EfficientNet-SAM: A Novel EffecientNet with Spatial Attention Mechanism for COVID-19 Detection in Pulmonary CT Scans

Ramy Farag[†] Parth Upadhay[†] Jacket Demby's Yixiang Gao Katherin Garces Montoya
Seyed Mohamad Ali Tousi Gbenga Omotara Guilherme DeSouza
Vision-Guided and Intelligent Robotics Lab - ViGIR Lab
Department of Electrical Engineering and Computer Science, University of Missouri-Columbia
{rmf3mc, pcuv48, udembys, yg5d6, kgxh2, stmmc, goowfd, desouzag}@missouri.edu,
[†] Corresponding authors. Equal Contribution.

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-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 one of the architectures in the EfficientNet "family", but with an added Spatial Attention Mechanism: EfficientNet-SAM. Also, unlike the traditional/past pipelines, which relied on a preprocessing 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 models nor combine RNN with a backbone, as other pipelines in the past did. Nevertheless, our pipeline outperformed all approaches presented by other teams in last year's instance of the same challenge using the validation subset. It also placed ***5th*** *in this year's competition, ranking less than 1.3% below the 1st place and close to 3.5% above the 6th place based on the macro-F1 score.*

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 [30]. 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 [29], 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 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) [16]. In the study by Shi et al.(2021), an attention mechanism encompassing both channel-wise attention (CA) and depthwise 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 [21].

Targeting the need for high-quality datasets, Kolliaz et al. introduced the COVID-19-CT-DB dataset [1, 2, 9–15], providing a substantial collection of labeled 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 system.

A few approaches have been suggested to tackle this issue. Chen et al.(2021) proposed two methods to assess CT scan image slice importance [4]. Their first method, a 2D approach known as 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. So, their subsequent 3D method incorporated self-attention structures (Within-Slice-Transformer, Between-Slice-Transformer) into a 3D CNN architecture. Still, the method faced the challenges of insufficient training samples and overfitting due to its large model architecture. Hsu et al.(2023) improved this method by incorporating a multi-model ensemble method [7], which we adopted in our approach in terms of the image selection step and the use of an Efficient CNN.

Our work consists of a pipeline involving a simple step for CT-slice selection, followed by feature extraction, and classification. The method employs a novel Efficient CNN with Spatial Attention Mechanism (EfficientNet-SAM). Our approach gets away with the need for segmentation of the lungs by enhancing and highlighting the regions of interest through a spatial attention mechanism, which allows the classification task to focus on a feature map extracted from the most representative regions of the selected CT slices, enhancing efficiency and accuracy. We have made the code available in the dedicated GitHub repository associated with this paper. [1].

2. Methods

This section provides an overview of the processing steps: image selection, feature extraction through an efficient CNN [3, 22], and the added Spatial Attention Mechanism. The model we have utilized has Efficient Channel Attention (ECA) [27] and employs Weight Standardized (WS) convolutions, incorporating extra scaling factors instead of utilizing traditional normalization layers [18]. In other words, besides the original channel attention, we integrated a spatial attention mechanism to highlight the areas of the lung relevant to the task at hand: COVID-19 detection. 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 [15]. Also, we followed [8] for partitioning the same datasets.

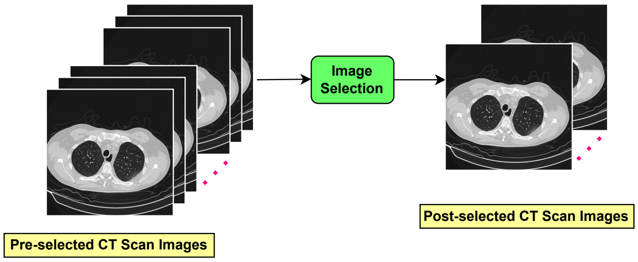

Figure 1. Image selection of pulmonary CT-scan images after detection and ranking by maximum lung area. All images from a given patient are analyzed, and their number is reduced to twelve or forty images for training and testing.

2.1. Image Selection Step

The motivation for this step was threefold: 1) to select CT images that display the maximum area of the lungs; 2) to reduce unnecessary redundancy that is expected in adjacent CT slices; and 3) to reduce the total number of images needed for training and evaluation. Our approach mirrors the technique described by Hsu et al. (2023) [7], and is illustrated in Figure 1. However, we did not perform any preprocessing, and the selected images were employed in their original (raw) format afterwards. In our methodology, we investigated the use of different values for the number N_{tr} of CT slices for the training phase and the number N_{vt} for the validation and testing phases. That is, initially, we select the 50% of CT scan slices that showcase the largest lung area. For this subset, we then varied N_{tr} and N_{vt} while ensuring that they are evenly distributed across the 50%, for training and testing purposes, respectively, and selected the best values obtained in terms of the F1 score. As we will explain later, the strategy that best performed was keeping all 50% of the CT scans for training, while randomly selecting N_{tr} of those for each epoch; and to evenly space N_{vt} CT slices for evaluation.

2.2. EfficientNet-SAM

EfficientNet represented a breakthrough in neural network architecture designed by emphasizing both performance and efficiency [3]. Developed by Google AI researchers, it introduced a scalable model that achieved 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 [22].

EfficientNet achieved 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 re-

[1] https://github.com/rmf3mc/covid-comp24

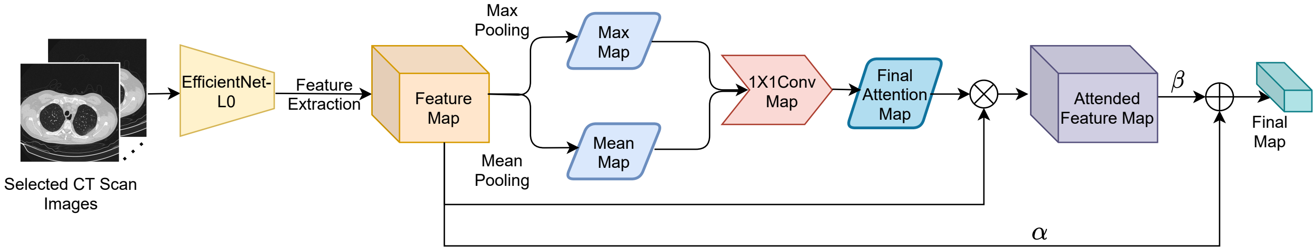

Figure 2. Flowchart of our novel EfficientNet-SAM method to extract features from CT scan images of lungs.

sources. 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 a member of the family of Efficient CNNs called Efficient Channel Attention (ECA) in place of the Squeeze-Excitation approach, while incorporating SiLU activations and diverging from the more commonly used GELU activations [28]. We also added a spatial attention mechanism, as it will be explained in the next subsections.

2.3. Attention Mechanism

Attention mechanisms (AM) in general have revolutionized the field of deep learning, particularly in natural language processing and computer vision [22, 25]. Spatial AM, in particular, enhance deep learning models by allowing them to focus on relevant components of the input data and their spatial relations. 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 [5, 6]. In our approach, we took inspiration from the study done by Wang et al.(2021) [26] to enhance that original model with the addition of spatial attention, and create the EfficientNet-SAM.

2.4. Pipeline

Figure 2 depicts the feature extraction step of our classification approach. Following the image-selection phase, an Efficient CNN 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 a spatial attention module, which identifies the most informative regions 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, as in:

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

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

The objective is to highlight the most informative regions of the image and to mitigate the potential of false positives. To achieve this, the model leverages (eq:2) the attended features f_{att} and the original image features f_{img}, resulting in a unified feature map $f_{merged} \subseteq R^{H\times W\times C}$. This integration ensures that both the specific areas of interest and the overall image features contribute to the final feature representation.

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

Equation 2 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 [26].

As previously mentioned, this combination of attended and original image features plays a crucial role in minimizing false positives. The final features used for classification, $f_{final} \subseteq R^C$, are condensed by performing global average pooling across each channel:

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

Finally, once our EfficientNet-SAM generates the final feature map f_{final} for the selected CT slice, the same feature map is 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 entire pipeline of the proposed EfficientNet-SAM for COVID-19 detection from pulmonary CT-scan images. Here, the longer, detailed pipeline in Figure 2 is condensed into the block ”EfficientNet-SAM”, after which the dense NN layer for the final classification of individual samples is appended. Next, one of

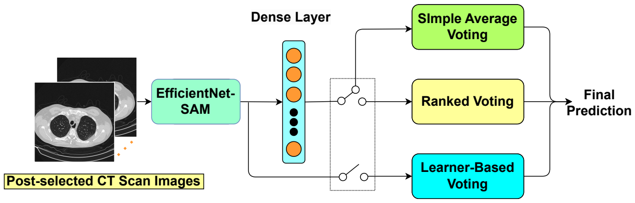

Figure 3. Complete pipeline for the proposed EfficientNet-SAM, for COVID-19 detection in pulmonary CT-scan images.

the three possible voting schemes (outlined by the dotted rectangle in the figure) is applied for the validation and testing phases– i.e. for the final diagnoses of the patients and their samples.

2.5. Voting Schemes

During testing, for the set of N_{vt} selected CT-slices from a single patient, 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_{vt} confidences and setting a threshold t for the COVID diagnoses. Consequently, if the average confidence exceeds this threshold, we categorize the entirety of the patient samples as COVID-19 *positive*; otherwise, it is considered *negative*.

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

$$C_{\text{avg}} = \frac{1}{n} \sum_{i=1}^{N_{vt}} C_i \tag{4}$$

The classification rule is defined by:

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

where $t = 0.5$.

We conducted tests using a variety of choices for N_{vt}: the number of CT-scan slice used for testing; and for different threshold levels as previously described. The final choice will be presented in the Results section.

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 $P\%$ of the positive diagnoses (i.e. the most decisively towards COVID) and the bottom $P\%$ of the negative diagnoses (i.e. the most decisively towards not COVID). We also conducted multiple experiments to choose the value of P, which will also be reported in the next section.

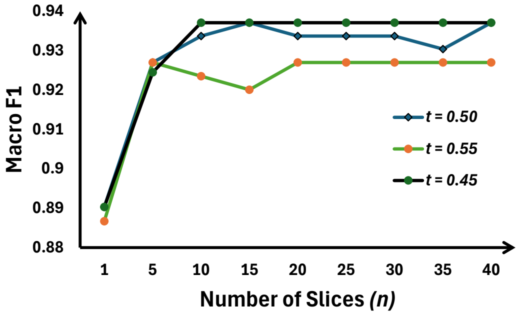

Figure 4. Impact of varying the threshold (t) values and the number N_{vt} of CT-scan slice used for validation and testing on the Macro F1 as a metric of performance

2.5.3 Learner-Based Voting

For this final scheme, we bypass the dense-layer NN (refer to Figure 3) and connect the EfficientNet-SAM directly to a Learner. That is, after obtaining the feature map from the EfficientNet-SAM 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-SAM pipeline. We also compare and evaluate how our proposed model performs against other teams' models from previous years on the validation dataset.

As previously mentioned, in order to decide the best parameters for the pipeline, we first conducted various experiments to evaluate the effects of the threshold levels (t) for deciding COVID, the number of CT slices (N_{vt}) used for validation and testing, and the percentage of top/bottom confidence votes (P) on the performance of the pipeline. As Figure 4 demonstrates, the optimal performance as measured by the Macro F1 score was achieved when choosing a threshold $t = 0.45$ and employing $N_{vt} = 10$ slices. This result is consistent with a second experiment, depicted in Figure 5, in which the same threshold $t = 0.45$ was obtained for different percentages P of top and bottom confidences used for voting – chosen to be $P = 25\%$. A third experiment on N_{tr}, the number of CT slices used for training, was

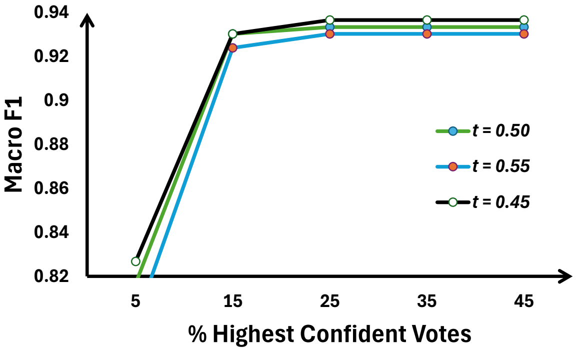

Figure 5. Impact on the model performance due to changes in threshold (t) and the percentage P of most confident votes for COVID and non-COVID cases.

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-SAM.

also performed. In this case, the best result was obtained for the entire 50% of the pre-selected CT-scan slices. However, a value of $N_{tr} = 40$ of randomly selected images was used for each epoch in training.

The results for the three voting schemes using the choices of parameters above 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.

Finally, we compare the performance of models proposed in previous years against the 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. In this year's competition, our method placed 5th – only 1.26% behind the first place, according to the macro F1 score, and 3.49% above the 6th place. It should be pointed out, however, that the team that ranked first relied on the STOIC dataset [24], which includes 2,000 CT scans for training and validation [19]. In other words, our approach was trained exclusively on the provided dataset – i.e. it was trained on less than half of the data used by this other team.

Method	Macro-F1
Eff-mix-conv-E [7]	0.922
EDPS-COVID-19-CT-LS [23]	0.932
IPSR-4L-CNN-C [17]	0.851
ResNet3D-18 + MHA [20]	0.9021
Ours	**0.9367**

Table 2. Comparison between our EfficientNet-SAM 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 Efficient CNN with Spatial Attention Mechanism (EfficientNet-SAM). Our framework aimed at enhancing the detection of COVID-19 from pulmonary CT images. The voting mechanisms included: a simple average voting, a ranked voting – both of which are based on the confidence levels from the EfficientNet-SAM classification module over a number (N_{vt}) 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 and this year's teams, we have demonstrated that our model is one of the best models for the challenge. In future work, we will include analysis and visualization tools to demonstrate the role of our spatial attention mechanism and how it highlights the important parts of the lungs related to the disease. This will allow us 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 into why the Learner-based voting could not outperform the simple voting scheme is warranted. Finally, the use of EfficientNet-SAM for other applications, such as leaf venation, has 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] 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. 2

[2] 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. 2

[3] 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. 2

[4] 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. 2

[5] 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. 3

[6] 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. 3

[7] 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. 2, 5

[8] Jesse. SBBT_COV19D_2023: Source code for COVID-19 Detection using Deep Learning. https://github.com/jesse1029/SBBT_COV19D_2023, 2023. 2

[9] 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. 2

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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. 2

[16] 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. 1

[17] Kenan Morani. Covid-19 detection using segmentation, region extraction and classification pipeline. *arXiv preprint arXiv:2210.02992*, 2022. 5

[18] Siyuan Qiao, Huiyu Wang, Chenxi Liu, Wei Shen, and Alan Yuille. Micro-batch training with batch-channel normalization and weight standardization. *arXiv preprint arXiv:1903.10520*, 2019. 2

[19] Marie-Pierre Revel, Samia Boussouar, Constance de Margerie-Mellon, Inès Saab, Thibaut Lapotre, Dominique Mompoint, Guillaume Chassagnon, Audrey Milon, Mathieu Lederlin, Souhail Bennani, et al. Study of thoracic ct in covid-19: the stoic project. *Radiology*, 301(1):E361–E370, 2021. 5

[20] 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. 5

[21] 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. 1

[22] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pages 6105–6114. PMLR, 2019. 2, 3

[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. 5

[24] Robert Turnbull and Simon Mutch. High-confidence pseudo-labels for domain adaptation in covid-19 detection. *arXiv preprint arXiv:2403.13509*, 2024. 5

[25] 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. 3

[26] 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. 3

[27] Qilong Wang, Banggu Wu, Pengfei Zhu, Peihua Li, Wangmeng Zuo, and Qinghua Hu. Eca-net: Efficient channel attention for deep convolutional neural networks. In *Proceed-*

ings of the IEEE/CVF conference on computer vision and pattern recognition, pages 11534–11542, 2020. 2

[28] Ross Wightman. Pytorch image models. https://github.com/rwightman/pytorch-image-models, 2019. 3

[29] 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. 1

[30] 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. 1