Comparative Analysis of Convolutional Neural Network Architectures for Lung Disease Detection in X-ray Images

Mahsa Shahbazi, Alois Chinyoka

***Abstract*—This study presents the development and evaluation of convolutional neural network (CNN) architectures to classify chest X-ray images into three categories: COVID-19, pneumonia, and normal. We initially built a Basic CNN model as a baseline and progressively introduced advanced features, including Depthwise Separable Convolutions, Attention Mechanisms, and DenseNet architectures, while maintaining consistent settings for a fair comparison. Each model was assessed using key performance metrics such as accuracy, precision, recall, F1-score, and area under the curve (AUC), as well as inference efficiency, memory usage, and confusion matrix.**

The Depthwise CNN consistently outperformed the other models, achieving high test accuracy (87.06%), precision (96.19%), recall (88.39%), and F1-score (92.12%). This model demonstrated a favorable balance between computational efficiency and performance, making it highly suitable for real-world deployment in medical diagnostics. The Attention CNN, while showing promise in reducing misclassification rates between COVID-19 and normal cases, lagged behind in overall performance metrics. The Basic CNN provided strong baseline results, but struggled with memory efficiency and COVID-19 misclassifications, while the DenseNet model displayed overfitting, leading to suboptimal generalization.

Overall, the Depthwise CNN emerged as the most effective model, combining high classification accuracy with resource efficiency, indicating its potential for clinical applications in disease diagnosis from X-ray images.

***Index Terms*—COVID-19, Pneumonia, X-ray Classification, Convolutional Neural Networks, Deep Learning, Medical Image Processing**

I. Introduction

The COVID-19 pandemic significantly impacted global health systems, underscoring the need for rapid and accurate diagnostic tools. Although the pandemic has subsided, the importance of efficient diagnostic methods for respiratory diseases remains critical. Chest X-ray imaging continues to be a vital tool for detecting lung diseases, including residual COVID-19 cases, pneumonia, and other pulmonary conditions. However, manual interpretation of X-ray images is time-consuming and subject to inter-observer variability, leading to potential diagnostic inaccuracies.

Automating the interpretation of medical images using artificial intelligence (AI) can address these challenges. Convolutional neural networks (CNNs), a class of deep learning models, have demonstrated exceptional performance in various image classification tasks. They are particularly suited for medical image analysis due to their ability to learn hierarchical feature representations directly from raw images.

In this study, we focus on the problem of classifying chest X-ray images into three categories: COVID-19, pneumonia, and normal. This problem remains highly relevant as efficient diagnostic tools for respiratory diseases are crucial in the post-pandemic era. While prior studies have employed various CNN architectures for similar tasks, achieving a balance between accuracy, computational efficiency, and generalizability remains a challenge. To address this, we explored different CNN architectures, progressively introducing advanced features to improve performance.

We propose a comprehensive processing pipeline and evaluate several CNN architectures, including a basic CNN, Depthwise Separable CNN, CNN with Attention Mechanisms, and DenseNet. Our contributions are as follows:

1) **Problem**: We tackle the challenge of classifying chest X-ray images into three clinically important categories: COVID-19, pneumonia, and normal.
2) **Relevance**: Our work addresses the critical need for fast, reliable, and scalable diagnostic tools for respiratory diseases.
3) **Approach**: We develop and systematically compare multiple CNN architectures, using techniques such as data augmentation, regularization, and feature extraction. We ensure a fair comparison by maintaining the same training pipeline across all models.
4) **Value**: The results demonstrate that the **Depthwise CNN** model consistently outperforms the other architectures. It achieves high test accuracy (87.06%), precision (96.19%), recall (88.39%), and F1-score (92.12%), offering a well-rounded balance between accuracy, computational efficiency, and confusion matrix.
5) **Applicability**: This study provides valuable insights into how different advanced features in CNN architectures perform in medical image classification. By comparing models with Basic CNN, Depthwise Separable Convolutions, Attention Mechanisms, and DenseNet architectures, we can identify which features are most suitable for different situations. For instance, the Depthwise CNN strikes an optimal balance between accuracy and computational efficiency, making it ideal for real-world clinical applications with resource constraints. On the other hand, models like the Attention CNN offer high precision and recall, which could be particularly useful in critical diagnostic tasks where minimizing false

negatives is crucial.

II. Related Work

Recent advancements in deep learning, particularly in convolutional neural networks (CNNs), have significantly enhanced the field of medical image analysis. Various studies have explored the application of CNNs for the classification of chest X-ray images, particularly for detecting lung diseases such as COVID-19 and pneumonia.

[1] presented a DenseNet-based architecture to predict the severity of COVID-19 from chest X-ray images. The study demonstrated that DenseNet could effectively capture the complex patterns associated with different severity levels of COVID-19. However, the approach primarily focused on severity prediction rather than classification into distinct categories such as COVID-19, pneumonia, and normal.

[2] introduced COVID-Net, a tailored deep convolutional neural network designed specifically for detecting COVID-19 cases from chest X-ray images. The network architecture combined standard convolutional layers with depth-wise separable convolutions to enhance feature extraction efficiency. While COVID-Net showed promising results in detecting COVID-19, its performance in distinguishing between other lung conditions, such as pneumonia and normal, was not extensively evaluated.

[3] proposed the use of transfer learning with pre-trained CNNs such as VGG19 and MobileNet for the classification of COVID-19, pneumonia, and normal chest X-ray images. Although transfer learning achieved high accuracy, the study relied heavily on pre-trained models, which may not capture domain-specific features as effectively as models trained from scratch on the target dataset.

[4] developed a deep learning model using a combination of CNN and recurrent neural network (RNN) layers for automated diagnosis of COVID-19 and pneumonia. This hybrid approach aimed to leverage the spatial features captured by CNNs and the temporal dependencies captured by RNNs. Despite its innovative architecture, the model's complexity resulted in longer training and inference times, making it less practical for real-time diagnostic applications.

[5] explored attention mechanisms within CNN architectures to improve the focus on relevant regions of chest X-ray images. The attention-based models demonstrated improved interpretability and diagnostic accuracy. However, the increased computational overhead associated with attention mechanisms posed challenges for deployment in resource-constrained environments.

In summary, the existing literature highlights the potential of CNNs in the classification of chest X-ray images for diagnosing lung diseases. However, challenges such as balancing model complexity, computational efficiency, and generalizability remain.

III. Processing Pipeline

The processing pipeline for this study is designed to classify chest X-ray images. This pipeline comprises several blocks, each contributing to the overall classification task. The goal is to ensure that all models, including the basic CNN and the advanced variants, are trained and evaluated under consistent settings for a fair comparison.

A. Data Collection and Preparation

The dataset consists of chest X-ray images categorized into three classes: COVID-19, pneumonia, and normal. The dataset is split into training, validation, and test sets, maintaining a consistent split ratio of 80% training, 10% validation, and 10% test across all models.

A total of 4575 images (1525 per class) are used in the experiments. The data is loaded using TensorFlow `image_dataset_from_directory` function, which infers the class labels based on the directory structure. The images are resized to 224×224 pixels, and a batch size of 16 is used for loading the data. Shuffling is enabled with a fixed random seed (123) for reproducibility.

B. Data Preprocessing

Preprocessing is crucial to ensure consistency and enhance the robustness of the models. The key steps include:

- **Normalization**: All images are normalized to a pixel range of [0, 1] using a `Rescaling` layer, standardizing the input data across all models to ensure fair comparison.
- **Augmentation**: Data augmentation techniques are applied to the training set to increase model robustness and prevent overfitting. The specific augmentations include:
  - `RandomRotation`: Randomly rotates images within a specified range.
  - `RandomZoom`: Applies random zooming to simulate varying distances.
  - `RandomTranslation`: Translates images along the height and width axes to mimic slight shifts in positioning.
- **Efficiency**: For efficient data processing, the training data is shuffled, cached, and prefetched to reduce bottlenecks during training. Validation and test datasets are only normalized without augmentation.

C. Feature Extraction

Feature extraction plays a critical role in the classification task, as it allows the models to learn and identify important spatial hierarchies and features from chest X-ray images. We experiment with four different architectures: Basic CNN, Depthwise CNN, CNN with Attention, and DenseNet. Each architecture progressively builds on the complexity of the baseline CNN model, introducing advanced techniques to optimize feature extraction and model performance.

- **Basic CNN**: The Basic CNN serves as the baseline model. It comprises three convolutional layers with increasing filter sizes to capture more complex features at each layer. The model structure is as follows:
  - `Input Layer`: Accepts input images of size 224x224x3, representing the RGB channels.

  - `Conv2D Layer 1`: 32 filters, each with a 3x3 kernel and ReLU activation, responsible for detecting low-level features such as edges and textures.
  - `MaxPooling Layer 1`: Reduces the spatial dimensions by downsampling, ensuring computational efficiency.
  - `Conv2D Layer 2`: 64 filters with a 3x3 kernel and ReLU activation. This layer extracts more abstract features like shapes.
  - `MaxPooling Layer 2`: Further reduces the spatial dimensions while retaining important feature information.
  - `Conv2D Layer 3`: 128 filters, designed to capture complex patterns and details from the X-ray images.
  - `MaxPooling Layer 3`: Continues downsampling the feature maps to minimize the feature space before fully connected layers.
  - `Flatten Layer`: Converts the 2D feature maps into a 1D vector to prepare for the fully connected (dense) layers.
  - `Dense Layer`: A dense layer with 128 units and ReLU activation extracts non-linear combinations of the features.
  - `Dropout Layer`: A dropout rate of 0.5 is applied to reduce overfitting by randomly setting 50% of the neurons to zero during training.
  - `Output Layer`: Softmax activation is used in the output layer, which has three units corresponding to the three classes (COVID-19, pneumonia, and normal).

- **Depthwise CNN**: The Depthwise CNN is designed to improve computational efficiency without sacrificing performance. It replaces standard convolutions with depthwise separable convolutions, which split the convolution operation into two steps: depthwise and pointwise convolutions.
  - `Depthwise Separable Conv2D`: Each convolutional block begins with a depthwise separable convolution to reduce computational cost. This separates the spatial filtering from the channel mixing, significantly reducing the number of parameters.
  - `Batch Normalization`: Applied after each convolutional layer to normalize the feature maps, improving training stability and speed.
  - `ReLU Activation`: Ensures non-linearity after the normalization, allowing the model to learn complex patterns.
  - `Residual Connections`: Shortcut connections are used to allow gradients to flow more freely through the network, mitigating the vanishing gradient problem.
  - `MaxPooling`: MaxPooling is applied after each convolution block to progressively reduce spatial dimensions.
  - `Global Average Pooling`: This operation reduces each feature map to a single value, further decreasing the number of parameters while retaining important global information.
  - `Dense Layer and Dropout`: Similar to the Basic CNN, a dense layer with 128 units is followed by dropout (0.5) to prevent overfitting.
  - `Output Layer`: Softmax activation is used to produce a probability distribution over the three classes.

- **CNN with Attention**: The Attention CNN introduces a spatial attention mechanism that allows the model to focus on the most important regions of the image. This is particularly useful in medical imaging, where specific areas of an X-ray may contain the most diagnostically relevant information.
  - `Conv2D Layers`: Three convolutional layers with increasing filter sizes (32, 64, 128), similar to the basic architecture.
  - `Spatial Attention Mechanism`: After the third convolutional layer, a spatial attention map is generated using a convolutional layer with a sigmoid activation. This attention map highlights the most important regions in the feature map.
  - `Element-Wise Multiplication`: The attention map is multiplied with the original feature map, allowing the network to focus on the regions with higher attention scores.
  - `Flatten and Dense Layers`: The resulting feature maps are flattened, followed by a dense layer with 128 units and ReLU activation. A dropout rate of 0.5 is applied before the output layer.
  - `Output Layer`: Softmax activation is used in the final layer for classification into the three categories.

- **DenseNet**: DenseNet introduces the concept of dense connectivity, where each layer is connected to every other layer in a feed-forward fashion. This encourages feature reuse and improves gradient flow.
  - `Initial Conv2D Layer`: The input is processed through a standard convolution layer (16 filters) followed by max-pooling to reduce spatial dimensions.
  - `Dense Blocks`: Each dense block contains several convolutional layers where each layer receives input from all previous layers. This allows for efficient feature reuse and reduces the number of parameters.
  - `Bottleneck Layers`: 1x1 convolutions are used as bottleneck layers to reduce the number of feature maps before the 3x3 convolutions, improving computational efficiency.
  - `Transition Layers`: Between dense blocks, transition layers (1x1 convolution followed by pooling) reduce the feature map size and prevent overfitting.
  - `Global Average Pooling`: Applied after the final dense block to reduce the feature maps to a

single value per feature.
  - `Dense Layer and Dropout`: A fully connected layer with 128 units is followed by a dropout layer with a rate of 0.5.
  - `Output Layer`: Softmax activation is used to classify the image into one of the three classes.

D. Training Process

The training process is identical across all models to ensure a fair comparison. Key elements of the training process include:

- **Loss Function**: All models are optimized using categorical cross-entropy loss, which is well-suited for multi-class classification tasks.
- **Optimizer**: The Adam optimizer is used with an initial learning rate of 0.0005, balancing fast convergence with stability.
- **Learning Rate Scheduling**: A learning rate scheduler is employed to reduce the learning rate after the initial epochs, allowing for finer updates as training progresses.
- **Early Stopping and Checkpoints**: Early stopping is implemented to prevent overfitting by halting training if the validation loss does not improve for 5 consecutive epochs. Model checkpoints are saved based on the best validation accuracy.

E. Evaluation Metrics

To ensure a comprehensive evaluation of the models, several metrics are used:

- **Accuracy**: Measures the overall percentage of correctly classified instances.
- **Precision**: Evaluates the number of true positives out of the total predicted positives, providing insight into the model's ability to avoid false positives.
- **Recall**: Assesses the proportion of true positives out of actual positives, focusing on the model's ability to detect relevant cases (e.g., COVID-19).
- **F1-Score**: The harmonic mean of precision and recall, offering a balanced metric when there is an uneven class distribution.
- **AUC**: Measures the ability of the model to distinguish between classes. A higher AUC indicates better model performance across all classification thresholds.
- **Confusion Matrix**: Provides a visual breakdown of the classification results for each class, displaying true positives, false positives, true negatives, and false negatives. This offers deeper insights into the model performance for each class and helps identify specific classes where the model might struggle.

F. Inference Efficiency

Inference efficiency is critical for real-world deployment, especially in medical applications where time and resource constraints are important considerations. Each model is evaluated for:

- **Average Inference Time**: The average time taken per image during the inference phase is measured, with lower times being favorable for real-time applications.
- **Memory Usage**: The memory footprint of each model is recorded to assess the computational resources required for deployment. Models like DenseNet, while effective, tend to have higher memory requirements compared to more lightweight architectures like the Basic CNN or Depthwise CNN.

IV. Results

In this section, we present the numerical and visual results obtained from our models.

A. Training Performance

The training performance of the models is evaluated using key metrics such as accuracy, loss, precision, recall, AUC, and F1 score. The following plots show the comparison of these metrics across different epochs for the Basic CNN, Depthwise CNN, Attention CNN, and DenseNet models.

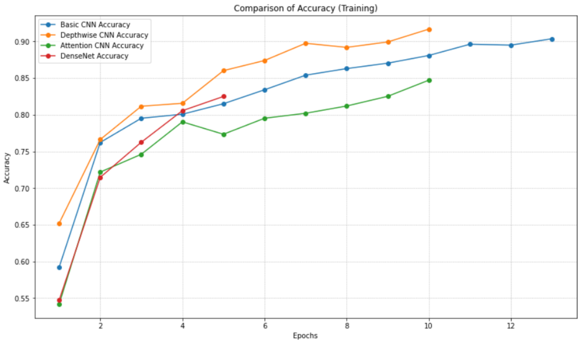

Fig. 1: Comparison of Accuracy (Training)

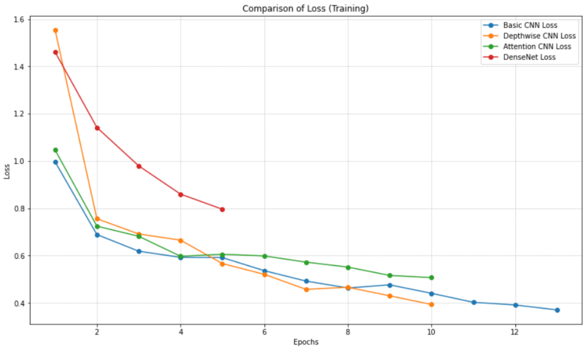

Fig. 2: Comparison of Loss (Training)

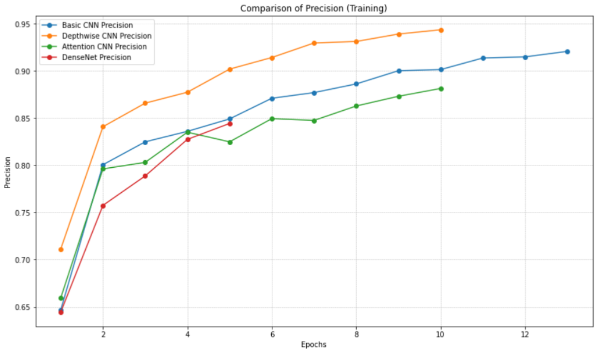

Fig. 3: Comparison of Precision (Training)

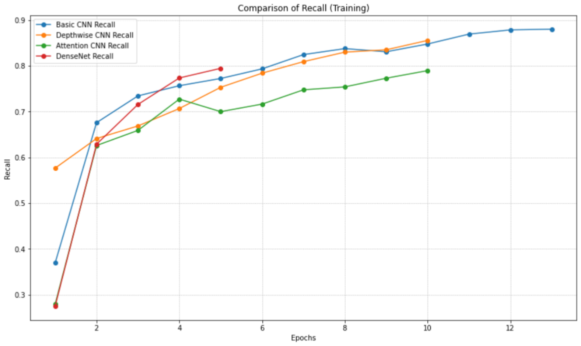

Fig. 4: Comparison of Recall (Training)

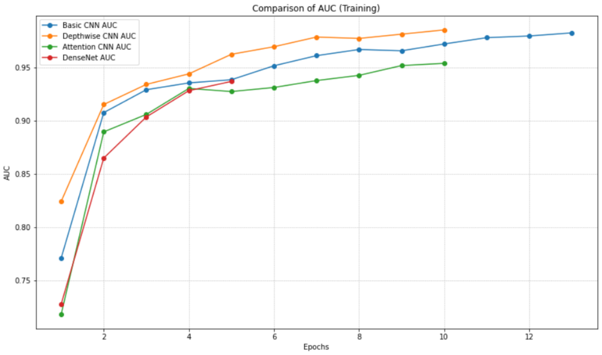

Fig. 5: Comparison of AUC (Training)

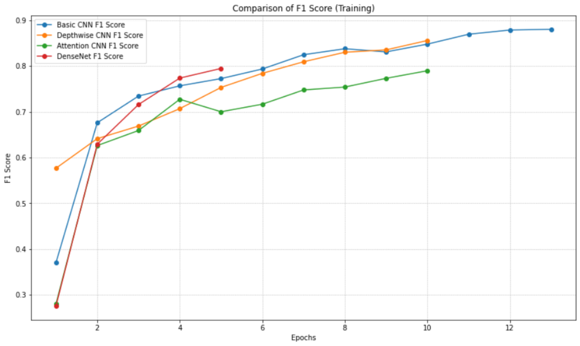

Fig. 6: Comparison of F1 Score (Training)

B. *Confusion Matrices*

To further evaluate the classification performance of each model, we present the confusion matrices for the Basic CNN, Depthwise CNN, Attention CNN, and DenseNet models.

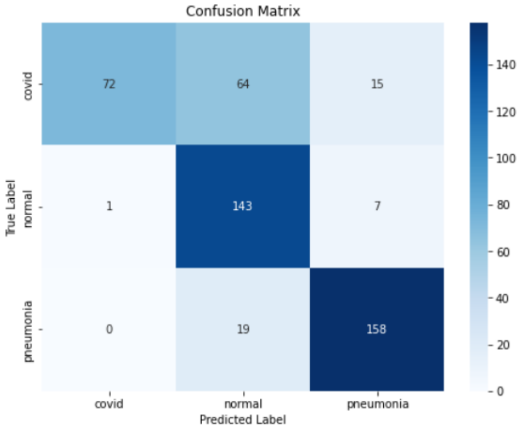

Fig. 7: Confusion Matrix for Basic CNN Model

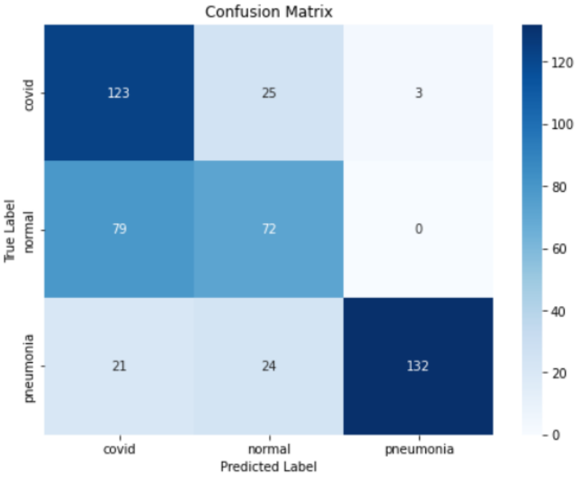

Fig. 8: Confusion Matrix for DenseNet Model

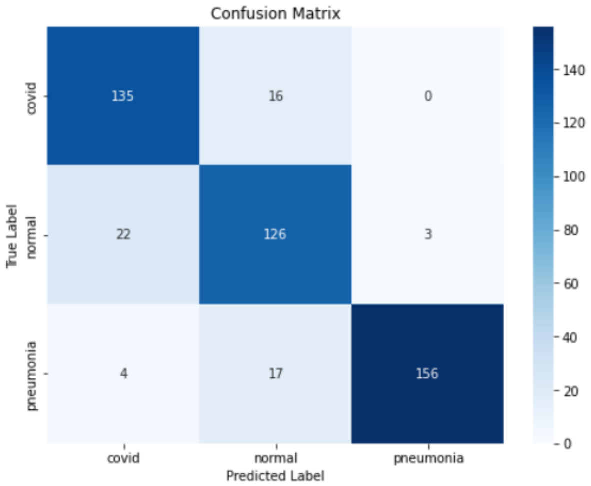

Fig. 9: Confusion Matrix for Depthwise CNN Model

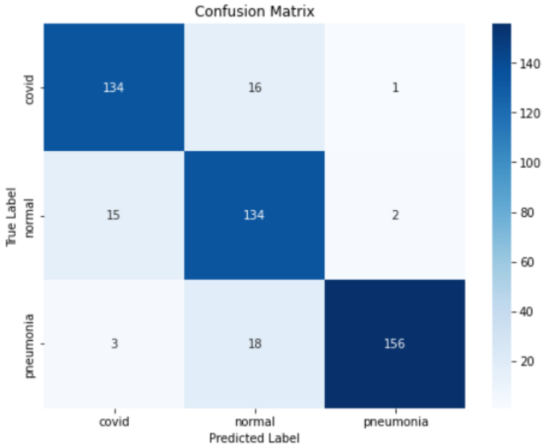

Fig. 10: Confusion Matrix for Attention CNN Model

C. Inference Performance

Inference performance is critical for real-time medical diagnostics. We evaluate each model based on average inference time per sample and memory usage, as shown in the figures below.

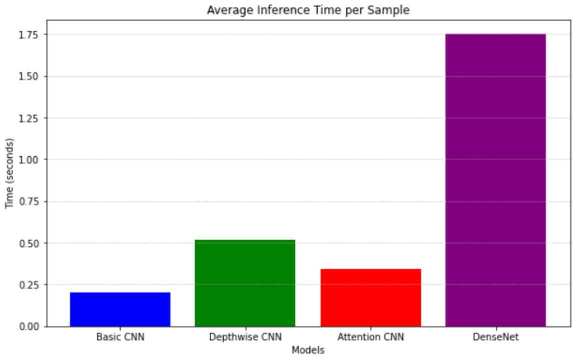

Fig. 11: Average Inference Time per Sample

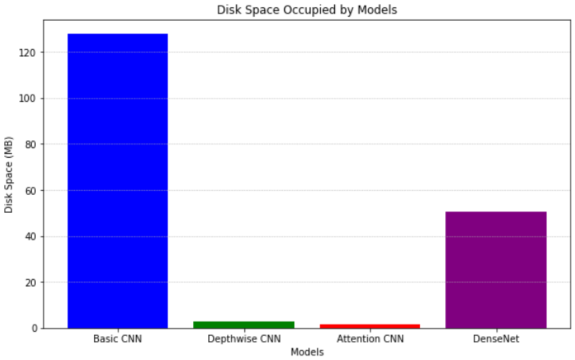

Fig. 12: Disk Space Occupied by Models

D. Analysis of Results

This section analyzes the tradeoffs between performance and complexity for each model, based on accuracy, F1 score, execution time, memory usage, and confusion matrices.

1) Accuracy and Loss: The accuracy and loss plots (Figures 1 and 2) reveal that **Depthwise CNN** consistently outperforms the other models, achieving the highest accuracy and the lowest loss throughout training. **Basic CNN** performs reliably, showing steady improvements in both accuracy and loss, though it does not reach the performance of the Depthwise CNN. **Attention CNN**, while demonstrating moderate improvements, lags behind in both accuracy and loss, indicating that it may require further tuning or longer training periods to fully leverage its capabilities. **DenseNet**, despite early improvements, shows the highest loss and struggles to optimize as effectively, indicating potential difficulties with optimization and generalization.

2) Precision and Recall: The precision and recall plots (Figures 3 and 4) indicate that **Depthwise CNN** provides the best balance between precision and recall, consistently achieving the highest values across training. **Basic CNN** also performs well, with strong recall and slightly lower precision compared to Depthwise CNN. **Attention CNN**, while demonstrating moderate precision and recall, lags behind in minimizing false positives and false negatives, indicating room for improvement. **DenseNet**, although showing early improvement, stabilizes at lower precision and recall levels, particularly struggling with false positives, leading to a reduction in overall performance.

3) AUC and F1 Score: The AUC and F1 score plots (Figures 5 and 6) indicate that **Depthwise CNN** and **Basic CNN** deliver the highest performance across both metrics, with Depthwise CNN achieving the highest AUC values, and Basic CNN slightly surpassing Depthwise CNN in F1 score by the final epochs. **Attention CNN** follows with slightly lower AUC and F1 scores, demonstrating moderate but consistent performance. **DenseNet**, while improving early on, shows the lowest AUC and F1 scores by the end of training, reflecting its difficulties in maintaining a strong balance between precision and recall, and in distinguishing between classes effectively.

4) Confusion Matrix Analysis: **Basic CNN**: The Basic CNN (Figure 7)demonstrates a reasonable performance in classifying pneumonia, with 158 correctly classified cases. However, it struggles significantly with COVID-19 and normal cases, correctly identifying only 72 COVID-19 cases and misclassifying 64 COVID-19 cases as normal. This indicates the modelâ€™s difficulty in distinguishing between COVID-19 and normal cases, though it performs well for pneumonia.

Depthwise CNN: Depthwise CNN (Figure 9)performs consistently well, correctly classifying 135 COVID-19, 126 normal, and 156 pneumonia cases. This model has fewer misclassifications overall, showing its strength in distinguishing between the three classes with high precision and recall. However, the confusion matrix reveals some confusion between COVID-19 and normal cases, with 16 COVID-19 cases misclassified as normal and 22 normal cases misclassified as COVID-19. Despite these errors, Depthwise CNN still outperforms the other models in overall classification accuracy.

Attention CNN: The Attention CNN (Figure 10)shows balanced classification across all classes, correctly identifying 134 COVID-19, 134 normal, and 156 pneumonia cases. It performs well in distinguishing normal and pneumonia cases but still misclassifies 16 COVID-19 cases as normal. The confusion matrix suggests that while the attention mechanism improves performance across all classes, it may require further tuning to fully optimize COVID-19 detection.

DenseNet: DenseNet (Figure 8) struggles with classifying normal cases, correctly identifying only 72, while misclassifying 79 normal cases as COVID-19. It shows relatively good performance in pneumonia classification, with 132 correct predictions, but suffers from a high rate of misclassification between normal and COVID-19. The confusion matrix reflects a lack of balance in performance across the three categories, particularly in distinguishing normal from disease states.

5) Inference Time and Memory Usage: The inference time and memory usage plots (Figures 11 and 12) reveal that **Basic CNN** is the fastest model, making it ideal for real-time applications. However, it has the highest memory usage, which may be a limitation in storage-constrained environments. **Depthwise CNN** offers a good balance, with moderate inference time and excellent memory efficiency, using minimal disk space. **Attention CNN** provides relatively fast inference times and also demonstrates strong memory efficiency, making it a viable choice for performance-critical tasks. **DenseNet**, while more memory-efficient than Basic CNN, suffers from the slowest inference time, limiting its practicality in real-world applications where both speed and resource efficiency are important considerations.

In conclusion, **Depthwise CNN** provides the most consistent and accurate classification across all categories, although it still exhibits some confusion between COVID-19 and normal cases. **Basic CNN** and **DenseNet** perform adequately but show significant misclassification of normal cases. **Attention CNN**, while strong in balancing the classification of all classes, may require additional tuning to enhance its COVID-19 detection performance.

E. Conclusion

The **Depthwise CNN** emerged as the best performer, achieving high accuracy, precision, and recall, while maintaining efficient inference time and low memory usage. However, it exhibited some misclassification between COVID-19 and normal cases.

The **Basic CNN** performed well in terms of inference speed but struggled with COVID-19 misclassifications, while **DenseNet** showed poor generalization and slow inference times. The **Attention CNN** provided balanced classification across all categories but needs further tuning for better COVID-19 detection.

Overall, **Depthwise CNN** offers the best trade-off between accuracy and efficiency, making it the most suitable model for real-world medical applications.

V. Concluding Remarks

In this study, we explored the application of convolutional neural networks (CNNs) to classify X-ray images into three categories: COVID-19, pneumonia, and normal. We implemented and compared several models, including a Basic CNN, Depthwise CNN, Attention CNN, and DenseNet, evaluating them on key metrics such as accuracy, precision, recall, AUC, F1-score, inference time, memory usage, and confusion matrix.

The results demonstrate that the **Depthwise CNN** model consistently outperforms the other architectures across most performance metrics. It achieves high test accuracy (87.06%), precision (96.19%), recall (88.39%), and F1-score (92.12%), offering an excellent balance between classification performance and computational efficiency. Additionally, it achieves the lowest loss and maintains a favorable trade-off between inference speed and memory usage, making it a practical solution for real-world deployment, particularly in medical diagnostics.

The **Attention CNN** model, while not the top performer overall, excels in the confusion matrix analysis, demonstrating the most balanced classification of COVID-19, normal, and pneumonia cases. It correctly identifies 134 COVID-19 cases, 134 normal cases, and 156 pneumonia cases, significantly reducing the confusion between COVID-19 and normal categories, a common challenge for other models. However, despite this improvement in specific areas, it lags behind Depthwise CNN in terms of accuracy, precision, recall, and F1-score, suggesting the need for further tuning.

The **Basic CNN** serves as a strong baseline, achieving competitive performance with high F1-score (88.16%) and recall (84.06%). However, it struggles with memory usage and misclassifies a higher number of COVID-19 cases as normal (64 instances), which impacts its overall utility in medical classification tasks. The **DenseNet** model, despite its promising potential, exhibits signs of overfitting, leading to poor generalization with lower accuracy (68.27%) and high misclassification rates, particularly between normal and COVID-19 cases.

In conclusion, the **Depthwise CNN** emerges as the best overall performer, striking an optimal balance between high performance and resource efficiency. While the **Attention CNN** shows promise in reducing misclassifications, particularly between COVID-19 and normal cases, its overall metrics suggest it needs further optimization to match the robust performance of the Depthwise CNN. These findings highlight the potential of CNN architectures in advancing medical image classification, particularly for diseases such as COVID-19 and pneumonia, where precise diagnosis is critical for effective treatment.

REFERENCES

[1] J. P. Cohen, P. Morrison, and L. Dao, "Covid-19 image data collection: Prospective predictions are the future," *arXiv preprint arXiv:2006.11988*, June 2020.

[2] L. Wang and A. Wong, "Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images," *arXiv preprint arXiv:2003.09871*, March 2020.

[3] I. D. Apostolopoulos and T. A. Mpesiana, "Covid-19: Automatic detection from x-ray images utilizing transfer learning with convolutional neural networks," *Physical and Engineering Sciences in Medicine*, vol. 43, pp. 635–640, June 2020.

[4] T. Ozturk, M. Talo, E. Yildirim, U. B. Baloglu, O. Y. Acharya, and U. R. Rajendra, "Automated detection of covid-19 cases using deep neural networks with x-ray images," *Computers in Biology and Medicine*, vol. 121, p. 103792, June 2020.

[5] M. Heidari, A. Heidari, M. Sidorov, and F. Heidari, "Improving the focus of attention mechanisms for medical image analysis," *arXiv preprint arXiv:2004.08973*, April 2020.