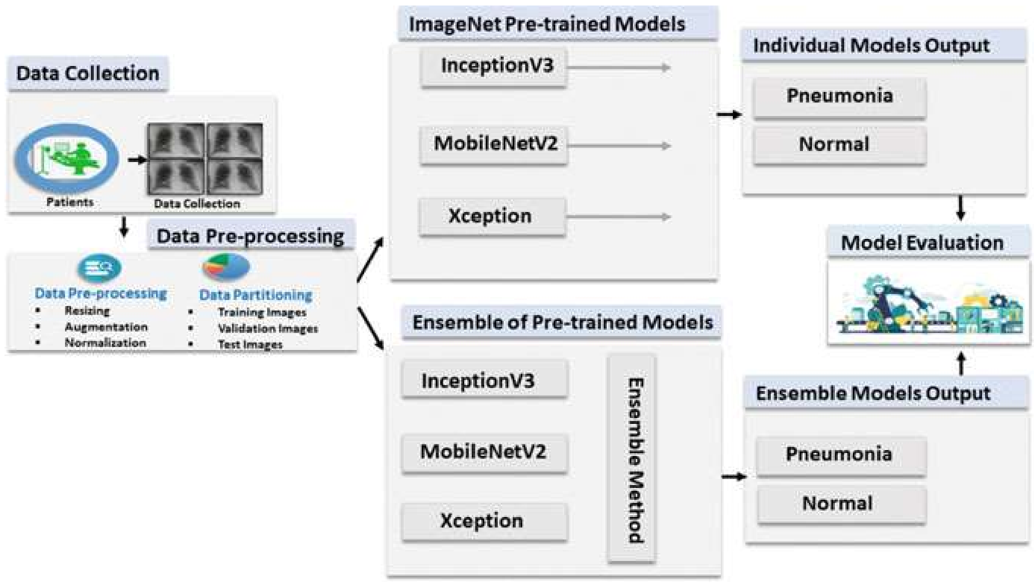

Fig. 1 The conceptual diagram of the proposed pre-trained models and ensemble models

3 Proposed Method

The fundamental objective of this research paper is to develop system that can accurately identify pneumonia using pre-trained models. We developed two frameworks, one composed of individual outputs from each pre-trained model and another consisting of an ensemble of these models. The pre-trained models used for pneumonia detection were InceptionV3, MobileNetV2, and Xception. Figure 1 provides an overview of our suggested methodology, which includes data collection, processing, individual pre-trained models, and ensemble models, each explained in detail.

4 Dataset

In this research, a collection of chest X-ray images was employed as the dataset, which originated from pediatric children between the ages of one and five, provided by Kermany and Goldbaum [8]. The dataset is now available on Kaggle and contains a total of 5856 images, divided into three subsets: validation, train, and test. Due to the employment of various scanning devices and imaging angles, the image resolution differs greatly. The training subset includes images of CXR from 4192 patients, of whom 3110 have pneumonia and 1082 are categorized as “normal.” The test subset includes CXR images of a total of 624 patients, 390 of whom have pneumonia and 234 of whom are categorized as “normal.” The validation dataset consists of 1040 images, of which 773 show pneumonia and 267 are normal (Figs. 2 and 3).

Fig. 2 The sample images of chest X-ray

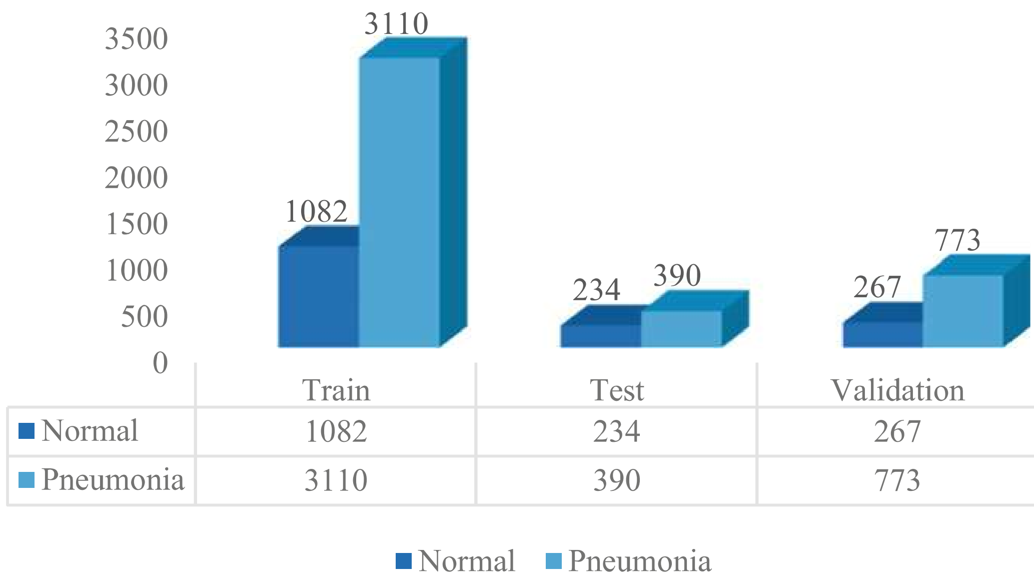

Fig. 3 Pediatric CXR dataset summary

5 Image Pre-processing

A model can be generated when the input data has been appropriately cleaned. By standardizing and optimizing the data through pre-processing, machine learning models can more easily find patterns and make accurate predictions.

5.1 Resizing

The CXR were initially RGB, but for the purpose of the research, it was converted to grayscale. A variety of sizes, including 1040 × 664, 1224 × 1000, and 1848 × 1632, were offered for the images in the database. We reduced its size to 224 × 224 pixels to make it simple to incorporate into our suggested CNN models.

Table 1 Augmentation technique used in this proposed methodology

Technique	Settings
Rescale	1/255
Width_shift_range	0.1
Height_shift_range	0.1
Brightness range	[0.2, 1.0]
Zoom range	0.2
Fill_mode	Nearest

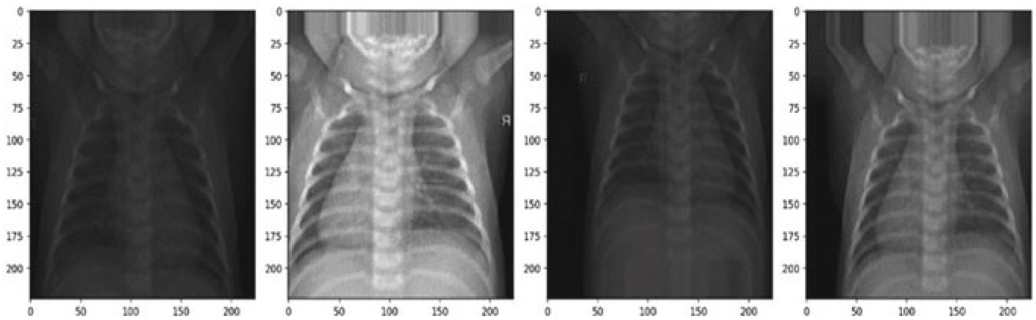

Fig. 4 The chest X-ray images after applying image augmentation technique

5.2 Augmentation

To ensure that a model performs well in real-world scenarios where lighting, rotation, scale, and other factors may differ, modified images can be added to the training data. This process, called image augmentation, exposes the model to diverse cases, prevents overfitting, and allows the model to generalize better. Table 1 outlines the image augmentation parameters that can significantly improve image-based models' performance and resilience in practice (Figs. 4 and 5).

5.3 Normalization

Normalization is a critical machine learning technique that scales features in a dataset to a common range, typically from 0 to 1, by dividing each pixel value by 255. Normalization solves the problem of significant scale disparities between features, making machine learning algorithms converge faster and reducing bias in the dataset. Moreover, normalization enhances the interpretability of machine learning model results by facilitating comparison of the relative importance of various features.

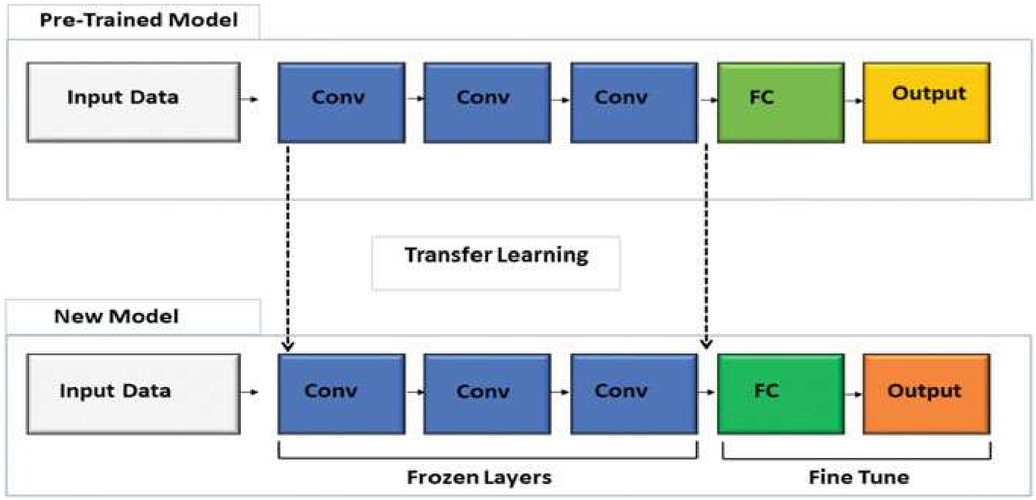

Fig. 5 Concept of transfer learning

6 Deep CNN Model Architectures Using Transfer Learning

6.1 Convolutional Neural Network Model Architectures

6.1.1 InceptionV3

InceptionV3 [9] builds on the success of GoogleNet [10] and has shown strong classification performance in biomedical applications with transfer learning [11, 12]. It introduced an innovative inception model that combines multiple convolutional filters of different sizes into a single filter, resulting in reduced computational complexity and fewer parameters to train.

6.1.2 MobileNetV2

The main framework of MobileNetV2 was built on MobileNetV1. MobileNetV2 employs the depthwise separable convolutions technique for probability. It addressed the issue of data loss in the non-linear layers of convolution blocks and stored the data by employing linear bottlenecks and inverted residuals [13].

- Depthwise Separable Convolution
- Linear Bottlenecks

6.1.3 Xception

François Chollet built the Xception model [14]. The input image is first processed by a single convolutional layer in the Xception model architecture. Three building blocks make up the Xception architecture. The input flow is the initial step in the data entry process. After that, the data pass via the middle flow, which is iterated eight times, before eventually passing through the exit flow. The depthwise separable convolutions principle is used in this model. Depthwise separable convolutions, which are a variation of regular convolutions, are a variation of regular convolutions, are anticipated to be more computationally efficient. It divides the convolution into two distinct operations: a depthwise convolution, which employs a single filter for each of the input channels, and a pointwise convolution, which combines the depthwise convolution's outputs using 1×1 convolutions. With fewer parameters and higher feature representation as a result, this enables the network to record spatial and channel-wise relationships separately. Its state-of-the-art performance on a number of computer vision tasks makes it a well-liked option for object identification, and image classification.

6.2 Transfer Learning: Fine Tuning

Transfer learning is a powerful technique that allows a model that has been trained for one task to be utilized to speed up learning for a different but related task [15]. In numerous applications where there is less training data, transfer learning enables a model to make use of the information gained from a related task to perform better on the intended task. Pre-trained models like InceptionV3, MobileNetV2, and Xception, which were trained on the 1.2 million images, 1000 class ImageNet dataset. Models' generalization abilities can be enhanced through transfer learning. The convolutional layer weights were obtained from pre-trained ImageNet weights before starting training. We trained the classifier part of the network from scratch. During model training and fine tuning, the different hyperparameters that we used were batch size, epoch, the number of dropouts, learning rate, etc. Using the procedure depicted in Fig. 6, which involves trial and error, we obtained the ideal hyperparameters.

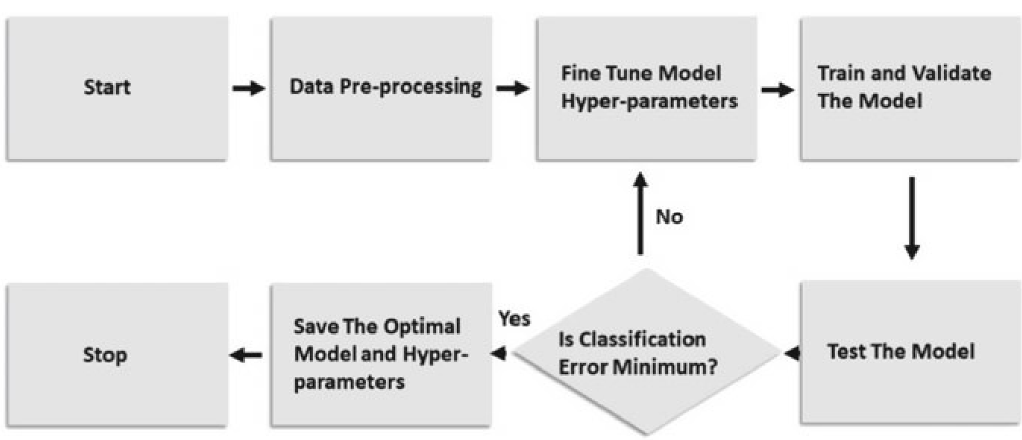

Fig. 6 Determining optimal hyperparameters

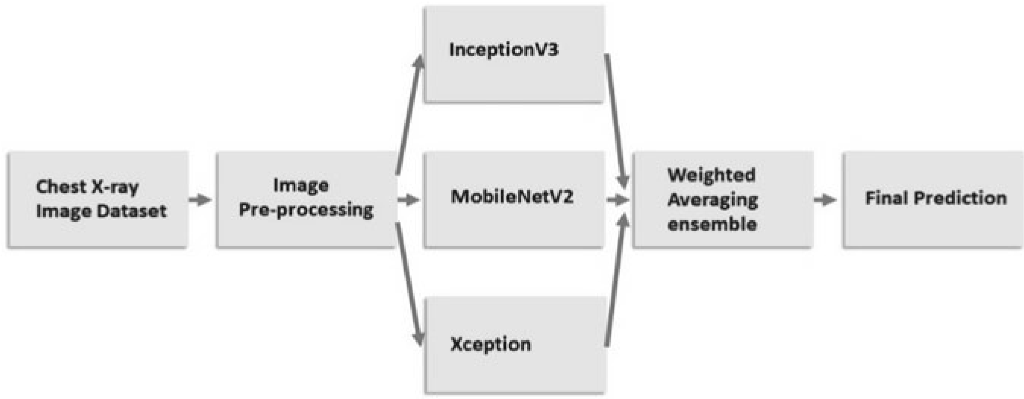

Fig. 7 Ensemble learning for pneumonia detection

7 Ensemble Learning

In this scenario, we employed the weighted average ensemble method for the InceptionV3, MobileNetV2, and Xception model architectures. This technique involves taking the predictions from all models, calculating their average score, and using that to determine the final prediction. Let's make the mathematical assumption that there are N distinct models, all of which were trained using the same dataset. Each ensemble model creates a prediction for a fresh input instance, denoted as $y1$, $y2$,..., yn, where $w1$, $w2$,..., and wn are the weights assigned to each model utilizing the weighted average approach, the final prediction is:

$$y = \frac{w1 * y1 + w2 * y2 + \ldots w_n * y_n}{w1 + w2 + \cdots + wn} \tag{1}$$

This approach can be applied to any machine learning model and is often paired with deep learning models. Notably, the average ensemble method is computationally efficient. It is a commonly employed technique for improving the performance of machine learning models and can significantly enhance their robustness and accuracy (Fig. 7).

8 Results and Discussion

8.1 Output of Single Model

Initially, we presented the accuracy, confusion matrix, and loss curves generated by three separate deep transfer learning models, namely InceptionV3, MobileNetV2, and Xception, which are illustrated in Figs. 8, 9, and 10. Subsequently, we compared the results of each model using various performance metrics (Table 2).

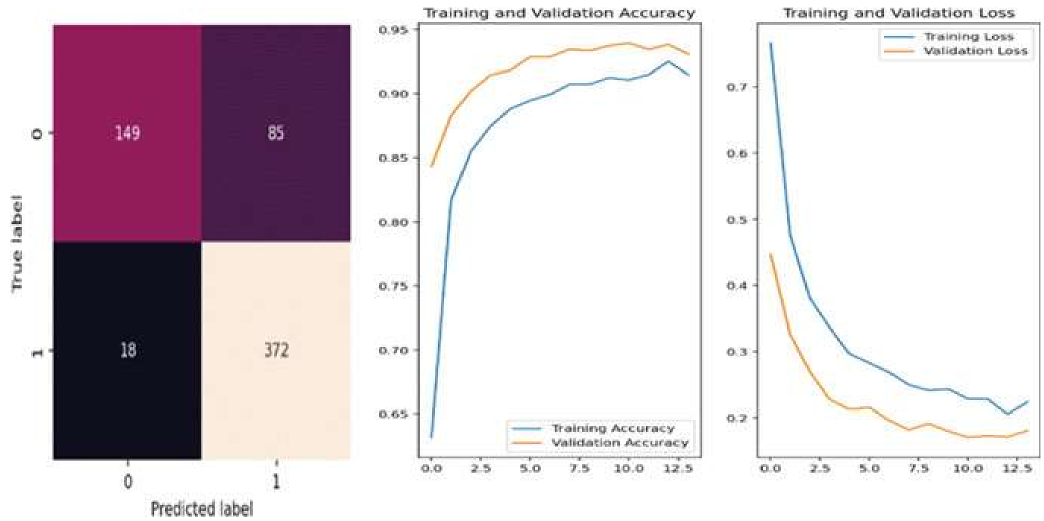

Fig. 8 Confusion matrix, accuracy, and loss curve of InceptionV3

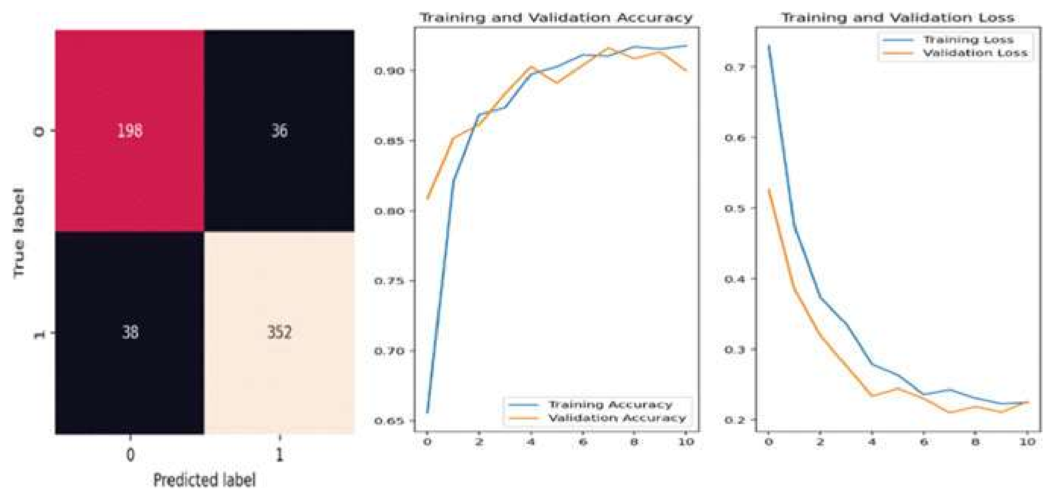

Fig. 9 Confusion matrix, accuracy, and loss curve of MobileNetV2

A. InceptionV3

We used a batch size of 32 to train the InceptionV3 model over several epochs during the training process. In each epoch, the training process consisted of 131 steps. To improve performance, we utilized the ReduceLROnPlateau callback, which monitors performance and decreases the learning rate by a specified factor if no progress is observed after a certain epoch number. After completing 20 epochs of training, the process was stopped at 14 epochs as the best results were obtained with an accuracy of 83%.

B. MobileNetV2

We used the same batch size for the MobileNetV2 model, but with different numbers of epochs. Since we also employed the ReduceLROnPlateau method, the training

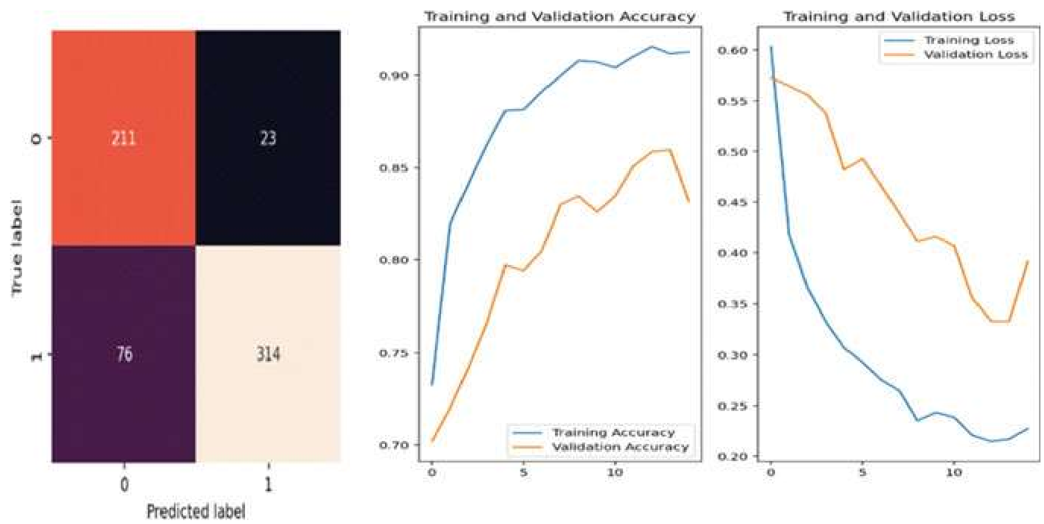

Fig. 10 Confusion matrix, accuracy, and loss curve of Xception

Table 2 Single model evaluation metrics

Model	Accuracy (%)	Precision (%)		Recall (%)		F1-score (%)	
		0	1	0	1	0	1
InceptionV3	83	89	81	64	95	74	88
MobileNetV2	88	84	91	85	90	84	90
Xception	84	74	93	90	81	81	86

process stopped after 11 epochs as there was no improvement in accuracy. The best results were achieved with an accuracy of 88%.

C. Xception

The training and validation accuracy of the Xception model changed across various epochs. After 15 epochs, since accuracy did not increase, the training process was terminated. The confusion metrics showed that the true positive (TP) and true negative (TN) values were 184 and 361, respectively. The overall accuracy of the model was 84%.

8.2 *Output of Ensemble Model*

The remainder of this study focuses on ensemble learning to evaluate whether performance metrics can be improved. We created three ensembles of various deep transfer learning models, as shown in the table below. From Figs. 11, 12, 13, 14 and 15, it is evident that the best ensemble model was created by combining Xception and MobileNetV2. This ensemble model gained accuracy of 90%, with precision, recall, and F1-scores of normal and pneumonia being 85, 94, 90, 91, 88, and 92%, respectively. On the other hand, the individual models produced unsatisfactory results. The accuracy of the InceptionV3, MobileNetV2, and Xception models was 83, 88, and 84%, respectively, as shown in Fig. 12. However, the ensemble of pre-trained models showed improved accuracy.

Here, in Table 3, normal = 0 and pneumonia = 1. In addition, it is evident that our proposed model has surpassed current models in terms of precision, F1-score, accuracy, and recall.

The maximum accuracy obtained by the combination of MobileNetV2 with Xception was 90%, while the accuracy obtained by the combination of MobileNetV2 with InceptionV3 and that obtained by the combination of Xception with InceptionV3 were both 88%.

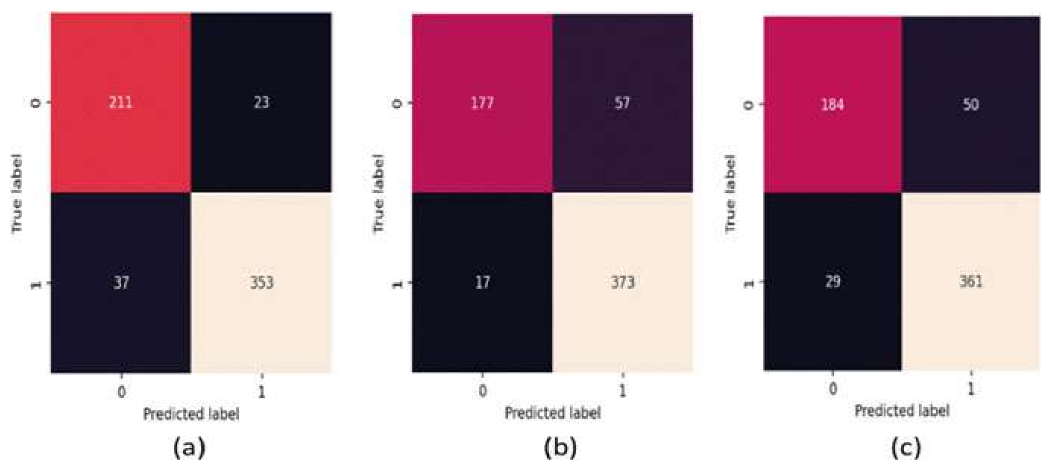

Fig. 11 Confusion metrics of **a** Xception with MobileNetV2, **b** MobileNetV2 with InceptionV3, **c** Xception with Inception

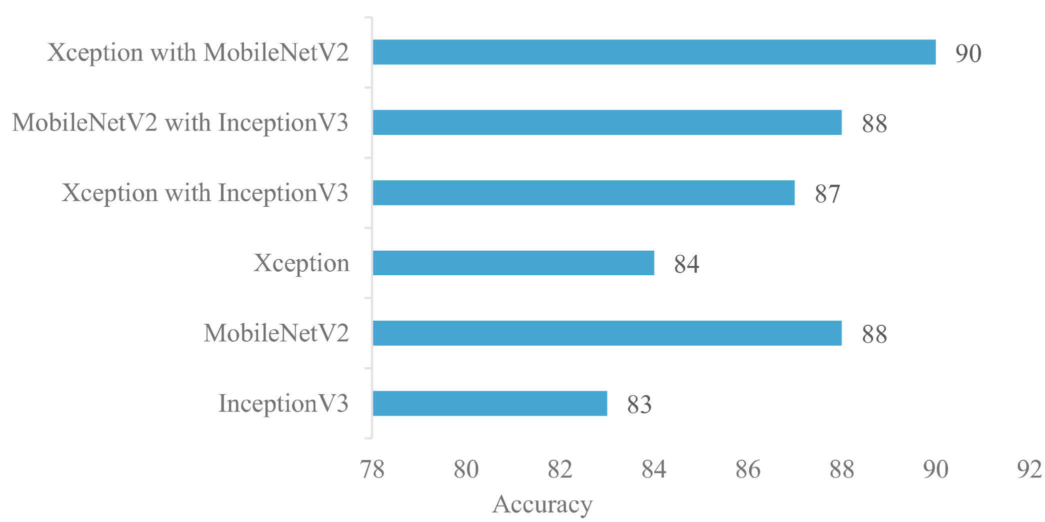

Fig. 12 Accuracy comparison of single and ensemble models

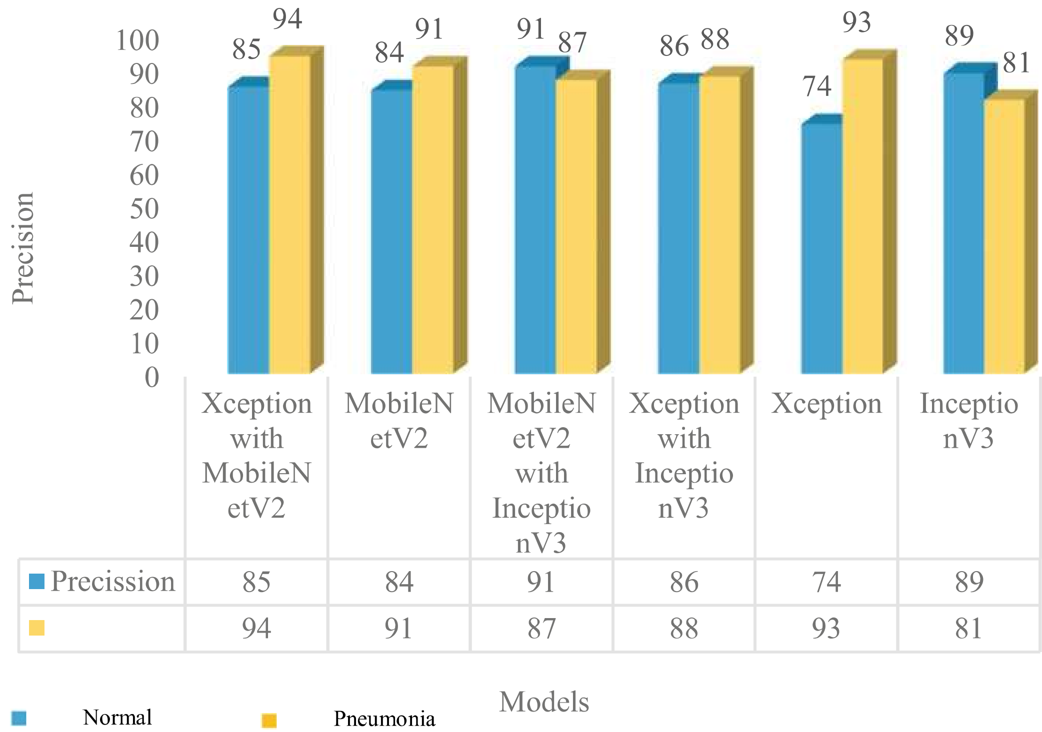

Fig. 13 Precision comparison of single and ensemble models

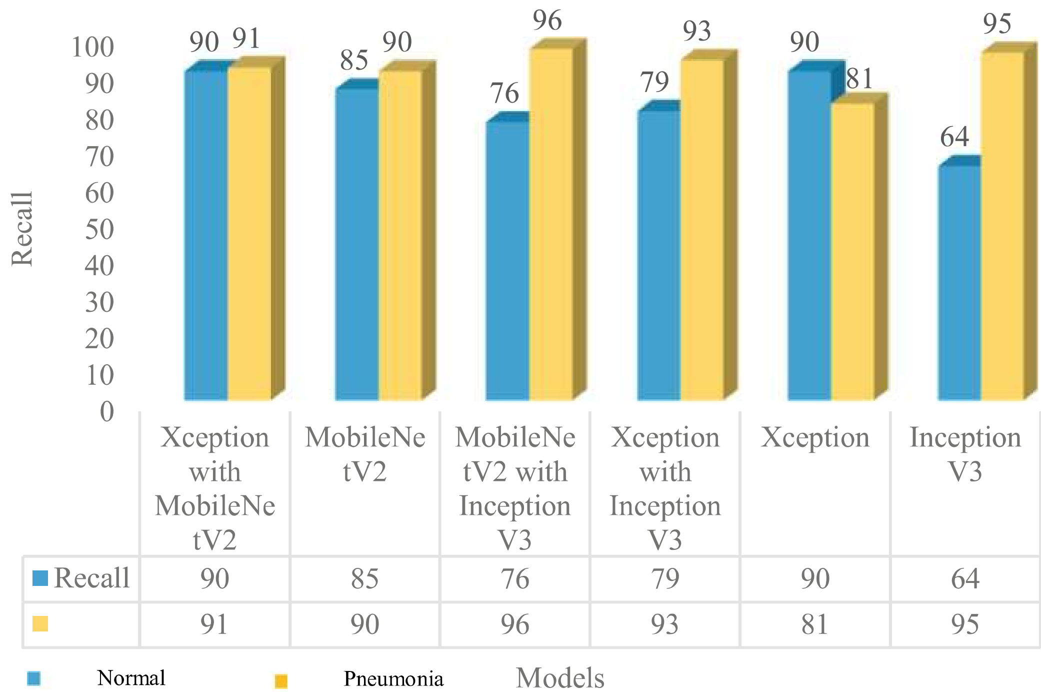

Fig. 14 Recall comparison of single and ensemble models

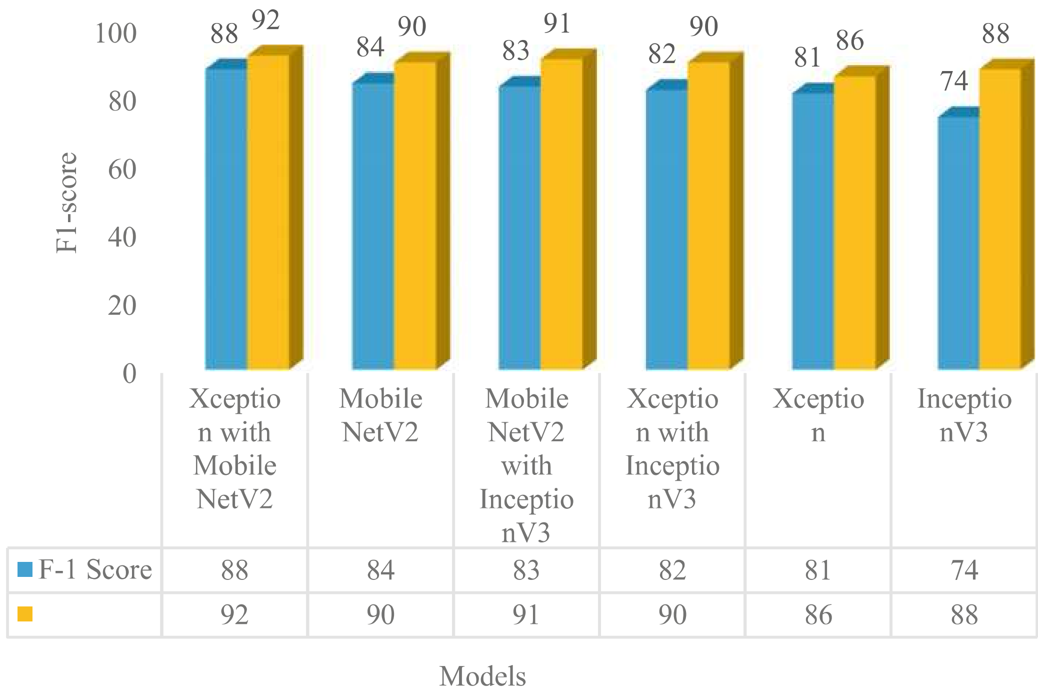

Fig. 15 F1-score comparison of single and ensemble models

Table 3 Ensemble model evaluation metrics

Model	Accuracy (%)	Precision (%)		Recall (%)		F1-score (%)	
		0	1	0	1	0	1
Xception with MobileNetV2	90	85	94	90	91	88	92
MobileNetV2 with InceptionV3	88	91	87	76	96	83	91
Xception with Inception	87	86	88	79	93	82	90
InceptionV3	83	89	81	64	95	74	88
MobileNetV2	88	84	91	85	90	84	90
Xception	84	74	93	90	81	81	86

Comparing the single and ensemble model precision values, the ensemble of MobileNetV2 with the Xception model for normal and pneumonia were 85 and 94%, respectively, which were the maximum.

Similarly, for performance metrics like recall and F1-score, the ensemble of MobileNetV2 with the Xception model outperformed the other ensembles and single models, and thus we got the best result from this ensemble of models (Table 4).

In addition, it is evident that our proposed model has surpassed current models in terms of precision, F1-score, accuracy, and recall.

Table 4 A comparative study of pneumonia diagnosis models

Authors	Methodology	Outcomes
Varshni et al. [7]	CNN models with DenseNet169 and SVM	Accuracy 80.02%
Omar et al. [4]	CNN architecture with 5 convolutional layers	Accuracy = 87.65%
Ayan et al. [2]	Xception and VGG16	For Xception and VGG16, the accuracy is 82 and 87%, respectively
Proposed model	Ensemble of Xception with MobileNetV2	Accuracy 90%

9 Conclusion

We have presented a method for accurately classifying pneumonia information using a collection of X-ray images. We have created an ensemble model of finely tuned InceptionV3, MobileNetV2, and Xception models, compared them with current methods, and evaluated their performance against individual model outcomes. Our proposed model has demonstrated accurate classification of CXR images, with increasing accuracy and decreasing loss with each training epoch. This model makes precise predictions with less computing complexity, which is particularly beneficial in the healthcare system for prompt and accurate pneumonia diagnosis. The authors intend to build a web application where users can upload chest X-ray images to the app, and it will automatically detect pneumonia, which can lead to effective treatment and potentially save lives.

References

1. Hug L, Alexander M, You D, Alkema L (2019) National, regional, and global levels and trends in neonatal mortality between 1990 and 2017, with scenario-based projections to 2030: a systematic analysis. Lancet Glob Health 7(6):e710–e720. https://doi.org/10.1016/S2214-109X(19)30163-9
2. Ayan E, Ünver HM (2019) Diagnosis of pneumonia from chest X-ray images using deep learning. In: Proceedings of the 2019 scientific meeting on electrical-electronics and biomedical engineering and computer science (EBBT), pp 1–5
3. Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T et al (2017) CheXNet: radiologist-level pneumonia detection on chest X-rays with deep learning. http://arxiv.org/abs/1711.05225
4. Omar H, Babalik A (2023) Proceedings Book, 2019: detection of pneumonia from X-ray images using convolutional neural network. In: International conference on data science, machine learning and statistics
5. Yao L, Poblenz E, Dagunts D, Covington B, Bernard D, Lyman K (2017) Learning to diagnose from scratch by exploiting dependencies among labels. http://arxiv.org/abs/1710.10501
6. Hasan MR, Ullah SMA, Karim ME (2023) An effective framework for identifying pneumonia in healthcare using a convolutional neural network. In: Proceedings of the 2023 international conference on electrical, computer and communication engineering (ECCE), pp 1–6. https://doi.org/10.1109/ECCE57851.2023.10101548

7. Varshni D, Nijhawan R, Thakral K, Mittal A, Agarwal L (2019) Pneumonia detection using CNN based feature extraction. In: Proceedings of the 2019 IEEE international conference on electrical, computer and communication technologies (ICECCT), Coimbatore, India, pp 1–7. https://doi.org/10.1109/ICECCT.2019.8869364
8. Kermany DS, Goldbaum M, Cai W, Valentim CCS (2018) Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell 172(5):1122-1131.e9. https://doi.org/10.1016/j.cell.2018.02.010
9. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2016) Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2818–2826. https://www.cvfoundtion.org/openaccess/content_cvpr_2016/html/Szegedy_Rethinking_the_Inception_CVPR_2016_paper.html. Accessed 21 Dec 2022
10. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D et al (2015) Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1–9. https://www.cvfoundation.org/openaccess/content_cvpr_2015/html/Szegedy_Going_Deeper_With_2015_CVPR_paper.html. Accessed 08 July 2023
11. Shin HR, Gao M, Lu L, Xu Z, Nogues I et al (2016) Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE Trans Med Imag 35(5):1285–1298. https://doi.org/10.1109/TMI.2016.2528162
12. Kumar A, Kim J, Lyndon D, Lyndon D, Fulham M, Feng D (2016) An ensemble of fine-tuned convolutional neural networks for medical image classification. IEEE J Biomed Health Inform 21(1):31–40. https://doi.org/10.1109/JBHI.2016.2635663
13. Dong K, Zhou C, Ruan Y, Li Y (2020) MobileNetV2 model for image classification. In: Proceedings of the 2020 2nd international conference on information technology and computer application, ITCA 2020, Institute of Electrical and Electronics Engineers Inc., pp 476–480. https://doi.org/10.1109/ITCA52113.2020.00106
14. Chollet F (2017) Xception: deep learning with depthwise separable convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1251–1258. http://openaccess.thecvf.com/content_cvpr_2017/html/Chollet_Xception_Deep_Learning_CVPR_2017_paper.html. Accessed 07 July 2023
15. Weiss K, Khoshgoftaar TM, Wang DD (2016) A survey of transfer learning. J Big Data 3(1):43. https://doi.org/10.1186/s40537-016-0043-6