Journal of Adaptive Learning Technologies
ISSN **2997-3902** (Online)
Vol. 1, No. 11, Nov 2024,
Available online at: https://scientificbulletin.com/index.php/JALT

Article

Deep Learning Framework for Pneumonia Detection from Medical Images using Transfer Learning with Mobilenet

Md. Soebur Rahman*[1], Md Ismail Hossain Siddiqui[2], Saeed Ur Rashid[3], Abdul Aziz Kabir[4], Farhad Uddin Mahmud[5], Ramisa Samin Shammah[6]
1,5. Master of Business Administration in Management Information Systems, International American University, Los Angeles, California, USA
2. Master of Science in Engineering/Industrial Management, Westcliff University, Irvine, California, USA
3,4. Master of Business Administration in Data Analytics, Westcliff University, Irvine, California, USA
6. Dept. College of Technology and Engineering, Westcliff University, Irvine, USA
*Corresponding Author: soebrahman10@gmail.com

Article information:
Manuscript received: 30 September 2024; **Accepted:** 27 October 2024; **Published:** 15 November 2024

Abstract: Pneumonia remains a leading cause of morbidity and mortality worldwide, particularly in regions with limited access to expert radiological assessment, highlighting the need for reliable automated diagnostic support systems. This study presents a deep learning–based framework for binary classification of chest X-ray images into pneumonia and normal categories using a pretrained convolutional neural network. A dataset comprising 5,856 chest X-ray images was utilized, including 4,273 pneumonia and 1,583 normal cases, with an imbalanced class distribution. The data were partitioned into training, validation, and held-out test sets with approximately 79%, 5%, and 16% of the data, respectively. To mitigate data scarcity and improve generalization, controlled data augmentation was applied during training.A pretrained MobileNet architecture was employed as the feature extractor, leveraging transfer learning to adapt to the medical imaging domain. The model was trained using mini-batch optimization with a batch size of 32 and input resolution of 224 × 224 pixels. Performance evaluation was conducted on an independent test set using multiple metrics, including accuracy, precision, recall, F1-score, and receiver operating characteristic area under the curve.The proposed framework achieved a test accuracy of 86.07%, with a precision of 83.88%, recall of 99.06%, and F1-score of 90.84%. The receiver operating characteristic analysis yielded an area under the curve of 0.96, indicating strong discriminative capability. The confusion matrix analysis revealed a low false negative rate, demonstrating the model's effectiveness in identifying pneumonia cases, although a relatively higher false positive rate suggests limitations in distinguishing normal cases. Overall, the results indicate that the model is well-suited for high-sensitivity screening scenarios, where minimizing missed pneumonia cases is critical.

Keywords: Pneumonia Detection, Chest X Ray Imaging, Deep Learning, Transfer Learning, Mobilenet, Medical Image Classification, Binary Classification, Computer Aided Diagnosis, Imbalanced Dataset, Receiver Operating Characteristic

1. Introduction

Pneumonia is a major world health issue and it is one of the major causes of morbidity and mortality especially in children, the aged and in immunocompromised persons. Early diagnosis followed by clinical intervention is necessary to achieve better patient outcomes and timely clinical intervention. One of the most common diagnostic techniques of detecting pneumonia is chest X-ray imaging because of its affordability, availability, and speed of acquisition. Nevertheless, the interpretation of the chest radiographs involves a substantial clinical expertise and in many instances, may be variably interpreted inter-observer particularly where there is a shortage of resources in healthcare facilities that may not have the experienced radiologists at hand.Over the last several years, deep learning algorithms, specifically convolutional neural networks, have shown good results in a range of medical image analysis scenarios and among them are disease classification, lesion detection and lesion segmentation. Pretrained architecture-based transfer learning has been a viable technique when limited annotated medical data are available, allowing models to take advantage of the knowledge gained on large-scale natural image datasets. In spite of these developments, there are a number of obstacles that still exist in the creation of pneumonia detection systems that can be trusted. These are: small size of the dataset, the imbalance on the classes between disease and normal cases and the probability of overfitting, which all may negatively impact the generalization ability of the models.Sensitivity and specificity: The balance between sensitivity and specificity is also another critical factor to consider in clinical applications. It is important to ensure high sensitivity to reduce the number of missed pneumonia cases which may have severe clinical implications. Nevertheless, high rates of false positives may decrease the clinical trust and result in unnecessary follow-up procedures. Thus, it is still a significant goal to design models that are both highly detectable and are not too specific.This paper presents a deep learning-based system to detect pneumonia automatically on chest X-rays with the help of pretrained MobileNet architecture. It is trained on a dataset of 5,856 images, which has an uneven distribution between classes pneumonia and normal. Training is performed with data augmentation to enhance the generalization and the model is tested with a held-out test set to determine real-world performance. Various measures of evaluation such as accuracy, precision, recall, F1-score, and receiver operating characteristic are employed to give a thorough examination of model behavior.The findings indicate high sensitivity of the model in identifying pneumonia cases, and high overall discriminative performance. Meanwhile, the results also indicate that there are issues with specificity, and specific attention should be paid to the choice of threshold and its analysis in clinical implementation cases. These findings give us a glimpse of the limitations and advantages of deep learning-based pneumonia screening systems in practice.

Related Work

Automated pneumonia detection from chest X-ray images has been extensively studied in recent years, driven by the increasing availability of medical imaging datasets and advances in deep learning [1,2]. Early approaches primarily relied on traditional machine learning methods combined with handcrafted feature extraction techniques. These methods typically used texture descriptors, edge-based features, or statistical representations, followed by classifiers such as support vector machines or random forests [3,4]. While these approaches demonstrated moderate performance, their reliance on manually designed features limited their ability to capture complex patterns present in medical images [5,6].With the emergence of deep convolutional neural networks, significant improvements have been achieved in medical image classification tasks. Architectures such as AlexNet, VGG, ResNet, and DenseNet have been widely adopted for pneumonia detection due to their ability to learn hierarchical feature representations directly from raw image data [7,8]. Several studies have reported high classification accuracy using these models, particularly when trained on large-scale datasets. However, training deep architectures from scratch often requires substantial computational resources and large annotated datasets, which are not always available in medical domains.To address data scarcity, transfer learning has become a widely used strategy, where models pretrained on large natural image datasets are fine-tuned for specific medical imaging tasks. Lightweight architectures such as MobileNet have gained attention due to their reduced computational complexity and suitability for deployment in resource-constrained environments [8]. These models offer a favorable trade-off between performance and efficiency, making them practical for real-world applications, including point-of-care diagnostics [9,11].Despite these advancements, several challenges persist in the literature. One of the primary issues is class imbalance, where pneumonia cases often outnumber normal cases or vice versa, leading to biased model predictions [12-14]. In addition, many studies report high overall accuracy without adequately addressing the trade-off between sensitivity and specificity, which is critical for clinical applicability[15-17]. High sensitivity is essential to ensure that pneumonia cases are not missed, but excessive false positives can reduce the usability of automated systems in practice.Another limitation observed in prior work is the variability in evaluation protocols[18-19]. Some studies rely solely on validation performance without using a separate held-out test set, which may lead to optimistic performance estimates. Furthermore, inconsistencies in data preprocessing, augmentation strategies, and dataset splitting make direct comparison across studies challenging. Robust evaluation using independent test data and comprehensive performance metrics remains essential for reliable assessment[20].

Dataset and Preprocessing

A. Dataset Description

The paper makes use of a chest X-ray dataset comprising of 5,856 images classified in two categories: pneumonia and normal. Out of these, 4,273 of these images are associated with cases of pneumonia, and 1,583 are associated with normal cases, which has an observable imbalance between classes (Figure 1). This skew represents the actual clinical distributions in the real world, where disease prevalence is frequently greater in screened populations, and should hence be cautiously taken into account when developing and testing the model. The data is divided into three discontinuous sets used to train, validate and test. The training set has 4,642 images, the validation set has 295 images and the held-out test set has 919 images (Table 1). This amounts to about a 79, 5 and 16 percent cutoff respectively. The test set is highly isolated and is only applied once to the final evaluation so that there is no bias in estimating model performance and generalization ability. The distribution of the classes is unbalanced in all splits (Figure 2). The training set has 3,418 pneumonia and 1,224 normal images. The validation set consists of 214 pneumonia, 81 normal images, and the test one consists of 641 pneumonia and 278 normal images. This overall imbalance across splits demands cautious consideration of evaluation metrics, especially focusing on recall and F1-score alongside accuracy.

Table 1. Dataset distribution and class composition across training, validation, and test splits

Dataset Split	Total Images	Pneumonia	Normal	Percentage (%)
Training	4,642	3,418	1,224	79.27%
Validation	295	214	81	5.04%
Test	919	641	278	15.69%
Total	5,856	4,273	1,583	100%

B. Image Preprocessing

Images are all re-scaled to a fixed size of 224 x 224 pixels to fit the input size of the pretrained convolutional neural network. The homogeneity among samples and the possibility to process them in large numbers is guaranteed by the resizing. The preprocessing pipeline does not apply any aggressive enhancement algorithm, like histogram equalization or denoising filters, and thus does not change the original radiographic features, and it does not introduce artificial patterns that may bias the learning process. Data augmentation is used to enhance the training set to overcome the constraint of small size of the dataset and enhance the generalization of the model. The augmentation strategy comprises of a set of controlled geometric transformations, rotation within the range of 7 degrees, horizontal and vertical translation within 5 percent, shear transformation of factor 0.2, zoom scaling to 45 percent and horizontal flipping. These augmentations are dynamically implemented in training to augment data diversity without affecting clinically relevant image structures.The validation set is generated with another data generator to be able to guarantee the same preprocessing when evaluating the model. A batch-based pipeline is used to load the dataset in 32 batches, which allows the efficient use of memory, and the optimization remains stable. Training data augmentation and structured iteration over validation and test samples can be performed on-the-fly with the use of generator-based loading.

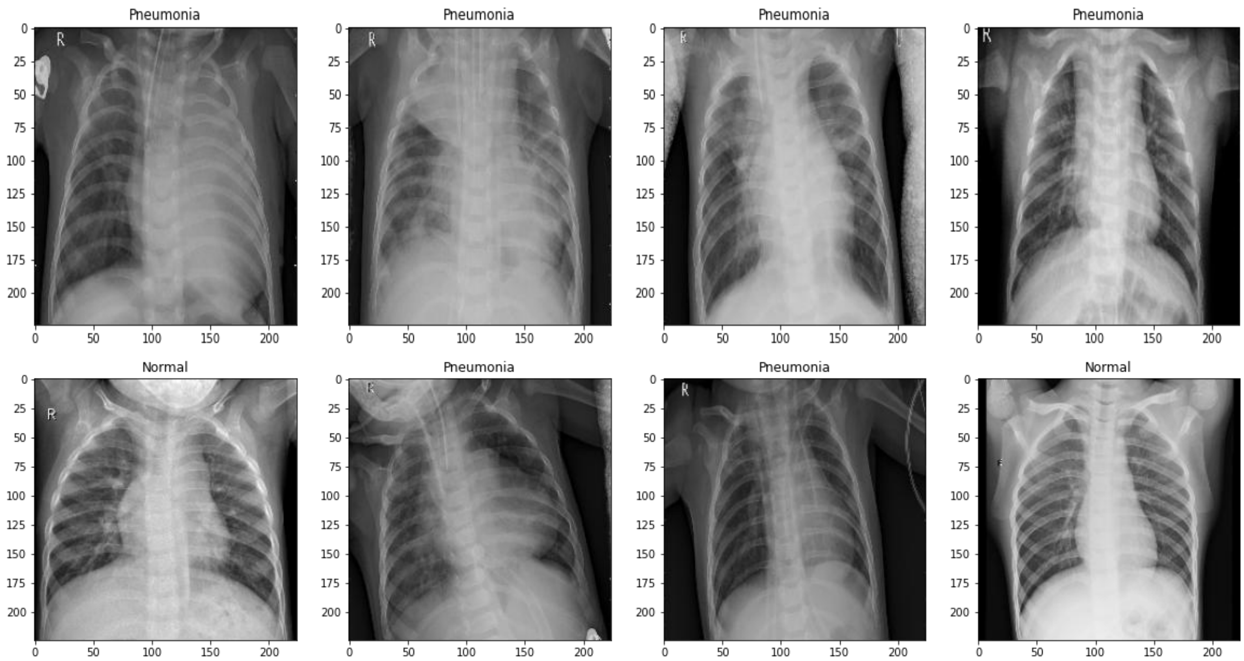

Figure 1. Representative chest X ray images of pneumonia and normal classes

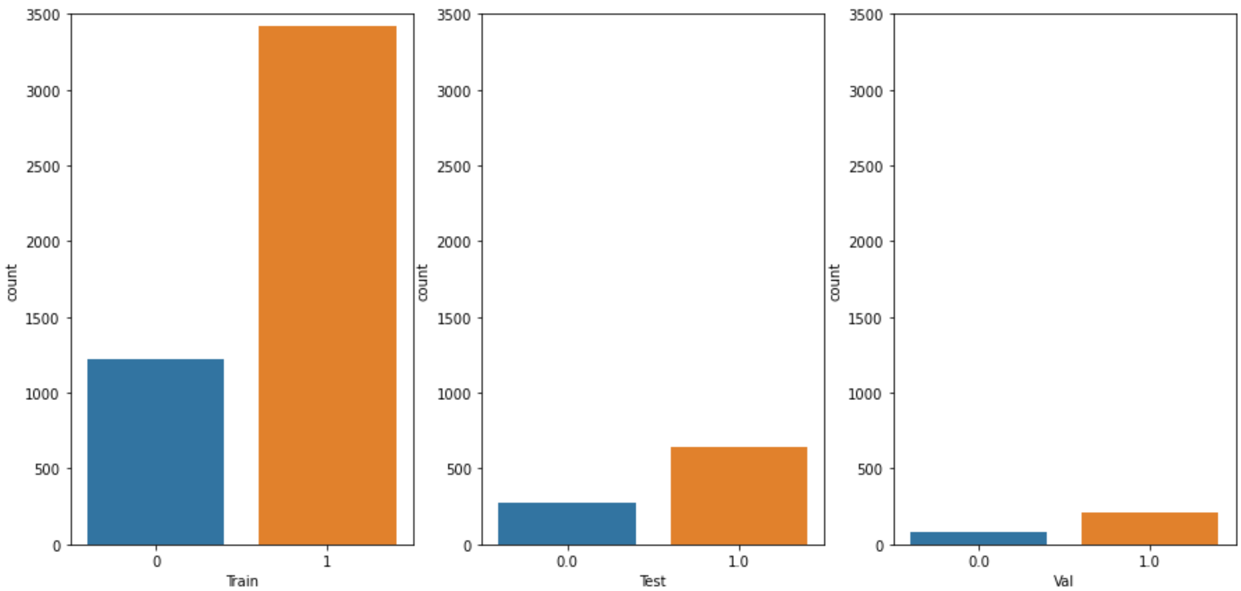

Figure 2. Class distribution across training, validation, and test datasets.

2. Materials and Method

Overall Framework

The proposed framework adopts a transfer learning–based approach for automated pneumonia detection from chest X-ray images. Instead of designing a convolutional neural network from scratch, a pretrained MobileNet model is utilized as the core feature extractor. This approach enables the model to leverage rich visual representations learned from large-scale image datasets and adapt them to the medical imaging domain. The overall pipeline consists of three main stages: image preprocessing, deep feature extraction using the pretrained network, and classification through a task-specific output layer.

Input Processing

Each chest X-ray image is resized to a fixed resolution of 224 × 224 pixels to ensure compatibility with the pretrained network architecture. The images are processed in a consistent format and fed into the model in mini-batches of size 32. This batch-based training strategy improves computational efficiency and stabilizes gradient updates during optimization. The preprocessing pipeline maintains the original structural characteristics of the radiographic images to preserve clinically relevant features.

Feature Extraction with MobileNet

The core of the suggested model lies in MobileNet, a lightweight convolutional neural network framework that is aimed at effective learning of features. MobileNet uses depthwise separable convolutions, a type of convolution operation that splits a standard convolution operation into depthwise and pointwise convolution. This design greatly minimizes the parameters and the cost of computation and also has good representational power. Consequently, the model can be effectively used in the settings where the scalability and the computational efficiency are relevant factors.The pretrained MobileNet model is trained over large-scale datasets of natural images. These features pretrained weights give a good starting point of extracting features, especially of low-level and mid-level visual patterns. The network is optimised during training to suit the particular details of the chest X-ray images. It is a very effective strategy especially in a relatively small medical dataset as it minimizes the chances of overfitting and converges quickly.A task-specific classification layer is placed on top of the pretrained feature extractor that will be used to perform binary classification. The obtained feature maps are fed to a fully connected layer, which is followed by a sigmoid activation function to obtain a probability score of whether a person is infected with pneumonia. This probability score is then used to predict an outcome by applying a decision threshold whereby each input image is classified as pneumonia or normal.

Training Configuration

The mini-batch optimization is based on a batch size of 32 to train the model (Table 2). The training involves the use of augmented data obtained as a result of controlled geometric transformations to enhance generalization. The input pipeline is generated based on the generators enabling them to dynamically augment during training without wasting memory. The training is performed on several epochs and the model performance is monitored on another validation set.The suggested framework aims to prioritize the sensitivity in identifying pneumonia cases, considering the clinical significance of reducing missed diagnoses. Simultaneously, the MobileNet architecture is lightweight, which guarantees effective computation and the possible applicability to the resource-constrained environment. Nonetheless, the design of the model also involves considerations of the trade-off between sensitivity and specificity, especially when dealing with imbalanced datasets where bias in prediction can be observed.

Experimental Setup

A. Implementation Environment

The suggested framework is deployed via the TensorFlow back-end and high-level model development and training APIs. Image data generator utilities are built-in to perform data preprocessing and augmentation. The training pipeline is implemented in a typical deep learning setup that supports the use of a GPU, which allows the computation of the training in batch and also converges quicker. Memory management techniques, such as explicit variable clearance and garbage collection, are used before training, so that the resources are used optimally. This model relies on a pretrained MobileNet model, modified to binary classification of chest X-ray images. The input images are downsampled to 224 224 pixels and batches of 32 images are used. The last classification layer is adjusted to produce a single probability score with the help of a sigmoid activation function, which is the probability of pneumonia. The pretrained backbone enables the feature extraction to be effective and less training compared to training on a large scale is necessary. Training is conducted with the help of a pipeline based on a generator which facilitates on-the-fly augmentation of data. The training data generator uses controlled geometric distortions, rotation, translation, shear, zoom, and horizontal flipping, to enhance the variety of the training data and minimize overfitting. A special generator is used in processing the validation data in order to ensure consistency in the evaluation. Mini-batch training enhances the efficiency of computation and stabilizes the updates in parameters across iterations.The model is trained to binary classification with a standard loss form that is applicable to probabilistic outputs. The optimization algorithm successively adjusts the model parameters to reduce the difference between the predicted probabilities and the ground truth labels. Although the specific optimizer setting is not explicitly defined, the training is a gradient-based optimization model typically applied in deep learning architectures. The batch-based updates are used in such a way that the convergence behavior is stable during the training.

Table 2. Image preprocessing and data augmentation configuration for model training

Step	Description
Input size	224 x 224 pixels
Batch size	32
Normalization	Standard resizing (no aggressive enhancement)
Rotation	±7 degrees
Width shift	0.05
Height shift	0.05
Shear	0.2
Zoom	0.45
Horizontal flip	Enabled
Valid preprocessing	Consistent preprocessing using generator
Augmentation applied to	Training set only

B. Evaluation Protocol

Independent held-out test set of 919 images is used to compute model performance, but not in training or validation. Various evaluation measures are used to give an all-inclusive evaluation, such as accuracy, precision, recall, F1-score, and receiver operating characteristic area under the curve. The confusion matrix is also examined to measure the true positive, true negative, false positive and false negative prediction. The imbalanced nature of the data makes this multi-metric analysis especially important. The dataset is divided into fixed training, validation, and test splits to achieve reproducibility and all experiments are performed with the same preprocessing and augmentation settings. A generator-based pipeline is used to make sure that transformations are implemented in a systematic manner when training. Experiments are held constant in terms of batch size, input resolution and model architecture. Moreover, attention is paid to making the test set immune to any training or validation processes and data leakage is avoided, which guarantees the performance estimation reliability.The experimental design is a realistic situation in which the size of the dataset is small and there is class imbalance. Although generalization is enhanced with data augmentation, no specific methods of imbalance handling are employed, potentially affecting the bias of the model in favor of the majority. Moreover, the trade-off between sensitivity and specificity may be influenced by using a fixed decision threshold in making a classification. These are some of the factors taken into consideration when interpreting the results and show areas where future work can be improved.

3. Results and Discussion

Overall Performance on Held-Out Test Set

The proposed MobileNet-based framework was evaluated on an independent held-out test set comprising 919 chest X-ray images. The model achieved an overall accuracy of 86.07%, with a precision of 83.88%, recall of 99.06%, and F1-score of 90.84%. These results indicate strong performance in detecting pneumonia cases, particularly reflected in the high recall and F1-score values. The high recall demonstrates that the model is highly effective in identifying positive pneumonia cases, which is critical in clinical screening scenarios where missed diagnoses can have severe consequences.

Confusion Matrix Analysis

A detailed analysis of the confusion matrix reveals that the model correctly classified 635 pneumonia cases and 156 normal cases, while misclassifying 6 pneumonia cases as normal and 122 normal cases as pneumonia (Figure 3). The very low number of false negatives confirms the model's strong sensitivity, ensuring that most pneumonia cases are successfully detected. However, the relatively high number of false positives indicates that a substantial portion of normal cases is incorrectly identified as pneumonia. This imbalance highlights a limitation in the model's ability to distinguish normal radiographic patterns from pathological features.

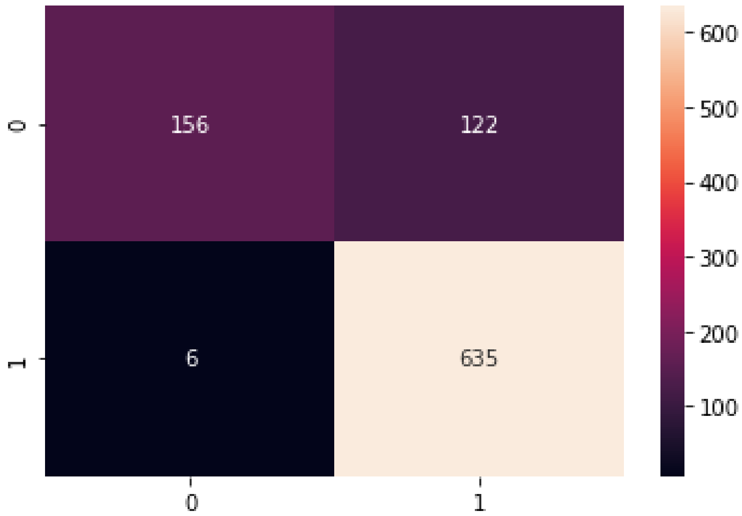

Figure 3. Confusion matrix of the proposed MobileNet model on the held out test set

The model exhibits a clear imbalance between sensitivity and specificity. While sensitivity (recall) reaches 99.06%, indicating near-complete detection of pneumonia cases, the specificity is considerably lower due to the high number of false positives. This trade-off suggests that the model is biased toward predicting the pneumonia class, which is consistent with the underlying class imbalance in the dataset. From a clinical perspective, such behavior may be acceptable in early screening systems where minimizing missed cases is prioritized. However, for diagnostic applications, the high false positive rate may lead to unnecessary follow-up procedures and reduced trust in automated predictions. The receiver operating characteristic curve demonstrates strong discriminative capability, with an area under the curve of 0.96 (Figure 4). This indicates that the model is highly effective at ranking pneumonia and normal cases across different decision thresholds. The high AUC value suggests that the learned feature representations are informative and capable of separating the two classes. However, despite this strong separability, the observed classification performance at the default decision threshold reveals suboptimal balance between false positives and false negatives. This indicates that the operating threshold plays a critical role in determining final classification behavior.

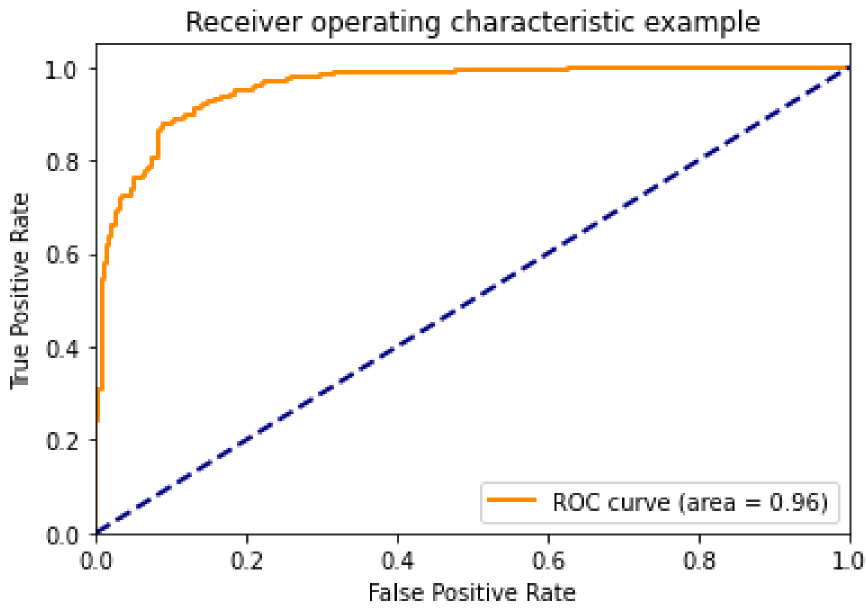

Figure 4: Receiver operating characteristic curve of the proposed model with area under the curve of 0.96

Training and Generalization Behavior

The model achieved a training accuracy of 96.27%, which is notably higher than the test accuracy of 86.07%. This gap suggests the presence of overfitting, where the model captures training data patterns more effectively than unseen data (Figure 5-6). This behavior is consistent with the relatively small dataset size and class imbalance, which can limit generalization. Although data augmentation is applied during training to mitigate overfitting, the results indicate that additional regularization or more balanced data handling strategies may further improve generalization performance.

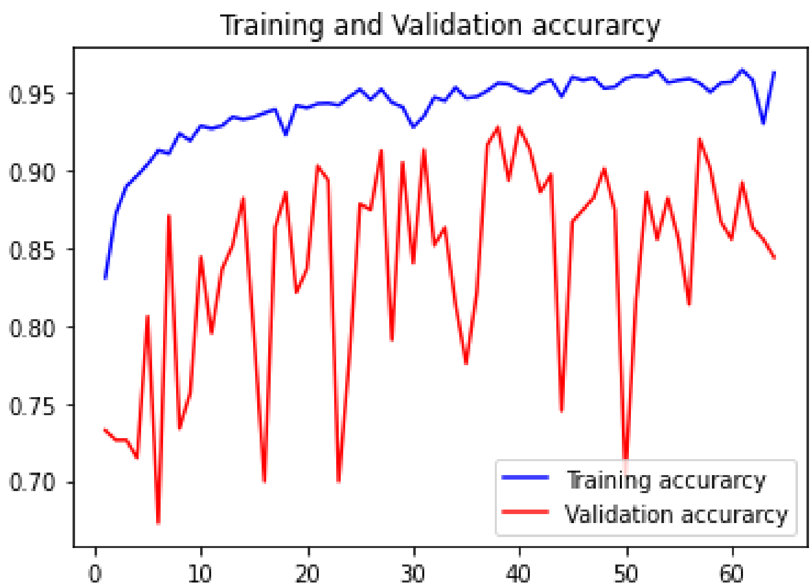

Figure 5. Training and validation accuracy curves over epochs

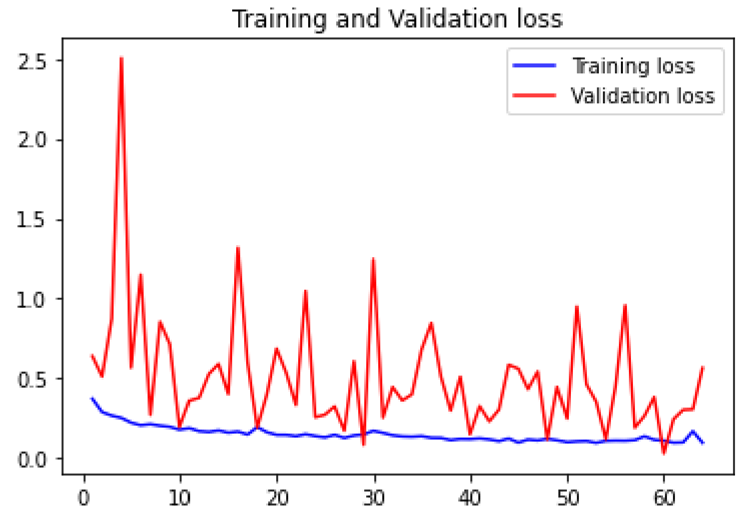

Figure 6. Training and validation loss curves over epochs

The imbalanced nature of the dataset plays a significant role in shaping model behavior. The higher proportion of pneumonia images encourages the model to favor positive predictions, contributing to the high recall and increased false positive rate. Additionally, variability in chest X-ray image quality, contrast, and anatomical presentation adds complexity to the classification task. These factors collectively influence the model's ability to generalize across diverse samples and highlight the importance of balanced data and robust preprocessing strategies.From an application standpoint, the proposed model demonstrates strong potential as a screening tool for pneumonia detection. The high sensitivity ensures that the majority of pneumonia cases are identified, which is critical in clinical workflows aimed at early detection. However, the elevated false positive rate suggests that the model should be used in conjunction with clinical assessment or as a preliminary screening step rather than a standalone diagnostic system. Adjusting the decision threshold or incorporating additional.

Limitations and Future Work

The current research is limited in a number of ways in terms of the properties of the dataset and generalizing the model. The data is adequate to perform preliminary experiments, but it is relatively small and is not representative of the range of clinical environments, such as differences in imaging equipment, acquisition guidelines, and patient populations. Moreover, the dataset is obviously skewed, with cases of pneumonia being much higher than normal cases. This disproportion leads to a bias in prediction to the majority class, as indicated by the high recall and comparatively low specificity found in the results. This is further enhanced by the lack of explicit imbalance-handling methods, like class weighting or resampling. In addition, the difference between training accuracy (96.27) and test accuracy (86.07), means that the model is overfitting and it is picking up training-specific patterns that do not necessarily apply to unseen data.

The proposed framework should be strengthened and become more clinically applicable in the future. The use of bigger and more multi-institutional data would contribute to generalization and diminish the bias within datasets. The combination of the strategies of class imbalance handling, including focal loss or cost-sensitive learning, can be used to obtain a more optimal balance between sensitivity and specificity. Also systematic threshold optimization which is dependent on the receiver operating characteristics may enhance the classification in the real world. Modeling-wise, the more complicated architectures or hybrid-style could be investigated, as well as more aggressive regularization strategies, which could further decrease overfitting. Lastly, the need to validate the proposed system on external data and test it in actual clinical settings are critical steps toward determining the reliability and deployment readiness of the proposed system.

4. Conclusion

This paper proposed a transfer learning-based system to detect pneumonia on chest X-ray images based on a pretrained MobileNet model. The model was trained on an imbalanced dataset consisting of 5,856 images and tested on a separate held-out test set to guarantee the ability to assess the performance without bias. The findings show good detection, high recall, and good F1-score, which means that pneumonia cases are detected effectively. The receiver operating characteristic analysis also substantiates the discriminative performance of the model to classify the pneumonia and normal classes.Although these strengths exist, the results also show that the sensitivity and specificity are imbalanced noteworthy with a comparatively high false positive rate that influences the overall accuracy. This is related to the imbalance of datasets and threshold-based decision settings. The results indicate that the suggested framework can be especially applicable to screening-based applications in which it is essential to reduce the number of missed cases of pneumonia. All in all, the research offers a systematic review of a lightweight deep learning model to classify pneumonia and its practical advantages as well as points at which the model needs to be improved to be trusted to be used in clinical practice.

REFERENCES

[1] H. E. Kim, A. Cosa-Linan, N. Santhanam, M. Jannesari, M. E. Maros, and T. Ganslandt, “Transfer learning for medical image classification: A literature review,” *BMC Medical Imaging*, vol. 22, no. 1, p. 69, 2022.

[2] L. Alzubaidi, M. A. Fadhel, O. Al-Shamma, J. Zhang, J. Santamaría, Y. Duan, and S. R. Oleiwi, “Towards a better understanding of transfer learning for medical imaging: A case study,” *Applied Sciences*, vol. 10, no. 13, p. 4523, 2020.

[3] H. Ravishankar, P. Sudhakar, R. Venkataramani, S. Thiruvenkadam, P. Annangi, N. Babu, and V. Vaidya, “Understanding the mechanisms of deep transfer learning for medical images,” in *Proc. Int. Workshop Deep Learning in Medical Image Analysis*, 2016, pp. 188–196.

[4] M. Raghu, C. Zhang, J. Kleinberg, and S. Bengio, “Transfusion: Understanding transfer learning for medical imaging,” in *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[5] Y. Yu, H. Lin, J. Meng, X. Wei, H. Guo, and Z. Zhao, “Deep transfer learning for modality classification of medical images,” *Information*, vol. 8, no. 3, p. 91, 2017.

[6] P. Dutta, P. Upadhyay, M. De, and R. G. Khalkar, “Medical image analysis using deep convolutional neural networks: CNN architectures and transfer learning,” in *Proc. Int. Conf. Inventive Computation Technologies (ICICT)*, 2020, pp. 175–180.

[7] H.-C. Shin *et al*., “Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning,” *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.

[8] S. Deepak and P. M. Ameer, “Brain tumor classification using deep CNN features via transfer learning,” *Computers in Biology and Medicine*, vol. 111, p. 103345, 2019.

[9] F. Zhuang *et al*., “A comprehensive survey on transfer learning,” *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, 2020.

[10] N. Tajbakhsh *et al*., “Convolutional neural networks for medical image analysis: Full training or fine tuning?,” *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1299–1312, 2016.

[11] S. M. Anwar *et al*., “Medical image analysis using convolutional neural networks: A review,” *Journal of Medical Systems*, vol. 42, no. 11, p. 226, 2018.

[12] S. Shao, S. McAleer, R. Yan, and P. Baldi, “Highly accurate machine fault diagnosis using deep transfer learning,” *IEEE Transactions on Industrial Informatics*, vol. 15, no. 4, pp. 2446–2455, 2018.

[13] S. S. Yadav and S. M. Jadhav, “Deep convolutional neural network based medical image classification for disease diagnosis,” *Journal of Big Data*, vol. 6, no. 1, pp. 1–18, 2019.

[14] J. Ker, L. Wang, J. Rao, and T. Lim, “Deep learning applications in medical image analysis,” *IEEE Access*, vol. 6, pp. 9375–9389, 2017.

[15] K. Suzuki, "Overview of deep learning in medical imaging," *Radiological Physics and Technology*, vol. 10, no. 3, pp. 257–273, 2017.

[16] M. H. Hesamian, W. Jia, X. He, and P. Kennedy, "Deep learning techniques for medical image segmentation: Achievements and challenges," *Journal of Digital Imaging*, vol. 32, no. 4, pp. 582–596, 2019.

[17] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," *Annual Review of Biomedical Engineering*, vol. 19, pp. 221–248, 2017.

[18] D. S. Kermany *et al*., "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, no. 5, pp. 1122–1131, 2018.

[19] H. A. Khan, J. Wu, M. Mushtaq, and M. U. Mushtaq, "Brain tumor classification in MRI image using convolutional neural network," *Mathematical Biosciences and Engineering*, vol. 17, no. 5, p. 6203, 2020.

[20] J.-G. Lee *et al*., "Deep learning in medical imaging: General overview," *Korean Journal of Radiology*, vol. 18, no. 4, pp. 570–581, 2017.