A Generalization Enhancement Approach for Deep Learning Segmentation Models: Application in COVID-19 Lesion Segmentation from Chest CT Slices

1st Nastaran Enshaei
Concordia Inst. for Inf. Sys. Eng.
Concordia University
Montreal, Canada
email address or ORCID

2nd Moezedin Javad Rafiee
Department of Medicine and Diagnostic Radiology
McGill University
Montreal, Canada
email address or ORCID

3rd Farnoosh Naderkhani
Concordia Inst. for Inf. Sys. Eng.
Concordia University
Montreal, Canada
email address or ORCID

***Abstract*—Deep learning has shown remarkable promise in medical imaging tasks, reaching an expert level of performance for some diseases. However, these models often fail to generalize properly to data not used during training, which is a major roadblock to successful clinical deployment. This paper proposes a generalization enhancement approach that can mitigate the gap between source and unseen data in deep learning-based segmentation models without using ground-truth masks of the target domain. Leveraging a subset of unseen domain's CT slices for which the model trained on the source data yields the most confident predictions and their predicted masks, the model learns helpful features of the unseen data over a retraining process. We investigated the effectiveness of the introduced method over three rounds of experiments on three open-access COVID-19 lesion segmentation datasets, and the results illustrate constant improvements of the segmentation model performance on datasets not seen during training.**

***Index Terms*—Deep learning, Generalization, Medical imaging, COVID-19 lesion segmentation.**

I. Introduction

Over recent years, advances in machine learning capabilities have sparked a surge in developing deep learning (DL)-based diagnostic/predictive models. Many areas of medicine have benefited from DL, including medical imaging, where sophisticated algorithms have achieved an expert level of performance. However, despite their high-level reported performance and rigorous evaluations during the development phase, most of the DL models for medical imaging tasks are not reliable to be deployed in real-world clinical settings [1]. One major roadblock to the successful deployment of medical imaging-based DL models in clinical practice is that trained models do not generalize well to unseen domains. Here, unseen domain refers to images from a different clinical center, probably taken with a different scanner and/or imaging setting. The reason is medical images from the same imaging modality may show significant visual differences in terms of quality, appearance, and spatial features when acquired by different scanners/vendors, scanning settings, and patient populations. A straightforward solution to improve DL models' generalizations in the field of medical imaging is using multi-center heterogeneous training sets [2], [3]. Indeed, DL medical imaging models can achieve a high level of generalizability when training datasets consist of an abundance of high-quality images from various health centers with different imaging vendors and acquisition settings. However, hospitals and healthcare institutions are prohibited from disclosing their in-house datasets due to data sharing regulations and patients' privacy protocols. As a result of these limitations, DL models trained on limited-diversity datasets, which work ideally well on in-domain evaluations, suffer from significant performance degradation on images from unseen domains [4]. Consequently, improving the DL model's generalization is crucial in medical imaging applications.

Prior work: One significant barrier in the successful deployment of DL medical imaging analysis models is performance degradation on unseen domains. In general, there are three popular approaches to enhance DL models' generalization trained on a limited source domain. The first is *transfer learning* that addresses the lack of training data by fine-tuning a partially pre-trained network [5]–[7]. Existing DL models have leveraged both natural (ImageNet) and medical images to improve performance. Adopting top-performing CNNs such as DenseNet, Inception-Net, and ResNet pre-trained on ImageNet as the encoding path can enhance the results of segmentation networks [8], [9]. Reference [6] benefits from a two-step transfer learning approach to cope with the lack of training data. First, they utilize ResNet50 pre-trained on ImageNet as the encoding path's basis. The next step involves pre-training the segmentation network using a large medical imaging dataset, compensating for the significant gap between natural and medical images and resulting in further improvements. The second approach is *domain generalization*, known as out-of-distribution generalization, which aims to learn a DL model

This work was partially supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada through the NSERC Discovery Grant RGPIN 2019 06966.

using a single or multiple different but related datasets that can be generalized on unseen test data [10], [11]. Data generalization techniques extract features specific to the source domain but invariant across domains [12]. As another example of data generalization, data augmentation strategies have shown effectiveness in reducing the shift between source and unseen domains. Reference [12] proposed a deep-stacked transformation method that could significantly improve the generalization of DL segmentation models in Magnetic Resonance (MR) and ultrasound images. Another proposed solution for boosting the generalization of DL methods is *domain adaption* that mitigates the gap between the source and a specific unseen domain without using its annotations during the training process. These techniques reduce the shift between the source and target domains in input space, feature space, or output space, mostly using generative adversarial network [13] or its variants [14], and have shown promising results in medical imaging [12], [15]–[17]. However, the need for reliable DL models in medical imaging to be deployed in clinical practice calls for developing novel techniques to achieve enhanced generalization on unseen domains.

Contributions: This study proposes a generalization enhancement approach specific to DL segmentation models that can reduce the shift between the source training set and test sets from unseen domains, addressing a critical challenge in deploying medical imaging DL models in clinical settings. Pseudo-labels have shown promising results in improving the performance of DL classifiers on in-domain and unseen data [18], [19] and DL segmentation models on in-domain data [20]. In this study, leveraging pseudo-labels, we can improve the DL models' performance in medical image segmentation tasks and on unseen data. We perform our experiments on three publicly available CT datasets for COVID-19 lesion segmentation. In the first step, our proposed method involves applying a DL segmentation model trained on a single source domain on test images from an unseen domain. Next, the model confidence on the predicted masks is measured using an introduced certainty index. The CT slices from the unseen domain with the highest certainty index and their predicted masks are used to retrain the previously trained model. The results indicate that the proposed method can enhance the model performance on unseen domains.

Following, we explain the proposed generalization enhancement approach and the implemented deep segmentation network in detail. Then the datasets used in our experiments are introduced. Next, we present the experimental setting, discuss the results, and conclude the paper.

II. Methodology

This study proposes a generalization enhancement approach for DL COVID-19 lesion segmentation models from CT slices, which mitigates the shift between the source and a target unseen test set. The goal is to increase the performance of a predictor (DL segmentation network), $F : X \rightarrow Y$, learned from a source domain, $S : (x_s, y_s)$, on a specific unseen domain, $E : (x_e, y_e)$. In other words, as illustrated in Figure 1, our proposed method automatically extracts and annotates a subset of the test set from an unseen domain utilizing a probabilistic selection index. The selected data samples and their associated predicted masks are then used to retrain and boost the previously trained model. For this purpose, first, we predict the infection masks of an unseen test set using a DL segmentation model trained on the source dataset. Then, we extract a portion of the test images for which the model has predicted the most certain infection masks. Assuming that we do not have the ground-truth masks of the test set, we propose a Certainty Index (CI) to quantify the DL segmentation model's confidence on predicted masks of test images, defined as follows

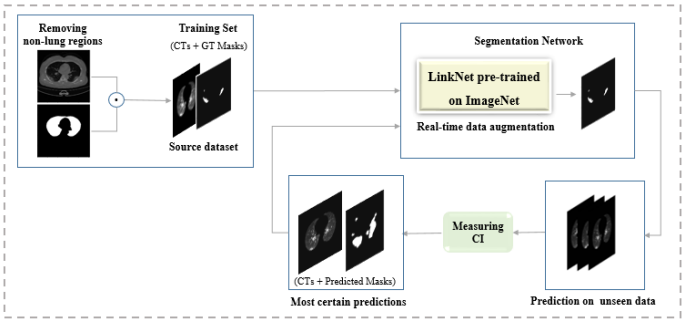

Fig. 1. The proposed generalization enhancement pipeline for DL COVID-19 lesion segmentation model.

$$CI(\hat{Y}) = \frac{\sum_{i=1}^{N} P_i}{N} \qquad for \;\; P_i > Cut-off \;\; threshold, \tag{1}$$

where $\hat{Y}$ is the predicted infection mask on a test CT slice, P_i denotes the predicted probability of pixel i belongs to the infection class, and N is the number of pixels in a CT slice where P_i is greater than a determined cut-off threshold. The pixels with P_i values higher than the cut-off threshold are assigned to the lesion class, and the rest are considered as the background class. The cut-off threshold is determined based on the in-domain test set and kept the same for external domains. In other words, the cut-off threshold that yields the best performance on the in-domain test set is considered the cut-off threshold for all the unseen domain test sets. We threshold the infection probabilities to consider pixels predicted as infection class for which the model is more confident and eventually find the most certain predicted masks. Indeed, for each test image, the CI determines the average of P_i over pixels belonging to the lesion class, quantifying the model's confidence in segmenting regions of infection. It should be noted that to get rid of the tiny fault predictions, we eliminate the test CT slices for which N is smaller than 50. It is worth noting that compared to the size of CT slices in our experiments (512×512 pixels) and the probable size of the lung region, an infection region with the size of 50 pixels is quite negligible. We then sort all predicted masks by their CI value and select the most certain predictions if their CI is not less than a determined limit. Next, we leverage the subset of CT slices with the highest CI value along with their predicted infection masks to retrain our initially

TABLE I
OPEN-ACCESS COVID-19 SEGMENTATION DATASETS USED IN OUR EXPERIMENTS.

	Number of infected CT slices	**Number of patients**	**Matrix size**
Dataset A	1351	10	512×512
Dataset B (COVID-CT-Rate)	433	82	512×512
Dataset C	373	9	630×630

trained segmentation model. It should be clarified that the number of most-certain predictions used for retraining and the CI limit are derived based on the size of test sets and measured CIs in our experimental settings. Our experimental results demonstrate that the network can learn informative features from the unseen domain dataset through the retraining process, leading to improved generalization. In other words, our proposed method can mitigate the domain shift between source and unseen data without having access to ground-truth masks of the unseen domain during training and retraining. We use LinkNet [21], which is an efficient semantic segmentation network in terms of computation and memory, as our DL segmentation model. As illustrated in Fig. 1, the model takes the lung region of a 2D CT slice as the input and generates a mask indicating the areas of infection. We remove non-lung regions from chest CT slices using an open-access pre-trained lung segmentation model [22] that can accurately detect lung regions from COVID-19 CT slices and has been used in a variety of COVID-19 studies such as outcome prediction [23], [24], diagnosis [25], [26], and lesion segmentation [27], [28]. Using the lung regions as the segmentation model's input decreases false-positive predictions that fall outside the lung area. Besides, rendering non-lung pixels to zero reduces the computational time, leading to faster convergence. Transfer learning is a powerful technique to resolve the shortage of data when deep learning models are designed for medical image analysis tasks with limited labeled datasets. Existing DL-based COVID-19 lesion segmentation studies have leveraged top-performing CNNs such as DenseNet, ResNet, and Inception-Net pre-trained on natural images (like the ImageNet dataset [29]) as the encoder's backbone to achieve improved results [8], [9]. Here, we adopt ResNet34 [30] pre-trained on the ImageNet dataset as the encoding path' backbone of the LinkNet. The pre-trained weights are used as the initial weights of the encoder, which will be updated during the training.

Furthermore, we leverage the data augmentation on the fly over training passes to improve the model's generalization on unseen data. Each mini-batch of input CT slices during the training is converted into synthetic images by applying conventional data transformation strategies, including zooming, shifting, flipping, and shearing. Indeed, during the training process, each synthetic image will be observed by the model only once, resulting in an improved generalization. It is noteworthy that the same data augmentation approach is used in the re-training process, meaning that different augmentation transformations are applied to high CI images from unseen domains to make the best use of them in the re-training process. We adopt a hybrid loss function, which is the sum of

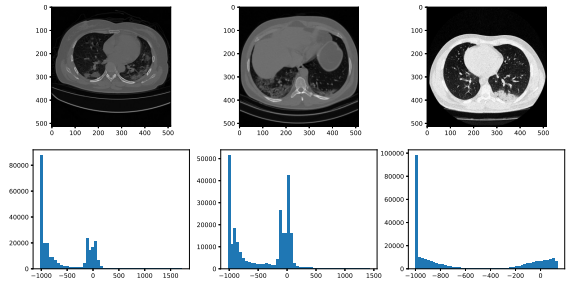

Fig. 2. Samples of COVID-9 chest CT slices from datasets A, B, and C used in our datasets and their intensity histograms.

Dice loss [31] and Binary Cross-Entropy (BCE). Dice loss balances the foreground and background pixels, and BCE provides image-level supervision for precise segmentation.

III. DATASET DESCRIPTION

As illustrated in Table I, we use three open-access COVID-19 lesion segmentation datasets in our experiments:

- ***Dataset A*** [32], containing whole lung CT volumes of 10 COVID-19 patients from one clinical center. It contains a total of $2,581$ 2D CT images with the size of 512×512 pixels. $1,351$ out of $2,581$ CT slices demonstrate traces of COVID-19 manifestations, which have been annotated by three expert radiologists. Lesion and lung regions' masks have been provided for all CT slices.
- ***Dataset B (COVID-CT-Rate)***, our previously introduced dataset [33], which contains 433 CT slices from 82 COVID-19 patients. The images' matrix size is 512×512. All images have been saved in the Digital Imaging and Communications in Medicine (DICOM) format and Hounsfield Unit. All the lesion masks have been carefully modified/verified by a thoracic radiologist with 20 years of experience. For more details regarding the COVID-CT-Rate dataset please refer to Ref [33].
- ***Dataset C*** [34], containing nine COVID-19 lung CT volumes, with a total of 829 CT slices. 373 out of 829 CT slices indicate infections that a radiologist has annotated. Besides the areas of infection, the lung regions' masks exist for all CT slices. The CT slices have a size of 630×630 pixels resized to 512×512 pixels for our study.

We perform three rounds of experiments. For each round, one dataset is used for training (source data). The other two are used as unseen test sets to evaluate the proposed method's performance in enhancing the segmentation network's generalization.

TABLE II

MODEL PERFORMANCE ON IN-DOMAIN AND UNSEEN TEST SETS BEFORE AND AFTER GENERALIZATION ENHANCEMENT. IN EACH EXPERIMENT, ONLY ONE DATASET WAS USED IN THE TRAINING PHASE, AND THE TWO OTHERS WERE USED AS UNSEEN TEST SETS.

Training Set	Test set		Model's performance on in- & and unseen domains			Model's performance on unseen domains after generalization enhancement		
	In-domain	Unseen	DSC	SPC	SEN	DSC	SPC	SEN
A	A		0.814 ± 0.05	0.997 ± 0.002	0.843 ± 0.05	-	-	-
		B	0.705 ± 0.06	0.996 ± 0.001	0.679 ± 0.09	0.73 ± 0.02	0.995 ± 0.001	0.769 ± 0.03
		C	0.788 ± 0.01	0.993 ± 0.0009	0.882 ± 0.01	0.794 ± 0.006	0.993 ± 0.0005	0.886 ± 0.01
B	B		0.832 ± 0.03	0.997 ± 0.001	0.861 ± 0.03	-	-	-
		A	0.684 ± 0.04	0.996 ± 0.002	0.679 ± 0.1	0.701 ± 0.03	0.996 ± 0.001	0.708 ± 0.09
		C	0.776 ± 0.02	0.993 ± 0.002	0.879 ± 0.07	0.778 ± 0.02	0.993 ± 0.002	0.887 ± 0.05
C	C		0.879 ± 0.02	0.997 ± 0.003	0.886 ± 0.02	-	-	-
		A	0.769 ± 0.012	0.997 ± 0.002	0.759 ± 0.11	0.778 ± 0.005	0.997 ± 0.002	0.79 ± 0.11
		B	0.794 ± 0.008	0.996 ± 0.002	0.816 ± 0.006	0.807 ± 0.005	0.996 ± 0.001	0.839 ± 0.002

TABLE III

THE AVERAGE OF MODEL PERFORMANCE ON IN-DOMAIN AND UNSEEN TEST SETS OVER THREE ROUNDS OF EXPERIMENTS

	DSC ($Ave \pm std$)	SPC ($Ave \pm std$)	SEN ($Ave \pm std$)
in-domain test set	0.842 ± 0.03	0.997 ± 0.002	0.886 ± 0.02
unseen test set (before enhancement)	0.753 ± 0.025	0.995 ± 0.0017	0.782 ± 0.064
unseen test set (after enhancement)	0.765 ± 0.014	0.995 ± 0.0013	0.813 ± 0.049

IV. EXPERIMENTAL RESULTS

In pre-processing, we normalize each CT slice based on the mean and standard deviation of its pixel intensities. The Adam optimization [35] with an initial learning rate equal to 0.001 is implemented to minimize the hybrid loss function during the training. The number of training epochs is set to 100. However, if the loss function on the validation set does not reduce after five epochs, the training process will stop to prevent the model from over-fitting. For training the segmentation network, we split the source dataset into three independent subsets for training (60%), validation (10%), and testing (30%). Datasets used in training, validation, and test sets are kept patient-independent to reduce the risk of information leakage, meaning that the CT slices of patients are not shared among underlying datasets. The model's weights associated with the highest Dice Similarity Coefficient (DSC) on the validation set are saved for the evaluation. We perform three rounds of experiments. In each round, we use one of the datasets as the source set and the other two as the external datasets. We train the segmentation network using source data and evaluate the performance of the trained network on in-domain test sets (a subset of the source data) and two other datasets as unseen domains. In each evaluation on unseen datasets, we measure the CI for the predicted lesion masks and sort the CT slices based on their CI to identify the most confident predictions. The CT slices with the highest CI value, if their measured CI is not less than a determined limit, are leveraged for retraining the initially trained segmentation model. The number of most-confident predictions and the value of the CI limit have been determined based on the size of the external test sets and measured CIs. As such, the most-certain predictions (CT slices and their predicted masks) are fed into the network for retraining. The implementation setting for retraining is kept the same as the initial training. Then, we evaluate the performance of the retrained model on the unseen test set. Fig. 2 presents random samples of datasets A, B, and C and their intensity distributions. As can be observed, CT slices acquired by different scanners, vendors, and from different populations differ in terms of quality, appearance, and spatial features.

We assess the model's performance in segmenting COVID-19 lesions by comparing the predicted and the ground-truth segmentation masks, using DSC, Sensitivity (SEN), and Specificity (SPC) as the evaluation metrics. Table II represents the average of the evaluation metrics on the in-domain and unseen test sets through a five-fold cross-validation approach. As illustrated, the proposed method can improve the performance of the DL segmentation network on unseen datasets, particularly the sensitivity metric, meaning that more lesion pixels have been labeled correctly. Besides, the results obtained by the retrained model show less variation on unseen data. Indeed, leveraging a subset of the target domain for which the model trained on the source set yields the most confident predictions and their predicted lesion masks, the network can extract informative features specific to the target domain during the retraining process. The average of model performance over three rounds of experiments on in-domain and unseen test sets have been demonstrated in Table III. As can be seen, although the enhancement approach can mitigate the gap between source and unseen data, there is still a significant decrease in segmentation results. One reason is that the most-confident predicted masks might be partially incorrect, which misleads the model during retraining, meaning that the introduced CI metric is not robust to possible outliers existing in the most confident predictions. One possible solution is using a human-in-loop approach to modify the predicted masks in a timely

fashion, leading to further improvements. Our findings indicate that the proposed generalization enhancement approach can improve the DL segmentation model's performance on CT slices acquired by different healthcare systems and from other patient populations.

V. Conclusion

This research proposes a generalization enhancement approach for DL segmentation models to improve model performance on an unseen target domain. Indeed, the proposed method mitigates the shift between the source and target datasets without having access to ground-truth pixel-level annotations of the unknown domain. We examined the effectiveness of our proposed method on three open-access datasets for COVID-19 lesion segmentation from chest CT slices. The results demonstrate that the enhancement method can improve the segmentation model's performance on unseen data, particularly the sensitivity metric, meaning that more lesion pixels have been labeled correctly. We leave the implementation of our method on other imaging modalities like MR and ultrasound images for the future. Exploring the efficacy of the proposed generalization enhancement models on multi-class segmentation tasks is another future direction of the present work.

References

[1] L. M. Prevedello, *et al.*, "Challenges related to artificial intelligence research in medical imaging and the importance of image analysis competitions," *Radiology: Artificial Intelligence*, vol. 1, no. 1, 2019.

[2] L. Yao, J. Prosky, B. Covington, and K. Lyman, "A strong baseline for domain adaptation and generalization in medical imaging," *arXiv preprint arXiv:1904.01638*, 2019.

[3] N. Kumar, *et al.*, "A dataset and a technique for generalized nuclear segmentation for computational pathology," *IEEE transactions on medical imaging*, vol. 36, no. 7, pp. 1550–1560, 2017.

[4] J. De Fauw, *et al.*, *et al.*, "Clinically applicable deep learning for diagnosis and referral in retinal disease," *Nature medicine*, vol. 24, no. 9, pp. 1342–1350, 2018.

[5] N. Karani, K. Chaitanya, C. Baumgartner, and E. Konukoglu, "A lifelong learning approach to brain mr segmentation across scanners and protocols," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 476–484, Springer, 2018.

[6] J. Liu, *et al.*, "Covid-19 lung infection segmentation with a novel two-stage cross-domain transfer learning framework," *Medical Image Analysis*, vol. 74, p. 102205, 2021.

[7] M. Ghafoorian, *et al.*, "Transfer learning for domain adaptation in mri: Application in brain lesion segmentation," in *International conference on medical image computing and computer-assisted intervention*, pp. 516–524, Springer, 2017.

[8] N. Enshaei, *et al.*, "An ensemble learning framework for multi-class covid-19 lesion segmentation from chest ct images," in *2021 IEEE International Conference on Autonomous Systems (ICAS)*, pp. 1–6, IEEE, 2021.

[9] B. Zheng, *et al.*, "Msd-net: Multi-scale discriminative network for covid-19 lung infection segmentation on ct," *Ieee Access*, vol. 8, pp. 185786–185795, 2020.

[10] D. Li, Y. Yang, Y.-Z. Song, and T. M. Hospedales, "Deeper, broader and artier domain generalization," in *Proceedings of the IEEE international conference on computer vision*, pp. 5542–5550, 2017.

[11] J. Wang, *et al.*, "Generalizing to unseen domains: A survey on domain generalization," *arXiv preprint arXiv:2103.03097*, 2021.

[13] I. Goodfellow, *et al.*, "Generative adversarial nets," *Advances in neural information processing systems*, vol. 27, 2014.

[12] L. Zhang, X. Wang, D. Yang, T. Sanford, S. Harmon, B. Turkbey, B. J. Wood, H. Roth, A. Myronenko, D. Xu, *et al.*, "Generalizing deep learning for medical image segmentation to unseen domains via deep stacked transformation," *IEEE transactions on medical imaging*, vol. 39, no. 7, pp. 2531–2540, 2020.

[14] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, pp. 2223–2232, 2017.

[15] K. Kamnitsas, *et al.*, "Unsupervised domain adaptation in brain lesion segmentation with adversarial networks," in *International conference on information processing in medical imaging*, pp. 597–609, Springer, 2017.

[16] M. A. Degel, N. Navab, and S. Albarqouni, "Domain and geometry agnostic cnns for left atrium segmentation in 3d ultrasound," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 630–637, Springer, 2018.

[17] J. Ren, I. Hacihaliloglu, E. A. Singer, D. J. Foran, and X. Qi, "Adversarial domain adaptation for classification of prostate histopathology whole-slide images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 201–209, Springer, 2018.

[18] D.-H. Lee *et al.*, "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks," in *Workshop on challenges in representation learning, ICML*, vol. 3, p. 896, 2013.

[19] S. Heidarian, *et al.*, "Robust automated framework for covid-19 disease identification from a multicenter dataset of chest ct scans," *arXiv preprint arXiv:2109.09241*, 2021.

[20] Y. Zou, *et al.*, "Pseudoseg: Designing pseudo labels for semantic segmentation," *arXiv preprint arXiv:2010.09713*, 2020.

[21] A. Chaurasia and E. Culurciello, "Linknet: Exploiting encoder representations for efficient semantic segmentation," in *2017 IEEE Visual Communications and Image Processing (VCIP)*, pp. 1–4, IEEE, 2017.

[22] J. Hofmanninger, F. Prayer, J. Pan, S. Röhrich, H. Prosch, and G. Langs, "Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem," *European Radiology Experimental*, vol. 4, no. 1, pp. 1–13, 2020.

[23] J. C. Quiroz, *et al.*, "Development and validation of a machine learning approach for automated severity assessment of covid-19 based on clinical and imaging data: retrospective study," *JMIR Medical Informatics*, vol. 9, no. 2, p. e24572, 2021.

[24] K. Grodecki, *et al.*, "Quantitative burden of covid-19 pneumonia at chest ct predicts adverse outcomes: A post hoc analysis of a prospective international registry," *Radiology: Cardiothoracic Imaging*, vol. 2, no. 5, p. e200389, 2020.

[25] B. Xiao, *et al.*, "Pam-densenet: a deep convolutional neural network for computer-aided covid-19 diagnosis," *IEEE Transactions on Cybernetics*, 2021.

[26] S. Heidarian, *et al.*, "Covid-fact: A fully-automated capsule network-based framework for identification of covid-19 cases from chest ct scans," *Frontiers in Artificial Intelligence*, vol. 4, 2021.

[27] Q. Dou, *et al.*, "Federated deep learning for detecting covid-19 lung abnormalities in ct: a privacy-preserving multinational validation study," *NPJ digital medicine*, vol. 4, no. 1, pp. 1–11, 2021.

[28] N. Paluru, *et al.*, "Anam-net: Anamorphic depth embedding-based lightweight cnn for segmentation of anomalies in covid-19 chest ct images," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 3, pp. 932–946, 2021.

[29] O. Russakovsky, *et al.*, "Imagenet large scale visual recognition challenge," *International journal of computer vision*, vol. 115, no. 3, pp. 211–252, 2015.

[30] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[31] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 fourth international conference on 3D vision (3DV)*, pp. 565–571, IEEE, 2016.

[32] J. Ma, *et al.*, "Towards efficient covid-19 ct annotation: A benchmark for lung and infection segmentation," 2020.

[33] N. Enshaei, *et al.*, "Covid-rate: an automated framework for segmentation of covid-19 lesions from chest ct images," *Scientific Reports*, vol. 12, no. 1, pp. 1–18, 2022.

[34] "Medseg, a free online segmentation tool, http://medicalsegmentation.com/covid19/, 2020."

[35] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.