International Journal of Recent Advances in Multidisciplinary Topics
Volume 6, Issue 9, September 2025
https://www.ijramt.com | ISSN (Online): 2582-7839

A Deep Learning Approach to Multiclass Pneumonia Detection in Chest X-ray Images

Timothy Karani Mwenda[1*], Stephen Titus Waithaka[2]
[1]*Student, Department of Computing and Information Science, Kenyatta University, Nairobi, Kenya*
[2]*Senior Lecturer, Department of Computing and Information Science, Kenyatta University, Nairobi, Kenya*

Abstract*: Pneumonia is a significant cause of mortality worldwide, particularly in children under five years, with detection from chest X-ray (CXR) images remaining challenging due to diagnostic errors common in manual radiographic analysis. This study develops a deep learning model for multiclass pneumonia classification in CXR images that is computationally inexpensive and suitable for resource-constrained settings. The proposed approach utilises pretrained models, including EfficientNet, MobileNet, RegNet, and ViT, fine-tuned using the PyTorch framework, with data augmentation and regularisation techniques applied to address class imbalance and overfitting. Using a dataset of 5,863 CXR images classified as normal, bacterial pneumonia, or viral pneumonia, the fine-tuned models achieved high classification accuracy, with ConvNeXt and EfficientNet attaining accuracy scores of 83% and 82% respectively. The findings demonstrate how data augmentation and regularisation significantly improved the models' generalisability, reducing overfitting and improving predictive performance. This work provides healthcare professionals with an efficient tool for multiclass pneumonia detection from CXR images suitable for resource-constrained settings.

***Keywords*: Artificial Intelligence (AI), class imbalance, multiclass detection, pneumonia detection, radiography, transfer learning.**

1. Introduction

Pneumonia is a respiratory infection that predominantly affects the lungs, causing inflammation of the alveoli and bronchial tree. It is commonly categorised into two types: community-acquired pneumonia (CAP) and hospital-acquired pneumonia (HAP) [1]. Globally, pneumonia remains one of the leading causes of death in children under the age of five, particularly in sub-Saharan Africa. In 2015, it accounted for 920,000 of the 5.9 million deaths in this age group, with countries such as Nigeria, the Democratic Republic of Congo, and Angola among the worst affected [2]. The number reduced to 740,000 in 2019 [3]. Although this represents significant progress, pneumonia continues to be a public health burden, especially in low-resource settings.

Traditionally, the diagnosis of pneumonia has relied on clinical evaluations and radiographic techniques such as chest X-ray (CXR) images. However, the manual interpretation of CXR images is susceptible to perceptual and interpretive errors, often resulting in inaccurate diagnoses. These errors are exacerbated by the high workload of radiologists and the limitations of human perception in detecting subtle variations in medical images [4]. Recent advancements in deep learning and computer vision have created new opportunities for automating the analysis of medical images to provide more accurate and timely diagnoses. However, significant obstacles persist, including class imbalance in medical datasets, overfitting due to large parameter counts in deep neural networks, and the high computational costs associated with training complex models [5]. Addressing these challenges is essential to developing reliable and efficient models that can be deployed in clinical settings, especially in low-resource environments where computational power and high-quality datasets may be limited.

This study aims to develop a deep learning model for multiclass pneumonia classification in CXR images. By employing transfer learning, data augmentation, and regularisation techniques, the model seeks to improve the accuracy and generalisability of pneumonia detection, particularly in differentiating between normal, bacterial pneumonia, and viral pneumonia cases. The research also investigates the performance of lightweight deep learning models that can be trained and deployed efficiently without sacrificing accuracy.

2. Literature Review

A. Overview of Deep Learning in Medical Imaging

Deep learning has revolutionised medical imaging, providing advanced methods for automated analysis and classification of complex medical data. Among the most effective models in this field are convolutional neural networks (CNNs), which have achieved significant success in tasks like detecting pneumonia, tumours, and other medical conditions. CNNs, by learning image features directly from raw pixel data, offer greater accuracy than traditional methods, particularly when analysing CXR images [6].

For pneumonia detection, CNNs have consistently demonstrated high accuracy in distinguishing between normal and infected lungs, addressing some of the challenges faced by radiologists. This capability is crucial in resource-constrained settings, where access to specialised healthcare professionals is

*Corresponding author: kartimothy@gmail.com

limited. Recent studies have shown that CNN-based models like DenseNet and ResNet can enhance diagnostic efficiency and minimise human error in interpreting CXR images [7].

Despite their success, deep learning models in medical imaging still face challenges, particularly in handling small datasets and imbalanced classes [8]. Overfitting is common, especially when training deep networks on limited data. To combat this, researchers often employ techniques like data augmentation and transfer learning to enhance model generalisation [5]. When integrated with CNNs, these approaches reduce model complexity and improve performance across various medical imaging tasks.

B. Deep Learning Models for Pneumonia Detection

Several state-of-the-art models, including ResNet, DenseNet, EfficientNet and MobileNet, have been widely adopted in pneumonia detection tasks. These models are often pretrained on large-scale datasets such as ImageNet, allowing them to learn to capture various image features before being fine-tuned on medical datasets. For example, Rahman et al. [9] successfully fine-tuned DenseNet201 for the multiclass classification of normal, bacterial pneumonia, and viral pneumonia, achieving an accuracy of 95%. Other studies have used models like MobileNet due to its higher computational efficiency, achieving the same accuracy score [10]. However, most research focuses on binary classification, which ignores the challenge of distinguishing between normal, bacterial pneumonia and viral pneumonia. This gap has motivated the current study to focus on multiclass classification for improved diagnostic accuracy.

C. Transfer Learning and Regularisation in Medical Imaging

Transfer learning is a widely used deep learning technique in which models trained on general datasets, such as ImageNet, are fine-tuned for specific tasks. This method can save time and computational resources and is especially useful in medical imaging due to the scarcity of large, labelled datasets. Hashmi et al. [11] demonstrated the effectiveness of fine-tuning pretrained models in pneumonia classification, achieving an accuracy as high as 98.43%.

In addition to transfer learning, regularisation techniques such as weight decay have been employed to prevent overfitting, particularly in deep models with numerous parameters. These techniques aim to mitigate overfitting by limiting the number of features that the model needs to learn. Dropout, along with L1 and L2 regularisation, are examples of such methods. Weight decay, a form of L2 regularisation, penalises large weights, thus promoting simpler models that generalise better to unseen data [12].

D. Data Augmentation for Addressing Class Imbalance

Class imbalance refers to a situation where the sizes of classes in a dataset differ significantly. It can be classified into multiclass and binary data imbalance. A binary dataset comprises only two classes, whereas a multiclass dataset includes more than two [13]. While researchers have proposed numerous solutions for the binary data imbalance problem, various issues related to multiclass data imbalance remain unresolved [14].

Class imbalance is a common issue in medical imaging datasets. If not addressed, it can bias models towards overrepresented classes [8]. Data augmentation stands out as one of the most effective techniques for tackling this problem. By applying transformations such as rotation, scaling, and flipping, the training dataset can be artificially expanded, thereby enhancing model generalisation.

E. Multiclass Pneumonia Classification

Although most existing research has focused on binary classification, this study aims to address the more complex task of multiclass pneumonia classification by distinguishing between normal, bacterial pneumonia, and viral pneumonia CXR images using deep learning. As mentioned previously, Rahman et al. [9] used DenseNet201 to achieve a multiclass classification accuracy of 95%. However, DenseNet201's high computational requirements make it impractical for real-time use in resource-constrained environments. In this research, we explore the use of more lightweight models such as EfficientNet and ConvNeXt, which require less computational power while still delivering high accuracy. Results from previous research suggest that smaller, less resource-intensive models can be just as effective in pneumonia detection, offering a more accessible solution for real-time clinical use [15].

F. Related Works

Various studies have explored the use of deep learning models for pneumonia detection, with some focusing on ensembles of multiple models to improve accuracy. Hashmi et al. [11] combined five state-of-the-art models in a weighted ensemble and achieved an accuracy of 98.43%. However, combining multiple models significantly increases computational requirements, making these solutions less feasible for real-time deployment. Varshni et al. [16] evaluated a DenseNet model for binary pneumonia classification and achieved an area under the curve (AUC) score of 80.02%. However, their study was limited to binary classification, and the model's complexity and computational cost limit its scalability. In contrast, this study focuses on optimising single models like ConvNeXt and EfficientNet for multiclass classification to ensure high accuracy and computational efficiency.

G. Gaps in Existing Research

Despite significant progress in using deep learning for pneumonia detection, several gaps remain. First, most studies have concentrated on binary classification, neglecting the need for models that distinguish between normal, bacterial pneumonia and viral pneumonia. Additionally, class imbalance has been inadequately addressed in many studies, leading to models that perform well on the majority class but poorly on minority classes.

This study seeks to fill these gaps by focusing on multiclass classification and addressing class imbalance through data augmentation and regularisation techniques. Using lightweight models like EfficientNet and ConvNeXt also addresses the need for efficient models that can be deployed in low-resource settings, providing a practical solution for real-time pneumonia detection in clinical environments.

3. Research Methods and Design

A. Study Design

This study employed a quantitative research design, using deep learning models to perform multiclass pneumonia classification of CXR images. The performance of each model was evaluated on the basis of established metrics such as accuracy, precision, recall, and specificity.

B. Study Setting

The research was conducted in a controlled computational environment with data sourced from the publicly available Labelled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification dataset, which contains 5,863 labelled CXR images across three classes: normal, bacterial pneumonia, and viral pneumonia [17]. No personal or identifying information is included in the dataset, ensuring confidentiality and compliance with ethical standards. The experiments were performed on an Nvidia RTX A6000 GPU with 10,752 CUDA cores and 49 GB of VRAM. This was installed in a machine with an Intel Xeon Gold 5315Y 3.20 GHz CPU on a machine with 45 GB of RAM.

C. Sampling

The CXR dataset has 5,863 labelled samples and was randomly split into training, validation, and test sets. Seventy percent of the images were used for training, 15% for validation, and 15% for testing. To address the class imbalance, in which the number of bacterial pneumonia images exceeds the number of both viral pneumonia and normal images, data augmentation techniques were applied to increase the number of samples of the minority classes in the training set.

D. Instruments and Data Collection

The primary instruments used in this research were deep learning algorithms implemented using the PyTorch machine learning framework. Training models from scratch, however, is expensive, so pretrained models were employed [18]. Eight pretrained models, namely versions of ConvNeXt, EfficientNet, MobileNet, RegNet, and ViT, were fine-tuned for pneumonia detection using transfer learning [19]. The dataset was augmented through rotation, horizontal and vertical flips, and scaling to increase the variability of the training data and help mitigate class imbalance.

E. Data Preprocessing

The training data were pre-processed by applying data augmentation techniques to alleviate class imbalance. Horizontal and vertical flips, rotation (up to 45°), and scaling were used to artificially increase the size of the minority classes (normal images and viral pneumonia images). The images were then scaled down and normalised to meet the specific input demands of the selected pretrained models.

F. Model Training and Evaluation

To build the DL model, one needs to select an appropriate algorithm. One identifies network architectures that are relevant to the domain space of the target problem, for example, CNNs for image-based problems or Long Short-Term Memory (LSTM) for text or sequenced data. Training models from scratch, however, is expensive, so the use of pretrained models can help in this regard [18].

The performance of the deep learning models is evaluated using the cross-entropy as the loss function and the Adam optimiser for optimisation. The models are fine-tuned for 32 epochs, with weight decay employed to prevent overfitting. Five key performance metrics are calculated to assess the models' ability to classify the normal, bacterial pneumonia, and viral pneumonia cases: accuracy, sensitivity, precision, specificity, and F1 score.

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

$$\text{Sensitivity} = \frac{\text{TP}}{(\text{TP+FP})} \tag{2}$$

$$\text{Precision} = \frac{\text{TP}}{(\text{TP+FP})} \tag{3}$$

$$\text{Specificity} = \frac{\text{TN}}{(\text{FP+TN})} \tag{4}$$

$$\text{F1 Score} = \frac{(2*\text{TP})}{(2*\text{TP} + \text{FN} + \text{FP})} \tag{5}$$

4. Results

This section presents the key findings from the experiments conducted with different deep learning models on the multiclass pneumonia classification task. The models were evaluated according to their performance in classifying normal, bacterial pneumonia, and viral pneumonia CXR images, as measured using the accuracy, precision, recall, specificity, and F1 score. The experiments were conducted across three distinct groupings, resulting in a total of eight possible combinations. The groupings are:

- *Dataset type*: base vs augmented data.
- *Model type*: feature extractor vs fine-tuned
- *Regularisation*: with or without weight decay.

A. Dataset and Augmentation

The dataset utilised in this study consists of 5,863 CXR images classified into three categories, as illustrated in Figure 1: normal (1,183 samples), bacterial pneumonia (2,380 samples), and viral pneumonia (1,093 samples). Augmentation techniques were applied to enhance the dataset, expanding each class to 5,000 samples to address class imbalance. The new images were obtained by applying horizontal and vertical flipping, rotation (up to 45°), and scaling to images from the original dataset. The data were then split into training (70%), validation (15%), and test (15%) sets. Figure 2 shows sample CXR images before and after augmentation. This augmented dataset improved model generalisation and helped mitigate overfitting.

B. Model Performance

Eight pretrained deep learning models – EfficientNet (B0, B2), MobileNet (v3 small, v3 large), RegNet (x1.6gf, y1.6gf), ConvNeXt Base, and ViT (B_16) – were fine-tuned on the augmented dataset with weight decay enabled. Table 1 presents the results obtained for each model, showing the accuracy, specificity, precision, recall, and F1 score.

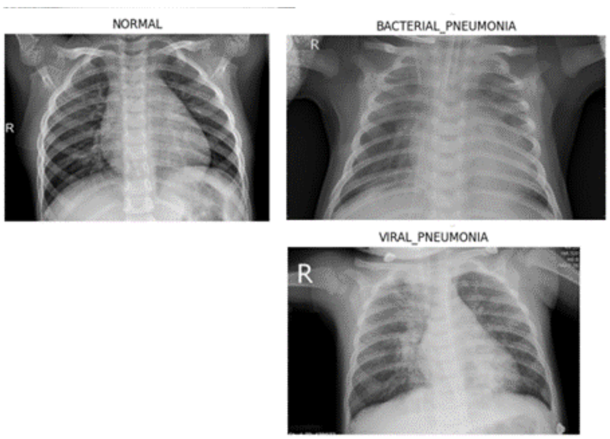

Fig. 1. Samples from the base dataset

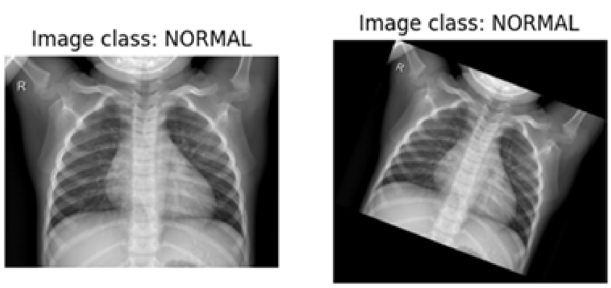

Fig. 2. Sample images after undergoing transformations

As shown in Table 2, feature-extractor models trained on the augmented dataset exhibited poor performance. This can be attributed to the limited number of trainable parameters, which led to overfitting on the dataset. Feature extractors performed optimally when trained on the base dataset. The highest performance was achieved by fine-tuning the models on the augmented dataset. Additionally, this performance was enhanced by incorporating regularisation in the form of weight decay, with EfficientNet B0 attaining an accuracy of 82% and ConvNeXt Base achieving 83%.

C. Loss and Accuracy Curves

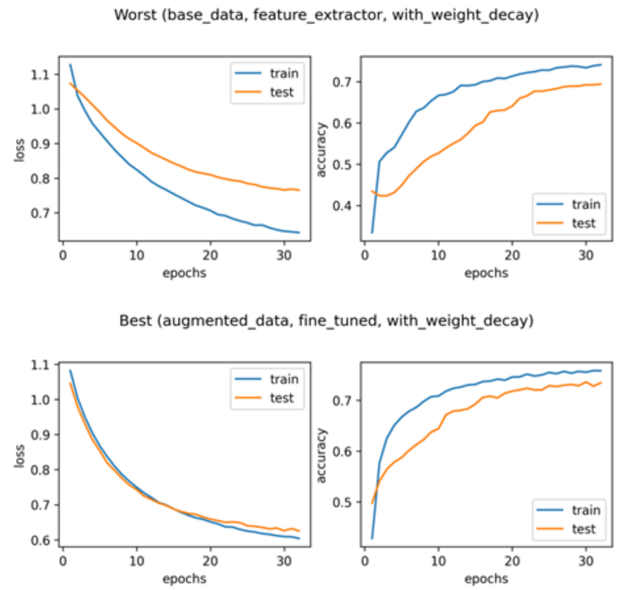

Fig. 3. Training and validation curves for the ConvNeXt Base model

Figure 3 shows training and validation loss and accuracy curves for the ConvNeXt Base model. These indicate that the model converged well, with some significant overfitting observed in the training and validation steps. Weight decay was employed during training to prevent overfitting, and the model that performed best on the validation set was selected as the final model.

EfficientNet B0 performed the worst when training a feature extractor on the base dataset with weight decay turned on. The best performance was achieved when fine-tuning the model on the augmented dataset with weight decay turned on. As demonstrated in Figure 4, neither configuration suffered from

Table 1
Model performance on the validation dataset

Model	Accuracy (%)	Specificity (%)	Precision (%)	Recall (%)	F1 Score (%)
EfficientNet B0	82.0	91.0	80.5	83.0	81.7
EfficientNet B2	81.8	90.9	80.3	82.8	81.5
MobileNet v3 small	81.0	90.5	79.8	81.0	80.4
MobileNet v3 large	80.0	90.0	79.5	80.0	79.7
RegNet x1.6gf	79.4	89.7	78.6	79.4	79.0
RegNet y1.6gf	77.6	88.8	77.0	77.6	77.3
ConvNeXt Base	83.0	91.5	82.0	83.5	82.7
ViT B_16	80.1	90.1	79.2	80.1	79.6

Table 2
The performance of EfficientNet B0 across the eight experiments

Dataset	Model Type	Regularisation (Weight Decay)	Accuracy (%)	Specificity (%)
augmented	fine tuned	on	82.0	91.0
augmented	fine tuned	off	80.8	90.4
base	fine tuned	off	77.4	88.7
base	fine tuned	on	77.2	88.6
base	feature extractor	off	76.8	88.4
base	feature extractor	on	76.8	88.4
augmented	feature extractor	off	75.6	87.8
augmented	feature extractor	on	75.5	87.8

significant overfitting.

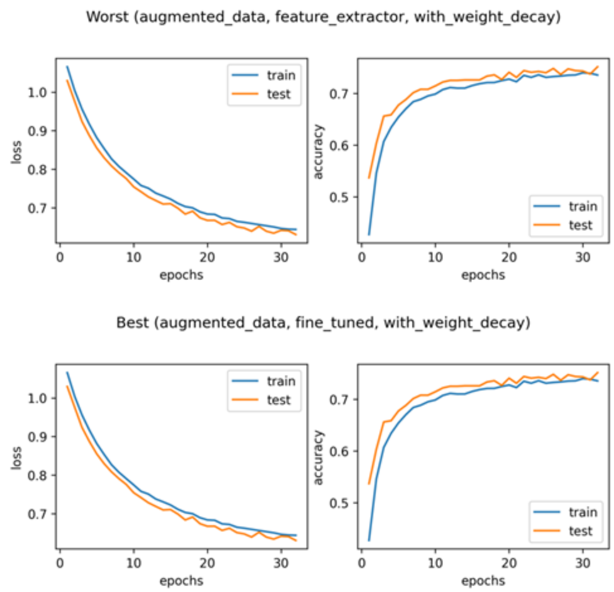

Fig. 4. Training and validation curves for EfficientNet B0

D. *Comparison with State-of-the-Art Models*

Table 3
Comparison with State-of-the-Art models

Reference	Architecture	Accuracy (%)
Hammoudi et al. [7]	DenseNet169	95.72
	VGG19	82.66
	ResNet + RNN1	78.16
	ResNet + RNN2	79.51
Dokur et al. [20]	CNN	77.72
	CNN x2 + InceptionV3	75.14
Ukwuoma et al. [21]	Transformer encoder	98.19
This research	ConvNeXt Base	83.00
	EfficientNet B0	82.00

The performance of the ConvNeXt Base model was compared with state-of-the-art models from related studies. Table 3 compares the best-performing model from this study with those in previous works. Although the ConvNeXt Base model did not achieve the same level of accuracy as some of the more complex models used in previous studies, it provides a balance between accuracy and computational efficiency. Thus, it is more suitable for deployment in low-resource settings with limited computational resources.

E. *Summary of Results*

The ConvNeXt Base model attained the highest accuracy (83.0%) and specificity (91.5%), proving a promising tool for multiclass pneumonia detection. The use of transfer learning, data augmentation, and regularisation techniques contributed to the improved generalisability and performance of the model, particularly in addressing class imbalance and reducing overfitting. The results also highlight the challenge of distinguishing between bacterial and viral pneumonia, suggesting that further model refinement is needed for more precise classification.

5. Discussion

A. *Transfer Learning*

The results of this study demonstrate the effectiveness of transfer learning in multiclass pneumonia detection. Fine-tuning pretrained models made it possible to achieve high accuracy scores without needing to train models from scratch, which is beneficial in medical domains where large labelled datasets are scarce. This approach consistently yielded better results than feature extraction alone. For instance, after fine-tuning, the ConvNeXt Base model achieved the highest accuracy (83.0%) and specificity (91.5%), outperforming the feature extractor configurations. These findings align with previous research that has demonstrated the advantages of fine-tuning in improving model generalisation, particularly when applied to small datasets [22].

B. *Data Augmentation*

Data augmentation proved crucial in addressing class imbalance and reducing overfitting in this study. As seen with models like EfficientNet B0, applying augmentation techniques such as flipping and rotation improved the performance metrics significantly. Without data augmentation, models such as MobileNet v3 large experienced overfitting, as evidenced by the lower test set accuracy and specificity scores. For EfficientNet B0, data augmentation increased its accuracy from 77.2% to 82.0% and specificity from 88.6% to 91.0%. This improvement aligns with the literature, in which data augmentation has been shown to enhance the robustness of deep learning models in imbalanced datasets [23].

C. *Regularisation and Weight Decay*

Regularisation techniques, specifically weight decay, were also explored in this study. The application of weight decay reduced overfitting, particularly in models with higher parameter counts. For example, when weight decay was applied, the accuracy of EfficientNet B0 improved from 80.8% to 82.0% and its specificity from 90.4% to 91.0%. Similarly, the ConvNeXt Base model exhibited a marked increase in performance with weight decay, demonstrating the importance of regularisation in optimising deep learning models for medical image classification. These results support the findings of previous studies that emphasise the role of weight decay in preventing overfitting in neural networks [12].

D. *Comparison with State-of-the-Art Models*

Although promising, the results achieved in this study did not surpass those of some state-of-the-art models, such as DenseNet169, which achieved an accuracy of 95.72 %.7 Nevertheless, regarding accuracy and specificity, the ConvNeXt Base model outperformed several other models, such as VGG19 and ResNet. Although ConvNeXt Base achieved the highest accuracy (83%), EfficientNet B0 was only slightly behind (82%) while being more computationally efficient. This finding supports the growing body of research suggesting that smaller models can be effective in medical image classification tasks, particularly in resource-constrained environments.

E. Limitations

Although this research achieved promising results, there are several limitations to consider. First, the dataset used in this study is relatively small compared with other large-scale datasets used in medical image analysis. Although data augmentation helped mitigate this problem, a more extensive and diverse dataset would provide more robust results. Additionally, the study focused on pretrained models and did not explore models trained from scratch, which could lead to different outcomes. Finally, although weight decay enhanced the performance of most models, further experimentation with other regularisation techniques, such as dropout, could lead to greater improvements.

6. Conclusion

This study explored the application of deep learning models to multiclass pneumonia classification in CXR images. We aimed to enhance model performance while addressing class imbalance and overfitting by incorporating transfer learning, data augmentation, and regularisation techniques such as weight decay. The best-performing model, ConvNeXt Base, achieved accuracy and specificity scores of 83.0% and 91.5%, respectively.

Data augmentation proved a crucial technique for improving model generalisation, particularly in mitigating the effects of overfitting. Fine-tuning the pretrained models on the augmented dataset led to significant improvements in accuracy and specificity. Weight decay greatly assisted in reducing overfitting, particularly in models with a higher number of trainable parameters.

This research confirmed the effectiveness of using pretrained models for multiclass pneumonia classification. Models like EfficientNet B0, despite being smaller, exhibited performance close to that of larger models like ConvNeXt Base, which suggests that model size does not necessarily correlate with performance. Pretrained models provide a cost-effective way to achieve high accuracy in medical imaging tasks, making them highly suitable for resource-constrained environments.

Future research could involve more detailed experiments tracking learning rate schedules and weight decay configurations. These parameters can have a significant impact on model performance, and studying their effects over time would provide deeper insights into the optimal configuration for multiclass pneumonia detection. Additionally, incorporating learning rate schedulers may improve training efficiency, reducing the time required to reach optimal performance.

References

[1] A. Torres et al., "Pneumonia," *Nat. Rev. Dis. Primer*, vol. 7, no. 1, pp. 25, Apr. 2021.

[2] L. Liu et al., "Global, regional, and national causes of under-5 mortality in 2000–15: an updated systematic analysis with implications for the Sustainable Development Goals," *The Lancet*, vol. 388, no. 10063, pp. 3027–3035, Dec. 2016.

[3] J. Perin et al., "Global, regional, and national causes of under-5 mortality in 2000–19: an updated systematic analysis with implications for the Sustainable Development Goals," Lancet Child Adolesc. Health, vol. 6, no. 2, pp. 106–115, Feb. 2022.

[4] G. Maskell, "Error in radiology—where are we now?," *Br. J. Radiol*., vol. 92, no. 1096, pp. 20180845, Apr. 2019.

[5] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep Learning for Computer Vision: A Brief Review," *Comput. Intell. Neurosci*., vol. 2018, pp. 1–13, 2018.

[6] W. G. Hatcher and W. Yu, "A Survey of Deep Learning: Platforms, Applications and Emerging Research Trends," *IEEE Access*, vol. 6, pp. 24411–24432, 2018.

[7] K. Hammoudi et al., "Deep Learning on Chest X-ray Images to Detect and Evaluate Pneumonia Cases at the Era of COVID-19," *J. Med. Syst.*, vol. 45, no. 7, pp. 75, June 2021.

[8] Z. Li, K. Kamnitsas, and B. Glocker, "Overfitting of neural nets under class imbalance: Analysis and improvements for segmentation," arXiv:1907.10982, Oct. 2019.

[9] T. Rahman et al., "Transfer Learning with Deep Convolutional Neural Network (CNN) for Pneumonia Detection Using Chest X-ray," *Appl. Sci.*, vol. 10, no. 9, pp. 3233, May 2020.

[10] M. S. A. Reshan et al., "Detection of Pneumonia from Chest X-ray Images Utilizing MobileNet Model*," Healthcare*, vol. 11, no. 11, pp. 1561, Jan. 2023.

[11] M. F. Hashmi, S. Katiyar, A. G. Keskar, N. D. Bokde, and Z. W. Geem, "Efficient Pneumonia Detection in Chest Xray Images Using Deep Transfer Learning," *Diagnostics*, vol. 10, no. 6, pp. 417, June 2020.

[12] X. Ying, "An Overview of Overfitting and its Solutions," *J. Phys. Conf.* Ser., vol. 1168, no. 2, pp. 022022, Feb. 2019.

[13] K. M. Hasib et al., "A Survey of Methods for Managing the Classification and Solution of Data Imbalance Problem," arXiv:2012.11870, 2020.

[14] N. Rout, D. Mishra, and M. Mallick, "Handling Imbalanced Data: A Survey," in *Proc. Adv. Comput. Commun*., 2018, pp. 431–443.

[15] J. León et al., "Deep learning for EEG-based Motor Imagery classification: Accuracy-cost trade-off," *PLOS ONE*, vol. 15, no. 6, pp. e0234178, June 2020.

[16] D. Varshni, K. Thakral, L. Agarwal, R. Nijhawan, and A. Mittal, "Pneumonia Detection Using CNN based Feature Extraction," in Proc. *IEEE Int. Conf. Electr. Comput. Commun. Technol*., 2019, pp. 1–7.

[17] D. Kermany, K. Zhang, and M. Goldbaum, "Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification," vol. 2, Jan. 2018.

[18] A. Burkov, Machine learning engineering. Quebec City, Canada: True Positive Inc, 2020.

[19] O. Russakovsky et al., "ImageNet Large Scale Visual Recognition Challenge," arXiv:1409.0575, Jan. 2015.

[20] Z. Dokur, T. Olmez, and M. B. Darici, "Pneumonia Detection and Classification Using Deep Learning on Chest X-Ray Images," *Int. J. Intell. Syst. Appl. Eng.*, vol. 8, no. 4, pp. 177–183, Dec. 2020.

[21] C. C. Ukwuoma et al., "A hybrid explainable ensemble transformer encoder for pneumonia identification from chest X-ray images," *J. Adv. Res.*, vol. 48, pp. 191–211, June 2023.

[22] A. Farahani, B. Pourshojae, K. Rasheed, and H. R. Arabnia, "A Concise Review of Transfer Learning," arXiv:2104.02144, Apr. 2021.

[23] S. Gautam, M. M.-C. Höhne, S. Hansen, R. Jenssen, and M. Kampffmeyer, "Demonstrating the Risk of Imbalanced Datasets in Chest X-ray Image-based Diagnostics by Prototypical Relevance Propagation," arXiv:2201.03559, Jan. 2022.