

A Novel Multi-branch ConvNeXt Architecture for Identifying Subtle Pathological Features in CT Scans

Irash Perera

Department of Computer Science and Engineering
University of Moratuwa
Colombo, Sri Lanka
irash.21@cse.mrt.ac.lk

Uthayasanker Thayasivam

Department of Computer Science and Engineering
University of Moratuwa
Colombo, Sri Lanka
rtuthaya@cse.mrt.ac.lk

Abstract—Intelligent analysis of medical imaging plays a crucial role in assisting clinical diagnosis, especially for identifying subtle pathological features. This paper introduces a novel multi-branch ConvNeXt architecture designed specifically for the nuanced challenges of medical image analysis. While applied here to the specific problem of COVID-19 diagnosis, the methodology offers a generalizable framework for classifying a wide range of pathologies from CT scans. The proposed model incorporates a rigorous end-to-end pipeline, from meticulous data preprocessing and augmentation to a disciplined two-phase training strategy that leverages transfer learning effectively. The architecture uniquely integrates features extracted from three parallel branches: Global Average Pooling, Global Max Pooling, and a new Attention-weighted Pooling mechanism. The model was trained and validated on a combined dataset of 2,609 CT slices derived from two distinct datasets. Experimental results demonstrate a superior performance on the validation set, achieving a final ROC-AUC of 0.9937, a validation accuracy of 0.9757, and an F1-score of 0.9825 for COVID-19 cases, outperforming all previously reported models on this dataset. These findings indicate that a modern, multi-branch architecture, coupled with careful data handling, can achieve performance comparable to or exceeding contemporary state-of-the-art models, thereby proving the efficacy of advanced deep learning techniques for robust medical diagnostics.

Index Terms—COVID-19, ConvNeXt, Transfer learning, Medical Image Analysis, Computer vision

I. INTRODUCTION AND BACKGROUND

The rapid global spread of Coronavirus Disease 2019 (COVID-19) highlighted the critical need for rapid and accurate diagnostic tools to manage the pandemic and mitigate its spread. While Reverse Transcription Polymerase Chain Reaction (RT-PCR) tests were the gold standard for confirmation, their shortage during peak outbreaks necessitated alternative diagnostic methods. [1] Computed Tomography (CT) scans emerged as a valuable tool for screening and diagnosing COVID-19, as they have been shown to be more sensitive than RT-PCR tests. [2] The analysis of CT scans is a time-intensive process that requires specialized medical expertise, which can be a significant bottleneck, especially for overwhelmed healthcare systems or in underdeveloped areas. As illustrated in Figure 1, the CT scans of COVID-19 and

non-COVID patients exhibit no readily apparent differences discernible without expert evaluation.

To address this challenge, artificial intelligence (AI) methods, particularly deep learning, have been developed to automate the screening of COVID-19 from CT images. [1] The development and validation of these AI models, however, were severely hampered by the scarcity of large, publicly available CT datasets due to patient privacy concerns. [3] In response, several pioneering datasets were created, including the **COVID-19 CT Lung and Infection Segmentation Dataset** [18], which was instrumental in advancing early AI-based diagnostic research. [1] The initial models developed using these datasets, while promising, achieved performance levels that have since been surpassed by more recent state-of-the-art (SOTA) approaches. [4] This disparity in performance underscores a critical research opportunity: to re-evaluate these foundational models and enhance their capabilities using contemporary deep learning techniques to achieve clinically robust, SOTA performance.

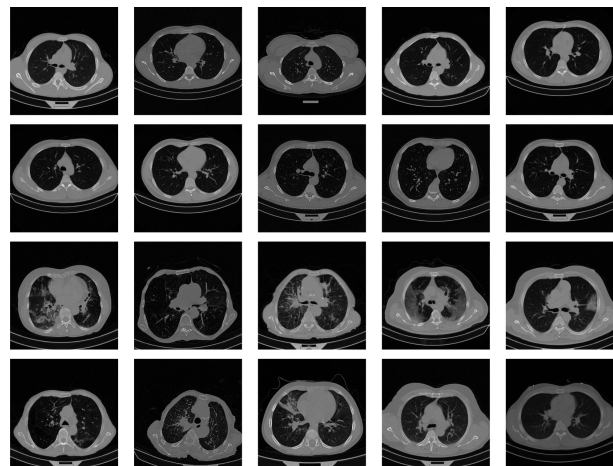


Fig. 1. Examples of CT scans, first two rows contain images from healthy subjects, whereas the last two rows contain images from COVID-19 patients. [4]

II. RELATED WORK

Research in automated diagnosis from medical imaging, particularly CT scans, has gained significant attention with the adoption of artificial intelligence and deep learning techniques. Grewal et al. [6] integrated DenseNet architecture with a recurrent neural network layer to analyze 77 brain CT scans, achieving a CT-level hemorrhage prediction accuracy of 81.82% through their RADnet model. Similarly, Song et al. [7] developed three deep learning models, convolutional neural networks (CNNs), deep neural networks (DNNs), and stacked autoencoders (SAEs), for lung cancer classification, where the CNN outperformed the others in terms of accuracy. In another study, Gonzalez et al. [8] applied CNN-based analysis to identify and stage chronic obstructive pulmonary disease (COPD), while also predicting acute respiratory disease (ARD) occurrences and mortality risks in smokers. Early CT diagnostic models primarily employed conventional CNN backbones. For classification, transfer learning with ResNet- and DenseNet-type networks proved effective. [9] These networks have been successfully leveraged to classify, segment, and detect abnormalities in various medical images, including CT scans, by learning and extracting hierarchical features from large datasets. However, the direct application of models pre-trained on natural image datasets to the medical domain proved challenging due to the inherent differences in image characteristics, such as lower contrast, noise, and artifacts, which provide limited tissue descriptions. This limitation necessitated the development of domain-specific enhancements.

During the COVID-19 outbreak, the critical need for rapid and accurate COVID-19 screening spurred a wave of research focused on automating the diagnosis from chest CT images. Key indicators observable in CT scans for COVID-19 detection include ground-glass opacities, consolidation, reticular patterns, and the crazy paving patterns. [10] Also there is a study to examine the relationship between these chest CT findings and the clinical manifestations of COVID-19 pneumonia. [11] When it comes to detect COVID-19 from CT scans, initial works served as a proof-of-concept, experimenting with well-known pre-trained CNN architectures including ResNet50, DenseNet169, EfficientNetB1 and VGG16. While these models showed the viability of deep learning for this task, their performance, while promising, often fell short of the clinical robustness required for widespread deployment. This early stage highlighted a critical research opportunity to re-evaluate and enhance these foundational models using contemporary techniques to achieve superior, state-of-the-art performance. [12]

In recent work, a Fine_DenseNet-based deep learning model combined with an Improved Generative Adversarial Network (IGAN_AHb) and optimized using the Artificial Hummingbird algorithm was proposed for automated multi-class COVID-19 detection from chest CT images, achieving a high classification accuracy of 95.73%. [13] Also, recent research has leveraged pre-trained deep neural networks combined with CycleGAN-based data augmentation for automated COVID-

19 detection from CT images with high accuracy, using a dataset of 3,163 images from 189 patients, while also providing interpretability through Grad-CAM visualization and calibration-based reliability assessment. [14] Another recent study proposed a hybrid model, ViTGNN, which combines Convolutional Neural Networks, Graph Neural Networks, and Vision Transformers for COVID-19 detection from CT scans, achieving high diagnostic performance with an accuracy of 95.98%, precision of 96.07%, recall of 96.01%, F1-score of 95.98%, and AUC of 98.69%. [15] Also, a three-layer stacked multimodal framework integrating eight pre-trained transfer learning models has been proposed for deep feature extraction from large COVID-19 chest radiographic datasets, achieving high performance with an accuracy of 95.79%, precision of 95.44%, and recall of 96.65%, providing an effective and computationally efficient approach for COVID-19 diagnosis. [16] Another dual-channel convolutional neural network (CNN) framework has been proposed for detecting diverse COVID-19 variants from lung CT scans, combining texture-based and spatial feature analysis with dynamic textural pattern learning, achieving high accuracy of 94.63% and 95.47% on the COVID-349 and Italian COVID-Set datasets, respectively, and demonstrating superior precision, recall, and diagnostic reliability compared to existing methods. [17]

III. METHOD

This section details the systematic and rigorous approach used to develop and train the proposed model. It begins with the initial steps of data acquisition and preprocessing, followed by image enhancement techniques and the extraction of the lung regions of interest. A robust data augmentation pipeline is then described, which addresses class imbalance by synthesizing new samples for the minority class. The core of the methodology is the introduction of a novel multi-branch ConvNeXt architecture designed to capture both global and fine-grained features. The section concludes with a detailed explanation of the two-phase training strategy, which effectively leverages transfer learning to adapt the model to the specific domain of medical imaging.

A. Data Acquisition and Preprocessing

For this study, two contemporary and higher-quality datasets from [5] and [18], were utilized to build a robust training and validation set.

- **COVID-19 CT Lung and Infection Segmentation**

Dataset: This dataset contains 20 labeled COVID-19 CT scans. Left lung, right lung, and infections are labeled by two radiologists and verified by an experienced radiologist. [18]

- **MedSeg Covid Dataset 2:** This contains 9 labeled axial volumetric CTs from Radiopaedia. Both positive and negative slices (373 out of the total of 829 slices have been evaluated by a radiologist as positive and segmented). [5]

A critical methodological choice in this work was to combine these two datasets, resulting in a total of 2,609 CT slices. This synthesis of data from different sources is a common

strategy in modern deep learning research to increase the sample size and improve the model's ability to generalize to variations in data source, a common challenge in medical imaging where datasets often originate from different scanners or clinical settings. The increase in sample size is a key factor in training a more robust model that is less prone to overfitting and can learn more complex patterns from the data.

The foundation of the proposed methodology is a robust data pipeline that handles the specific characteristics of CT scans. The process begins with loading the volumetric NIfTI files. This pipeline is designed to load each volumetric scan and perform an initial set of transformations. The image array is first rotated to the correct orientation, and then a subset of slices, specifically from 20% to 80% of the total, is selected. This step is crucial for discarding uninformative slices at the top and bottom of the scan, which typically do not contain the primary lung regions of interest. Each selected slice is then resized to a uniform dimension of 512x512 pixels and normalized to a range of 0 to 1.

B. Image Enhancement and Region of Interest Extraction

To prepare the images for classification, two key enhancement and extraction steps are applied. The first is Contrast Limited Adaptive Histogram Equalization (CLAHE). This technique improves the local contrast of the CT images, which is essential for highlighting subtle pathological features like ground-glass opacities. Unlike standard histogram equalization, CLAHE avoids over-amplifying noise in homogeneous regions by operating on small, distinct tile grids and limiting the contrast enhancement.

Figure 2 illustrates the effect of CLAHE on CT scans. The original CT image (*top left*) and its corresponding histogram (*top right*) show that most pixel intensities are clustered in narrow low-value ranges, limiting contrast between anatomical structures. After applying CLAHE (*bottom left*), the enhanced image exhibits improved visibility of lung details, while the corresponding histogram (*bottom right*) demonstrates a more balanced distribution of pixel intensities across the full dynamic range. This redistribution enhances local contrast and highlights subtle structural variations that are less discernible in the original scan.

The second step is a meticulous region-of-interest (ROI) extraction and cropping process. This process is specifically designed to isolate the lung regions from the rest of the CT scan. It operates by first creating a binary mask of the lungs and then using contour finding algorithms to identify the two largest contours, representing the left and right lungs. Bounding boxes are then drawn around each lung. A key design choice is to crop each lung region separately and then resize both to a consistent size of 125x250 pixels before horizontally concatenating them into a single 250x250 pixel image. As shown in figure 3, this approach focuses the model exclusively on the diagnostically relevant pulmonary regions, removing extraneous information such as ribs, the heart, or background noise. By providing the model with a consistent, focused input, the feature learning process becomes more

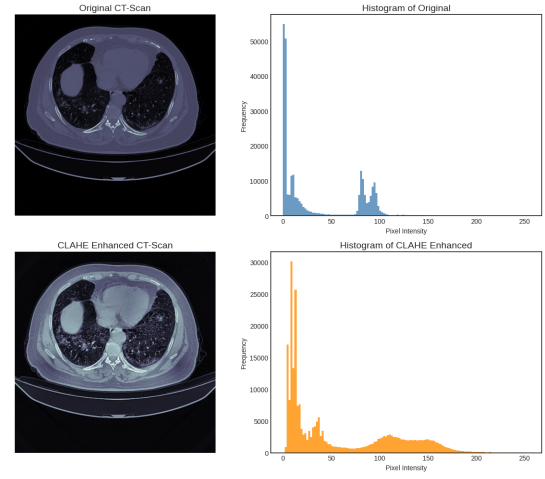


Fig. 2. Comparison of original and CLAHE-enhanced CT scans with corresponding histograms

efficient and effective, which is a strong contributing factor to the final high performance.

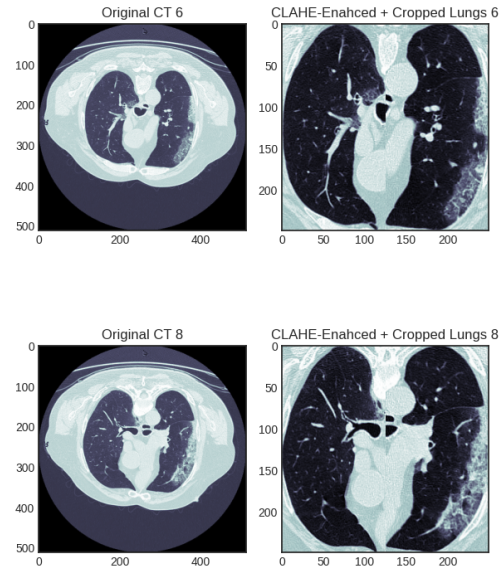


Fig. 3. CT Scans and Cropped Lung Regions of COVID-19 CT segmentation dataset

C. Data Augmentation for Dataset Expansion and Class Balancing

The raw data, even when combined, presents a class imbalance, with 1,724 COVID-19 cases and 885 non-COVID cases. A comprehensive data augmentation strategy was implemented to address this imbalance and to significantly increase the overall size of the training dataset. The augmentation pipeline applies a variety of transformations, including rotation, horizontal and vertical flipping, shifting, gamma correction, and the addition of slight noise. These transformations artificially increase the diversity of the training data, helping the model

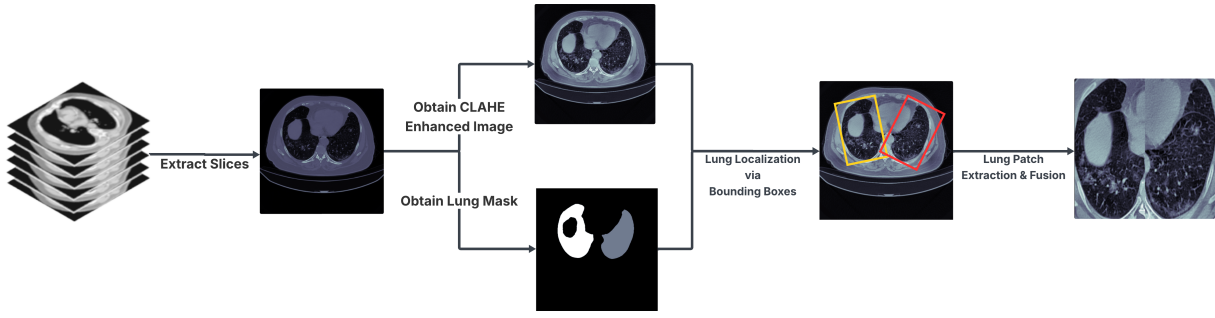


Fig. 4. Pre-processing steps done for the CT scans

learn features that are invariant to minor variations in image orientation or quality. The augmentation successfully balanced the dataset, resulting in a training set with 2,500 samples for each class, for a total of 5,000 images. This balancing obviated the need for a specialized loss function like Focal Loss, which was initially considered in the project's planning stages. The balanced dataset simplifies the training process, allowing for the use of a standard binary cross-entropy loss with equal class weights, which leads to a more robust and reliable model, as evidenced by the high F1-score.

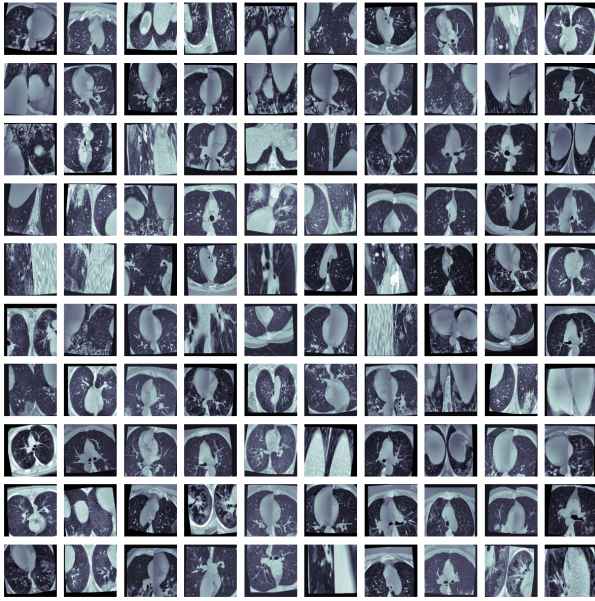


Fig. 5. Generated CT scans from data augmentation

D. The Proposed Multi-branch ConvNeXt Architecture

The proposed model builds on the ConvNeXtSmall architecture (Figure 6), a high-performing convolutional model pre-trained on the ImageNet dataset. To enhance its feature extraction capabilities for the specific task of medical image diagnosis, a novel multi-branch structure was designed and implemented. This architecture processes the feature maps from the base ConvNeXt model through three parallel, distinct pathways before fusing them for final classification.

- 1) **Global Average Pooling Branch:** This branch uses a global average pooling operation to condense the feature maps into a single vector, capturing the overall, holistic features of the image. This pathway is effective for learning general texture and context.
- 2) **Global Max Pooling Branch:** This branch employs a global max pooling operation to identify the most salient or prominent features within the feature maps. This is particularly useful for detecting strong signals, such as large, high-intensity lesions that may indicate severe pathology.
- 3) **Attention-weighted Pooling Branch:** This novel branch is designed to dynamically focus the model on the most diagnostically relevant regions. It learns an attention mask that is then multiplied with the base feature maps, effectively weighting the importance of each feature channel and spatial location. A global average pooling is then applied to this attention-weighted output, compelling the model to learn where to look for critical clinical markers, such as subtle ground-glass opacities.

The outputs from these three branches are concatenated and processed through a feature selection layer, a dense layer with a sigmoid activation, which learns to weigh the importance of the combined feature vector. This enriched feature representation is then passed through a final classification head composed of dense, normalization, and dropout layers, culminating in a single-neuron output with a sigmoid activation for binary classification. This synergistic combination of different pooling strategies allows the model to gain a comprehensive understanding of the input image, capturing global patterns, strong local signals, and the most critical regions, which is a significant factor in its superior discriminative power.

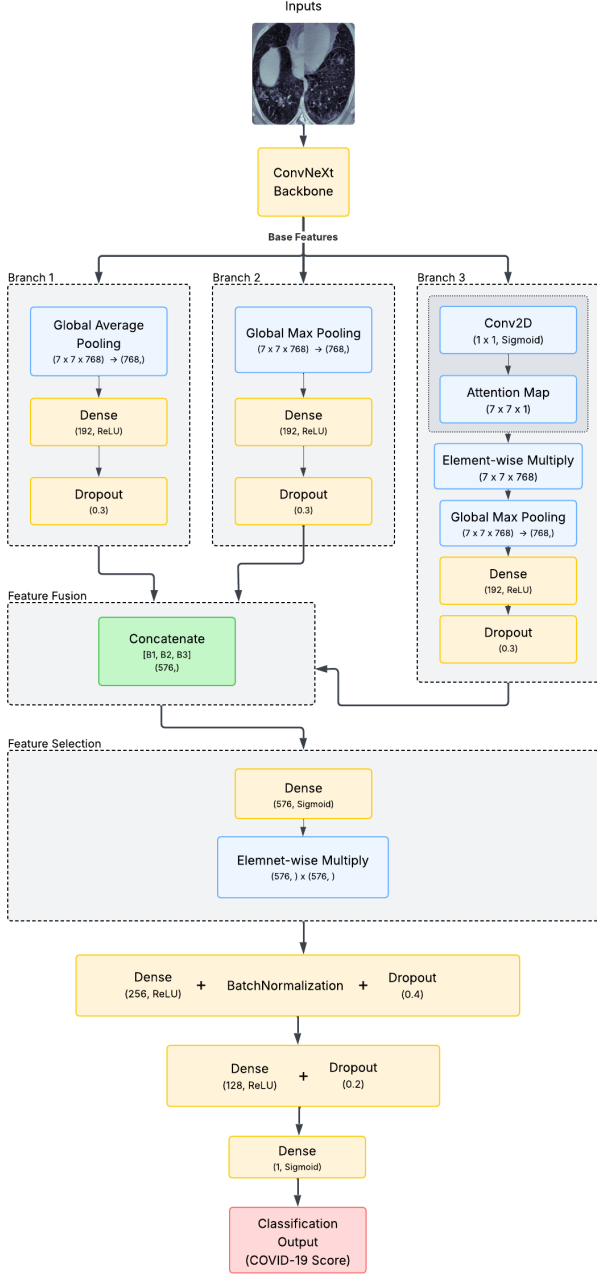


Fig. 6. Multi-branch ConvNeXt Architecture

E. Two-Phase Training Strategy

As shown in Figure 7 two-phase training strategy was employed to leverage the benefits of transfer learning while adapting the model to the specific domain of CT scans.

1) Phase 1: Training with a Frozen Base Model:

The initial phase focused on training only the newly added classification head while keeping the pre-trained ConvNeXt base model frozen. This phase was conducted for 12 epochs with a relatively high learning rate of 1×10^{-3} and an adaptive optimizer. This approach allows the new layers to quickly

learn how to interpret the features extracted by the pre-trained ConvNeXt model, preventing "catastrophic forgetting" of the general visual features learned from ImageNet.

2) **Phase 2: Fine-tuning with Unfrozen Layers:** Following the initial phase, a portion of the base ConvNeXt model was unfrozen to allow for fine-tuning. Specifically, half of the base model's layers were made trainable. The model was recompiled with a much lower learning rate of 1×10^{-6} and trained for an additional 8 epochs. This step is critical for adapting the pre-trained weights to the unique visual patterns of CT scans, enabling the model to learn the subtle, domain-specific features of pulmonary pathologies.

Both phases of training utilized callbacks to save the best model weights based on validation loss and AUC. Other callbacks were used to prevent overfitting and to automatically reduce the learning rate when validation performance plateaued. This structured training methodology ensures the model is both efficient to train and highly specialized for the final task.

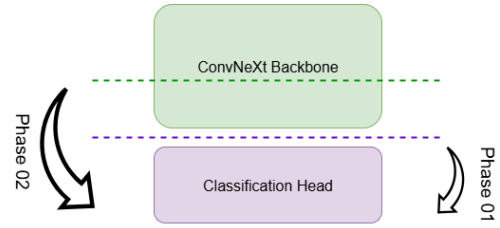


Fig. 7. Two-phase training strategy

IV. RESULTS

A. Experimental Setup

The combined dataset of 2,609 CT slices was split into training and validation sets using a stratified `train_test_split` with a 70/30 ratio, which maintained the original class distributions in both sets. After data augmentation, the final training set comprised 5,000 images, evenly split between COVID-19 and non-COVID-19 cases. The validation set consisted of 701 images after cleaning, with 486 COVID-19 cases and 215 non-COVID-19 cases. The models were trained using a batch size of 32, and an adaptive optimizer was used for both training phases.

B. Performance Evaluation

The final performance of the trained model was evaluated on the unseen validation dataset. The model achieved a high level of performance across multiple key metrics, which is a direct consequence of the robust methodology and sophisticated architecture. The final evaluation yielded a validation loss of 0.0912 and AUC of 0.9937. A detailed breakdown of the performance is provided in the classification report and the performance metrics table I below.

TABLE I
EVALUATION METRICS ON THE TEST SET

Metric	Value
Loss	0.0912
AUC	0.9937
F1-Score	0.9825
Precision	0.9835
Recall	0.9815
Accuracy	0.9757

The confusion matrix, figure 8 further illustrates the model’s high accuracy, with only 8 false positives and 9 false negatives out of 701 validation images. Reported precision, recall, and F1-score are calculated from it, and this provides a transparent view of class-wise performance and ensures reproducibility of the reported metrics.

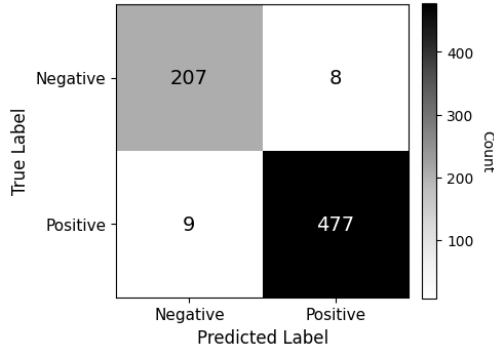


Fig. 8. Confusion matrix of the classification results

Given the inherent class imbalance in the original dataset, with a greater number of COVID-19 cases, metrics beyond simple accuracy are crucial for a comprehensive evaluation. Accuracy can be misleading because a model can achieve a high score by simply predicting the majority class, which would be a poor outcome for a clinical diagnostic tool. The AUC and F1-score provide a more reliable assessment. AUC measures the model’s ability to distinguish between the two classes across all possible classification thresholds, which is a vital indicator of its overall discriminating power. The F1-score is the harmonic mean of precision and recall, providing a balanced measure that is particularly useful for imbalanced datasets, as it penalizes models that favor one class over the other.

Therefore, during training and evaluation, our models were optimized to maximize AUC rather than raw accuracy. This choice ensures that the reported performance better reflects real-world diagnostic needs, where sensitivity (minimizing missed COVID cases) and specificity (avoiding false alarms) are both critical. The source code is publicly available online [24]

Key Performance Metrics Explained

The performance of a classification model is measured using a variety of metrics, each providing a different perspective on

its effectiveness. These metrics are derived from the values in the confusion matrix: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).

- **Accuracy:** The ratio of correctly predicted observations to the total observations.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

- **Precision:** The ratio of correctly predicted positive observations to the total predicted positive observations. It answers the question: “Of all the cases we predicted as positive, how many were actually positive?”

$$Precision = \frac{TP}{TP + FP}$$

- **Recall:** The ratio of correctly predicted positive observations to all observations in the actual class. It answers the question: “Of all the actual positive cases, how many did we correctly identify?” Also known as Sensitivity or True Positive Rate (TPR).

$$Recall = \frac{TP}{TP + FN}$$

- **F1-Score:** The weighted average of Precision and Recall. It is a single metric that balances both concerns and is especially useful in cases with uneven class distribution.

$$F_1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}$$

- **Area Under the Curve (AUC):** A measure of the model’s ability to distinguish between classes. The AUC for the Receiver Operating Characteristic (ROC) curve plots the True Positive Rate (Recall) against the False Positive Rate (FPR), which is the ratio of incorrectly predicted positive observations to all actual negative observations. A higher AUC indicates a better model.

$$FPR = \frac{FP}{FP + TN}$$

C. Comparative Analysis

The results of the proposed model were evaluated against contemporary state-of-the-art benchmarks reported in the literature. Importantly, all comparisons were performed on the same datasets, namely the **COVID-19 CT Lung and Infection Segmentation Dataset** [18] and the **MedSeg Covid Dataset 2** [5], ensuring fairness and consistency in evaluation. As shown in the table II below, the **multi-branch ConvNeXt** model achieves competitive performance, demonstrating its effectiveness in comparison to the best-performing models in the field.

D. Qualitative Analysis and Visual Insights

Beyond the quantitative metrics, a qualitative analysis was performed to gain a deeper understanding of the model’s behavior. The AUC evolution plot (Figure 9) shows a steep increase during the initial training phase, indicating that the classification head rapidly adapted to the pre-trained ConvNeXt features. After fine-tuning began at epoch 12, both

TABLE II
PERFORMANCE COMPARISON OF DIFFERENT METHODS

Methods	Acc	Sen	Spe	AUC
CNN 8-layers [23]	0.7467	0.8	0.70	0.78
InceptionV3 [23]	0.8267	0.88	0.78	0.82
Efficient-Net [23]	0.9067	0.91	0.85	0.93
ResNet+SE [21]	0.8707	0.9322	0.8030	0.9557
ResNet [19]	0.8890	0.9253	0.8439	0.9649
ResNet+CBAM [20]	0.9162	0.8808	0.9552	0.9784
MTL [23]	0.9467	0.96	0.92	0.97
MA-Net [22]	0.9588	0.9512	0.9672	0.9885
Ours (MB-ConvNeXt)	0.9757	0.9815	0.9835	0.9937

training and validation curves stabilized, with validation AUC peaking at 0.9937. The close alignment between the two curves suggests strong generalization and minimal overfitting.

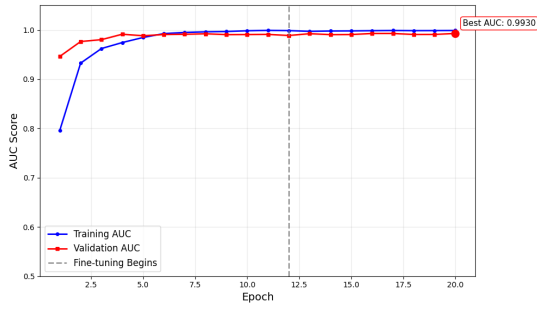


Fig. 9. AUC evolution plot

The final ROC curve plot (Figure 10 also confirms the model's excellent performance, with a curve that closely follows the top-left corner of the graph, indicating a strong ability to distinguish between the two classes.

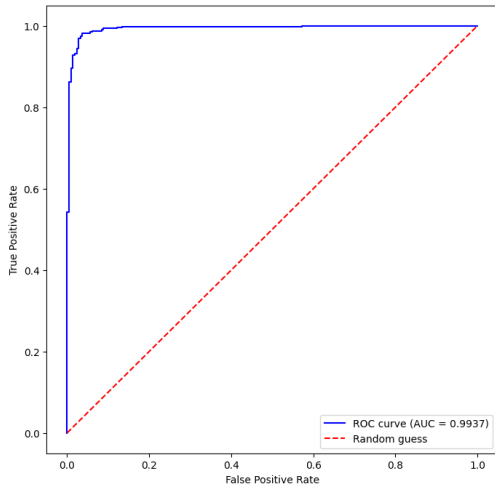


Fig. 10. ROC curve

An analysis of the predicted probability distributions using a violin plot (Figure 11 reveals that the model is highly confident in its predictions for both classes, with the distributions for COVID and non-COVID cases showing clear separation.

This visualization confirms the model's robust discriminative capabilities.

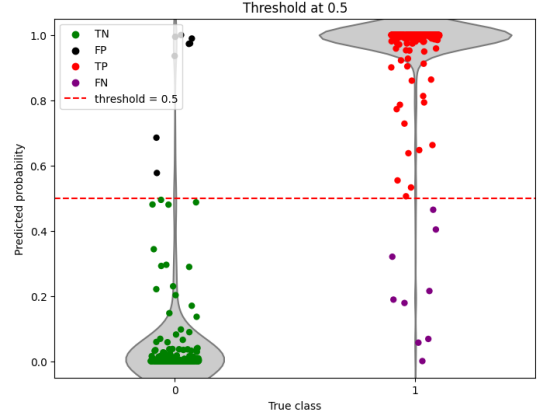


Fig. 11. Predicted probability distributions

V. DISCUSSION

The results demonstrate that the proposed multi-branch ConvNeXt model, in combination with a detailed preprocessing and training pipeline, is a powerful tool for COVID-19 diagnosis from CT scans. The high precision (0.9835), recall (0.9815), and F1-score (0.9825) for the COVID-19 class are particularly significant for a diagnostic application, as they indicate that the model is both highly effective at identifying true positive cases and robust in avoiding false positives. The high ROC-AUC of 0.9937 confirms the model's excellent ability to discriminate between the two classes across various classification thresholds.

The success of this approach is attributed to several key methodological choices. First, the combination of data from two distinct sources allowed for a larger training set, which is fundamental for learning robust features. Second, the meticulous preprocessing steps, CLAHE enhancement and the two-lung cropping and concatenation, acted as a form of a manual attention mechanism, directing the model's focus to the diagnostically relevant regions and thereby simplifying the learning task. Third, the data augmentation pipeline effectively solved the class imbalance problem, which allowed for a more straightforward training process with binary cross-entropy loss and equal class weights. This avoided the complexities of a specialized loss function and ensured the model learned a balanced representation of both classes.

Finally, the novel multi-branch architecture itself is a core factor in the model's performance. By combining features from global average, global max, and a learned attention-weighted pooling branch, the model is able to synthesize a more comprehensive understanding of the input image. This design allows it to capture a wider array of visual cues, from general texture to the most salient features and the most critical regions, which are all important for an accurate diagnosis. The two-phase training strategy, which first allowed the classification head to learn and then fine-tuned the base

model, was also crucial for adapting the pre-trained weights to the unique visual patterns of CT scans without over-fitting.

Despite the high performance, certain limitations should be noted. The dataset, while sufficient for a robust proof-of-concept, is still relatively small compared to what might be available in a large-scale clinical setting. The model's generalizability could be further improved by training on a larger, more diverse dataset from multiple hospitals and with varied scanner parameters. Future work could also involve exploring alternative architectures, such as a full Vision Transformer, or a hybrid model that combines convolutional and transformer layers, to see if they can achieve even higher performance.

VI. CONCLUSION

This research successfully demonstrates a multi-faceted approach to automated COVID-19 diagnosis from CT scans. The proposed methodology, which includes a novel multi-branch ConvNeXt architecture and a rigorous end-to-end training pipeline, significantly improves upon the performance of early deep learning models and achieves results that are competitive with contemporary state-of-the-art benchmarks.

The project's success is a testament to the power of modern deep learning techniques when applied systematically and with careful consideration of domain-specific challenges. The use of data augmentation to balance the training set, the meticulous preprocessing to focus the model on the region of interest, and the multi-branch architecture to capture a comprehensive feature representation were all key to achieving the final high performance. The culminating model, with an ROC-AUC of 0.9937 and an F1-score of 0.9825, represents a significant step forward in developing robust and clinically useful AI tools for medical diagnostics. The principles of this multi-faceted approach, including advanced preprocessing, architectural enhancements, and a structured training regimen, are broadly applicable to a wide range of medical image classification challenges beyond COVID-19. The findings from this research provide a valuable contribution to the field and offer a clear path for future work aimed at translating these models into practical clinical applications.

REFERENCES

- [1] X. Yang, X. He, J. Zhao, Y. Zhang, S. Zhang, and P. Xie, "COVID-CT-Dataset: A CT Scan Dataset about COVID-19," June 17, 2020, arXiv: arXiv:2003.13865. doi: 10.48550/arXiv.2003.13865.
- [2] Deep learning approach for classifying CT images of COVID-19: A Systematic Review, accessed on August 22, 2025, https://www.researchgate.net/publication/364545817_Deep_learning_approach_for_classifying_CT_images_of_COVID-19_A_Systematic_Review
- [3] COVID-CT-MD, COVID-19 computed tomography scan dataset applicable in machine learning and deep learning, accessed on August 22, 2025, <https://pmc.ncbi.nlm.nih.gov/articles/PMC8085195/>
- [4] Automatic diagnosis of COVID-19 from CT images using CycleGAN ..., accessed on August 22, 2025, <https://pmc.ncbi.nlm.nih.gov/articles/PMC10263244/> COVID-19 - Medical segmentation, accessed on August 22, 2025,
- [5] "MedSeg Covid Dataset 2." figshare, Jan. 05, 2021. doi: 10.6084/m9.figshare.13521509.v2.
- [6] M. Grewal, M. M. Srivastava, P. Kumar, and S. Varadarajan, "RAD-net: Radiologist level accuracy using deep learning for hemorrhage detection in CT scans," in 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), Apr. 2018, pp. 281–284. doi: 10.1109/ISBI.2018.8363574.
- [7] Q. Song, L. Zhao, X. Luo, and X. Dou, "Using Deep Learning for Classification of Lung Nodules on Computed Tomography Images," *J Healthc Eng*, vol. 2017, p. 8314740, 2017, doi: 10.1155/2017/8314740.
- [8] G. González et al., "Disease Staging and Prognosis in Smokers Using Deep Learning in Chest Computed Tomography," *Am J Respir Crit Care Med*, vol. 197, no. 2, pp. 193–203, Jan. 2018, doi: 10.1164/rccm.201705-0860OC.
- [9] Y.-M. Chen, Y. J. Chen, W.-H. Ho, and J.-T. Tsai, "Classifying chest CT images as COVID-19 positive/negative using a convolutional neural network ensemble model and uniform experimental design method," *BMC Bioinformatics*, vol. 22, no. 5, p. 147, Nov. 2021, doi: 10.1186/s12859-021-04083-x.
- [10] Z. Ye, Y. Zhang, Y. Wang, Z. Huang, and B. Song, "Chest CT manifestations of new coronavirus disease 2019 (COVID-19): a pictorial review," *Eur Radiol*, vol. 30, no. 8, pp. 4381–4389, Aug. 2020, doi: 10.1007/s00330-020-06801-0.
- [11] "Relation Between Chest CT Findings and Clinical Conditions of Coronavirus Disease (COVID-19) Pneumonia: A Multicenter Study - PubMed." Accessed: Sept. 06, 2025. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/32125873/>
- [12] X. (Freddie) Liu, G. Karagoz, and N. Meratnia, "Analyzing the Impact of Data Augmentation on the Explainability of Deep Learning-Based Medical Image Classification," *Machine Learning and Knowledge Extraction*, vol. 7, no. 1, p. 1, Mar. 2025, doi: 10.3390/make7010001.
- [13] H. P. Sahu and R. Kashyap, "Fine_Denseganet: Automatic Medical Image Classification in Chest CT Scan Using Hybrid Deep Learning Framework," *International Journal of Image and Graphics*, July 2023, doi: 10.1142/S0219467825500044.
- [14] N. Ghassemi et al., "Automatic diagnosis of COVID-19 from CT images using CycleGAN and transfer learning," *Applied Soft Computing*, vol. 144, p. 110511, Sept. 2023, doi: 10.1016/j.asoc.2023.110511.
- [15] K. Amuda, A. Wakili, T. Amoo, L. Agbetu, Q. Wang, and J. Feng, "Detecting SARS-CoV-2 in CT Scans Using Vision Transformer and Graph Neural Network," *Algorithms*, vol. 18, no. 7, p. 413, July 2025, doi: 10.3390/a18070413.
- [16] B. Kaushik, A. Chadha, A. Mahajan, and M. Ashok, "A three layer stacked multimodal transfer learning approach for deep feature extraction from Chest Radiographic images for the classification of COVID-19," *Engineering Applications of Artificial Intelligence*, vol. 147, p. 110241, May 2025, doi: 10.1016/j.engappai.2025.110241.
- [17] S. Fekri-Ershad and K. B. Dehkordi, "A Flexible Multi-Channel Deep Network Leveraging Texture and Spatial Features for Diagnosing New COVID-19 Variants in Lung CT Scans," *Tomography*, vol. 11, no. 9, p. 99, Sept. 2025, doi: 10.3390/tomography11090099.
- [18] M. Jun et al., "COVID-19 CT Lung and Infection Segmentation Dataset," Zenodo, Apr. 20, 2020. Accessed: Sept. 04, 2025. [Online]. Available: <https://zenodo.org/records/3757476>
- [19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.
- [20] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional Block Attention Module," July 18, 2018, arXiv: arXiv:1807.06521. doi: 10.48550/arXiv.1807.06521.
- [21] J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation Networks," in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 2018, pp. 7132–7141. doi: 10.1109/CVPR.2018.00745.
- [22] B. Zheng, Y. Zhu, Q. Shi, D. Yang, Y. Shao, and T. Xu, "MA-Net: Mutex attention network for COVID-19 diagnosis on CT images," *Appl Intell*, vol. 52, no. 15, pp. 18115–18130, Dec. 2022, doi: 10.1007/s10489-022-03431-5.
- [23] A. Amyar, R. Modzelewski, H. Li, and S. Ruan, "Multi-task deep learning based CT imaging analysis for COVID-19 pneumonia: Classification and segmentation," *Computers in Biology and Medicine*, vol. 126, p. 104037, Nov. 2020, doi: 10.1016/j.combiomed.2020.104037.
- [24] "In21-S7-CS4681-AML-Research-Projects/projects/210471F-Healthcare-AI-Medical-Imaging at main · aaivu/In21-S7-CS4681-AML-Research-Projects." Accessed: Oct. 18, 2025. [Online]. Available: <https://github.com/aaivu/In21-S7-CS4681-AML-Research-Projects/tree/main/projects/210471F-Healthcare-AI-Medical-Imaging>