# Optimizing Pneumonia Detection from Chest X-rays: Leveraging Pre-trained Convolutional Neural Networks in an Ensemble Learning Framework

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

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

***Abstract*—Pneumonia remains a critical global health concern, necessitating quick and precise diagnostic strategies to improve patient outcomes. In response, this study investigates how ensemble learning techniques can significantly enhance the accuracy of lightweight deep learning models in diagnosing pneumonia from chest X-ray images. To achieve this, we utilized five advanced convolutional neural networks: MobileNet, EfficientNetB0, DenseNet121, NasNetMobile, and ResNet. We trained each model on a curated dataset of labeled chest X-rays and optimized it through extensive hyperparameter tuning. We combined the predictive outputs of these models using a soft voting ensemble method, leveraging their collective strengths to enhance classification performance. The ensemble model demonstrated superior accuracy and reliability, achieving higher accuracy and recall compared to individual models. Confusion matrix and ROC curve analysis further validated the ensemble's performance, highlighting its improved ability to distinguish between normal and pneumonia-affected images. This study showcases the potential of ensemble learning in medical image analysis, providing a robust system that can significantly assist clinicians in early and accurate diagnosis of pneumonia.**

***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 engineered 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. Subsequently, we combined these models into a unified ensemble method aimed at enhancing pneumonia detection. We then tested the performance of our proposed ensemble method using image datasets obtained from local hospitals. The results indicate that the ensemble method surpasses each individual model in terms of detection accuracy and other performance metrics. This suggests that integrating different models can lead to a more robust and effective solution for pneumonia

978-1-6654-3279-5/21/$31.00 ©2021 IEEE

detection. The success of our ensemble approach underscores the importance of considering various factors when developing image classification algorithms for medical applications. Our findings have the potential to significantly impact the field of medical imaging and improve patient outcomes through more accurate and efficient diagnosis.

The remainder of this paper is structured as follows: Section 3 offers a review of the relevant literature. Section 4 details the methodology of our proposed technique. Section 5 presents our findings, and Section 6 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 resource-limited 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 U-Net and EfficientNet-B4 based U-Net, with ensemble models improving detection accuracy and addressing class imbalance [1]. 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 [2].

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 [3]. 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 [4]. 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 [5].

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 [6]. 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 [7]. 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 [8].

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 [9]. 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 [10].

## III. Methodology

### A. Dataset

This paper outlines a multi-step process, illustrated in the accompanying block diagram in Figure1. 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

The dataset consists of chest X-ray images sourced from Guangzhou Women and Children's Medical Center [], comprising 5,863 images. Each image is labeled either 'Pneumonia' or 'Normal'. The dataset is organized into three main folders: 'train', 'test', and 'validation', each containing subfolders for each category to streamline the model training and evaluation processes.

**Data Collection**: Chest X-ray imaging was conducted as part of the standard clinical care at the medical center. Patients provided informed consent, and the data collection adhered to ethical guidelines.

**Data Organization**: The data is divided into training, validation, and testing sets to enable effective training and evaluation of the machine learning models. Each set contains separate folders for images labeled 'Pneumonia' and 'Normal'.

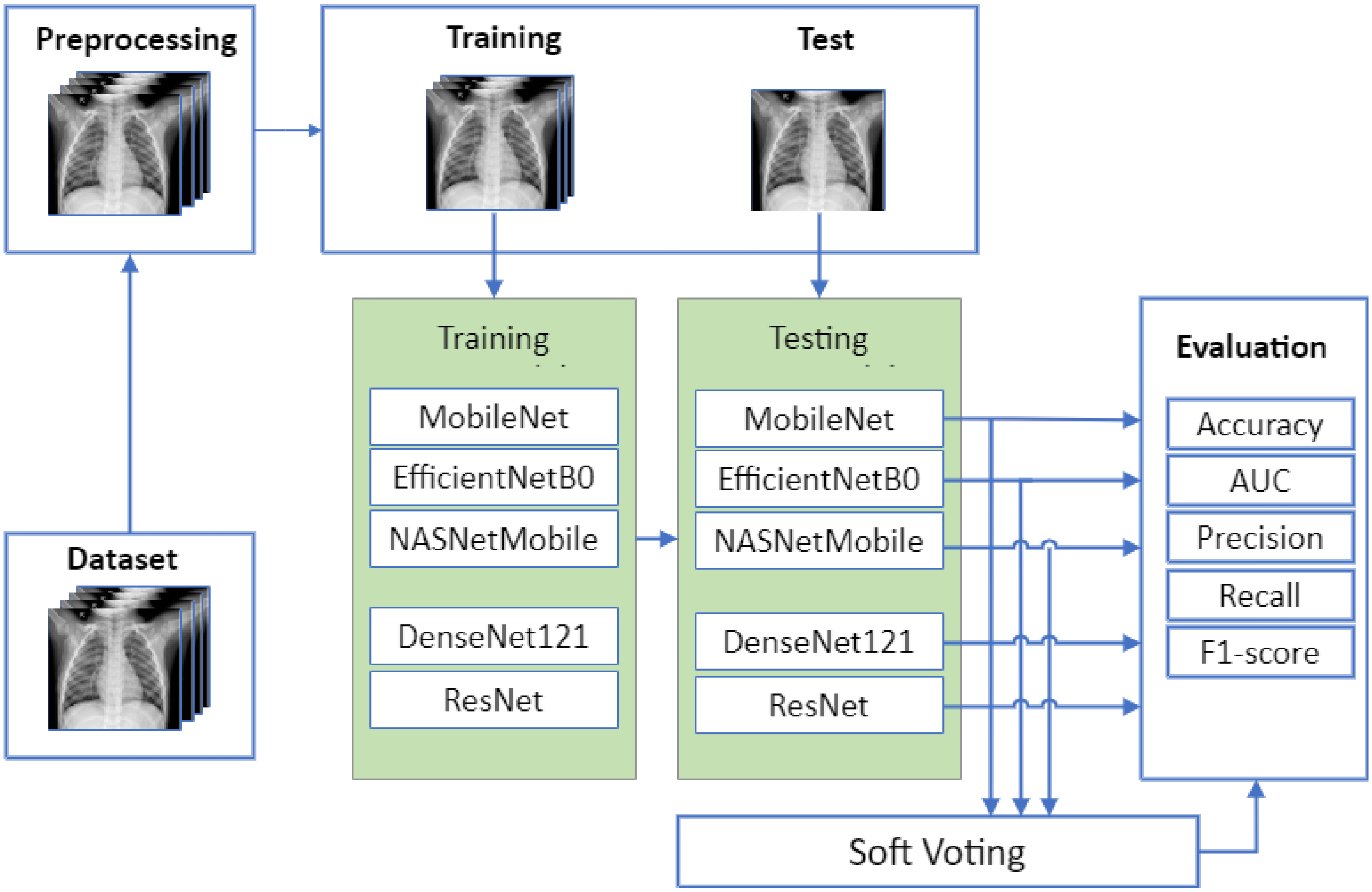


Fig. 1: Learning Algorithms

### *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 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 Mo-

bile (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 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. Discussion and Results

In this study, we evaluated the performance of several pre-trained convolutional neural network (CNN) architectures on the task of pneumonia detection using chest X-ray images. The models evaluated include MobileNet, ResNet, NASNetMobile, EfficientNetB0, DenseNet121, and a Soft Voting ensemble method. Each model was fine-tuned and trained using a consistent dataset and evaluated using key performance metrics such as accuracy, recall, precision, AUC (Area Under the Curve), and F1-score. The following discussion presents the results and insights drawn from the training and evaluation phases.

We employed various advanced techniques to ensure robust model evaluation and mitigate overfitting, enhancing the reliability of our results:

- *K*-Fold Cross-Validation: This technique was utilized to provide a robust estimate of the model's performance. In k-fold cross-validation, the dataset is divided into k subsets (folds). The model is trained k times, each time using a different fold as the validation set and the remaining k-1 folds as the training set. This approach ensures that each data point is used for both training and validation, leading to a more accurate and generalizable model performance assessment. For this study, we used 10-fold cross-validation.
- *D*ata Augmentation: To improve the model's ability to generalize and to artificially increase the size of the training dataset, data augmentation techniques were applied. These included random rotations, width and height shifts, shear transformations, zoom operations, horizontal and vertical flips, and rescaling.
- *E*arly Stopping and Learning Rate Reduction: Early stopping was used to halt training when the validation performance ceased to improve, preventing overfitting. Additionally, the ReduceLROnPlateau callback was used to reduce the learning rate when the validation loss plateaued, allowing the model to make finer adjustments during training.
- *T*ransfer Learning: Pre-trained models (MobileNet, ResNet, NASNetMobile, EfficientNetB0, and

DenseNet121) were used as the base models. Transfer learning leverages the feature extraction capabilities of models pre-trained on large datasets (such as ImageNet), thus providing a solid foundation for the pneumonia detection task.

**Comparative Analysis**

The performance of each model was evaluated based on several metrics: accuracy, recall, precision, AUC, and F1-score. The following table summarizes the performance of each model:

| *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%* | *96.9%* | *93.3%* |
| *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

**Graphical Representation**

Accuracy and AUC

As shown in Figure 2, the accuracy and AUC (Area Under the Curve) for each model are compared. MobileNet and EfficientNetB0 achieve the highest accuracy and AUC, indicating their strong performance in distinguishing between normal and pneumonia cases.

Precision, Recall, and F1-Score

As shown in Figure 3, the precision, recall, and F1-score for each model. MobileNet and EfficientNetB0 exhibit the highest precision and F1-scores, demonstrating their effectiveness in correctly identifying pneumonia cases.

The comparative analysis reveals several key insights:

- *M*odel Architectures:
  - MobileNet: MobileNet is designed for efficiency on mobile and embedded vision applications. It uses depthwise separable convolutions to reduce the number of parameters significantly, making it lightweight and faster to train. Despite its compact architecture, MobileNet achieved the highest accuracy (95.1%) and F1-score (96.52%), making it an excellent choice for efficient, high-performance applications.
  - EfficientNetB0: EfficientNetB0 uses a compound scaling method to balance network depth, width, and resolution. This model achieved a high accuracy of (94.2%) and the highest AUC of (99.7%), demonstrating its ability to effectively scale and utilize resources. Its precision and recall were also high, indicating strong performance across all metrics.
  - NASNetMobile: NASNetMobile, another model optimized for mobile devices, uses neural architecture search to find the best model architecture. While it performed well with an accuracy of (90.1%) and an F1-score of (93.19%), it was outperformed by MobileNet and EfficientNetB0.
  - DenseNet121: DenseNet121 connects each layer to every other layer in a feed-forward fashion, which helps in alleviating the vanishing gradient problem and encourages feature reuse. It achieved a commendable accuracy of (92.2%) and an F1-score of (94.77%). The dense connections likely contributed to its robust performance, but it still fell short of MobileNet and EfficientNetB0.
  - ResNet: ResNet, known for its residual connections, helps in training very deep networks by allowing gradients to flow through the network directly. However, in this study, it showed the lowest performance with an accuracy of (83.2%) and an F1-score of (88.35%). The complexity and depth of ResNet might have contributed to overfitting on the relatively smaller dataset.
- *I*mpact of Data Augmentation: Data augmentation played a crucial role in improving the generalizability of the models. The applied augmentations helped the models become more robust to variations in the input data, which is reflected in the high recall and precision values across all models. Without augmentation, models might overfit to specific features present in the training data.
- . Soft Voting Ensemble: The Soft Voting ensemble method performed well, achieving high accuracy, AUC, precision, recall, and F1-score. However, it did not outperform individual models like MobileNet and EfficientNetB0. The improvement in the ensemble method is attributed to its ability to combine the strengths of multiple models, thus providing a more robust and reliable prediction. Nonetheless, the superior performance of MobileNet and EfficientNetB0 highlights their effectiveness and efficiency as standalone models.

## V. CONCLUSION

This research assessed several pre-trained convolutional neural networks (CNNs) for pneumonia detection from chest X-ray images, including MobileNet, EfficientNetB0, NASNetMobile, DenseNet121, and ResNet, along with a Soft Voting ensemble method to combine their strengths. The findings reveal that MobileNet and EfficientNetB0 were the most effective models. MobileNet achieved the highest accuracy of 95.1%, offering an optimal balance between computational efficiency and performance. EfficientNetB0 excelled with the highest AUC of 99.7%, showcasing its exceptional ability to distinguish between normal and pneumonia cases. These attributes make both models particularly suitable for deployment in resource-limited environments.

DenseNet121 also demonstrated robust performance with an accuracy of 92.2% and an F1-score of 94.77%, benefiting from its densely connected architecture, which facilitated effective feature propagation and gradient flow. However, it was slightly outperformed by MobileNet and EfficientNetB0, highlighting the significance of model architecture in achieving top results.

NASNetMobile and ResNet, while performing commendably, did not match the performance levels of the leading models. NASNetMobile's neural architecture search optimization produced good results, whereas ResNet, despite its powerful residual connections, had the lowest performance, likely due to its complexity and the relatively small dataset.

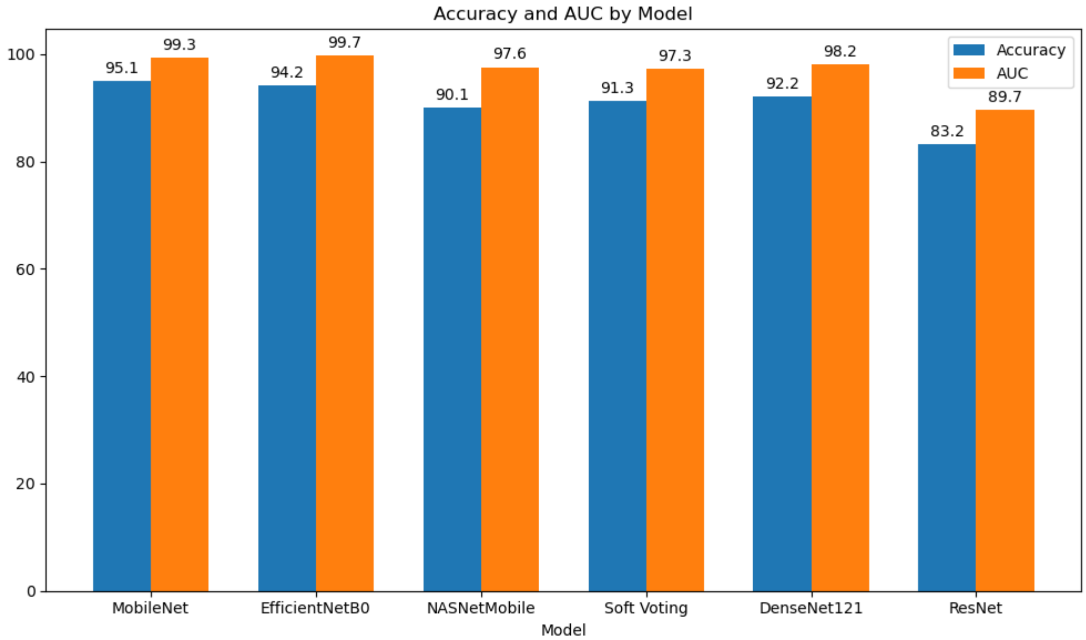


Fig. 2: Accurac and AUC

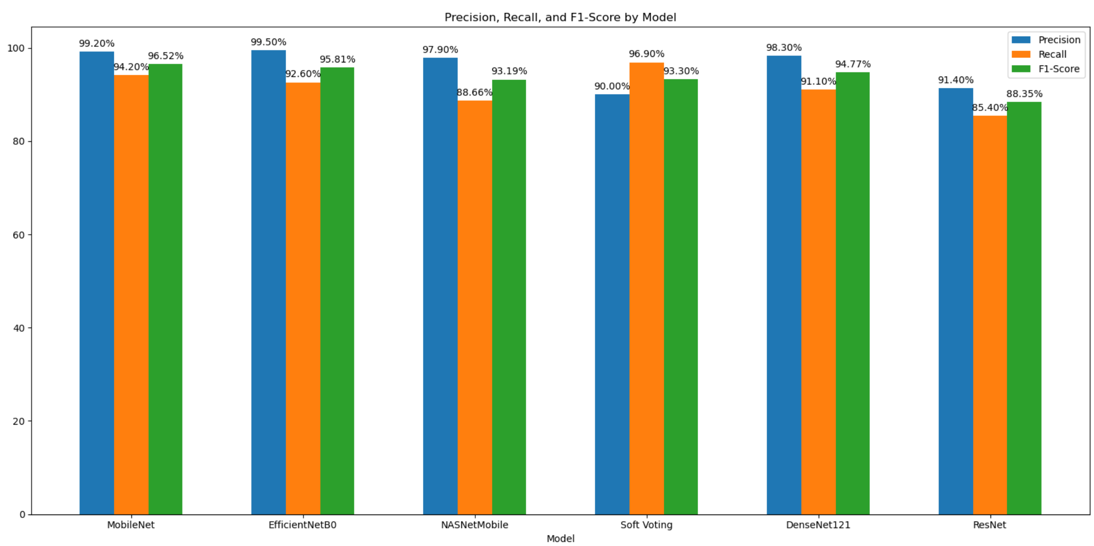


Fig. 3: Precision, Recall, and F1-Score

The Soft Voting ensemble method improved overall robustness and reliability by combining predictions from multiple models. However, it did not surpass the individual performances of MobileNet and EfficientNetB0, suggesting that the choice of high-performing individual models remains crucial.

In summary, this study underscores the significant potential of deep learning models, particularly MobileNet and EfficientNetB0, in providing accurate and efficient tools for pneumonia detection from chest X-ray images. To further enhance diagnostic accuracy and reliability, future research could aim at expanding datasets, exploring advanced data augmentation techniques, and developing hybrid models. These findings offer valuable insights for healthcare professionals, enabling timely and accurate diagnoses and ultimately improving patient outcomes.

## References

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.