Domain adaptation, Explainability & Fairness in AI for Medical Image Analysis: Diagnosis of COVID-19 based on 3-D Chest CT-scans

Dimitrios Kollias
Queen Mary University of London, UK
d.kollias@qmul.ac.uk

Anastasios Arsenos
National Technical University of Athens, Greece
anarsenos@gmail.com

Stefanos Kollias
National Technical University of Athens, Greece
National Infrastructures for Research and Technology, Greece
stefanos@cs.ntua.gr

Abstract

The paper presents the DEF-AI-MIA COV19D Competition, which is organized in the framework of the 'Domain adaptation, Explainability, Fairness in AI for Medical Image Analysis (DEF-AI-MIA)' Workshop of the 2024 Computer Vision and Pattern Recognition (CVPR) Conference. The Competition is the 4th in the series, following the first three Competitions held in the framework of ICCV 2021, ECCV 2022 and ICASSP 2023 International Conferences respectively. It includes two Challenges on: i) Covid-19 Detection and ii) Covid-19 Domain Adaptation. The Competition use data from COV19-CT-DB database, which is described in the paper and includes a large number of chest CT scan series. Each chest CT scan series consists of a sequence of 2-D CT slices, the number of which is between 50 and 700. Training, validation and test datasets have been extracted from COV19-CT-DB and provided to the participants in both Challenges. The paper presents the baseline models used in the Challenges and the performance which was obtained respectively, together with the best corresponding performances of the methods submitted and evaluated in the Challenges.

1. Introduction

In the past few years, Deep Learning (DL) techniques have made rapid advances in many medical image analysis tasks. In pathology and radiology applications, they managed to increase the accuracy and precision of medical image assessment, which is often considered subjective and not optimally reproducible. This is due to the fact that they can extract more clinically relevant information from medical images than what is possible in current routine clinical practice by human assessors. Nevertheless, considerable development and validation work lies ahead before AI-based methods can be fully integrated ad used in routine clinical tasks.

Of major importance is research on domain adaptation, fairness and explainability in AI-enabled medical image analysis. This research constitutes the main target of the Domain adaptation, Explainability and Fairness in AI for Medical Image Analysis (DEF-AI-MIA) Workshop, held in the 2024 Computer Vision and Pattern Recognition (CVPR) International Conference. The DEF-AI-MIA workshop aims to foster discussion and presentation of ideas to tackle these challenges in the field, as well as identify research opportunities in this context. It is the fourth in the AI-MIA series of Workshops, which includes the Workshops held at IEEE ICASSP 2023, ECCV 2022 and ICCV 2021 Conferences.

This Workshop's focus is also motivated by recent actions and regulatory policies developed in Europe and considered worldwide. GRNET, the Greek National Infrastructures for Research and Technology, has implemented the integration of public hospital units in GRNET academic network, to support research and clinical activities in medicine and biology, also providing an archiving service for data produced by the imaging devices of the hospitals at the GRNET health data centers. At the European level, EU has been regulating a European Health Data Space, which: a) fosters a genuine single market for electronic health record systems, relevant medical devices and high risk AI systems (primary use of health data), b) generates a consistent, trustworthy and efficient set-up for the use of health data for research and innovation (secondary use of health data; GRNET is involved in the implementation of this set-up). The above are linked to the recent EU AI-Act regulatory framework for AI, which classifies AI systems used in different applications according to the risk they pose to users. These

are under consideration, by the public and the private sector, in Europe, USA and other countries all over the world.

Topics covered in the workshop are domain adaptation, explainability, fairness, for trustworthiness in AI-enabled medical imaging which include a digital pathology and radiology images; use of self-supervised and unsupervised methods to enforce shared patterns emerging directly from data, develop strategies to leverage few (or partial) annotations, promoting interpretability in both model development and/or results obtained, ensure generalizability to data coming from multi-centers, multi-modalities or multi-diseases, in edge, or cloud frameworks, and robustness to out of distribution data.

Technologies and topics addressed in the DEF-AI-MIA Workshop include the following: explainable 2-D & 3D-CNN, CNN-RNN, transformer, foundation models, multimodal Large Language Models, unsupervised, self-supervised Machine Learning (ML) models for medical diagnosis; sensing "salient features" of AI/ML models related to decision-making, in spatial (images), temporal (video), volumetric (3-D) data; optimal visualization of salient features and areas in the input data; Low/Middle/High level feature extraction & analysis for model interpretatability and explainability; explanation of which features and at what time, or slice, or respective intervals, are the most prominent for the provided decision in temporal and 3-D data; explainable data correlations for predictions in data streams of multimodal data; joint optimization of positive and negative saliencies; global and local models for prediction or classification; attention and self-attention mechanisms in DL/AI approaches; interpretability at training time through adversarial regularization; learning new data (from multiple sources) by leveraging knowledge already extracted and codified, through domain adaptation; generalizable ML/DL methods when the training medical image datasets are small; generalizable ML/DL methods in cases of images with potential domain shift; unsupervised, weakly supervised and semi-supervised model adaptation; uncertainty estimation and quantification, self-training; adaptation and prompt engineering in Foundation Models (e.g., LLMs) for explainable decisions and prediction; algorithmic fairness; zero/one shot learning, avoidance of catastrophic forgetting.

2. The 4th COV19D Competition

A variety of technologies have been developed for early diagnosis of Covid-19, based on medical image analysis, especially focusing on 3-D chest CT scans. Special interest has been given to combined segmentation and classification approaches [23], targeting detection of abnormalities, including consolidation, ground-glass opacities, interlobular septal lung thickening, mostly under pleura.

The 4th COV19D Competition is the 4th in the series of COV19D Competitions following the first 3 Competitions we organized in the framework of ICCV 2021 [13], ECCV 2022 [16] and ICASSP 2023 [15] Workshops respectively. It includes two Challenges: i) Covid-19 Detection Challenge and ii) Covid-19 Domain Adaptation Challenge.

Both Challenges are based on the COV19-CT-DB database, briefly described next, including 3-D chest CT scan series. Each chest CT scan series consists of a sequence of 2-D CT slices, the number of which is between 50 and 700.

2.1. Covid-19 Detection Challenge

Many CT scans have been aggregated, each one of which has been manually annotated in terms of Covid-19 and non-Covid-19 categories. The resulting dataset is split into training, validation and test partitions. The training and validation sets along with their annotations have been provided to the Competition participating teams to develop AI/ML/DL models for Covid-19 and non-Covid-19 prediction. Performance of the different approaches have been evaluated on the test set in terms of the 'macro' F1 score.

2.2. Covid-19 Domain Adaptation Challenge

CT scans have been aggregated from various hospitals and medical centres. Each CT scan has been manually annotated with respect to Covid-19 and non-Covid-19 categories. The resulting dataset is split into training, validation and test partitions. Participants have been provided with a training set that consists of: i) the annotated data of the 1st Challenge which are aggregated from some hospitals and medical centres (case A); ii) a small number of annotated data and a larger number of non-annotated data (case B), all of which are aggregated from other hospitals and medical centres and their distribution is different from that of case A. Participants have been also provided with a validation set that consists of a small number of annotated data of case B. The participating teams have developed AI/ML/DL models for Covid-19 prediction. Performance of the different approaches have been evaluated on a test set (that contains data of case B) in terms of the 'macro' F1 score.

3. The COV19-CT-DB Database

COV19-CT-DB [14], which we have developed, contains 3-D chest CT scans, collected in various medical centers. The database includes 7,756 3-D CT scans; 1,661 are COVID-19 samples, whilst 6,095 refer to non COVID-19 ones. There are about 2,500,000 images included in these datasets. All have been anonymized. 724,273 images refer to the COVID-19 class, whilst 1,775,727 slices belong to non COVID-19 class [1].

Table 1 presents a summary of the main elements of COV19-CT-DB.

Figure 1 analyzes the length of the CT scan series, presenting their histogram. This shows the differences regarding the length of 3-D CT scans in COV19-CT-DB; these are caused by various reasons, including the requested resolution analysis, or the specific features of the used equipment.

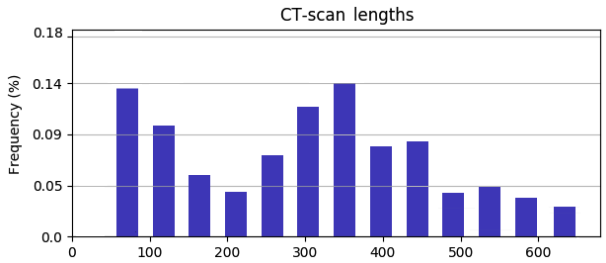

Figure 1. COV19-CT-DB: 3-D scan length histogram

It should be mentioned that for explainability purposes [9, 11, 12], an anchor set was generated for the COV19-CT-DB database [14]. This included 11 anchors, each representing a respective 3-D CT scan obtained through an appropriate clustering procedure. 7 of them corresponded to COVID-19 cases, with the rest corresponding to non COVID-19 cases. Justification is provided for the respective diagnosis, as shown in Table 2.

Figure 2 shows a series of slices from a COVID-19 case, whereas Figure 3 shows a series of slices from a non COVID-19 case.

The first Challenge on COVID-19 detection is based on extract of this database. The training set contains, in total, 1358 3-D CT scans. The validation set consists of 326 3-D CT scans. The number of COVID-19 and of Non-COVID-19 cases in each set are shown in Table 3.

The second Challenge on COVID-19 Domain Adaptation is also based on extract from this database. The CT scans utilized have been sourced from a variety of hospitals and medical centers, providing a diverse range of data for analysis. The dataset has been partitioned into distinct training, validation and test subsets.

239 3-D CT scans have been annotated and provided as training set to the participants, with 178 3-D CT scans constituting the validation set. In addition, 494 3-D CT scans have been provided without annotations, as shown in 4 so that they can be used by the participants in the adaptation process.

Table 1. COV19-CT-DB: main elements

Elements	Values
number of 3-D CT scans	1,661 COVID 6,095 non-COVID
number of 2-D images	724,273 COVID 1,775,727 non-COVID
number of images in scan series	50 - 700
size of images	512×512

4. The baseline configurations

4.1. COVID-19 detection & domain adaptation baselines

The baseline architecture adopted for both Challenges, namely the COVID-19 Detection Challenge and the Covid-19 Domain Adaptation Challenge, is a CNN-RNN architecture [2, 10, 14, 17].

The input 3-D CT scans have been padded to achieve a uniform length t, ensuring that every 3-D CT scan contains t slices. The entire unsegmented sequence [21] of 2-D slices from a CT scan is then fed into the CNN component. This CNN component conducts localized analysis on a per-2D-slice basis, primarily extracting features from the lung regions. The objective is to facilitate diagnosis using the entire 3-D series of CT scans, mirroring the annotations provided by medical experts.

Subsequently, the RNN component analyzes the CNN features of the complete 3-D CT scan, sequentially traversing from slice 0 to slice $t-1$. The outputs of the RNN component are forwarded to a Fully Connected layer and subsequently to an output layer utilizing a softmax activation function to provide the COVID-19 diagnosis. We also include a Dropout layer before the Fully Connected one.

Table 2. Description of COV19-CT-DB anchors

Cluster ID	Description
0	Bilateral shadows ground-glass that become more compact locally in lower lung lobes with an image of pneumonia due to COVID-19; severe category
1	Bilateral shadows ground-glass as in pneumonia due to COVID-19; moderate category
2	Minimal shadows ground-glass in left upper lung lobe. Severe thickening shadows, dense atelectasis of lower lung lobes. Minimal pleural fluid on the right. Possible microbial cause; critical category
3	Bilateral shadows ground-glass mainly in lower lung lobes as in pneumonia due to COVID-19 in rather mild condition; mild category
4	Bilateral shadows ground-glass that occupy more than 75 % of the pulmonary parenchyma as in pneumonia COVID-19 of critical condition; critical category
5	Bilateral shadows ground-glass that occupy about 50 % of the pulmonary parenchyma as in pneumonia COVID-19 of critical condition; severe category
6	Bilateral shadows ground-glass that occupy more than 75 % of the pulmonary parenchyma as in pneumonia COVID-19 of critical condition; critical category
7	Bilateral emphysematous lesions as in chronic obstructive pulmonary disease. Dense atelectasis in paravertebral right lung; mild category
8	Normal CT scan
9	Normal CT scan
10	Normal CT scan

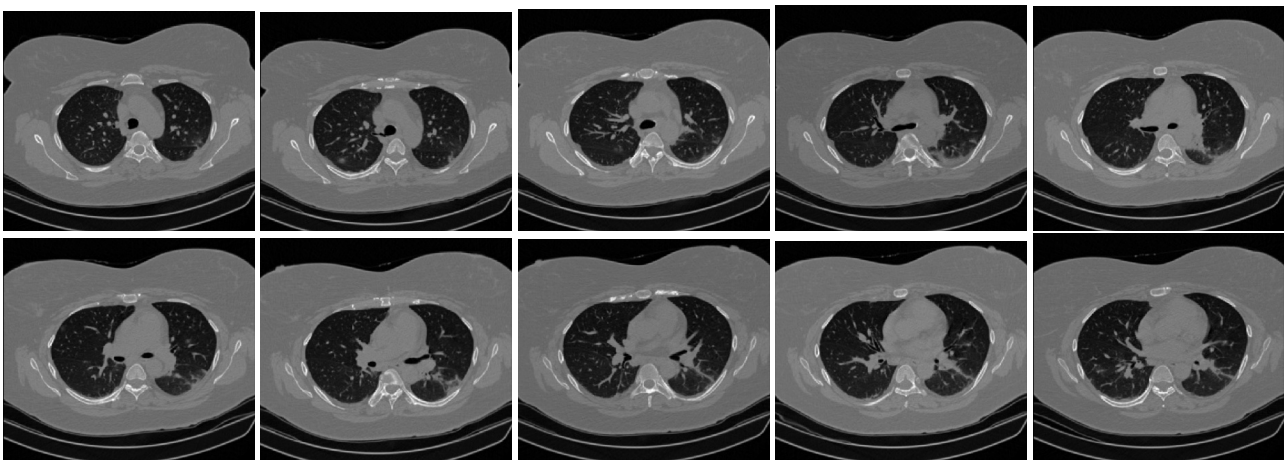

Figure 2. Slices from a COVID-19 case in COV19-CT-DB

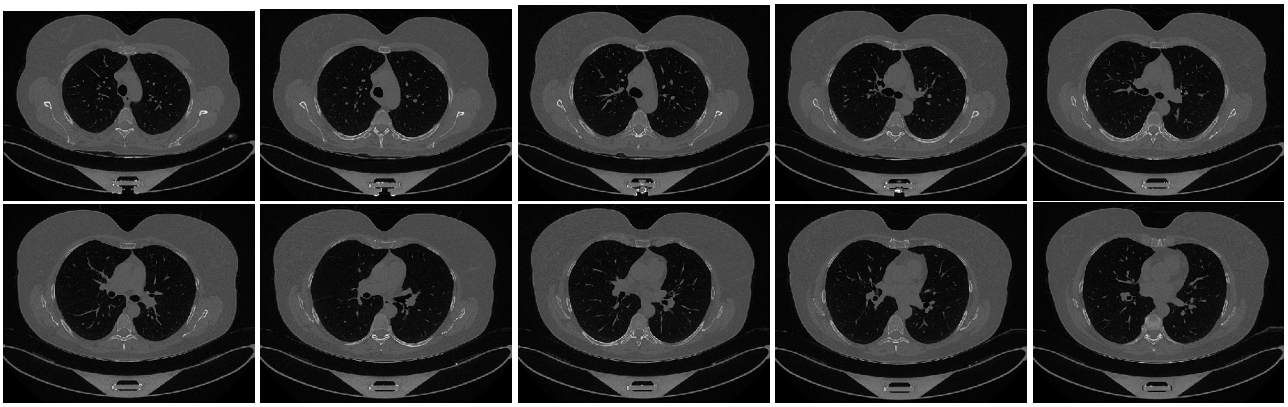

Figure 3. Slices from non COVID-19 case in COV19-CT-DB

Table 3. Data samples in each Set in Covid-19 Detection Challenge

Set	Training	Validation
COVID-19	703	170
Non-COVID-19	655	156

Table 4. Data samples in each Set in Covid- 19 Domain Adaptation Challenge

Set	Training	Validation
COVID-19	120	65
Non-COVID-19	119	113
Non-annotated	494	-

In the second Challenge (Covid-19 Domain Adaptation), we employed Monte Carlo Dropout to assess uncertainty while training the CNN-RNN architecture using data from both case A (annotated) and case B (annotated). Monte Carlo Dropout is a technique that involves performing multiple forward passes through the network with dropout activated during inference, allowing us to capture the model's inherent uncertainty. Subsequently, we annotated the non-annotated data from case B based on the model's predictions, specifically considering COVID instances where the model exhibited a high confidence level. This approach enabled us to leverage the model's uncertainty estimates to adapt to the non-annotated data of case B.

4.2. Pre-Processing & Implementation Details

In the pre-processing stage, all 2-D CT slices have been extracted from respective DICOM images. Next, voxel intensity values were computed through a window of 350 Hounsfield units (HU)/-1150 HU; they were then normalized in the range $[0, 1]$. Data augmentation was also performed, including random rotation in [-10°, 10°] and horizontal flip [8, 25] to extract region of interests, such as lung areas in the 2-D images.

As far as implementation of the baseline approach is concerned, the following models have been used: i) we adopted

Table 5. Performance of baseline models over Validation sets

Challenge	'macro' F1 Score
COVID-19 Detection	0.78
COVID-19 Domain Adaptation	0.73

the CNN ResNet50 model; on top of it we included a global average pooling, as well as a batch normalization layer and dropout (with keep probability 0.8), ii) we used a single one-directional GRU RNN layer comprising 128 neurons. The model input consisted of the 3-D CT scans. Each 2-D image was resized from its size of $512 \times 512 \times 3$ to $224 \times 224 \times 3$. We selected a confidence threshold of 70% to determine high-confidence annotations for non-annotated data in the Domain Adaptation Challenge.

Batch size was equal to 5 (i.e, at each iteration our model processed 5 CT scans) and the input length 't' was 700 (the maximum number of slices found across all CT scans). We utilized the softmax cross entropy as loss function for training both baseline methods. Adam optimizer was used with learning rate 10^{-4}. Training was performed on a Tesla V100 32GB GPU.

5. Experimental Results

5.1. Baseline Model Performance

This section describes a set of experiments evaluating the performance of the baseline configurations.

Table 5 shows the performance of the network over the validation sets in both Challenges, after training with the training datasets, taking into account that there exists only a single label for the whole CT scan and no labels for each CT scan slice [14].

In both Challenges the performance of the baseline methods were evaluated in terms of the macro F1 score. The macro F1 score is defined as the unweighted average of the class-wise/label-wise F1-scores, i.e., the unweighted average of the COVID-19 class F1 score and of the non-COVID-19 class F1 score.

5.2. The COV19D Competition Results

21 Teams participated in the 4th COV19D Competition (COVID19 Detection Challenge and COVID19 Domain Adaptation Challenge). 12 Teams submitted their results to the COVID19 Detection Challenge; 10 Teams submitted their results to the COVID19 Domain Adaptation Challenge. 6 and 4 Teams scored higher than the baseline and made valid submissions, respectively; their results are shown in the leaderboards presented in Table 6 and Table 7.

It is worth mentioning that in the COVID-19 Detection Challenge, the top four performing teams had up to 0.65 % difference in their performance.

Moreover, in the COVID-19 Domain Adaptation Challenge, the top two performing teams had only 0.34 % difference in their performance.

In particular, the MDAP approach, which was developed and used in both Challenges, preprocessed the CT scans to segment the lungs, as well asthe output volumes with the lungs, individually and together. It then trained 3D ResNet and Swin Transformer models on these data and used them for COVID-19 detection. It then annotated the unlabeled CT scans using an ensemble of these models and chose the high-confidence predictions, as pseudo-labels for fine-tuning the models and apply them for COVID-19 Domain Adaptation.

The Deep adaptation method centered on lung segmentation and COVID-19 infection segmentation, employing the CNN-based segmentation architecture PDAtt-Unet, which simultaneously segmented lung regions and infections. It then concatenated the input slice (grayscale) with segmented lung and infection, generating three input channels akin to color channels. In addition, it employed three 3-D CNN backbones—customized Hybrid-DeCoVNet, along with pretrained 3D-Resnet-18 and 3D-Resnet-50 models—to train COVID-19 detection and adaptation models for both challenges. In this framewerk, it explored ensemble approaches and testing augmentation to enhance performance.

The FDVTS approach used different methods in the two Challenges. In the Detection Challenge, firstly, it analyzed the characteristics of the 3D CT scans and removed the non-lung parts, facilitating the model to focus on lesion-related areas and reducing computational cost. It then used ResNeSt50 as a strong feature extractor, initializing it with pretrained weights from COVID-19-specific prior knowledge. In the Domain Adaptation Challenge, it applied a two-stage framework that leveraged pseudo labels for domain adaptation so as to enhance the detection of COVID-19 from CT scans. By utilizing annotated data from one domain and non-annotated data from the other, the model overcame the problems of data scarcity and variability, generating pseudo labels and iteratively refined its learning process, thereby improving its accuracy and adaptability.

The ACVLAB approach adopted an advanced Spatial-Slice Feature Learning framework, specifically tailored for 3-D CT scans. It targeted to filter out out-of-distribution (OOD) data within the entire 3-D CT scan, allowing the selection of essential spatial-slice features for analysis by reducing data redundancy by 70 %. Additionally, it introduced a Kernel-Density-based slice Sampling method to enhance stability during training and inference phases, thereby accelerating convergence and enhancing overall performance, while using a simple EfficientNet-2D model.

The ViGIRLab method used EfficientNet, but with an added Attention Mechanism, resulting in EfficientNet-AM.

Table 6. COVID19 Detection Challenge Results: F1 Score in %; the best performing submission is in bold

Teams	Submission #	Macro F1	F1 (NON-COVID)	F1 (COVID)	Github
MDAP [22]	1	93.80	95.06	92.53	link
	2	92.56	94.21	90.91	
	3	94.30	95.50	93.09	
	4	94.84	95.88	93.79	
	5	**94.89**	**95.97**	**93.81**	
Deep-Adaptation [4]	1	93.30	94.40	92.20	link
	2	93.50	94.63	92.37	
	3	92.75	93.86	91.64	
	4	93.67	94.71	92.63	
	5	**94.60**	**95.53**	**93.66**	
ACVLAB [7]	1	93.88	95.10	92.66	link
	2	94.19	95.32	93.06	
	3	92.96	94.27	91.65	
	4	94.37	**95.57**	93.17	
	5	**94.39**	95.52	**93.26**	
FDVTS [19]	1	93.59	94.86	92.32	link
	2	93.60	94.86	92.33	
	3	**94.24**	**95.41**	**93.07**	
	4	93.11	94.39	91.84	
	5	93.67	94.91	92.44	
ViGIR Lab [5]	1	92.91	94.36	91.45	link
	2	92.91	94.36	91.45	
	3	93.12	94.55	91.70	
	4	93.21	94.58	91.84	
	5	**93.63**	**94.97**	**92.29**	
M2@Purdue [20]	**1**	**90.14**	**92.06**	**88.22**	link
baseline [18]	**1**	**85.11**	**87.48**	**82.74**	

Table 7. COVID19 Domain Adaptation Challenge Results: F1 Score in %; the best performing submission is in bold

Teams	Submission #	Macro F1	F1 (NON-COVID)	F1 (COVID)	Github
FDVTS [24]	**1**	**77.55**	**96.97**	**58.14**	link
	2	76.27	96.57	55.97	
	3	76.96	96.79	57.14	
	4	77.07	96.75	57.39	
	5	76.05	96.42	55.68	
MDAP [22]	1	70.47	94.76	46.17	link
	2	76.58	96.78	56.37	
	3	74.22	96.09	52.36	
	4	75.91	96.56	55.27	
	5	**77.21**	**96.82**	**57.60**	
Deep-Adaptation [4]	**1**	**74.96**	**96.52**	**53.39**	link
	2	73.67	96.10	51.25	
	3	64.74	92.48	37.00	
	4	74.33	96.23	52.44	
	5	63.23	91.50	34.97	
M2@Purdue [20]	**1**	**65.79**	**91.92**	**39.66**	link
baseline [18]	**1**	**60.16**	**86.67**	**33.65**	

Unlike other pipelines, which relied on a pre-processing step, the used pipeline used the raw input images, only applying an image-selection step to reduce the number of CT images required for training and/or testing. A computationally efficient pipeline was used, without incorporating a decoder to segment the lungs, nor combining different backbones, or an RNN with a backbone.

The M2@Purdue approach was applied to both Challenges. It used a lightweight detector, leveraging a frozen CLIP image encoder and a trainable multilayer perception. This was enhanced with Conditional Value at Risk for robustness and a loss landscape flattening strategy for improved generalization. Furthermore, it integrated a teacher-student framework to capitalize on the vast amounts of unlabeled data.

6. Conclusions and Future Work

In this paper we presented the 4th COV19D Competition and particularly the two Challenges that it contained: the first on COVID-19 detection and the second on COVID-19 domain adaptation. We provided a short description of the COV19-CT-DB, extracts from which were used in the two Challenges. We also presented the developed baseline approaches and their performance in the Challenges.

We also provided a comparison of the performance of all methods that outperformed the baselines in both Challenges. It can be seen that four approaches provided a Macro F1 score higher than 94 % in the COVID-19 Detection Challenge. Moreover two of them also provided a performance with Macro F1 score higher than 77 % in the COVID-19 Domain Adaptation Challenge.

These results illustrate the ability of the deep learning enabled methods to detect COVID-19 based on 3-D chest CT scans with high accuracy. Moreover, they illustrate that Domain Adaptation and Generalization [3] can be a valuable approach for tackling the diversity of datasets obtained across different hospitals and medical centers; they show that this research direction is of high importance and further improvement can be achieved in the future. Deployment of a recently developed solution has been developed for medical usage [6].

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)*, pages 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)*, pages 1–5. IEEE, 2023. 3

[3] Anastasios Arsenos, Dimitrios Kollias, Evangelos Petrongonas, Christos Skliros, and Stefanos Kollias. Uncertainty-guided contrastive learning for single source domain generalisation. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6935–6939. IEEE, 2024. 7

[4] Fares Bougourzi, Feryal Windal Moula, Halim Benhabiles, Fadi Dornaika, and Abdelmalik Taleb-Ahmed. Ensembling and test augmentation for covid-19 detection and covid-19 domain adaptation from 3d ct-scans. *arXiv preprint arXiv:2403.11338*, 2024. 6

[5] Ramy Farag, Parth Upadhyay, and Guilhermen DeSouza. Covid-19 detection from ct scans using efficientnet and attention mechanism. *arXiv preprint arXiv:2403.11505*, 2024. 6

[6] Demetris Gerogiannis, Anastasios Arsenos, Dimitrios Kollias, Dimitris Nikitopoulos, and Stefanos Kollias. Covid-19 computer-aided diagnosis through ai-assisted ct imaging analysis: Deploying a medical ai system. *arXiv preprint arXiv:2403.06242*, 2024. 7

[7] Chih-Chung Hsu, Chia-Ming Lee, Yang Fan Chiang, Yi-Shiuan Chou, Chih-Yu Jiang, Shen-Chieh Tai, and Chi-Han Tsai. Simple 2d convolutional neural network-based approach for covid-19 detection. *arXiv preprint arXiv:2403.11230*, 2024. 6

[8] Lu Huang, Rui Han, Tao Ai, Pengxin Yu, Han Kang, Qian Tao, and Liming Xia. Serial quantitative chest ct assessment of covid-19: a deep learning approach. *Radiology: Cardiothoracic Imaging*, 2(2):e200075, 2020. 4

[9] Ilianna Kollia, Andreas-Georgios Stafylopatis, and Stefanos Kollias. Predicting parkinson's disease using latent information extracted from deep neural networks. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2019. 3

[10] Dimitrios Kollias, Athanasios Tagaris, Andreas Stafylopatis, Stefanos Kollias, and Georgios Tagaris. Deep neural architectures for prediction in healthcare. *Complex & Intelligent Systems*, 4(2):119–131, 2018. 3

[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. 3

[12] Dimitris Kollias, Y Vlaxos, M Seferis, Ilianna Kollia, Levon Sukissian, James Wingate, and S Kollias. Transparent adaptation in deep medical image diagnosis. In *International Workshop on the Foundations of Trustworthy AI Integrating Learning, Optimization and Reasoning*, pages 251–267. Springer, 2020. 3

[13] 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*, pages 537–544, 2021. 2

[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. 2, 3, 5

[15] 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)*, pages 1–5. IEEE, 2023. 2

[16] Dimitrios Kollias, Anastasios Arsenos, and Stefanos Kollias. Ai-mia: Covid-19 detection and severity analysis through medical imaging. In *Computer Vision–ECCV 2022 Workshops: Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part VII*, pages 677–690. Springer, 2023. 2

[17] Dimitrios Kollias, Karanjot Vendal, Priyankaben Gadhavi, and Solomon Russom. Btdnet: A multi-modal approach for brain tumor radiogenomic classification. *Applied Sciences*, 13(21):11984, 2023. 3

[18] 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. 6

[19] Qingqiu Li, Runtian Yuan, Junlin Hou, Jilan Xu, Yuejie Zhang, Rui Feng, and Hao Chen. Advancing covid-19 detection in 3d ct scans. *arXiv preprint arXiv:2403.11953*, 2024. 6

[20] Li Lin, Yamini Sri Krubha, Zhenhuan Yang, Cheng Ren, Xin Wang, and Shu Hu. Robust covid-19 detection in ct images with clip. *arXiv preprint arXiv:2403.08947*, 2024. 6

[21] Natalia Salpea, Paraskevi Tzouveli, and Dimitrios Kollias. Medical image segmentation: A review of modern architectures. In *Computer Vision–ECCV 2022 Workshops: Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part VII*, pages 691–708. Springer, 2023. 3

[22] Robert Turnbull and Simon Mutch. High-confidence pseudo-labels for domain adaptation in covid-19 detection. *arXiv preprint arXiv:2403.13509*, 2024. 6

[23] Shuai Wang, Bo Kang, Jinlu Ma, Xianjun Zeng, Mingming Xiao, Jia Guo, Mengjiao Cai, Jingyi Yang, Yaodong Li, Xiangfei Meng, et al. A deep learning algorithm using ct images to screen for corona virus disease (covid-19). *European radiology*, pages 1–9, 2021. 2

[24] Runtian Yuan, Qingqiu Li, Junlin Hou, Jilan Xu, Yuejie Zhang, Rui Feng, and Hao Chen. Domain adaptation using pseudo labels for covid-19 detection. *arXiv preprint arXiv:2403.11498*, 2024. 6

[25] Chuansheng Zheng, Xianbo Deng, Qing Fu, Qiang Zhou, Jiapei Feng, Hui Ma, Wenyu Liu, and Xinggang Wang. Deep learning-based detection for covid-19 from chest ct using weak label. *MedRxiv*, 2020. 4