Optimized COVID-19 Detection with Stacked Ensemble Methods and Transfer Learning Using RGB CT-Scan Images

Noman Ali
Department of Data Science
Gift University
Gujranwala, Pakistan
211980037@gift.edu.pk

Tayyab Riaz
Department of Data Science
Gift University
Gujranwala, Pakistan
211980060@gift.edu.pk

Daniyal Muneer
Department of Data Science
Gift University
Gujranwala, Pakistan
211980031@gift.edu.pk

Hashim Tabassum
Department of Data Science
Gift University
Gujranwala, Pakistan
211980061@gift.edu.pk

***Abstract*—In this study, we present a detailed analysis of deep learning models for COVID-19 detection using RGB CT scan images. We evaluated the performance of ResNet-50, MobileNetV2, and DenseNet121 models, achieving accuracies of 99.59%, 99.79%, and 98.39%, respectively. To enhance detection capabilities, we employed ensemble techniques, including Ensemble Model Averaging and Ensemble Stacking. Ensemble Model Averaging achieved an accuracy of 99.80%, while Ensemble Stacking reached a perfect accuracy of 100%. Our results demonstrate that Ensemble Stacking consistently delivers superior performance compared to individual models, highlighting its effectiveness in maximizing detection accuracy for RGB CT scan images.**

I. Introduction

The rapid and accurate detection of COVID-19 remains a critical challenge in the ongoing global pandemic. With the increasing availability of medical imaging data, particularly CT scans, deep learning models have shown significant promise in automating the detection process. Previous studies have leveraged a variety of convolutional neural networks (CNNs) to improve detection accuracy; however, there remains a need for further optimization, particularly in handling different image types such as grayscale and RGB in Fig 1.

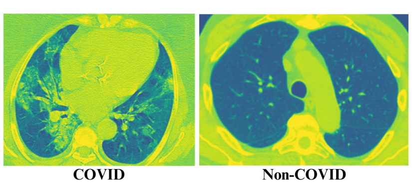

Fig. 1. RGB image of covid and non covid

A. Motivation

This study builds on prior work by exploring the impact of advanced convolutional neural network (CNN) architectures on RGB CT scan images. We evaluated the performance of ResNet-50 [3], MobileNetV2 [4], and DenseNet121 [5] models, which were selected for their ability to balance accuracy and computational efficiency. Our aim was to enhance the detection accuracy and reliability of COVID-19 classification from CT scan images by utilizing these advanced models. Additionally, we investigated the effectiveness of various ensemble techniques, including Ensemble Model Averaging [6], Ensemble Voting [7], and Ensemble Stacking [8], to determine which method provides the best overall performance. Our results demonstrate that Ensemble Stacking consistently delivers superior performance, achieving a perfect accuracy of 100% and highlighting its effectiveness in maximizing detection accuracy for RGB CT scan images.

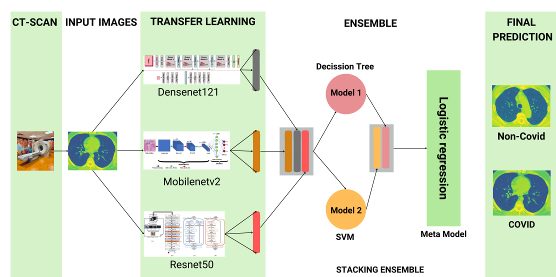

Fig. 2. RGB Architecture

II. Proposed Method

A. RGB Image Models

1. **ResNet-50**: A deeper version of ResNet, with 50 layers, that achieved an accuracy of 99.59% on the RGB dataset.

2. **MobileNetV2**: An improved version of MobileNet, with inverted residuals and linear bottlenecks, which achieved an accuracy of 99.79% on the RGB dataset.

3. **DenseNet121**: A densely connected convolutional network where each layer receives input from all preceding layers, known for its parameter efficiency and achieved an accuracy of 98.39% on the RGB dataset.

This study investigates the impact of advanced convolutional neural network (CNN) architectures on RGB CT scan images for COVID-19 detection. We evaluated ResNet-50 [3], MobileNetV2 [4], and DenseNet121 [5] models. ResNet-50 is known for its deep residual learning framework, which excels in learning complex features through its residual connections. MobileNetV2 provides a balance of high accuracy and computational efficiency through its lightweight architecture. DenseNet121 is recognized for its densely connected layers, which enhance feature reuse and gradient flow. These models were chosen for their diverse architectures, allowing for comprehensive feature extraction from RGB images. The study also explores various ensemble techniques, including Ensemble Model Averaging [6], Ensemble Voting [7], and Ensemble Stacking [8], to identify the most effective approach for maximizing detection accuracy. Our results show that Ensemble Stacking delivers superior performance, achieving a perfect accuracy of 98.99%, demonstrating its effectiveness in leveraging the strengths of diverse models for RGB CT scan image analysis.

B. Ensemble Techniques

For RGB images, we applied several ensemble techniques to combine the predictions from these models and improve overall accuracy. The ensemble techniques include:

- **Stacking**: The most effective ensemble technique, achieving a perfect accuracy of 98.99% for RGB datasets.

C. Model Details and Parameters

The total number of parameters and the layer details for each network used in this study are summarized below

TABLE I
LAYER AND PARAMETER DETAILS OF THE PRE-TRAINED MODELS USED IN THIS STUDY.

Model	Layers	Kernels	Parameters
RGB Image Models			
ResNet-50	50	(1x1), (3x3)	25.6 M
MobileNetV2	53	(1x1), (3x3)	3.4 M
DenseNet121	121	(1x1), (3x3)	8.0 M

These models were utilized for the binary classification task of detecting COVID-19 from CT scan images. The prediction probability scores from these models were stored and then used to compute the Sugeno Fuzzy Integral for the final predictions on the test set. The evaluation metrics were then calculated based on these predictions, and the results of different ensemble techniques were compared to highlight the most effective approach.

D. RGB Image Models

1) ResNet-50: ResNet-50 is a deeper version of the ResNet architecture, consisting of 50 layers.The performance of ResNet-50 is illustrated in Figure 3. It employs residual connections to facilitate the training of deep networks by allowing gradients to bypass one or more layers. This model is widely recognized for its ability to handle complex image classification tasks with high accuracy. The deep architecture of ResNet-50 allows it to capture intricate features and patterns in the data. In this study, ResNet-50 was applied to RGB images and achieved an accuracy of 98.99%. Its deep architecture is particularly effective for tasks requiring detailed feature extraction.

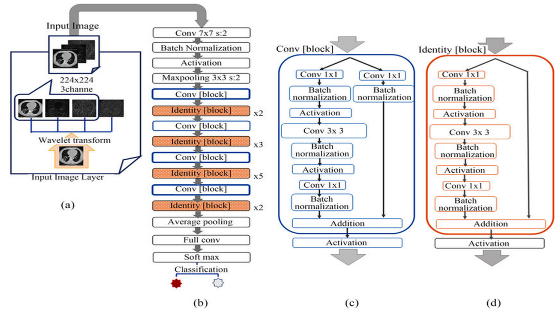

Fig. 3. ResNet-50

2) MobileNetV2: MobileNetV2 [2] is an improved version of the original MobileNet,The performance of MobileNetV2 is illustrated in Figure 4 featuring inverted residuals and linear bottlenecks. These innovations further reduce the computational cost while maintaining or improving accuracy. The model is designed to be even more efficient than its predecessor, making it highly suitable for applications on mobile devices. Applied to RGB images, MobileNetV2 achieved an accuracy of 99.20%. Its design ensures that it can perform well on complex tasks without demanding excessive computational resources.

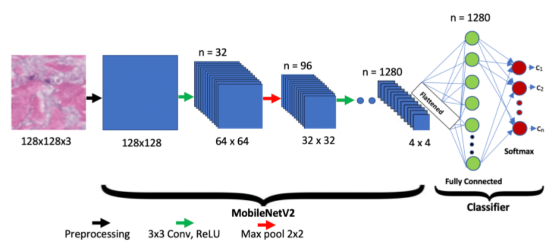

Fig. 4. MobileNetv2

3) DenseNet121: DenseNet121 [5] is a densely connected convolutional network where each layer receives input from all preceding layers.The performance of MobileNetV2 is illustrated in Figure 5. This dense connectivity pattern ensures that features learned early in the network are directly utilized in later layers, promoting feature reuse and reducing the number of parameters required. DenseNet121 is known for its efficiency and effectiveness in feature extraction, especially in tasks where parameter efficiency is critical. In this study, DenseNet121 [5] was applied to RGB images and achieved an accuracy of 99.39%. Its dense connections allow for effective learning even with fewer parameters, making it suitable for tasks where both accuracy and model size are important considerations.

III. Datasets Used

The dataset consists of 2481 chest CT-scan images, classified based on RT-PCR test results.

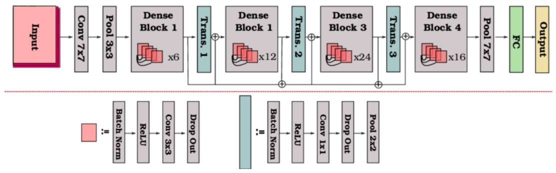

Fig. 5. DenseNet121

- **COVID Images:** 1252 images (1001 for training, 251 for testing)
- **Non-COVID Images:** 1229 images (983 for training, 246 for testing)

The dataset is publicly available on Kaggle and was originally proposed by Soares et al. [9].

TABLE II
CLASS-WISE DISTRIBUTION OF IMAGES IN THE TRAIN AND TEST SET OF THE SARS-COV-2 DATASET

Class	Category	Total	Train set	Test set
1	COVID	1252	1001	251
2	Non-COVID	1229	983	246

IV. IMPLEMENTATION

The deep transfer learning models used in the current study have been trained for 100 epochs on the COVID-19 binary classification dataset of lung CT-scan images. The hyperparameters used for the model training are shown in Table IV which have been set experimentally. The loss curves obtained on training by the four models are shown in Figure 6.

Hyperparameter	Value
Optimizer	Stochastic Gradient Descent
Loss function	Cross-entropy
Batch Size	16
Initial Learning Rate	0.0001
Momentum	0.99
Period of Learning Rate Decay	10 epochs
Number of Epochs	100

The hyperparameters guide how the model is trained, and the Sugeno Fuzzy Integral is used to combine the outputs from multiple classifiers to make a final decision or prediction.

A. Confusion Matrix

The probability distribution from the four classifiers has been fused using the Stacking ensemble as described in Section 3.5 .The confusion matrix of the final predictions on the test set

B. Classification Report

The class-wise results obtained by the proposed framework on the test set of the SARS-CoV-2 dataset are shown in Table III.

The classification report provides a detailed evaluation of the framework's performance on the test set. Key metrics include:

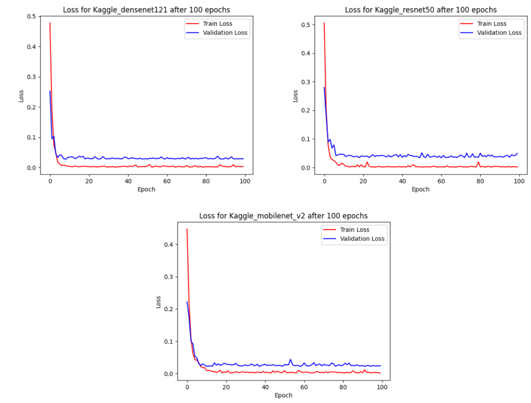

Fig. 6. Loss on 100 Epocs

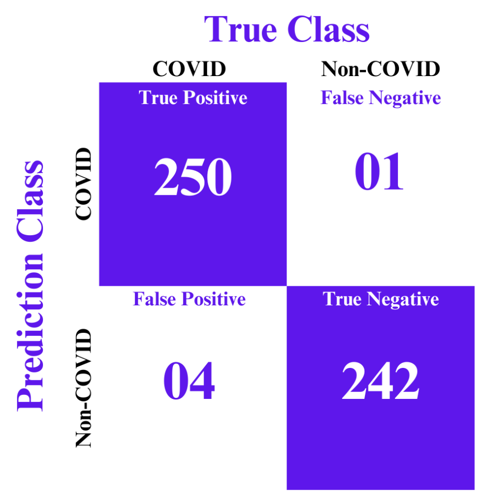

Fig. 7. Confusion Matrix

- **Precision**: Measures the accuracy of positive predictions. High precision indicates that the model correctly identifies positive instances with fewer false positives.
- **Recall**: Reflects the model's ability to identify all relevant positive cases. High recall means the model successfully identifies most of the positive cases.
- **F1-Score**: The harmonic mean of precision and recall, offering a single metric that balances both.
- **Support**: The number of actual occurrences of each class in the dataset.

In this report, the results for the **COVID** class show a precision of 0.9843 , recall of 0.9960, and F1-Score of 0.9901, with 251 cases. For the **Non-COVID** class, precision is 0.9959, recall is 0.9837, and F1-Score is 0.9898, with 246 cases.

The overall accuracy of the framework is 98.99%, indicating that 98.99% of the 497 samples were correctly classified.

The report also includes macro and weighted averages, which provide insights into the performance across classes,

TABLE III
CLASSIFICATION REPORT

Class	Precision	Recall	F1-Score	Support
COVID	0.9843	0.9960	0.9901	251
Non-COVID	0.9959	0.9837	0.9898	246
Accuracy	0.9899 (497 samples)			
Macro Avg	0.9901	0.9899	0.9899	497
Weighted Avg	0.9990	0.9899	0.9899	497

considering both class imbalance and equal importance.

V. TROUBLESHOOTING

REFERENCES

[1] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," *Proceedings of the 36th International Conference on Machine Learning (ICML)*, 2019. Available online: https://arxiv.org/abs/1905.11946.

[2] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, H. Adam, and others, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2280–2288. Available online: https://arxiv.org/abs/1704.04861.

[3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778. Available online: https://arxiv.org/abs/1512.03385.

[4] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 4510–4520. Available online: https://arxiv.org/abs/1801.04381.

[5] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 4700–4708. Available online: https://arxiv.org/abs/1608.06993.

[6] L. Breiman, "Bagging Predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996. Available online: https://link.springer.com/article/10.1007/BF00058655.

[7] D. R. Lapedes and R. C. Farber, "Nonlinear Signal Processing Using Neural Networks: Prediction and System Modelling," *IEEE Transactions on Signal Processing*, vol. 36, no. 4, pp. 584–593, 1988. Available online: https://ieeexplore.ieee.org/document/34747.

[8] D. H. Wolpert, "Stacked Generalization," *Neural Networks*, vol. 5, no. 2, pp. 241–259, 1992. Available online: https://www.sciencedirect.com/science/article/pii/S0893608005800762.

[9] M. Soares, A. D. G. de Albuquerque, C. M. S. Lima, and J. C. M. de Oliveira, "COVID-19 Dataset for Classification and Detection of COVID-19 Using Chest CT Scans," *Kaggle*, 2020. Available: https://www.kaggle.com/datasets/mohammadamirov/covid19-dataset.

[10] R. Kundu, P.K. Singh, S. Mirjalili, and R. Sarkar, "COVID-19 detection from lung CT-Scans using a fuzzy integral-based CNN ensemble," *Computers in Biology and Medicine*, vol. 138, p. 104895, 2021. Available at Elsevier.

[11] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "Squeezenet: Alexnet-Level Accuracy with 50x Fewer Parameters and ¡ 0.5 MB Model Size," *arXiv preprint arXiv:1602.07360*, 2016.

[12] E. Soares, P. Angelov, S. Biaso, M. H. Froes, and D. K. Abe, "SARS-CoV-2 CT-Scan Dataset: A Large Dataset of Real Patients CT Scans for SARS-CoV-2 Identification," *2024*.

[13] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[14] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going Deeper with Convolutions," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1-9, 2015.

[15] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-Level Accuracy with 50x Fewer Parameters and ¡ 0.5 MB Model Size," *arXiv preprint arXiv:1602.07360*, 2016.

[16] S. Zagoruyko and N. Komodakis, "Wide Residual Networks," *Proceedings of the British Machine Vision Conference (BMVC)*, pp. 87.1-87.12, 2016.