
COVID-19 DETECTION FROM PULMONARY CT SCANS USING A NOVEL EFFICIENTNET WITH ATTENTION MECHANISM

Ramy Farag^{1,*}, Parth Upadhyay^{2,*}, Yixiang Gao, Jacket Demby's,
 Katherin Garces Montoya, Seyed Mohamad Ali Tousi, Gbenga Omotara, G. N. DeSouza³
 Vision-Guided and Intelligent Robotics (ViGIR) Laboratory
 Department of Electrical Engineering and Computer Science (EECS)
 University of Missouri- Columbia
 Columbia, MO, 65211
 {rmf3mc¹, desouzag³}@missouri.edu; pcuv48@mail.missouri.edu²

ABSTRACT

Manual analysis and diagnosis of COVID-19 through the examination of Computed Tomography (CT) images of the lungs can be time-consuming and result in errors, especially given high volume of patients and numerous images per patient. So, we address the need for automation of this task by developing a new deep learning model-based pipeline. Our motivation was sparked by the CVPR Workshop on "Domain Adaptation, Explainability and Fairness in AI for Medical Image Analysis", more specifically, the "COVID-19 Diagnosis Competition (DEF-AI-MIA COV19D)" under the same Workshop. This challenge provides an opportunity to assess our proposed pipeline for COVID-19 detection from CT scan images. The same pipeline incorporates the original EfficientNet, but with an added Attention Mechanism: EfficientNet-AM. Also, unlike the traditional/past pipelines, which relied on a pre-processing step, our pipeline takes the raw selected input images without any such step, except for an image-selection step to simply reduce the number of CT images required for training and/or testing. Moreover, our pipeline is computationally efficient, as, for example, it does not incorporate a decoder for segmenting the lungs. It also does not combine different backbones nor combine RNN with a backbone, as other pipelines in the past did. Nevertheless, our pipeline still outperforms all approaches presented by other teams in last year's instance of the same challenge, at least based on the validation subset of the competition dataset.

Keywords EfficientNet · Attention Mechanism · Transformer · COVID-19 Detection

1 Introduction

The recent COVID pandemic has affected people differently in terms of severity. In order to move from suspected to confirmed COVID-19 cases, initial diagnosis and management often relied on pulmonary images from a Computed Tomography (CT) scan, as it remains the most effective diagnostic tool. However, the inefficacy of manual analysis becomes obvious when thousands of individual CT scans need to be processed in short periods of time, which happens especially during a pandemic, with physicians having to assess numerous patients daily. In that case, the need to automate the diagnostic process becomes as obvious as the potential for mistakes to happen. Thankfully, the emergence of deep learning models in recent years has enabled their use as supportive tools in clinical diagnosis and examination.

Numerous studies in recent years have applied deep learning to CT scan image analysis. Zhang et al.(2022) utilized the VGG19 architecture. They incorporated the GlobalMax-Pool 2D Layer, achieving good performance in various metrics like sensitivity, specificity, and accuracy relative to traditional CNN models and vision transformer (ViT) models [1]. Zhang et al.(2021) introduced a transformer-based framework for automated COVID-19 diagnosis, comprising two primary stages: initial lung segmentation using UNet, followed by classification [2], which performed better with respect to other state-of-the-art methods. Li et al.(2020) introduced a 2D CNN for extracting features from individual slices in a

*These authors contributed equally to this work.

CT scan, with subsequent fusion of slice-level features through a max-pooling layer and achieved improved performance in classifying COVID-19 from CT scan images and achieved 0.95 area under the receiver operating characteristic curve (AUC) [3]. In the study by Shi et al.(2021), an attention mechanism encompassing both channel-wise attention (CA) and depth-wise attention (DA) was incorporated into a modified 3D Resnet18, achieving 0.99 AUC. These studies underscore the effectiveness of deep learning models in COVID-19 CT image classification tasks [4].

Kolliaz et al. introduced the COVID-19-CT-DB dataset [5, 6, 7, 8, 9, 10, 11, 12, 13], providing a substantial collection of labelled COVID-19 and non-COVID-19 data to tackle the demand for extensive training data in deep learning models. However, the challenge to the architecture of deep learning models arises from the diverse resolutions and numbers of slices in CT images, which depend on the imaging machine.

A few approaches have been suggested to tackle this issue. Chen et al.(2021) proposed two methods to assess CT scan image slice importance [14]. First, the 2D method, Adaptive Distribution Learning with Statistical Hypothesis Testing (ADLeaST), integrates statistical analysis with deep learning for COVID-19 CT scan image classification, ensuring stable predictions by mapping images to a specific distribution. However, 2D detection is influenced by positive slices without clear symptoms during training. The 3D method incorporates self-attention structures (Within-Slice-Transformer, Between-Slice-Transformer) into the 3D CNN architecture. Still, it faces the challenges of insufficient training samples and overfitting due to its large model architecture. Hsu et al.(2023) improved this method by incorporating multi-model ensemble method [15], and we adopted their approach in terms of the image selection step and the use of the EfficientNet in our proposed method.

Our work consists of a pipeline involving a simple step for CT-slice selection, followed by feature extraction, and classification using a novel EfficientNet with Attention Mechanism (EfficientNet-AM). Our approach gets away with the need for segmentation of the lungs by enhancing and highlighting the regions of interest through the attention mechanism, which allows the classification task to focus on a feature map extracted from the most representative regions of the selected CT slices.

2 Methods

This section provides an overview of the processing steps: image selection, feature extraction through EfficientNet, and the added Attention Mechanism. Subsequently, the entire proposed pipeline for COVID-19 detection from pulmonary CT scan is discussed. We used the provided challenge1 and challenge2 datasets to train, but we kept the challenge1 validation set for testing. Also, we followed [15] for partitioning the same datasets.

2.1 Image Selection Step

The goal of this step is simply to select images with maximum lung area and to reduce the number of images required for training and testing. We followed similar steps as those presented in Hsu et al. (2023) [15], and as depicted in Figure 1. However, in our case, during this step, twelve CT slices are kept for training, and forty are kept for testing/validation based on the area occupied by the lungs in the images. Initially, we kept the top 50% of CT scan slices with the largest lung area. Subsequently, we preserved twelve or forty CT scan slices, spaced equally along the already selected top 50%, for training and testing, respectively. The rationale for keeping equally spaced CT slices is to reduce unnecessary redundancy that is expected in adjacent CT slices.

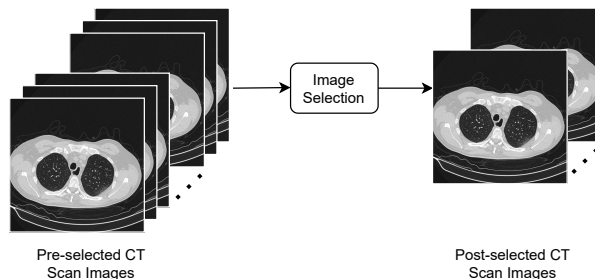


Figure 1: Image selection of pulmonary CT-scan images after detection and ranking by maximum lung area. All images from a given patient are analyzed and their number is reduced to twelve or forty images for training and testing, respectively.

2.2 EfficientNet

EfficientNet represents a breakthrough in neural network architecture design by emphasizing both performance and efficiency [16]. Developed by Google AI researchers, it introduces a scalable model that achieves state-of-the-art accuracy on various computer vision tasks while maintaining a smaller number of parameters and computational cost compared to other models like ResNet or Inception [17].

EfficientNet achieves this by using a compound scaling method that efficiently balances model depth, width, and resolution, resulting in models that are significantly more efficient in terms of both accuracy and computational resources. This innovative approach has made EfficientNet a popular choice for various applications, from image classification to object detection and segmentation, enabling more practical deployment on resource-constrained devices without sacrificing performance. Considering these advantages, we chose to use the ECA-NFNet-L0 pre-trained model in our approach [18].

2.3 Attention Mechanism

The attention mechanism has revolutionized the field of deep learning, particularly in natural language processing and computer vision [17, 19]. This mechanism enhances deep learning models by allowing them to focus on relevant parts of the input data. Inspired by human cognition, it assigns varying weights to different elements based on their importance for the task. This approach helps in improving accuracy and context understanding [20, 21]. In our approach, we took inspiration from the study done by Wang et al.(2021) [22] to enhance the original EfficientNet and create the EfficientNet-AM.

2.4 Pipeline

Figure 2 depicts the feature extraction step of our classification approach. Following the image-selection phase, an EfficientNet is utilized to extract image features, denoted as $f_{img} \subseteq R^{H \times W \times C}$, where H and W represent the image height and width, and C is the number of channels. Next, an attention map $A_m \subseteq R^{H \times W}$ is generated by an attention module, which identifies informative parts in images. Using this attention map A_m and the image features f_{img} , attended features $f_{att} \subseteq R^{H \times W \times C}$ are computed using spatial-wise multiplication.

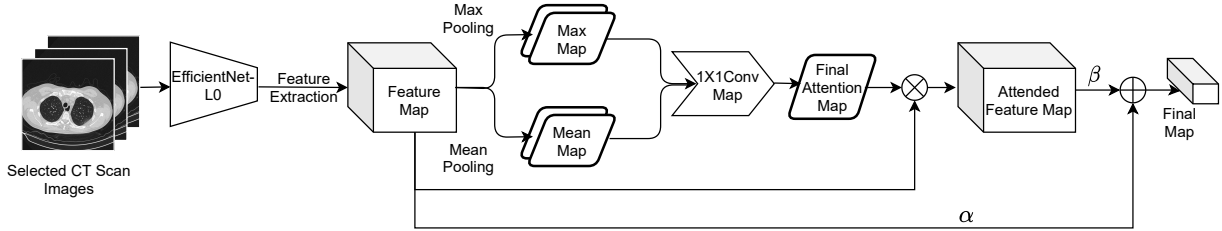


Figure 2: Flowchart of our novel EfficientNet-AM method to extract features from CT scan images of lungs.

$$f_{att}(i) = f_{img}(i) \circ A_m \quad (1)$$

where i is the channel index in f_{img} and f_{att} , while $\circ$ represents spatial-wise multiplication.

The aim is to get the most informative parts of the image to mitigate the potential influence of false positives. Both the attended features f_{att} and the original image features f_{img} are leveraged to obtain merged feature $f_{merged} \subseteq R^{H \times W \times C}$.

$$f_{merged}(i) = \alpha f_{img} + \beta f_{att} \quad (2)$$

The equation is subject to the constraint $\alpha + \beta = 1$, where α and β are hyperparameters responsible for balancing the contribution between the two feature maps. In this study, after empirical analysis, we set $\alpha = 0.5$ and $\beta = 0.5$ to equally support the importance of the original and the attended features in the Final Map.

As mentioned earlier, this combination of attended and original image features helps mitigate false positives. The final features $f_{final} \subseteq R^C$ are used for classification and are obtained by performing global average pooling across each channel.

$$f_{final} = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W f_{merged}(i, j, :) \quad (3)$$

Finally, once our EfficientNet-AM generates the final feature map f_{final} for the selected slice, the same f_{final} features are forwarded to a dense-layer NN with a sigmoid activation function. This last-stage NN is appended at the end of the original pipeline in Figure 2. This final-stage NN is also responsible for calculating the confidence of each selected CT-slice in being classified as *positive* or *negative* for COVID-19.

Figure 3 depicts the complete pipeline of the proposed EfficientNet-AM for COVID-19 detection from CT-scan images. Here, the longer, detailed pipeline in Figure 2 is condensed into the block "EfficientNet-AM", after which the dense NN layer for the final classification of individual samples is appended. Next, one of the three possible voting schemes (the dotted rectangle in the figure) is applied for testing/validation – i.e. for the final diagnoses of the patients and their samples.

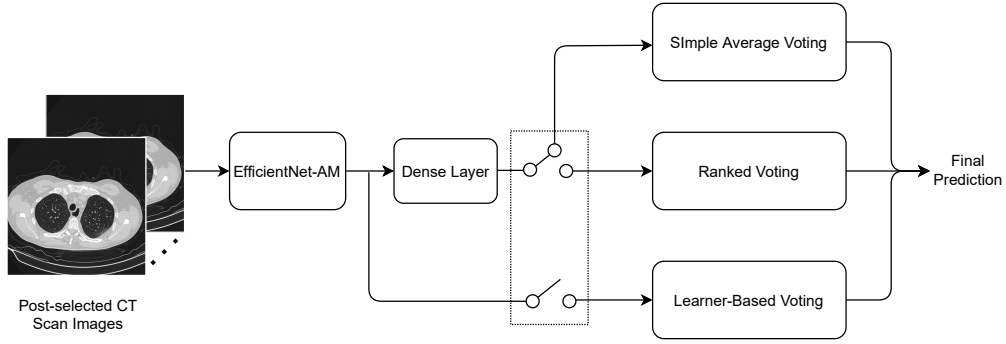


Figure 3: Complete pipeline for the proposed EfficientNet-AM, for COVID-19 detection in pulmonary CT-scan images.

2.5 Voting Schemes

During testing, for the set of n selected CT-slices from a single patient (here, $n = 40$), a final diagnosis for the patient is produced based on one of three possible voting schemes: simple voting, ranked voting, or learner voting. This "Decision Block", i.e. the block that yields the final patient-wise prediction, is shown in Figure 3. The following subsections detail each voting scheme.

2.5.1 Simple Average Voting

In this scheme, the final decision involves averaging all n confidences and setting a threshold t at 0.5 for the COVID diagnoses. Consequently, if the average confidence exceeds this threshold, we categorize the entirety of the patient samples as COVID-19 *positive*; if not, it is considered *negative*.

The average confidence, C_{avg} , of classifying the images as COVID-19 positive is calculated as follows:

$$C_{avg} = \frac{1}{n} \sum_{i=1}^n C_i \quad (4)$$

The classification rule is defined by:

$$\text{Classification} = \begin{cases} \text{COVID-19 Positive} & \text{if } C_{avg} > t \\ \text{COVID-19 Negative} & \text{otherwise} \end{cases} \quad (5)$$

where $t = 0.5$.

2.5.2 Ranked Voting

For ranked voting, we first sort the confidences associated with each CT-slice according to the most likely COVID-positive diagnoses. Then, ranked voting is applied to the top 2.5% of the positive diagnoses (i.e. the most decisively towards COVID) and the bottom 2.5% of the negative diagnoses (i.e. the most decisively towards not COVID).

2.5.3 Learner-Based Voting

For this final scheme, we bypass the dense NN (refer to Figure 3) and connect the EfficientNet-AM directly to a Learner. That is, after obtaining the feature map from the EfficientNet-AM for each selected CT-slice of the same patient, the

same features are input directly into a Single-Head Attention (SHA) transformer. This SHA-Transformer, i.e. the Learner-Based Voting, will then determine whether all patient samples should be classified as COVID-19 positive or negative.

3 Results

In this section, we report the results from all three voting schemes derived from the EfficientNet-AM pipeline. We also compare and evaluate how our proposed model performs against other teams’ models from previous years on the validation dataset. The results of the three voting schemes are presented in Table 1. Here, the first column contains the type of voting, and the second and third columns provide the AUC and Macro-F1 scores obtained using the validation dataset for the respective voting scheme.

Method	AUC	Macro-F1
Simple Average Voting	0.9713	93.67
Ranked Voting	0.9717	93.35
Learner-Based Voting	0.9645	93.67

Table 1: Comparative results from three voting schemes derived from the EfficientNet-AM.

Finally, we compare the performance of models proposed in previous years against the best of our three models (Simple Average Voting). These results are presented in Table 2. The Macro-F1 scores listed in Table 2 are the ones reported by the respective teams’ publications (see citations on the table). As the reader will notice, our model outperformed all other teams’ competing models on the validation dataset.

Method	Macro-F1
Eff-mix-conv-E [15]	0.922
EDPS-COVID-19-CT-LS [23]	0.932
IPSR-4L-CNN-C [24]	0.851
ResNet3D-18 + MHA [25]	0.9021
Ours	0.9367

Table 2: Comparison between our EfficientNet-AM method, using simple average voting, and previous year’s approaches, using Macro-F1 score on the validation set.

4 Conclusion

In this study, we presented a pipeline with three different voting schemes after the feature extraction and classification using a novel EfficientNet with Attention Mechanism (EfficientNet-AM). Our framework aimed at enhancing the detection of COVID-19 from pulmonary CT images. The voting mechanism included: a simple average voting, a ranked voting – both of which are based on the confidence levels from the EfficientNet-AM classification module over a number (40) of CT scan slices – and a Learner-based voting, which learns directly from the feature map from the same CT slices. Based on the Macro-F1 scores reported by last year’s teams, we have demonstrated that our model is a likely contender for the best model for this year’s challenge. In the future, we intend to study further the balance between attended (from the attention mechanism) versus the original features (from EfficientNet) and how to provide more attention to more relevant areas of the lungs. Also, further investigation on why the Learner-based voting could not outperform the simple voting scheme is warranted. Finally, the use of EfficientNet-AM for other applications, such as leaf venation, have already started and will be expanded in order to evaluate the usefulness of this network beyond what has been demonstrated so far.

Acknowledgements

Computational resources for this research have been supported by the NSF National Research Platform, as part of GP-ENGINE (award OAC #2322218).

References

- [1] Sai Zhang and Guo-Chang Yuan. Deep transfer learning for covid-19 detection and lesion recognition using chest ct images. *Computational and mathematical methods in medicine*, 2022, 2022.
- [2] Lei Zhang and Yan Wen. A transformer-based framework for automatic covid19 diagnosis in chest cts. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 513–518, 2021.
- [3] Lin Li, Lixin Qin, Zeguo Xu, Youbing Yin, Xin Wang, Bin Kong, Junjie Bai, Yi Lu, Zhenghan Fang, Qi Song, et al. Using artificial intelligence to detect covid-19 and community-acquired pneumonia based on pulmonary ct: evaluation of the diagnostic accuracy. *Radiology*, 296(2):E65–E71, 2020.
- [4] Jun Shi, Huite Yi, Xiaoyu Hao, Hong An, and Wei Wei. Dual-attention residual network for automatic diagnosis of covid-19. <https://europepmc.org/article/ppr/ppr343319>, 2021.
- [5] Dimitrios Kollias, Anastasios Arsenos, and Stefanos Kollias. Ai-enabled analysis of 3-d ct scans for diagnosis of covid-19 & its severity. In *2023 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW)*, page 1–5. IEEE, 2023.
- [6] Anastasios Arsenos, Andjoli Davidhi, Dimitrios Kollias, Panos Prassopoulos, and Stefanos Kollias. Data-driven covid-19 detection through medical imaging. In *2023 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW)*, page 1–5. IEEE, 2023.
- [7] Dimitrios Kollias, Anastasios Arsenos, and Stefanos Kollias. A deep neural architecture for harmonizing 3-d input data analysis and decision making in medical imaging. *Neurocomputing*, 542:126244, 2023.
- [8] Dimitrios Kollias, Anastasios Arsenos, and Stefanos Kollias. Ai-mia: Covid-19 detection and severity analysis through medical imaging. In *European Conference on Computer Vision*, page 677–690. Springer, 2022.
- [9] Anastasios Arsenos, Dimitrios Kollias, and Stefanos Kollias. A large imaging database and novel deep neural architecture for covid-19 diagnosis. In *2022 IEEE 14th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP)*, page 1–5. IEEE, 2022.
- [10] Dimitrios Kollias, Anastasios Arsenos, Levon Soukissian, and Stefanos Kollias. Mia-cov19d: Covid-19 detection through 3-d chest ct image analysis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, page 537–544, 2021.
- [11] Dimitrios Kollias, N Bouas, Y Vlaxos, V Brillakis, M Seferis, Ilianna Kollia, Levon Sukissian, James Wingate, and S Kollias. Deep transparent prediction through latent representation analysis. *arXiv preprint arXiv:2009.07044*, 2020.
- [12] Dimitris Kollias, Y Vlaxos, M Seferis, Ilianna Kollia, Levon Sukissian, James Wingate, and Stefanos D Kollias. Transparent adaptation in deep medical image diagnosis. In *TAILOR*, page 251–267, 2020.
- [13] Dimitrios Kollias, Anastasios Arsenos, and Stefanos Kollias. Domain adaptation, explainability & fairness in ai for medical image analysis: Diagnosis of covid-19 based on 3-d chest ct-scans. *arXiv preprint arXiv:2403.02192*, 2024.
- [14] Guan-Lin Chen, Chih-Chung Hsu, and Mei-Hsuan Wu. Adaptive distribution learning with statistical hypothesis testing for covid-19 ct scan classification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 471–479, 2021.
- [15] Chih-Chung Hsu, Chih-Yu Jian, Chia-Ming Lee, Chi-Han Tsai, and Sheng-Chieh Dai. Strong baseline and bag of tricks for covid-19 detection of ct scans. *arXiv preprint arXiv:2303.08490*, 2023.
- [16] Andy Brock, Soham De, Samuel L Smith, and Karen Simonyan. High-performance large-scale image recognition without normalization. In *International Conference on Machine Learning*, pages 1059–1071. PMLR, 2021.
- [17] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pages 6105–6114. PMLR, 2019.
- [18] Ross Wightman. Pytorch image models. <https://github.com/rwightman/pytorch-image-models>, 2019.
- [19] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.
- [20] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.
- [21] Ramy Farag, Jacket Demby’s, Muhammad Arifuzzaman, and G. N. DeSouza. Xmnet: Xgboost with multitasking network for classification and segmentation of ultra-fine grained datasets. In *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, Yokohama, Japan, 2024.

- [22] Jun Wang, Xiaohan Yu, and Yongsheng Gao. Mask guided attention for fine-grained patchy image classification. In *2021 IEEE International Conference on Image Processing (ICIP)*, pages 1044–1048. IEEE, 2021.
- [23] Robert Turnbull. Enhanced detection of the presence and severity of covid-19 from ct scans using lung segmentation. *arXiv preprint arXiv:2303.09440*, 2023.
- [24] Kenan Morani. Covid-19 detection using segmentation, region extraction and classification pipeline. *arXiv preprint arXiv:2210.02992*, 2022.
- [25] Alessia Rondinella, Francesco Guarnersa, Oliver Giudice, Alessandro Ortis, Francesco Rundo, and Sebastiano Battiato. Unict dmi solution for 3rd cov19d competition on covid-19 detection trough attention deep learning for ct scan. *arXiv preprint arXiv:2303.08728*, 2023.