Optimizing Pneumonia Detection from Chest X-rays: A Performance Comparison of Lightweight CNNs and Ensemble Methods

1st Yousef Taha
Computer Science Department
An-najah National University
Nablus, Palestine
tahay6700@gmail.com

2nd Yazeed Rashed
Computer Science Department
An-najah National University
Nablus, Palestine
yazeedrashed6@gmail.com

3rd Adnan Salman
Computer Science Department
An-najah National University
Nablus, Palestine
aalshaikh@najah.edu

***Abstract*—Pneumonia remains a critical global health concern, necessitating rapid and accurate diagnostic strategies to improve patient outcomes. This study evaluates lightweight deep learning models and ensemble learning techniques for pneumonia detection from chest X-ray images. Five fine-tuned convolutional neural networks (CNNs)—MobileNet, EfficientNetB0, DenseNet121, NASNetMobile, and ResNet—were optimized using a curated dataset. A soft voting ensemble was also employed to combine model predictions. However, the best-performing individual models, MobileNet and EfficientNetB0, outperformed the ensemble across key metrics, achieving accuracies of 95.1% and an AUC of 99.7%, respectively. These results highlight the robustness of lightweight CNNs as standalone diagnostic tools, with ensemble learning providing limited added value. This study offers practical insights into model selection for early and accurate pneumonia diagnosis in medical imaging.**

***Index Terms*—Pneumonia Detection, CNNs, Deep Learning, MobileNet, EfficientNet, NASNetMobile, DenseNet, ResNet**

I. Introduction

Pneumonia, an inflammatory infection of one or both lungs, leads to fluid or pus filling the alveoli, the lungs' air sacs. Various agents, including bacteria, viruses, and fungi, can cause this serious condition. Symptoms range from mild to severe, typically featuring coughing—sometimes with mucus—fever, chills, and breathing difficulties. The severity of pneumonia largely depends on the affected individual's age, overall health, and the specific pathogen involved. Recognized as a significant global health threat, pneumonia causes approximately 14% of all deaths in children under five years old worldwide, according to the World Health Organization (WHO). Prompt detection of the disease is crucial for effective treatment and improved outcomes. Diagnosis often involves a chest X-ray, and treatment generally includes antibiotics, rest, and adequate hydration. Given its potential severity, immediate medical attention for pneumonia is essential to prevent serious complications and ensure a better chance of recovery

Chest X-rays are both cost-effective and efficient for detecting pneumonia. Recent advances in deep learning and computer vision have spurred the development of models capable of identifying pneumonia from these images with high accuracy. Such models have demonstrated promising potential in aiding healthcare professionals to diagnose pneumonia more swiftly and precisely, thus improving patient outcomes. Furthermore, implementing artificial intelligence in pneumonia detection can relieve some of the pressure on healthcare systems by enhancing diagnostic efficiency and reducing errors. The ongoing research and development in this field suggest a bright future for the diagnosis and treatment of pneumonia using AI technologies. Nevertheless, concerns remain regarding the consistency of AI performance across diverse patient demographics and varying imaging techniques. There is also a risk that an overreliance on AI could lead to complacency among healthcare workers, potentially causing them to overlook vital clinical information

In this study, we assessed the efficacy of five convolutional neural network (CNN) models, each designed with unique architectures to excel in various aspects of image classification, including computational efficiency, accuracy, model size, and performance on specialized data types. Initially, we evaluated these models' individual abilities to detect pneumonia, with MobileNet and EfficientNetB0 emerging as the top-performing architectures. Subsequently, we combined the predictive outputs of all models into a soft voting ensemble method, aiming to enhance pneumonia detection by leveraging their collective strengths. However, our results revealed that the ensemble method did not surpass the best individual models in terms of detection accuracy and key performance metrics. Instead, MobileNet and EfficientNetB0 demonstrated superior standalone performance, highlighting their robustness and efficiency for this task. These findings underscore the importance of careful model selection and suggest that lightweight CNNs, when well-optimized, can provide reliable and effective solutions for pneumonia detection without requiring ensemble techniques. This work offers valuable insights for developing practical image classification algorithms in medical imaging, with the potential to improve patient outcomes through accurate and efficient diagnosis. The implementation of our models and experiments is available at GitHub link [1].

The remainder of this paper is structured as follows: Section II offers a review of the relevant literature. Section III details the methodology of our proposed technique. Section IV presents our findings, and Section V concludes the paper with a summary of our results and their implications.

II. Literature Review

The application of machine learning, particularly convolutional neural networks (CNNs), in medical imaging has seen significant advancements in recent years. CNNs have been widely adopted for their ability to automatically learn and extract features from images, making them highly effective for various image classification tasks. One prominent area of research is the use of CNN models to predict pneumonia in chest X-ray images. Pneumonia, a potentially life-threatening infection that inflames the air sacs in one or both lungs, often leads to substantial morbidity and mortality worldwide. Early and accurate detection is crucial for effective treatment and patient outcomes. Traditional methods of diagnosing pneumonia involve manual interpretation of chest X-rays by radiologists, which can be time-consuming and prone to human error. Leveraging CNNs for this task has the potential to enhance diagnostic accuracy, reduce the workload on healthcare professionals, and provide timely diagnosis in resourcelimited settings. This literature review explores various CNN architectures and their applications in the automated detection of pneumonia from chest X-ray images, highlighting their performance, challenges, and advancements in the field.

Studies have presented methods for detecting pneumonia using deep learning models such as ResNet-34 based UNet and EfficientNet-B4 based U-Net, with ensemble models improving detection accuracy and addressing class imbalance [2]. Neural network models including VGG16, VGG19, and InceptionV3 have been used to detect pneumonia in pediatric chest X-rays, achieving over 97% accuracy, illustrating the potential of CNNs and transfer learning to improve early detection and reduce mortality rates [3].

Research comparing lightweight CNN models like MobileNetV3 and ShuffleNetV2 with more complex architectures has shown that lightweight models achieve high accuracy with reduced computational requirements, making them suitable for deployment on mobile devices in resource-limited settings [4]. Various CNN models such as InceptionResNetV2, Xception, DenseNet201, and VGG19 have been evaluated for pneumonia detection, with InceptionResNetV2 achieving the highest accuracy, highlighting the importance of data preprocessing, augmentation, and fine-tuning to enhance model performance [5]. Combining image processing techniques with models like VGG-16 and VGG-19 has achieved high accuracy through effective preprocessing, underscoring its importance in enhancing deep learning model performance for pneumonia detection [6].

Ensemble models using RetinaNet and Mask R-CNN have demonstrated robustness and accuracy in pneumonia detection, highlighting the potential of deep learning for reliable medical image analysis [7]. Studies evaluating six CNN models, including GoogLeNet and LeNet, for detecting pneumonia have shown high accuracy rates, contributing to improved diagnostic processes and aiding medical professionals [8]. The novel 'ResNetChest' model, a modified ResNet-50 architecture, achieved superior accuracy of 97.65% compared to models like CheXnet and traditional CNN architectures, emphasizing its effectiveness for pneumonia detection from chest X-rays and potential to aid radiologists by providing accurate diagnostic support [9].

An innovative CNN framework leveraging Process Convolution and Dual Feedback blocks demonstrated a mean accuracy of 97.78%, with sensitivity of 98.84% and specificity of 95.04%, outperforming traditional models like VGG19 and ResNet50 while operating with fewer parameters, making it suitable for memory-constrained mobile platforms [10]. The effectiveness of CNNs combined with data augmentation techniques was also highlighted, with geometric data augmentation achieving the highest performance, underscoring the potential of CNNs in medical image analysis and the significant impact of data augmentation in improving model performance [11].

III. Methodology

This paper outlines a multi-step process, illustrated in the accompanying block diagram in Figure 1. Initially, we discuss the collection and preprocessing of the dataset. Subsequently, the dataset is partitioned into training and test subsets. The next phase involves training the five models on the training set and individually assessing their performance on the test set. Following this, an ensemble of these models is formed and evaluated. Finally, the performance of the ensemble model, as well as that of each individual model, is comprehensively analyzed on the test data

A. Dataset

The dataset utilized in this study consists of chest X-ray images sourced from the Guangzhou Women and Children's Medical Center [12], comprising a total of 5,863 images. Each image is labeled as either 'Pneumonia' or 'Normal'. Additionally, to supplement the dataset, we incorporated a subset of infected images obtained from a local hospital, which were specifically used for testing purposes. This inclusion helps evaluate the model's robustness across diverse clinical environments. Figure 2 illustrates sample images from the dataset, showcasing a normal chest X-ray image alongside one depicting pneumonia.

Data Collection: Chest X-ray imaging was conducted as part of routine clinical care at the Guangzhou Women and Children's Medical Center. Ethical guidelines were strictly adhered to throughout the data collection process, and informed consent was obtained from all patients involved. The imaging protocol ensured consistency across the dataset, capturing high-quality scans suitable for machine learning analysis.

Data Organization: The dataset is meticulously organized into three primary folders: 'train', 'test', and 'validation'. Each of these folders is further divided into subfolders for the two categories, 'Pneumonia' and 'Normal', to streamline

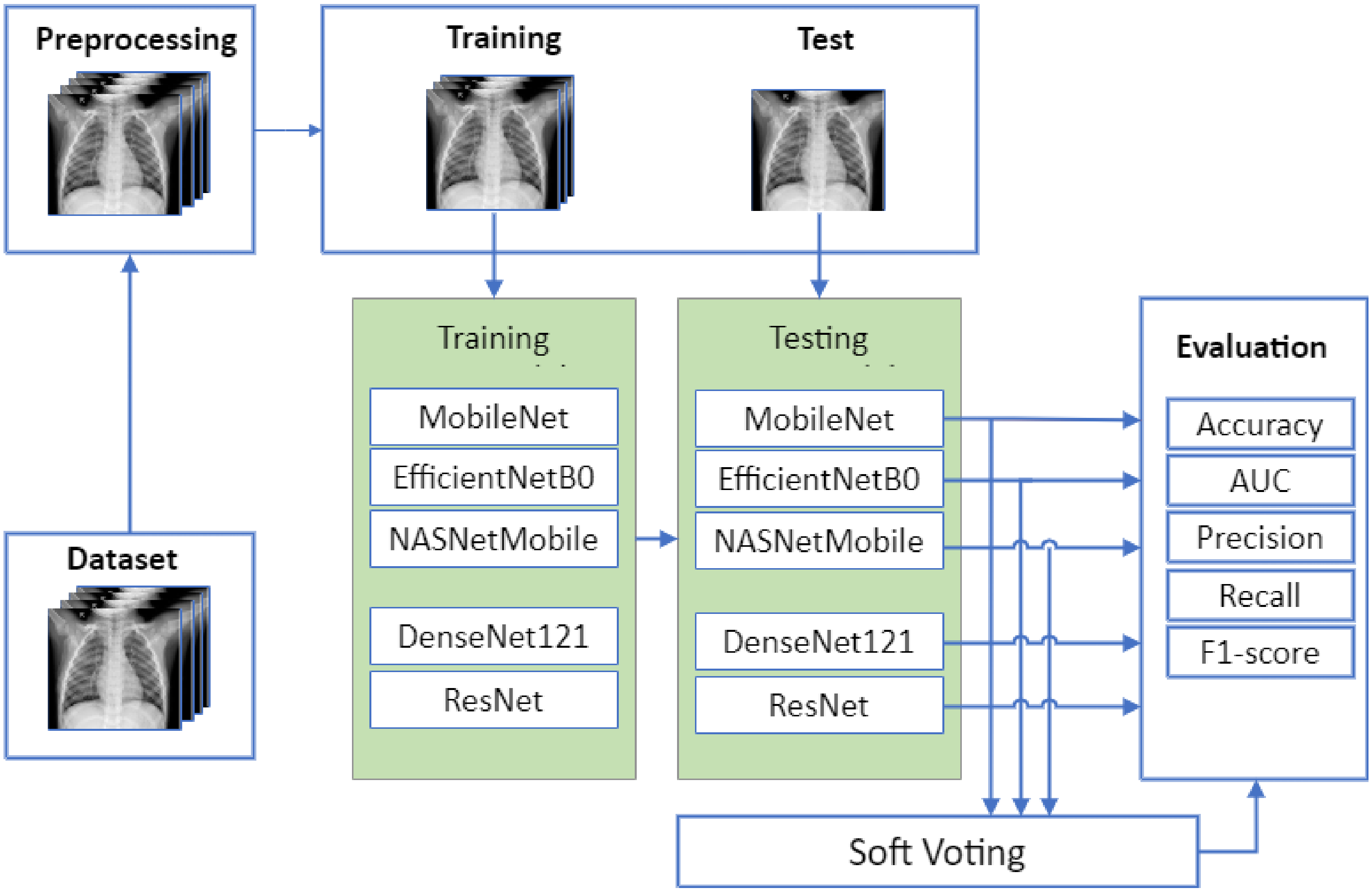

Fig. 1: Overview of the pneumonia detection framework: The dataset undergoes preprocessing before being divided into training and test sets. Five convolutional neural network (CNN) models (MobileNet, EfficientNetB0, NASNetMobile, DenseNet121, and ResNet) are trained and evaluated individually on key performance metrics (accuracy, AUC, precision, recall, and F1-score). The soft voting ensemble method combines the predictions of the models for final evaluation

the workflow for training, validation, and evaluation of machine learning models. This structure allows for efficient data handling and prevents data leakage between the training and testing phases, ensuring the reliability of model performance metrics.

- Training Set: Contains the majority of the images and is used to train the machine learning models. This set includes both normal and pneumonia-labeled images in sufficient quantity to enable robust learning of patterns.
- Validation Set: This subset is used for hyperparameter tuning and performance monitoring during training. It ensures that the models generalize well to unseen data.
- Testing Set: Comprises images exclusively reserved for evaluating the final model. This set includes both the Guangzhou dataset and the additional infected images sourced from the local hospital. Testing with the local hospital data highlights the model's generalizability to real-world scenarios and diverse clinical sources.

Data Augmentation and Preprocessing: To enhance the performance of the models, standard data augmentation techniques were applied to the training set. These included random rotations, flips, and contrast adjustments, which improve the model's robustness by simulating variations encountered in real-world settings. All images were resized to a consistent resolution and normalized to ensure compatibility with the machine learning pipeline.

By organizing the data and supplementing it with additional infected images, the dataset provides a comprehensive and diverse foundation for developing and testing machine learning models capable of distinguishing pneumonia from normal cases effectively.

B. Learning Algorithms

MobileNet, NasNetMobile, EfficientNetB0, DenseNet121, and ResNet are all convolutional neural networks (CNNs) designed for various tasks in computer vision, most commonly used for image classification. Each of these models introduces unique architectures or techniques to optimize certain factors, such as computational efficiency, accuracy, and model size. MobileNet, NasNetMobile, and EfficientNetB0 are known for their lightweight and efficient designs, making them ideal for deployment on mobile devices or in situations with limited computational resources. On the other hand, DenseNet121 is

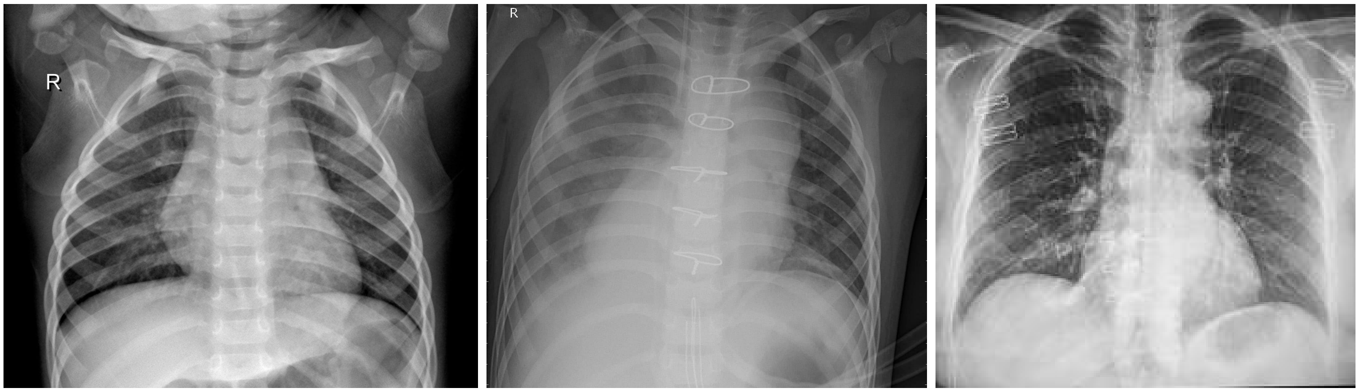

Fig. 2: Examples of chest X-ray images: normal image from the dataset (left), infected image from the dataset (middle), and infected image from the local hospital (right)

favored for its better accuracy and performance on challenging datasets. Depending on the specific requirements of a computer vision task, developers choose the CNN model that best suits their needs in terms of speed, accuracy, and resource constraints.

The objective of this study is to compare the performance of these models on a chest pneumonia classification task. Additionally, we investigated the performance of an ensemble with soft voting, where the ensemble consisted of lightweight models, and then compared the performance of each model with the performance of the ensemble and the more complex networks: ResNet and DenseNet. Below is a brief overview of each:

- *M*obileNet: MobileNet is designed for use in mobile and embedded vision applications. It prioritizes efficiency to achieve low latency and low power consumption. It uses depthwise separable convolutions to reduce the number of parameters and computational complexity. MobileNet is designed for use in mobile and embedded vision applications. It prioritizes efficiency to achieve low latency and low power consumption. MobileNet utilizes depthwise separable convolutions to significantly reduce the number of parameters and computational complexity compared to traditional convolutional networks. This architecture makes it particularly suitable for devices with limited processing power and memory. However, despite its efficiency, MobileNet models can still exhibit reduced accuracy compared to larger, more complex models. For instance, MobileNetV1 has around 4.2 million parameters, which allows for a good balance between performance and computational efficiency. While MobileNet is efficient in terms of inference time, it may not achieve the highest possible accuracy in comparison to more computationally intensive models. We consider MobileNet in this study to evaluate its trade-offs between efficiency and performance relative to other networks.
- *N*asNetMobile: Neural Architecture Search Network Mobile (NasNetMobile) is a model designed for mobile and embedded vision applications. It leverages automated architecture search to create an optimized network that balances accuracy and computational efficiency. By utilizing depthwise separable convolutions and a cell-based structure, NasNetMobile significantly reduces the number of parameters and computational complexity, making it suitable for devices with limited processing power and memory. However, the intricate architecture search process can increase the initial computational cost, which is a trade-off for its high performance in visual recognition tasks.
- *E*fficientNetB0: EfficientNetB0 is designed for both mobile and high-performance computing applications. It prioritizes model efficiency and scalability, utilizing a compound scaling method to balance network depth, width, and resolution uniformly. This approach results in a highly efficient model that delivers state-of-the-art accuracy while minimizing computational resources and power consumption. EfficientNetB0 achieves remarkable performance by combining depthwise separable convolutions with a novel architecture search technique, making it ideal for various visual recognition tasks.
- *R*esNet: Residual Network (ResNet), is a popular deep learning model known for its effectiveness in handling deep networks through the use of residual connections. These connections help solve the vanishing gradient problem and allow the network to be deeper without training difficulties. However, ResNet models can be quite large. For example, ResNet-50 has around 25 million parameters, and larger versions like ResNet-101 and ResNet-152 have even more. The depth and complexity of ResNet make it computationally expensive, which become a limiting factor for mobile devices that have constraints on memory, processing power, and battery life. ResNet models, especially the deeper variants, may not be the fastest in terms of inference time. This is a crucial factor for real-time applications on mobile devices. We consider ResNet in this study as base for comparison with the other networks.
- *D*enseNet121: DenseNet121 is a Densely Connected

Model	*Loss*	*Accuracy*	*AUC*	*Precision*	*Recall*	*F1-Score*
MobileNet	*0.135*	*95.1%*	*99.3%*	*99.2%*	*94.2%*	*96.52%*
EfficientNetB0	*0.148*	*94.2%*	*99.7%*	*99.5%*	*92.6%*	*95.81%*
NASNetMobile	*0.240*	*90.1%*	*97.6%*	*97.9%*	*88.66%*	*93.19%*
Soft Voting	*—*	*91.3%*	*97.3%*	*90.0%*	*96.9%*	*93.30%*
DenseNet121	*0.191*	*92.2%*	*98.2%*	*98.3%*	*91.1%*	*94.77%*
ResNet	*0.370*	*83.2%*	*89.7%*	*91.4%*	*85.4%*	*88.35%*

TABLE I: Performance comparison of different models

Convolutional Networks (DenseNet) family, introduces a unique architecture that differs from traditional CNNs like VGG or ResNet. The core innovation in DenseNet is that each layer is connected to every other layer in a feed-forward fashion, which means it has a very dense connectivity pattern. In DenseNet, each layer receives additional inputs from all preceding layers and passes on its own feature-maps to all subsequent layers. This dense connectivity pattern encourages feature reuse, significantly improves the flow of information and gradients throughout the network, and substantially reduces the number of parameters. Due to the reuse of features, DenseNet requires fewer parameters than an equivalently performing architecture with traditional sequential connections. This efficiency can make it lighter and faster for inference, once trained. Unlike in ResNets where features are added, DenseNet concatenates features from previous layers. This method preserves the original features through the network and contributes to improved feature propagation, which is advantageous for gradient flow during training. Despite its parameter efficiency, the DenseNet121 model is still quite substantial in size (around 8 million parameters) and be too computationally expensive for direct deployment on mobile devices without optimizations.

As shown in Figure 1, we first trained each model using transfer learning, and then we evaluated the performance of each model. Then we used we used ensemble learning with softvoting where the ensemble consist of the mobile-based model

IV. Results and Discussion

In this study, we evaluated the performance of several pre-trained convolutional neural network (CNN) architectures for pneumonia detection using chest X-ray images. The models assessed include MobileNet, ResNet, NASNetMobile, EfficientNetB0, DenseNet121, and a Soft Voting ensemble method. Each model was fine-tuned and trained on a consistent dataset and evaluated using key performance metrics: accuracy, recall, precision, AUC (Area Under the Curve), and F1-score.

To ensure robust performance and mitigate overfitting, we applied several strategies. 10-fold cross-validation was used to provide more reliable performance estimates by ensuring each data point contributed to both training and validation. Data augmentation techniques—such as random rotations, shifts, shear transformations, zoom operations, flips, and rescaling—were incorporated to improve generalization and expand the dataset artificially. Additionally, early stopping and learning rate reduction were employed to optimize training and prevent overfitting, while transfer learning allowed the models to leverage pre-trained weights for more effective feature extraction.

The results, summarized in Table I and visualized in Figures 3 and 4, reveal that MobileNet and EfficientNetB0 consistently outperformed the other models. MobileNet achieved the highest F1-score (96.52%) and accuracy (95.1%), highlighting its suitability for deployment in mobile and resource-constrained environments due to its lightweight and efficient design. EfficientNetB0 achieved the highest AUC (99.7%), confirming its strength in distinguishing between pneumonia and non-pneumonia cases. Both models demonstrated a strong balance between efficiency and performance, making them robust standalone diagnostic tools.

In contrast, NASNetMobile showed slightly lower performance, with an F1-score of 93.19%, despite being optimized for mobile deployment. DenseNet121 performed well with an accuracy of 92.2% and an F1-score of 94.77%, benefiting from its dense connectivity, which encourages feature reuse. However, its larger size and increased computational demands limit its practicality for deployment in low-resource environments. ResNet, despite its deep architecture, yielded the lowest performance among the models, with signs of overfitting likely due to the limited dataset size. This emphasizes the need for careful model selection tailored to the dataset and application context.

The Soft Voting ensemble method, although effective (F1-score: 93.3%), failed to surpass the best individual models. This outcome illustrates that while ensemble methods typically enhance generalization and robustness, they may not always outperform well-optimized standalone models. The limited performance gain from the ensemble can be attributed to the similarity in architecture and performance levels among the base models, reducing the potential benefits of combining their predictions. This finding highlights the importance of including more diverse models in an ensemble to capture complementary strengths, particularly in medical imaging tasks.

One notable observation was the significant impact of data augmentation on model performance. The models demonstrated high recall values, indicating their improved ability to correctly identify pneumonia cases. This suggests that the applied augmentation techniques allowed the models to learn more robust and invariant features, improving their generalization to unseen data.

Despite these promising results, the study has some limitations. The relatively small dataset size of 5,863 chest X-ray images may limit the generalizability of the models.

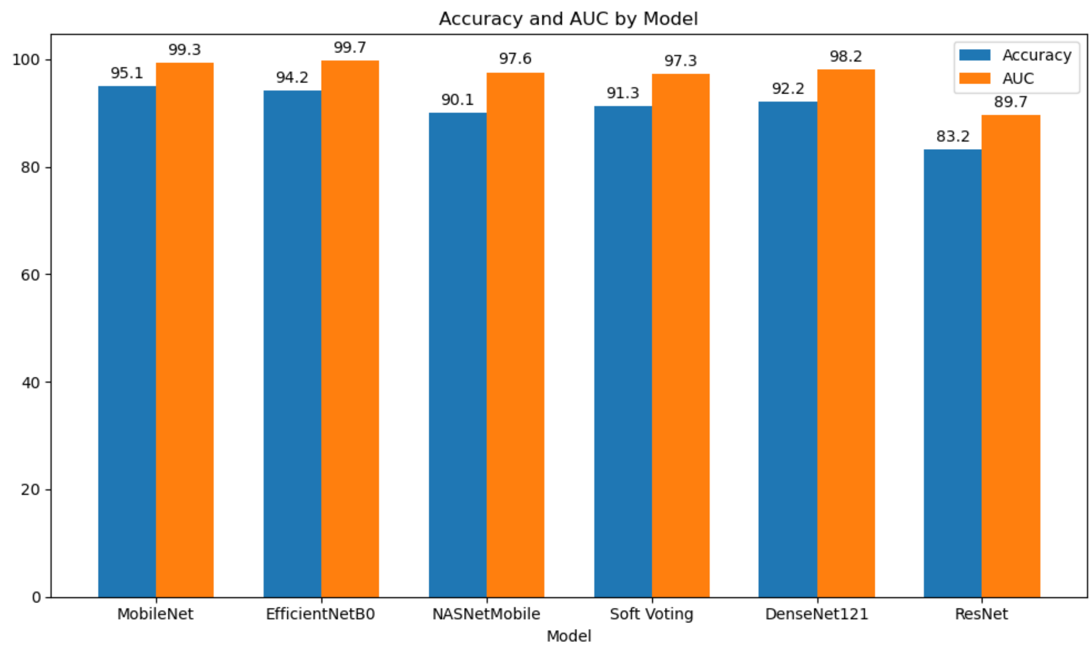

Fig. 3: Accuracy and AUC comparison across models.

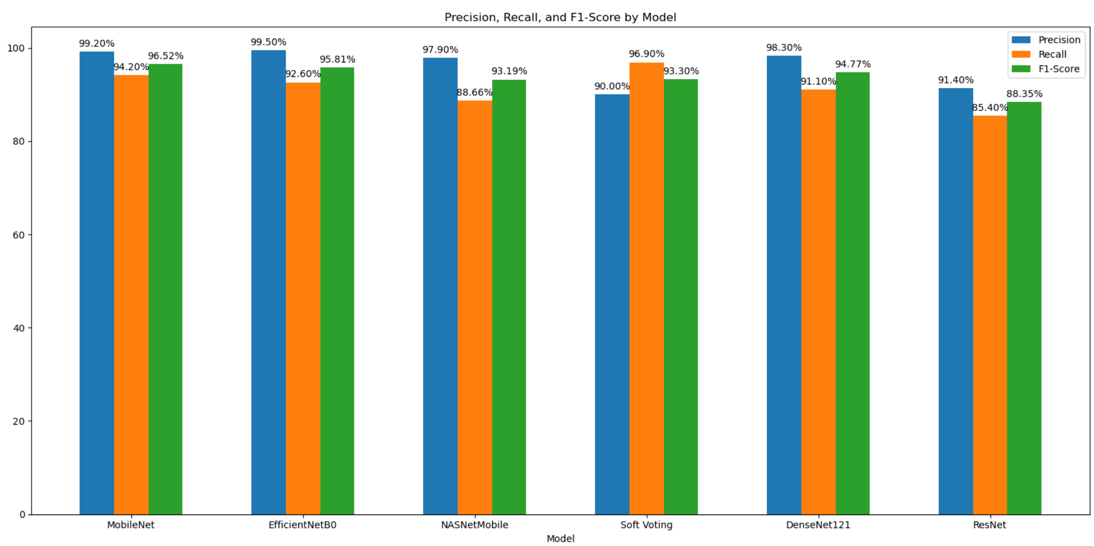

Fig. 4: Precision, Recall, and F1-Score comparison across models.

Future research will focus on validating these models on larger and more diverse datasets, such as NIH ChestX-ray14, to improve robustness and reliability. Additionally, exploring explainability techniques, such as Grad-CAM and LIME, could provide greater transparency in model decision-making, fostering clinical trust and facilitating integration into real-world diagnostic workflows. Further improvements could also be achieved by incorporating more diverse ensemble strategies and experimenting with hybrid models that combine CNNs with advanced mechanisms like attention layers or transformer-based architectures to enhance feature extraction and classification.

In summary, this study demonstrates the potential of lightweight CNN models, particularly MobileNet and EfficientNetB0, as efficient and accurate diagnostic tools for pneumonia detection. However, careful consideration of model diversity and explainability is essential for maximizing performance and ensuring successful clinical adoption.

V. CONCLUSION

This study assessed pre-trained CNNs, including MobileNet, EfficientNetB0, NASNetMobile, DenseNet121, and ResNet, for pneumonia detection from chest X-ray images. MobileNet achieved the highest accuracy (95.1%), while EfficientNetB0 excelled with the highest AUC (99.7%), making both models ideal for deployment in resource-constrained environments due to their balance of performance and computational efficiency. DenseNet121 also performed well, with an accuracy of 92.2% and an F1-score of 94.77%, though it was outperformed by MobileNet and EfficientNetB0. NASNetMobile and ResNet showed respectable results but were less effective, likely due to their higher complexity and the limited dataset.

The Soft Voting ensemble method enhanced robustness but did not surpass the individual performance of MobileNet and EfficientNetB0, underscoring the importance of selecting optimal base models. Future work should focus on larger datasets, advanced data augmentation, and hybrid model development to further improve diagnostic accuracy. These findings highlight the potential of MobileNet and EfficientNetB0 as accurate and efficient tools for pneumonia detection, ultimately supporting better patient outcomes.

References

[1] Y. Taha and Y. Rashed, "Pneumonia-prediction-project-," 2024, accessed: 2024-11-30. [Online]. Available: https://github.com/YousefTaha6700/Pneumonia-Prediction-Project-

[2] A. Pant, A. Jain, K. C. Nayak, D. Gandhi, and B. G. Prasad, "Pneumonia detection: An efficient approach using deep learning," in *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 2020, pp. 1–6.

[3] G. Labhane, R. Pansare, S. Maheshwari, R. Tiwari, and A. Shukla, "Detection of pediatric pneumonia from chest x-ray images using cnn and transfer learning," in *2020 3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE)*, 2020, pp. 85–92.

[4] A. Cococi, I. Felea, D. Armanda, and R. Dogaru, "Pneumonia detection on chest x-ray images using convolutional neural networks designed for resource constrained environments," in *2020 International Conference on e-Health and Bioengineering (EHB)*, 2020, pp. 1–4.

[5] Z. Jiang, "Chest x-ray pneumonia detection based on convolutional neural networks," in *2020 International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE)*, 2020, pp. 341–344.

[6] M. M. Hasan, M. Md. Jahangir Kabir, M. R. Haque, and M. Ahmed, "A combined approach using image processing and deep learning to detect pneumonia from chest x-ray image," in *2019 3rd International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE)*, 2019, pp. 89–92.

[7] L. mao, T. Yumeng, and C. Lina, "Pneumonia detection in chest x-rays: a deep learning approach based on ensemble retinanet and mask r-cnn," in *2020 Eighth International Conference on Advanced Cloud and Big Data (CBD)*, 2020, pp. 213–218.

[8] S. V. Militante, N. V. Dionisio, and B. G. Sibbaluca, "Pneumonia detection through adaptive deep learning models of convolutional neural networks," in *2020 11th IEEE Control and System Graduate Research Colloquium (ICSGRC)*, 2020, pp. 88–93.

[9] T. A. Youssef, B. Aissam, D. Khalid, B. Imane, and J. E. Miloud, "Classification of chest pneumonia from x-ray images using new architecture based on resnet," in *2020 IEEE 2nd International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS)*, 2020, pp. 1–5.

[10] M. J. Alam, S. N. Ali, and M. Z. Hasan, "A robust cnn framework with dual feedback feature accumulation for detecting pneumonia opacity from chest x-ray images," in *2020 11th International Conference on Electrical and Computer Engineering (ICECE)*, 2020, pp. 77–80.

[11] J. Garstka and M. Strzelecki, "Pneumonia detection in x-ray chest images based on convolutional neural networks and data augmentation methods," in *2020 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)*, 2020, pp. 18–23.

[12] Guangzhou Women and Children's Medical Center, "Guangzhou women and children's medical center," https://www.gwcmc.com, n.d., accessed: 2024-11-29.