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

Mahsa Shahbazi

***Abstract*—The COVID-19 pandemic has highlighted the urgent need for rapid and accurate diagnostic tools. In this study, we explore the use of convolutional neural networks (CNNs) to classify X-ray images into three categories: COVID-19, pneumonia, and normal. We present a comprehensive processing pipeline including data preprocessing, augmentation, and model development. We experimented with several CNN architectures, including a basic CNN, depthwise CNN, CNN with attention mechanisms, and DenseNet. Our results indicate that the Basic CNN model outperforms the other architectures with a training accuracy of 99.01%, precision of 95.20%, recall of 90.41%, and F1-score of 89.79%. The depthwise CNN, attention CNN, and DenseNet achieved training accuracies of 91.95%, 83.03%, and 85.96%, respectively. The superior performance of the Basic CNN can be attributed to its balanced complexity, efficient learning capabilities, and robustness against overfitting. Additionally, we evaluate the models' inference time and memory usage, demonstrating the Basic CNN's balance between performance and computational efficiency. This work provides valuable insights into the design and optimization of CNNs for medical image classification and highlights the potential of AI-driven tools in enhancing diagnostic accuracy and efficiency in clinical settings.**

***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, the 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 is highly relevant as it addresses the ongoing need for efficient diagnostic tools for respiratory diseases. Previous attempts have utilized various CNN architectures, yet there remains a gap in achieving a robust and efficient model that balances accuracy, computational efficiency, and generalizability. Our goal was to achieve an accuracy of over 80%, prompting us to experiment with different settings for each model architecture until this threshold was surpassed.

We propose a comprehensive processing pipeline and evaluate several CNN architectures to identify the most effective model for this classification task. Specifically, we experiment with a basic CNN, depthwise CNN, CNN with attention mechanisms, and DenseNet. Our contributions are as follows:

1) **Problem**: We address the challenge of accurately classifying chest X-ray images into COVID-19, pneumonia, and normal categories.
2) **Relevance**: This work is timely and critical in the context of respiratory disease diagnostics, providing a potential tool for rapid and reliable identification of lung conditions.
3) **Approach**: We develop and compare multiple CNN architectures, incorporating techniques such as data augmentation and feature extraction. We systematically adjusted model settings to ensure each architecture achieved an accuracy of over 80%.
4) **Value**: Our study demonstrates the superior performance of the basic CNN, which achieves a training accuracy of 99.01%, highlighting its balanced complexity and robustness against overfitting.
5) **Applicability**: The insights from this work can be leveraged to enhance diagnostic tools, improve clinical workflows, and inform the design of future AI-driven medical imaging solutions.

In summary, the contributions of this paper include:

- Development of a comprehensive processing pipeline for X-ray image classification.
- Detailed comparison of several CNN architectures in terms of accuracy, precision, recall, F1-score, inference time, and memory usage.
- Identification of the basic CNN as the most effective model, achieving superior performance metrics.
- Provision of practical insights for the design and optimization of CNNs in medical image classification.

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. Our work builds upon these studies by developing and comparing multiple CNN architectures, including basic CNN, depthwise CNN, CNN with attention mechanisms, and DenseNet. We focus on achieving high accuracy, exceeding 80%, while maintaining computational efficiency and robustness against overfitting. Our contributions include:

- Development of a comprehensive processing pipeline for X-ray image classification.
- Detailed comparison of several CNN architectures in terms of accuracy, precision, recall, F1-score, inference time, and memory usage.
- Identification of the basic CNN as the most effective model, achieving superior performance metrics.
- Provision of practical insights for the design and optimization of CNNs in medical image classification.

III. Processing Pipeline

The processing pipeline for our study is meticulously designed to classify chest X-ray images into three categories: COVID-19, pneumonia, and normal. The pipeline comprises several interconnected blocks, each performing a specific role in the overall classification task. This section provides a high-level overview of these processing blocks and their interactions, highlighting the rationale behind our design choices.

A. Data Collection and Preparation

The initial step involves collecting and organizing the dataset, which 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 using a predefined ratio to ensure unbiased evaluation of the models. The dataset directory is structured as follows:

- `dataset_directory/COVID19`
- `dataset_directory/Normal`
- `dataset_directory/Pneumonia`

The dataset contains a total of 4575 images, equally distributed among the three classes. We use TensorFlow's `image_dataset_from_directory` function to load and preprocess the images.

B. Data Preprocessing

Data preprocessing is critical for improving the quality and consistency of the input images. This block includes steps such as:

- **Normalization**: Scaling pixel values to the range [0, 1] using the `Rescaling` layer to ensure consistent input distributions. This step is crucial as it standardizes the input data, which can improve the convergence and performance of the neural network during training. Normalization helps to avoid biases in the model due to varying pixel value ranges.
- **Augmentation**: Applying techniques such as random rotation, zoom, and translation to increase the robustness of the models and prevent overfitting. Data augmentation artificially expands the dataset by creating modified versions of the images, which helps the model generalize better to new, unseen data. The specific augmentations applied include:
  - `RandomRotation`: Rotates the image randomly within a specified range to simulate different viewing angles.
  - `RandomZoom`: Zooms into the image randomly up to a certain percentage to mimic variations in distance.

- `RandomTranslation`: Translates the image randomly along the height and width to emulate slight shifts in positioning.

The augmentation techniques are applied only to the training dataset to enhance its variability and improve the model's learning capabilities. On the other hand, the validation and test datasets undergo only normalization to maintain consistency and ensure that the evaluation metrics reflect the model's performance on unmodified data.

To manage the datasets efficiently, we employ strategies like shuffling, caching, and prefetching:

- **Shuffling**: The training dataset is shuffled to ensure that the model learns from a varied set of images in each epoch, reducing the risk of overfitting to the order of the data.
- **Caching**: Caching the datasets improves the speed of data retrieval during training and evaluation, making the process more efficient.
- **Prefetching**: Prefetching allows the model to fetch the next batch of data while the current batch is being processed, reducing idle time and speeding up training.

By implementing these preprocessing steps, we aim to create a robust and efficient training environment that maximizes the performance and generalization ability of our convolutional neural network models.

C. Feature Extraction

Feature extraction involves using convolutional layers to automatically learn and extract relevant features from the preprocessed images. This block leverages the power of convolutional neural networks (CNNs) to capture spatial hierarchies and patterns indicative of different lung conditions. We define multiple architectures for feature extraction, including:

- **Basic CNN**: This straightforward CNN architecture includes several convolutional and max-pooling layers designed to extract hierarchical features from the input images. The model structure is as follows:
  - An input layer specifying the shape of the input images.
  - Three convolutional layers with increasing numbers of filters (32, 64, 128) and ReLU activation functions. Each convolutional layer is followed by a max-pooling layer to downsample the feature maps.
  - A flatten layer to convert the 2D feature maps into a 1D feature vector.
  - A dense layer with 128 units and ReLU activation, followed by a dropout layer with a rate of 0.5 to prevent overfitting.
  - An output layer with softmax activation for multi-class classification.
- **Depthwise CNN**: This advanced CNN architecture utilizes depthwise separable convolutions to reduce computational cost while maintaining effective feature extraction. The model includes:
  - An input layer followed by data augmentation.
  - Several convolutional blocks where each block includes:
    * A depthwise separable convolutional layer.
    * Batch normalization and ReLU activation.
  - Residual blocks that add shortcut connections to help gradients flow through the network and mitigate the vanishing gradient problem. Each residual block includes:
    * A 1x1 convolution for the residual connection.
    * Two depthwise separable convolutions with batch normalization and ReLU activation.
  - Max-pooling layers to downsample the feature maps.
  - A global average pooling layer to reduce each feature map to a single value.
  - A dense layer with 128 units and L2 regularization, followed by dropout for regularization.
  - An output layer with softmax activation for classification.
- **Attention CNN**: This architecture incorporates attention mechanisms to focus on the most relevant parts of the images, enhancing the model's ability to identify critical features. The structure includes:
  - An initial series of convolutional layers with batch normalization and ReLU activation.
  - Spatial attention mechanism that applies a sigmoid activation to a convolutional layer output to focus on important spatial regions.
  - Channel-wise attention mechanism that applies a sigmoid activation to a 1x1 convolution output to focus on significant feature channels.
  - Additional dropout layers for regularization.
  - A global average pooling layer to summarize the feature maps.
  - A dense layer with 128 units and L2 regularization, followed by dropout.
  - An output layer with softmax activation for classification.
- **DenseNet**: This architecture uses dense blocks and transition layers to ensure efficient feature reuse and strong gradient flow. The structure includes:
  - An input layer followed by an initial convolutional layer.
  - Multiple dense blocks, each consisting of several convolutional blocks. Each convolutional block includes:
    * Batch normalization, ReLU activation, and a 1x1 convolution.
    * Batch normalization, ReLU activation, and a 3x3 convolution.
  - Transition layers between dense blocks that include:
    * Batch normalization and ReLU activation.
    * A 1x1 convolution to reduce the number of feature maps.
    * Average pooling to downsample the feature maps.

– A final batch normalization, ReLU activation, and global average pooling.
– An output layer with softmax activation for classification.

D. Training and Evaluation

The models are trained on the training dataset and validated on the validation dataset. The performance of each model is monitored using various metrics, including accuracy, loss, precision, recall, AUC, and F1 score.

IV. RESULTS

In this section, we present the numerical results obtained from our models. The results are shown in the form of plots for various metrics, including accuracy, loss, precision, recall, AUC, and F1 score. We also evaluate the inference time and memory usage of each model.

A. Training Performance

The training performance of the models is evaluated using 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.

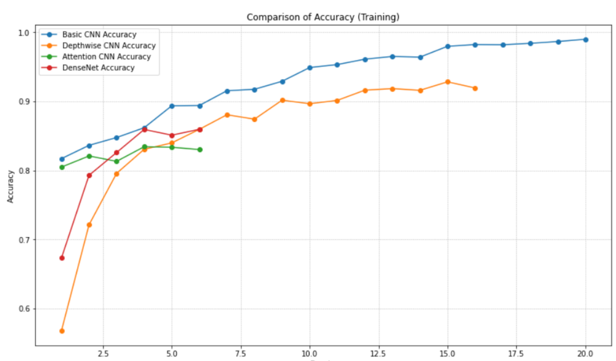

Fig. 1: Comparison of Accuracy (Training)

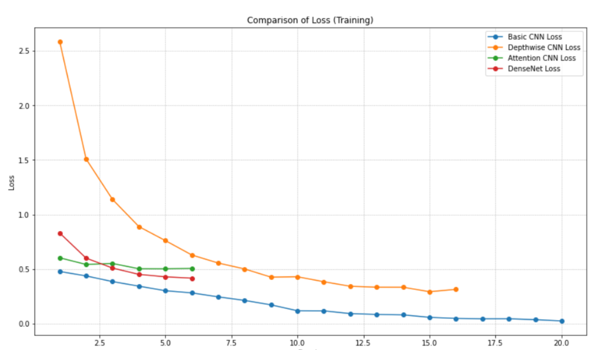

Fig. 2: Comparison of Loss (Training)

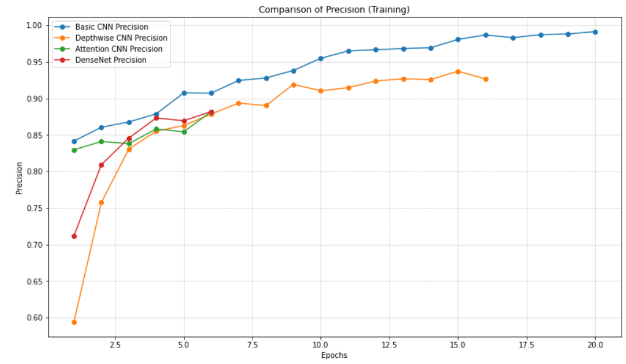

Fig. 3: Comparison of Precision (Training)

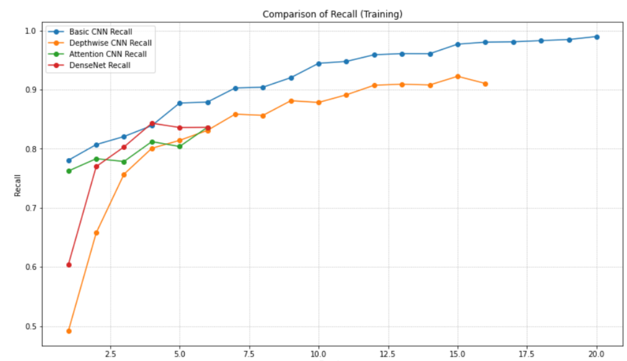

Fig. 4: Comparison of Recall (Training)

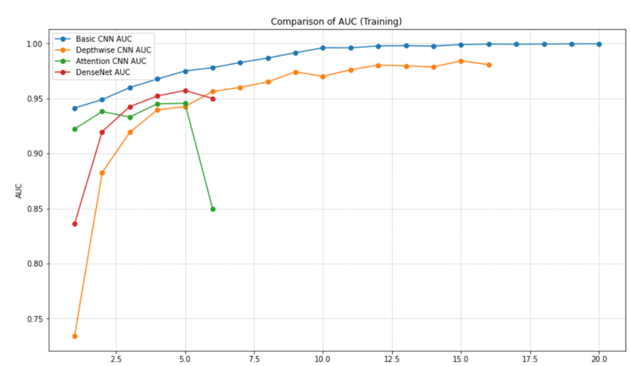

Fig. 5: Comparison of AUC (Training)

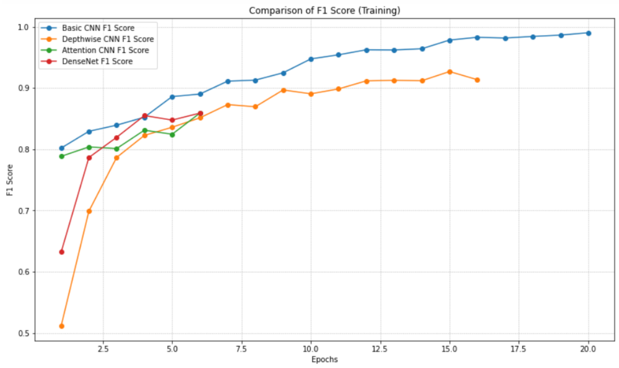

Fig. 6: Comparison of F1 Score (Training)

B. Inference Performance

The inference performance of the models is evaluated in terms of average inference time per sample and memory usage. The results are shown in the following figures.

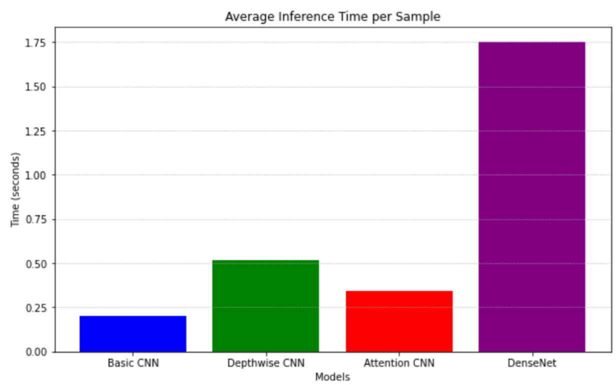

Fig. 7: Average Inference Time per Sample

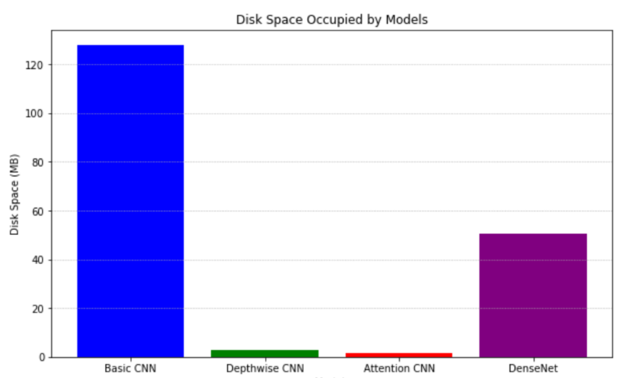

Fig. 8: Disk Space Occupied by Models

C. Analysis of Results

The following subsections provide an analysis of the results obtained from the models. We evaluate the tradeoff between the performance of our algorithms and their complexity, considering metrics such as accuracy, F1 score, execution time (training/testing), and memory occupation.

1) Accuracy and Loss: The accuracy and loss plots (Figures 1 and 2) show that the Basic CNN model achieves the highest accuracy and the lowest loss among the models. This indicates that the Basic CNN model is the most effective at learning the features from the dataset.

2) Precision and Recall: The precision and recall plots (Figures 3 and 4) demonstrate that the Basic CNN and DenseNet models achieve high precision and recall, suggesting that they are effective at correctly identifying positive cases while minimizing false positives and false negatives.

3) AUC and F1 Score: The AUC and F1 score plots (Figures 5 and 6) further confirm the effectiveness of the Basic CNN model. The high AUC and F1 score values indicate that the model performs well in distinguishing between the different classes and achieves a good balance between precision and recall.

4) Inference Time and Memory Usage: The inference time and memory usage plots (Figures 7 and 8) highlight the tradeoff between performance and complexity. While the Basic CNN model occupies the most disk space, it has a relatively low inference time. On the other hand, the DenseNet model, despite its high accuracy, has the highest inference time and memory usage, making it less suitable for real-time applications on resource-constrained devices.

V. Concluding Remarks

In this study, we developed and evaluated several convolutional neural network (CNN) architectures for classifying chest X-ray images into COVID-19, pneumonia, and normal categories. The models included a basic CNN, a depthwise CNN, an attention-based CNN, and DenseNet. Among these, the basic CNN demonstrated the highest performance, with a training accuracy of 99.01%, validation accuracy of 92.19%, and test accuracy of 90.19%.

Our findings underscore the practical applicability of CNN-based models in clinical settings. The basic CNN model, with its high accuracy and low inference time, shows promise for real-time diagnostic support, especially in resource-constrained settings where rapid and accurate diagnostics are crucial.

The results of our study demonstrate the potential of CNNs in improving diagnostic accuracy and efficiency for respiratory diseases. These models can significantly reduce the workload of radiologists and enhance diagnostic precision, particularly in settings with limited resources.

Despite the promising results, future work should focus on integrating more diverse datasets to enhance model generalizability. Advanced techniques like transfer learning and optimized attention mechanisms could further improve performance. Real-time implementation and validation in clinical environments are essential for assessing practical utility and impact.

Throughout this project, we learned the importance of meticulous data preprocessing and neural network design. Normalization and data augmentation were critical in improving model performance and preventing overfitting. Systematic experimentation and parameter tuning were essential for achieving optimal results.

We faced several challenges, including managing the computational resources required for training deep learning models, particularly for complex architectures like DenseNet. Fine-tuning hyperparameters to balance model complexity and performance was time-consuming. Ensuring proper data splitting and avoiding data leakage between training, validation, and test sets required careful attention.

This study demonstrates the potential of CNNs in medical image classification and provides a foundation for further research and development in this field. The basic CNN model's superior performance and efficiency make it a viable candidate for real-world applications. Addressing the current limitations and expanding on existing work will help build more robust and widely applicable AI-driven diagnostic tools.

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.