A Deep Neural Architecture for Harmonizing 3-D Input Data Analysis and Decision Making in Medical Imaging

Dimitrios Kollias[a], Anastasios Arsenos[b], Stefanos Kollias[b,*]

[a]*School of Electronic Engineering & Computer Science, Queen Mary University London, Mile End Rd, London, E14NS, United Kingdom*
[b]*School of Electrical & Computer Engineering, National Technical University Athens, Polytechnioupoli, Zografou, 15780, Greece*

Abstract

Harmonizing the analysis of data, especially of 3-D image volumes, consisting of different number of slices and annotated per volume, is a significant problem in training and using deep neural networks in various applications, including medical imaging. Moreover, unifying the decision making of the networks over different input datasets is crucial for the generation of rich data-driven knowledge and for trusted usage in the applications. This paper presents a new deep neural architecture, named RACNet, which includes routing and feature alignment steps and effectively handles different input lengths and single annotations of the 3-D image inputs, whilst providing highly accurate decisions. In addition, through latent variable extraction from the trained RACNet, a set of anchors are generated providing further insight on the network's decision making. These can be used to enrich and unify data-driven knowledge extracted from different datasets. An extensive experimental study illustrates the above developments, focusing on COVID-19 diagnosis through analysis of 3-D chest CT scans from databases generated in different countries and medical centers.

Keywords: RACNet, harmonization, routing, latent variables, COV19-CT-DB, COVID-19 diagnosis.

*Corresponding author
Email address: stefanos@cs.ntua.gr (Stefanos Kollias)

1. Introduction

In a variety of applications, input data are in the form of 3-D volumetric images, i.e., two dimensional image sequences which include different number of frames, or slices, and which are annotated in terms of a single label per sequence. Such applications are, for example, 3-D chest CT scan analysis for pneumonia, COVID-19, or Lung cancer diagnosis [1], [2]; 3-D magnetic resonance image (MRI) analysis for Parkinson's, or Alzheimer's disease prediction [3], [4]; 3-D subject's movement analysis for action recognition & Parkinson's detection [5]; analysis of audiovisual data showing subject's behaviour for affect recognition [6]; anomaly detection in nuclear power plants [7]. Dealing with a single annotation per volumetric input and harmonizing the input variable length constitutes a significant problem when training Deep Neural Networks (DNNs) to perform the respective prediction, or classification task.

Furthermore, in each of the above application fields, public, or private datasets are produced in different environments and contexts and are used to train deep learning systems to successfully perform the respective tasks. Extensive research is currently made on using data-driven knowledge, extracted from a single, or from multiple datasets, so as to deal with other datasets. Transfer learning, domain adaptation, meta-learning, domain generalization, continual or life long learning are specific topics of this research, based on different conditions related to the considered datasets [8]. An additional condition can be that some, or all of the datasets may not be available during continual learning, due for example to privacy, or General Data Protection Regulation (GDPR) issues. In such cases it can be possible to perform diagnosis by only sharing some data-driven knowledge, like the weights of independently trained DNNs.

COVID-19 diagnosis based on medical image analysis is the application domain examined in this paper. Various methods have been proposed to diagnose COVID-19, using analysis of chest x-rays, or CT scans. In particular, chest 3-D CT images can be used for precise COVID-19 early diagnosis. Recent approaches target segmentation and automatic detection of the pneumonia region

in lungs and subsequent prediction of anomalies related to COVID-19 [9]. Common anomalies are multiple ground-glass opacity, consolidation, and interlobular septal thickening in both lungs, which are mostly distributed under the pleura.

Such approaches require large training datasets. A few databases with CT scans have been recently developed [10], [11]. However, a rather fragmented approach is followed: research is based on specific datasets, provided by small, or larger numbers of hospitals, with no proof of good performance generalization over different datasets and clinical environments. Moreover, many datasets are small, in terms of total CT scans, or scan slices, or COVID-19 annotated CT scans, or number of patients [12].

In this paper we present and use a new very large database, COV19-CT-DB, which we have developed, including chest 3-D CT scans, aggregated from different hospitals. In particular, it includes 7,756 3-D CT scans, annotated for COVID-19 infection; 1,661 are COVID-19 cases and 6,095 are non-COVID-19 cases. The 3-D CT scans consist of different numbers of CT slices, ranging from 50 to 700, totalling around 2,500,000 CT slices. The whole database is being currently made available to the research community through our website.

We also present a deep neural architecture able to: i) analyze 3-D CT scan inputs, ii) effectively handle the problem that each CT scan consists of a different number of CT slices and iii) provide a very high performance, when used on COV19-CT-DB and on other public datasets for COVID-19 diagnosis. RACNet is a CNN-RNN architecture that is modified to include routing and feature alignment steps which dynamically select the specific RNN outputs to be fed to the dense layers for decision making, i.e., COVID-19 diagnosis.

In addition, we extract latent variables from the trained RACNet and derive a set of anchors that provide insight into the network's data driven knowledge. These anchors are further used for unification with other public datasets, developing a continual learning framework that does not require sharing of datasets.

The novel contributions of the paper include: i) presentation and analysis of the COV19-CT-DB, which is a new very large database of 3-D Chest CT scans, annotated for COVID-19 detection and provided to the public for research

purposes ii) development and evaluation of RACNet which can effectively and efficiently analyze 3-D CT scans aggregated from different sources, with different characteristics and lengths, and perform highly accurate COVID-19 diagnosis iii) extensive evaluation of RACNet performance over six different datasets, showing that it provides a greatly improved performance when compared with the state-of-the-art iv) unification of COVID-19 detection via extracted RACNet anchor representations and evaluation over 3 different datasets.

The rest of the paper is organized as follows. Related work is presented in Section 2. Section 3 provides a short description of COV19-CT-DB. The RACNet is described in Section 4. Section 5 includes the experimental study presenting the evaluation of the performance of RACNet when trained on COV19-CT-DB and several other public databases, as well as the validation of the unification approach. Conclusions and future work are presented in Section 6.

2. Related Work

A variety of 3-D CNN models have been used for detecting COVID-19 and distinguishing it from other common pneumonia (CP) and normal cases, using volumetric 3-D CT scans. In [13] a pretrained DenseNet-201 was trained to classify CT scan images to COVID-19, or non-COVID-19 category. The network's performance was compared to that of fine-tuned VGG16, ResNet152V2, and Inception-ResNetV2. 3-D CNN models have also been used in [14], [15].

In [2], a weakly supervised deep learning framework was presented using 3-D CT volumes for COVID-19 classification and lesion localization. A pre-trained UNet was utilized for segmenting the lung region of each CT scan slice; the latter was fed into a 3-D CNN that provided the classification outputs; the COVID-19 lesions were localized by combining the activation regions in the CNN and some connected components in unsupervised way.

In [16] and [17] 3-D models, such as ResNet3D101 and MC3_18 [18], were used to perform CT scan classification. These 3-D CNN models provided moderate results and were quite large in terms of model size compared to CNN-RNN

models. In addition, both papers proposed a neural network search framework to automatically search the 3-D models for the classification task. In [16] a differentiable neural architecture search method is developed combined with the Gumbel Softmax technique; the training, validation and test datasets used have been published for future research. In [17] an Evolutionary Multi-objective Architecture Search algorithm called EMARS was introduced, which takes into consideration the size of the model in the search space. These auto-generated models developed in [16] and [17] outperformed the previously manually constructed 3-D CNNs in terms of accuracy and model size, but they lack practicality, since a different model must be constructed for every analyzed dataset.

[19] used a CNN plus RNN network, receiving input CT scan images at its input and discriminating between COVID-19 and non-COVID-19 cases. In [1], the authors developed a multi-task architecture consisting of a (common) encoder that takes a 3-D CT scan as input and i) a decoder that reconstructs it; ii) a second decoder that performs COVID lesion segmentation; and iii) a multilayer perceptron for classification between COVID and non-COVID categories.

[20] and [21] used large 2D CNN models (Inception-ResNet-V2 and VGG16) to classify the 3D CT-scan volume through Ensemble models. Despite their good performance, these approaches are very difficult to contribute to real life environments, due to very high computational complexity and inference time.

3. The COV19-CT-DB Database

COV19-CT-DB includes 3-D chest CT scans annotated for existence of COVID-19. Data collection was conducted in the period from September 1 2020 to November 30 2021. It consists of 1,661 COVID and 6,095 non-COVID chest CT scan series, which correspond to a high number of patients (more than 1150) and subjects (more than 2600). These are anonymous; in general more than one CT scan series refer to each person, usually taken in different time instances. In total, 724,273 slices correspond to the CT scans of the COVID-19 category and 1,775,727 slices correspond to the non COVID-19 category.

Annotation of each CT slice has been performed by 4 very experienced (each with over 20 years of experience) medical experts; two radiologists and two pulmonologists. Labels provided by the 4 experts showed a high degree of agreement (around 98%). Each of the 3-D scans includes different number of slices, ranging from 50 to 700. Figure 1 shows the histogram of the 3-D CT scan lengths, where the length denotes the number of CT scan slices. This Figure illustrates the variety of 3-D CT scan lengths in the COV19-CT-DB database. This variation in number of slices is due to the context of CT scanning. The context is defined in terms of various factors, such as the accuracy asked by the doctor who ordered the scan, the characteristics of the CT scanner that is used, or specific subject's features, e.g., weight and age.

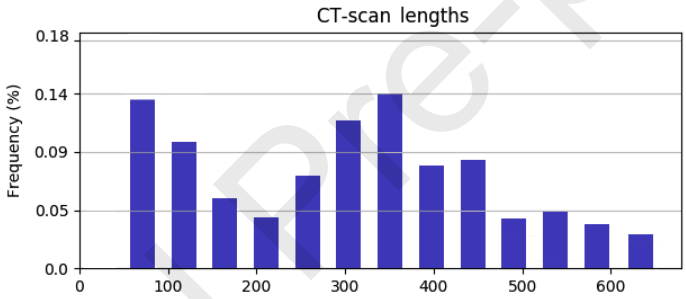


Figure 1: Histogram of CT scan lengths in COV19-CT-DB

Figure 2 shows slices from a CT scan series from a non-COVID case. Similarly, Figure 3 shows slices from a CT scan series from a COVID case. Figure 4 shows four CT scan slices, two from a non-COVID-19 CT scan, on the left and two from a COVID-19 scan, on the right. Bilateral ground glass regions are seen especially in lower lung lobes in the COVID-19 slices.

The database has been split in training, validation and test sets. The training set contains, in total, 1991 CT scans, 882 of which are labeled as COVID-19 and 1109 are labeled as non COVID-19. The validation set contains, in total, 484 CT scans, 215 of which are labeled as COVID-19 and 269 are labeled as non COVID-19. The test set contains, in total, 5281 CT scans, 564 of which

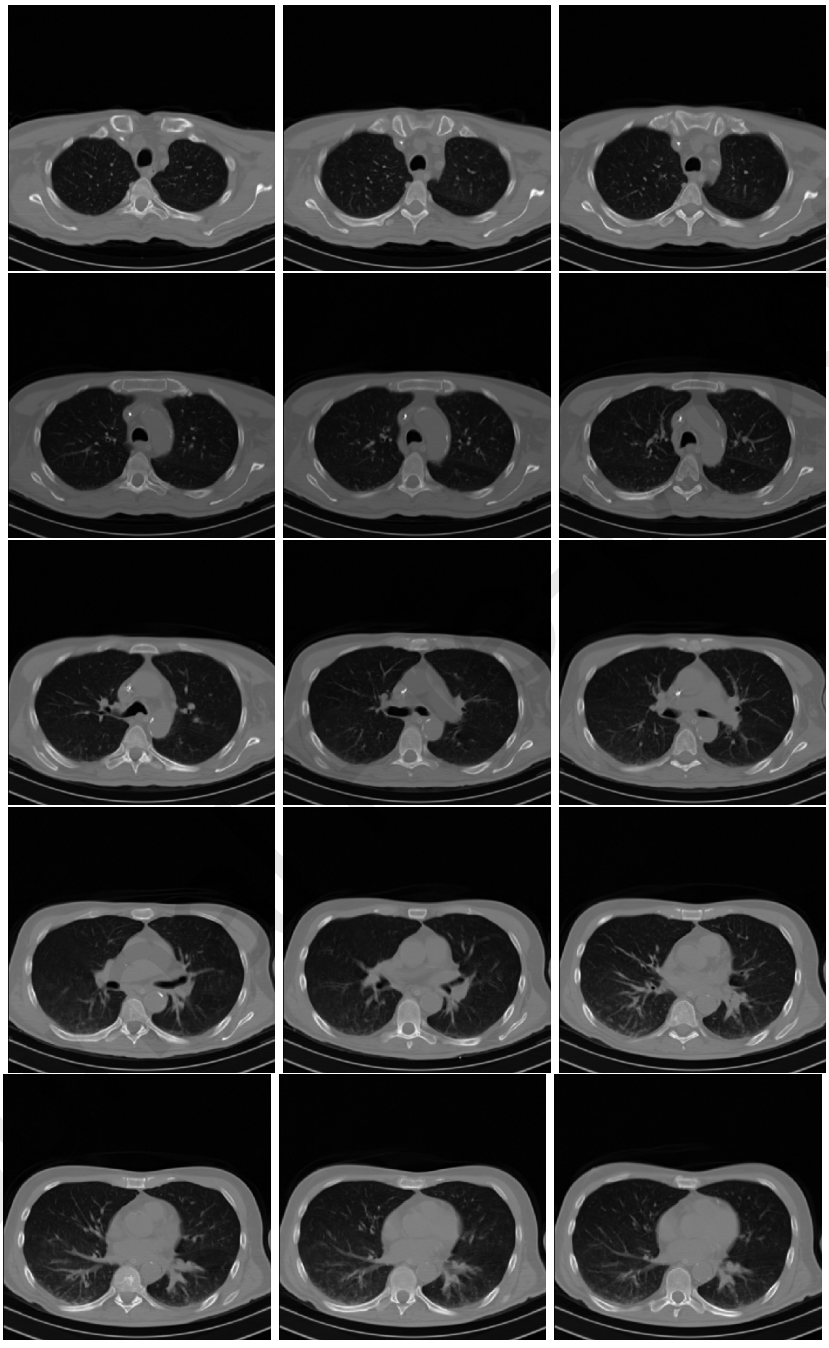


Figure 2: Slices from a non COVID-19 CT scan in COV19-CT-DB

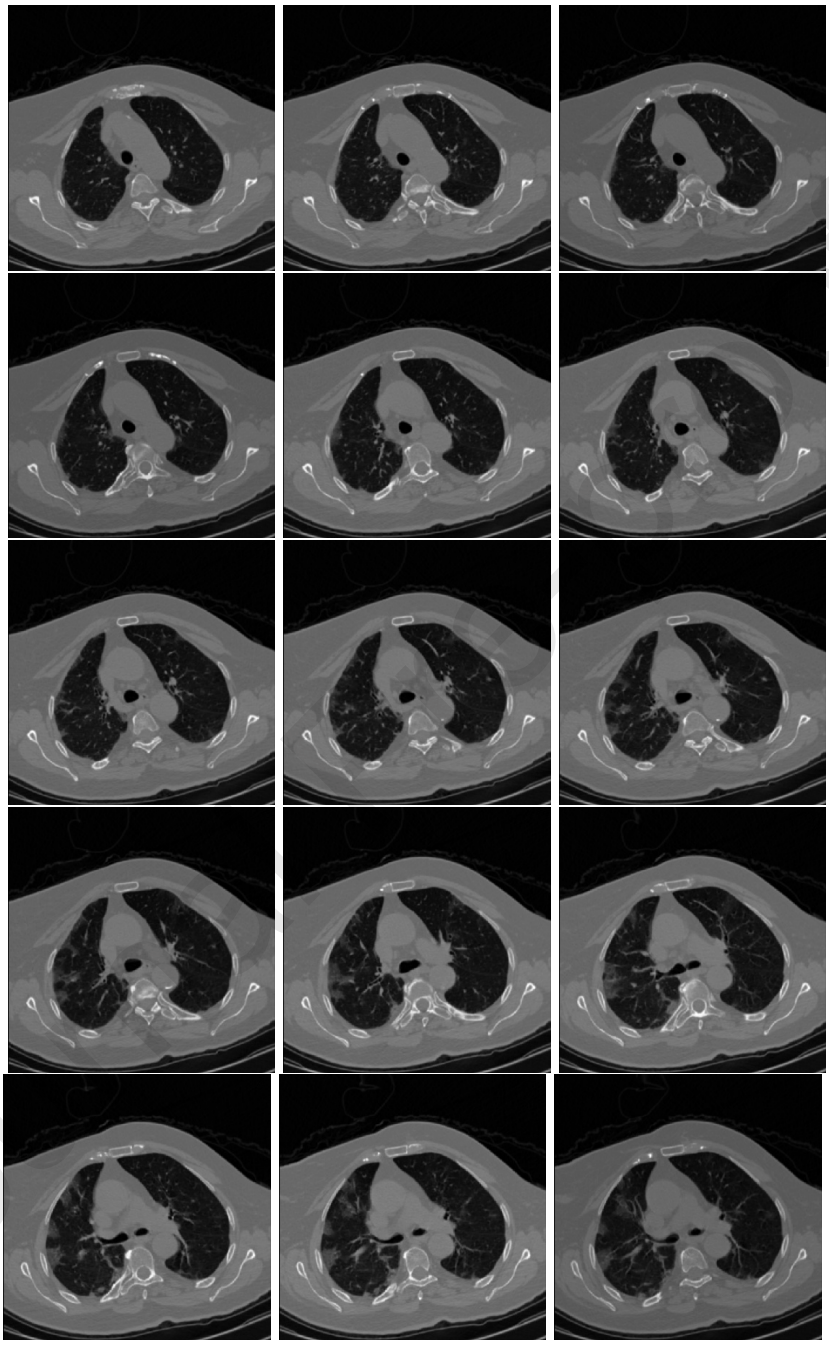


Figure 3: Slices from a COVID-19 CT scan in COV19-CT-DB

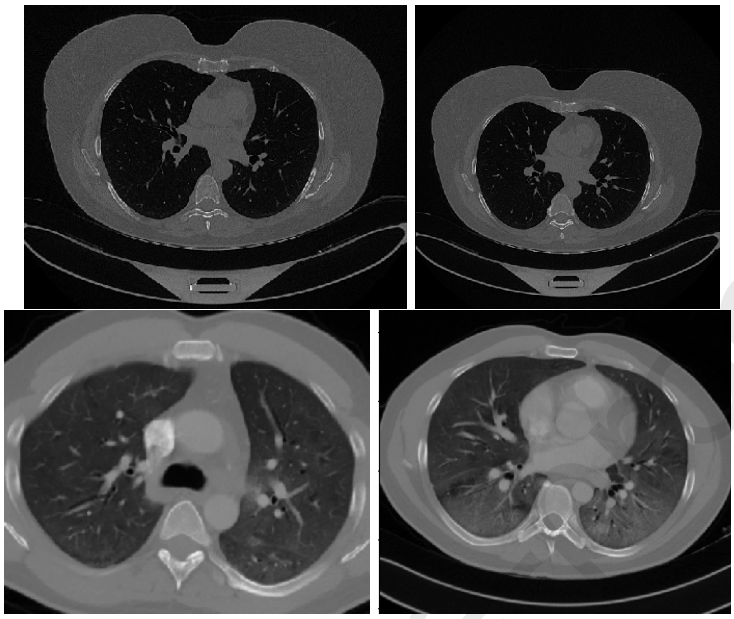


Figure 4: Four CT scan slices in COV19-CT-DB, the top 2 from a non-COVID-19 scan and the bottom 2 from a COVID-19, including bilateral ground glass regions in lower lung lobes.

are labeled as COVID-19 and 4717 are labeled as non COVID-19.

Some CT series of the same person have been taken at different time instances. In order to verify that there is no leakage of data from the training to the test set, we compared each 3-D CT scan in the test set with each 3-D CT scan in the training set. At first, we compared each 3-D CT scan in the test set with all 3-D CT scans in the train set which had the same length (in terms of the number of their CT slices). Then, we compared each 3-D CT scan in the test set with all 3-D CT scans in the train set that did not have the same length. In this case we made the comparison over the minimum CT scan length, i.e., over the first 50 slices of each 3-D CT scan. We found that there was no 3-D CT scan in the test set which was identical or almost identical to a 3-D CT scan of the training set.

Finally, Table 1 summarizes the main attributes of COV19-CT-DB that are presented above.

Table 1: Attributes of the COV19-CT-DB

Attribute	Description
total # of CT scans	1,661 COVID 6,095 non-COVID
total # of slices	724,273 from CT scans of COVID 1,775,727 from CT scans of non-COVID
# of slices per CT scan	50 - 700
# Patients	>1150
# Subjects	>2600
slice image resolution	512×512
# Annotators	4 medical experts (2 radiologists & 2 pulmonologists)

4. RACNet: The Proposed Architecture

Let us specifically define the input data characteristics for the COVID-19 diagnosis problem. The input sequence is a 3-D signal, consisting of a series of chest CT slices, i.e., 2-D images, the number of which is varying. The 3-D signal can be handled using a 3-D CNN architecture, such as a 3-D ResNet. However, handling the different input lengths, i.e., the different number of slices that each CT scan contains, can only be tackled in some ad-hoc way; by selecting a fixed input length and removing slices when a larger length is met, or by duplicating slices when the input contains a smaller number of slices.

In this paper, we propose a CNN-RNN architecture, RoutingAlignCluster-Net (RACNet), instead of a 3-D CNN one. RACNet consists of three components: the 3D Analysis component, the Routing one and the Classification one. At first, all input CT scans are padded to have length t (i.e., consist of t slices). For example, let us consider a specific scan series with 50 slices and a total length t of 700. The series will be duplicated 13 times, so as to reach the total input length. Then, there are two modes of feeding input data to our model. In the first, segmentation of each 2-D slice is performed so as to detect the lung regions; then the resulting segmented image constitutes the input to the CNN.

In the second, the whole unsegmented 2-D slices are fed as input to the CNN part. Both modes are studied in the presented experimental study of the paper. In the following each component of RACNet is presented.

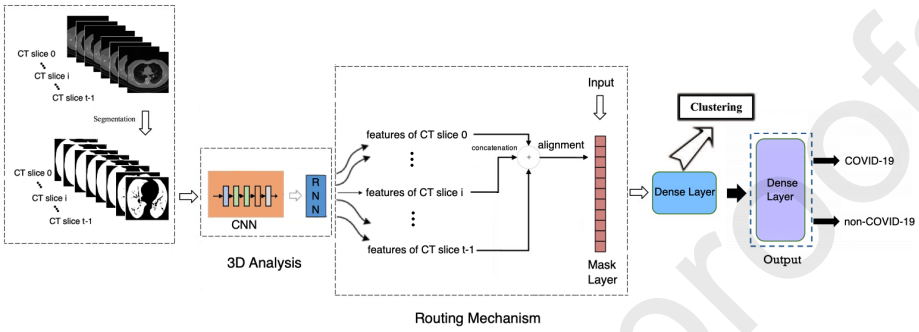


Figure 5: The proposed Pipeline: A 3-D input composed of, up to t chest CT slices is processed for COVID-19 diagnosis; 3-D analysis is performed by a CNN-RNN architecture, while a routing mechanism including an 'alignment' step and a Mask Layer handles the varying input length t. A dense layer follows, preceding the output layer that provides the COVID-19 diagnosis; the neuron outputs of the dense layer are further analyzed through clustering to derive a latent variable model and a related set of anchors that provide further insight into the achieved decision making.

4.1. 3-D Analysis Component

The input data are fed into the 3D Analysis component of RACNet. This component consists of a CNN network followed by an RNN one. The CNN network performs local, per 2-D slice, analysis, extracting features mainly from the lung regions. It should be mentioned that the target is to make diagnosis using the whole 3-D CT scan series, similarly to the way medical experts provide the annotation. The RNN part provides this diagnosis, by analyzing the CNN features of the whole 3-D CT scan, sequentially moving from slice 0 to slice $t-1$. This is depicted in Figure 5, where t denotes the maximum number of slices that appear among all available chest CT scans.

4.2. Routing Component

As shown in Figure 5, we get RNN features corresponding to each CT slice, from 0 to $t-1$. These constitute the input of the Routing Component of RACNet. At first, these features are concatenated and are then fed to the Mask Layer. The original (before padding) length l of the input series is transferred from the input to the Mask Layer to inform the routing process. By including a Mask Layer after the RNN part, RACNet dynamically selects RNN outputs taking into account the input length, i.e., the number of slices of the analyzed CT scan. This is depicted in Figure 5, where the Mask Layer performs a dynamic routing procedure, as is explained next.

During RACNet training, the routing mechanism performs dynamic selection of the RNN outputs/features. In particular, it selects as many of them as denoted by the length l of the input series, keeping their values, while zeroing the values of the rest RNN outputs. In this way, it is routing only the selected ones into the following dense layer. This can be done: a) by selecting the first l RNN outputs, or, b) through an 'alignment' step, i.e., by first placing the l original RNN outputs in equidistant positions in $[0, t-1]$ and by then placing the remaining outputs in the in-between positions; the Mask gets their positions and performs routing of the respective RNN outputs to the Classification module.

The 'alignment' step. As before, let us assume, e.g., that a maximum input length of 700 CT scan slices is considered. For a specific input CT scan consisting of 50 slices, 650 duplicate slices are inserted so that the scan is made to contain 700 slices in total. During training, all 700 slices are fed to the CNN-RNN.

In the case where no 'alignment' is performed, the network's output is fed to the Mask layer which: i) zeroes the features corresponding to the 650 duplicate slices, ii) lets the first 50 features keep their values. This is shown in Fig. 6 (a).

In the case where 'alignment' is performed, the features extracted from the CNN-RNN part are re-positioned as follows. The features corresponding to the 50 original slices are placed in equidistant positions in $[0, 699]$. The rest features corresponding to the 650 duplicate slices are placed in the in-between positions,

following the respective duplicated slice; i.e., all slice 0 duplicates are placed between slice 0 and slice 1 and so on. In Figure 6 (b), between 'features of CT slice 0 and slice 1' we place 13 features corresponding to the duplicates of slice 0; between 'features of CT slice 1 and slice 2' we place 13 features corresponding to the duplicates of slice 1 and so on. After the features corresponding to the last original slice (49), we place the features corresponding to the final 13 duplicates of slice 49. The operation of the Mask Layer is the same as when no 'alignment' is performed; it i) zeroes the features corresponding to the duplicate slices, ii) lets the 50 features (corresponding to original slices) keep their values, iii) it forwards these values to the following Classification module.

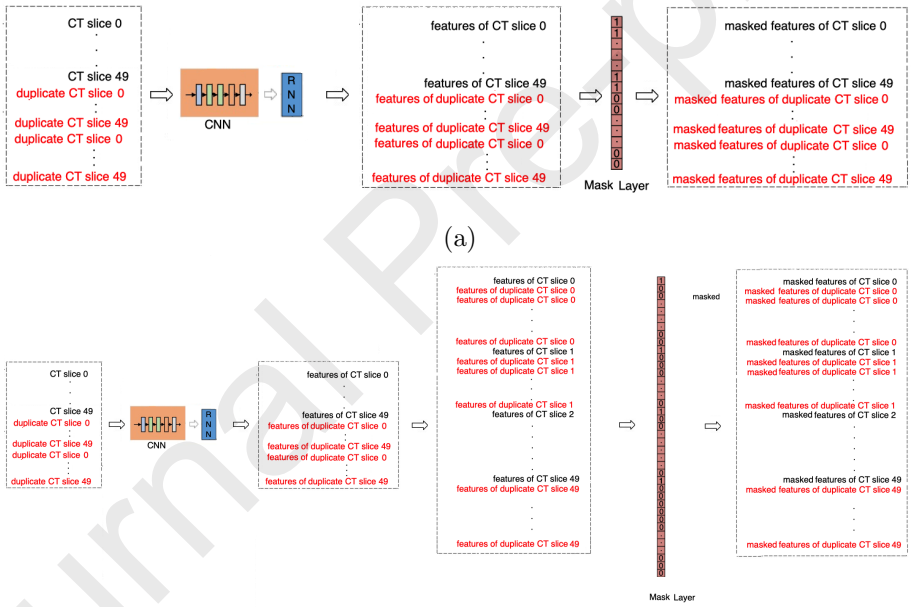


Figure 6: The Routing Mechanism: (a) without and (b) with the 'alignment' step

Let us clarify the significance of the 'alignment' step and the Mask layer via an example, by looking at the weights between the neurons of the dense layer and e.g. the 45th RNN output. When, during training, RACNet's input is a CT scan with length 700, then these weights will be updated with slice

information from the beginning (i.e., initial slices) of the CT scan; when its input is a CT scan with length 90 then these weights will be updated with slice information from the middle slices of the CT scan; when its input is a CT scan with length 50 then these weights will be updated with slice information from the ending slices of the CT scan. This makes the weights' training harder, requiring large amounts of data to work effectively. This problem is handled through the alignment step. In this step, the features extracted from the CNN-RNN part up to length l are re-positioned in equidistant positions in [0, 699] and the rest features corresponding to the duplicate slices are placed in the in-between positions; the former features are kept, whereas the latter are zeroed out by the Mask Layer. Due to the alignment step, each weight is updated during training with CT slice information corresponding to a similar part of the 3-D CT scan, for example, its beginning, middle, or end. This can make training more efficient and improve the performance, especially when a limited number of training data is available.

4.3. Classification Component

As mentioned above, the concatenated RNN outputs are fed to the Classification Component that consists of a dense layer followed by the output layer which uses a softmax activation function and provides the final COVID-19 diagnosis. The dense layer learns to extract high level information from the concatenated RNN outputs. During training, we update only the weights that connect the dense layer neurons with the RNN outputs routed in the concatenated vector by the Mask layer. The remaining weights are updated whenever (i.e., in another input CT scan) respective RNN outputs are selected in the concatenated vector by the Mask Layer. Loss function minimization is performed, as in networks with dynamic routing, by keeping the weights that do not participate in the routing process constant, and ignoring links that correspond to non-routed RNN outputs.

4.4. Latent Variable Analysis and Anchor Set Generation

After RACNet is trained, we extract and further analyze, through clustering, the, say L, neuron outputs of the dense layer of the trained RACNet. These latent variables carry high level, semantic information, which is used to generate the final classification at the output layer. We choose to discard the output layer and perform unsupervised analysis of these variables, so as to generate a representation that can provide further insight into the achieved decision making ability.

Let us assume that we feed the presented architecture with a training dataset. For each 3-D CT scan input k, we extract the L neuron outputs of the dense layer, forming a vector $\mathbf{v}(k)$. In total, we get:

$$\mathcal{V} = \left\{(\mathbf{v}(k) \in \mathbb{R}^L,\ k = 1, \ldots, N\right\} \tag{1}$$

where N is the number of available training data.

We aim to generate a concise representation of the $\mathbf{v}$ vectors, that can be used as a backward model, to trace the most representative CT scan inputs for the performed diagnosis. This is achieved using a clustering algorithm, e.g., k-means++ [22], which generates, say, M clusters $Q = \{\mathbf{q}_1, \ldots, \mathbf{q}_M\}$, where $\mathbf{q}_i \in \mathbb{R}^L$ for $i = 1, \ldots, M$ by minimizing the following criterion:

$$\widehat{Q}_{k\text{-means}} = \arg\min_{Q} \sum_{i=1}^{M} \sum_{\mathbf{v} \in V} \|\mathbf{v} - \mu_i\|^2 \tag{2}$$

where $\mu_i \in \mathbb{R}^L$ is the mean of $\mathbf{v}$ values included in cluster i.

Then, we compute each cluster center $\mathbf{c}(i) \in \mathbb{R}^L$, generating the set C, which constitutes the targeted concise representation:

$$\mathcal{C} = \left\{(\mathbf{c}(i) \in \mathbb{R}^L,\ i = 1, \ldots, M\right\} \tag{3}$$

The CT scan inputs corresponding to the cluster centers can be then examined by medical experts, who can add semantic information related to each one of them.

The generated C set can be used as an anchor set model assisting COVID-19 diagnosis in new subject cases. By testing the trained RACNet model on a new input CT scan case, we will extract the corresponding $\mathbf{v}$ vector of latent variables and will compute the euclidean distance between this vector and each cluster center, i.e., anchor, in C. As a consequence, the new input case is linked to the cluster center with the minimum euclidean distance and is annotated with the label of this center.

The presented latent variable extraction and anchor generation can, therefore, be used to assist COVID-19 diagnosis in a rather efficient way; by computing M distances between L-dimensional vectors and selecting their minimum value. Moreover, by computing the respective cluster radii, we can provide confidence levels in addition to the confidence levels provided by the RACNet output layer.

An additional advantage of this approach, when added to the main RACNet architecture, is that the latter needs retraining, or fine-tuning with new datasets, whenever such datasets become available, e.g., are generated by another hospital, or in another country. Due to privacy purposes, it is highly probable that it is not possible for different medical centers to share their datasets for retraining with all of them. It can, however, be possible for different medical centers to share, or find on github, the best performing networks trained on others' data, as well as the respective anchor derived information. By continual aggregation of older and newer anchor sets, together with the respective trained (RACNet) networks, they can generate common enriched data-driven representations.

5. Experimental Study

This section describes a set of experiments that evaluate the performance of the proposed approach (RACNet) for COVID-19 detection when applied to a variety of databases. At first, we trained RACNet on the COV19-CT-DB database and compared its performance to that of other types of networks. We further performed an ablation study that verified the contribution of each

component of RACNet. Next, we considered five more databases, applied and retrained RACNet on them and compared its performance to the state-of-the-art methods for COVID-19 detection applied to these databases. It is shown that in all cases, RACNet outperformed the state-of-the-art methods. Finally, we performed latent variable extraction from the trained RACNet, deriving a set of anchors and subsequently used them (in addition to RACNet) to derive a unified model over three different databases.

5.1. The Chest 3-D Scan Databases used in the Experimental Study

In the experimental study, we utilized at first the COV19-CT-DB database which has been described in Section 3.

We also utilized the database that was shared in the COV19D Competition of the Workshop "AI-enabled Medical Image Analysis Workshop and Covid-19 Diagnosis Competition" held in conjunction with the International Conference on Computer Vision (ICCV) 2021 [23, 24, 25, 26]. The database, annotated with respect to COVID-19/non-COVID-19 diagnosis, includes a total of 1405 COVID-19 and 4066 non-COVID-19 3-D CT scans; its training set includes 1552 3-D CT scans (707 COVID-19 and 845 non-COVID-19 ones) and its validation set contains 374 3-D CT scans (165 COVID-19 and 209 non-COVID-19 ones). Thereafter we refer to this database as COV19D-ICCV2021.

Furthermore, we utilized the database that was shared in the COVID-19 Detection Challenge of the 2nd COV19D Competition of the the Workshop "AI-enabled Medical Image Analysis – Digital Pathology & Radiology/COVID19" held in conjunction with the European Conference on Computer Vision (ECCV) 2022 [27, 28, 23, 24, 25, 26]. The database is annotated with respect to COVID-19/non-COVID-19 diagnosis and includes a total of 1550 COVID-19 and 5044 non-COVID-19 3-D CT scans; its training set containeS 1992 3-D CT scans (882 COVID-19 and 1110 non-COVID-19 ones) and its validation set contains 504 3-D CT scans (215 COVID-19 and 289 non-COVID-19 ones). Thereafter we refer to this database as COV19D-ECCV2022.

We also utilized the CC-CCII [10] database that includes 3-D CT scans of

3 classes: novel coronavirus pneumonia (NCP), common pneumonia (CP), and Normal. The original CC-CCII database contained a total number of 617,775 slices of 6,752 CT scans obtained from 4,154 patients. However, there were some problems with it (i.e., damaged data, non-unified data type, repeated and noisy slices, disordered slices, and non-segmented slices). The authors of [16] published training and test partitions that do not include damaged data, naming this version of CC-CCII as 'Clean CC-CCII'. The training partition consists of 3,195 3-D CT scans (1,213 NCP, 1,210 CP and 772 Normal) and the test one consists of 798 3-D CT scans (302 NCP, 303 CP and 193 Normal). We utilized this version in our experimental study. We refer to this database as CC-CCII.

Another database that we utilized was the MosMedData database [11], annotated with respect to COVID-19/non-COVID-19 diagnosis. It contains 1,110 3-D CT-scans, with the two classes being rather imbalanced. The COVID class consists of 856 CT-scans, whereas the Normal class consists of 254. The training set consists of 601 COVID-19 and 178 non-COVID-19 CT scans and the testing set consists of 256 COVID-19 and 76 non-COVID-19. We refer to this database as MosMedData.

Finally we utilized the CT-image Database [29] annotated with respect to COVID-19/non-COVID-19 diagnosis. This dataset does not include 3-D CT-scans, but 408 non-COVID-19 and 349 COVID-19 CT-scan slices, obtained from different patients. The training set consists of 279 COVID-19 and 326 non-COVID-19 CT-scan slices and the training set consists of 70 COVID-19 and 82 non-COVID-19 CT-scan slices. To create respective 3-D CT-scans, we augmented each slice using random rotation with an angle ranging from -20 to 20 degrees, random noise, and horizontal flip. Following this procedure, we constructed 3-D CT-scans from the 2-D slices, which we utilized in our experiments. We refer to this database as CT-image DB.

5.2. RACNet Training: Implementation Details

Regarding implementation of RACNet: i) we used EfficientNetB0 as CNN model, stacking a global average pooling layer on top, a batch normalization

layer and dropout (with keep probability 0.8); ii) we used a single one-directional GRU layer consisting of 128 units as RNN model; iii) the first dense layer consisted of 128 hidden units.

Regarding implementation details of RACNet training, batch size was equal to 5 (i.e, at each iteration our model processed 5 CT scans) and the input length 't' (see Figure 5) was 700 (the maximum number of slices found across all CT scans). Our model was fed with 3-D CT scans composed of CT slices; each slice was resized from its original size of 512×512 to 256×256. Loss function was cross entropy. Adam optimizer was used with learning rate 10^{-4}. In k-means clustering, we tried values of $k \in \{2, .., 25\}$, whilst evaluating the performance of our model. Training was performed on a Tesla V100 32GB GPU.

5.3. Experiments & Ablation Study on COV19-CT-DB

In this subsection we compare the performance of RACNet to that of 3-D CNN and CNN-RNN models when trained and tested on the COV19-CT-DB.

According to [30] 3-D CNNs require a large number of labeled image sequences to learn the 3-D kernels. Therefore, we used a pre-trained for action recognition 3-D ResNet-50 [31] and further trained it on COV19-CT-DB. We also considered MedicalNet [32], a 3-D ResNet-34 network trained for pulmonary nodule classification. MedicalNet was trained with CT scans (i.e., the same input considered in this paper). We used it as a pre-trained network and further trained it on COV19-CT-DB.

We also utilized a conventional CNN-RNN that provided a probability for each CT scan slice at its output and was followed by a voting scheme that made the final decision; the voting scheme was either a majority voting, or an at-least one voting (i.e., if at least one slice in the scan was predicted as COVID-19, then the whole CT scan was diagnosed with COVID-19, otherwise it was diagnosed with non-COVID-19); the CNN and RNN parts of this network were the same as the respective parts of RACNet, i.e., EfficientNetB0 and single-directional GRU (we also considered a dense layer between the RNN and the output layer, but the achieved performance was worse).

Table 2 shows a comparison of the performance of RACNet to the performance of these networks over the non-segmented COV19-CT-DB database. It can be seen that RACNet outperformed all of them, in terms of both accuracy and F1 score, for both COVID-19 and non COVID-19 categories. Particularly, RACNet -although a lighter model- outperformed the 3-D ResNet-50 that had been pre-trained on a large action database. It also outperformed MedicalNet that had been pre-trained, not on action recognition (which is another task), but on a task similar to ours. This shows that the proposed model structure, although lighter than the 3-D ones (around 4.4M parameters versus about 63M and 46M respectively) and although not pre-trained on another task, achieved the best performance.

In addition, RACNet outperformed the CNN-RNN network; the main downside of that model is that there existed only one label for the whole CT scan and there were no labels for each CT scan slice. In contrast to this, RACNet analyzed the whole CT scan, based on information extracted from each slice, as was described in the former Section.

Table 2: Performance comparison between RACNet and other state-of-the-art networks on the test set of COV19-CT-DB database (non-segmented data)

Method	Accuracy		F1
	COVID	non-COVID	
3D ResNet-50 [31]	0.74	0.80	0.82
MedicalNet [32]	0.78	0.83	0.86
EfficientNet-GRU	0.73	0.80	0.82
RACNet	**0.82**	**0.86**	**0.90**

Ablation Study on the effect of the routing mechanism: We compared the performance of RACNet when the mask was used (and thus, only the RNN outputs corresponding to true input slices were routed to the dense and output layers) and when the mask was not used (and thus, inserted duplicates of input slices were also fed to the dense layers). The last two rows of Table 3 illustrate the

efficacy of the proposed mask, compared to the case of no masking. The mask effectively filtered out unnecessary and repeating information.

We could also start with a zeroed 3-D scan input; then put the existing CT slices in equidistant positions and use slice duplication, so as to increase the varying input length reaching a constant maximum value. We have performed experiments with this configuration without masking the outputs. The obtained performance was about 8% lower to that shown in Table 3, providing an F1 score of 0.82.

Ablation Study on the effect of the 'alignment' step: We compared the performance when the 'alignment' step was included, or not. In rows 6 and 8 (i.e., the last row) of Table 3 one can see that the 'alignment' case provided the best results. This result was expected, since the 'alignment' step managed to better align the features of the important slices in each CT scan series.

Ablation Study on the effect of various CNNs: Table 3 shows that when EfficientNetB0 was used as the CNN part of RACNet (shown on last row of the Table), it outperformed the cases when other state-of-the-art CNNs were used, in particular ResNet-50 and DesneNet-121 (shown on rows 4 and 5 of the Table). We also studied the effect of using 3-D convolutions instead of the CNN-RNN part in RACNet. Table 3 shows that using the CNN-RNN provided a better performance, with the model being also lighter (this is indicated on rows 3 and 8 -the last one- of the Table). Finally, we studied the effect of using different numbers of hidden units in the first dense layer of RACNet. Table 3 illustrates that 128 units provided the best performance compared to the cases when 16 or 64 units were utilized (shown on rows 1, 2 and 8 of the Table).

5.4. Experiments on COV19D ICCV & ECCV Competitions

Until this point, the presented results refer to the case where no segmentation has been performed on the 3-D CT scan inputs; we did that as we did not want the specific selection of the lung segmentation method to affect the presented analysis and obtained results. From here and thereafter, we examined the obtained performance of RACNet -for COVID-19 detection- when its

Table 3: Ablation Study: Performance comparison on the test set of COV19-CT-DB database (non-segmented data)

RACNet	Accuracy		F1
	COVID	non-COVID	
64 units in dense layer	0.79	0.83	0.86
16 units in dense layer	0.78	0.82	0.85
3-D conv instead of CNN-RNN	0.79	0.84	0.87
ResNet-50 as CNN	0.80	0.85	0.88
DenseNet-121 as CNN	0.79	0.84	0.87
without 'alignment'	0.80	0.85	0.88
without mask	0.78	0.84	0.87
whole (EfficientNetB0, GRU, 128 units in dense layer)	**0.82**	**0.86**	**0.90**

input was the segmented CT scans and compared its performance to the best performing methods in the ICCV 2021 and ECCV 2022 Competitions, on the COV19D-ICCV2021 and COV19D-ECCV2022 databases, respectively.

In Table 4, the performance of RACNet is compared to that of the best performing method on COV19D-ICCV2021 database (winner of the COV19D ICCV 2021 Competition), FDVTS_COVID, and to the performance of the methods that ranked in the second and third places, SenticLab.UAIC and ACVLab, respectively. It can be observed that RACNet outperformed them by 3.4%-5.09% in terms of the macro F1 score, which was the criterion used in the Competition. Moreover, it can be observed that RACNet has produced a much larger increase in diagnosis of COVID-19 cases, which is very important; it has outperformed the state-of-the-art by more than 10%.

Table 4 also illustrates a comparison of the performance of RACNet to the performance of the best performing methods on COV19D-ECCV2022 database (winners of the 2nd COV19D ECCV 2022 Competition), ACVLab and FDVTS, and to the performance of the method that ranked in the second place, MDAP. It can be observed that RACNet outperformed them by 5.95%-7.19% in terms of the macro F1 score, which was the criterion used in the Competition. More-

over, it can be observed that RACNet has produced a much larger increase in diagnosis of COVID-19 cases, which is very important; it has outperformed the state-of-the-art by more than 13%.

Table 4: Performance comparison between RACNet and the state-of-the-art on the test set of COV19D-ICCV2021 and COV19D-ECCV2022 databases of the respective ICCV and ECCV Competitions; F1 Score presented in %

Databases	Methods	F1		
		Macro	COVID	non-COVID
COV19D-ICCV2021	ACVLab [33]	88.74	80.63	96.84
	SenticLab.UAIC [34]	90.06	82.96	97.17
	FDVTS_COVID [35]	90.43	83.60	97.27
	RACNet	**93.83**	**93.62**	**94.04**
COV19D-ECCV2022	MDAP [36]	87.87	78.80	96.95
	FDVTS [37]	89.11	80.92	97.31
	ACVLab [38]	89.11	80.78	97.45
	RACNet	**95.06**	**94.18**	**95.95**

5.5. Experiments on CC-CCII, MosMedData and CT-image Databases

In the following, we examine the obtained performance of RACNet on three databases (CC-CCII, MosMedData and CT-image Database) and compare its performance to the state-of-the-art.

Table 5 illustrates the performance of RACNet trained for 3-class classification (to distinguish novel coronavirus pneumonia from common pneumonia and Normal) on the CC-CCII database according to various metrics (accuracy, precision, sensitivity, F1 score); Table 5 also illustrates a performance comparison of RACNet and the state-of-the-art, which have been reported in Section 2.

RACNet outperformed by a large margin all methods on all metrics. In particular, RACNet outperformed the best performing method, EMARS-APS by 3.28% in terms of F1 score and 4.33% in terms of accuracy; it also outperformed the other state-of-the-art methods by 4.74%-9.11% in terms of F1 score

and 5.25%-7.7% in terms of accuracy. In this case, RACNet was pre-trained on COV19-CT-DB and then fine-tuned on the CC-CCCII's training set. We exploited COV19-CT-DB for learning features, which were then used as priors for achieving best performance. When RACNet was not pre-trained (but trained from scratch) it still outperformed EMARS-APS by 1.28% in terms of F1 score and 2.39% in terms of accuracy. It also outperformed the other state-of-the-art by 2.76%-7.11% in terms of F1 score and 3.31%-5.84% in terms of accuracy.

Table 5 also shows a comparison in terms of model size (in MB) between RACNet and the state-of-the-art. It can be observed that RACNet is a way lighter model than all state-of-the-art but EMARS-APS; the latter, however, is a neural architecture search method which constructs a different model for each analyzed dataset, as can be seen comparing the obtained sizes in the cases of CC-CCII and MosMedData databases (presented in the following). Let us also report that the computational complexity of our model is about 112 BN FLOPs and the number of parameters about 4.4 M.

Table 5 further shows the performance of RACNet trained for COVID-19 vs non-COVID-19 diagnosis on the MosMedData database according to various metrics and compares its performance to the state-of-the-art, explained in Section 2. RACNet outperformed all methods on all metrics, although RACNet is a lighter in size model than all of them as can be verified in Table 5. In particular, the pre-trained RACNet on COV19-CT-DB outperformed the best performing method, EMARS-APS by 1.73% in terms of F1 score and 1.79% in terms of accuracy; it also outperformed the other state-of-the-art methods by 5.65%-5.84% in terms of F1 score and 7.58%-9.83% in terms of accuracy. When RACNet was not pre-trained (but trained from scratch) it still outperformed EMARS-APS by 1.02% in terms of F1 score and 1.05% in terms of accuracy. It also outperformed the other state-of-the-art by 4.94%-5.13% in terms of F1 score and 6.84%-9.09% in terms of accuracy.

Finally, Table 5 shows the performance of RACNet trained for COVID-19 vs non-COVID-19 diagnosis on the CT-image database and compares its performance to the state-of-the-art, explained in Section 2. RACNet outperformed

VGG19 and ResNet50 by 10.47% and 19.95%, respectively, in terms of accuracy, although it is a lighter model in size by at least 10 times order of magnitude.

To sum up, through this extensive comparison analysis we demonstrate the superior performance and efficiency of RACNet in terms of both various classification metrics and model size. RACNet is a computationally efficient approach.

Table 5: Performance comparison between RACNet and the state-of-the-art on the test set of CC-CCII, MosMedData, and CT-image Database; Acc stands for Accuracy metric

Dataset	Method	Size (MB)	Acc	Precision	Sensitivity	F1
CC-CCII	MC3_18 [18]	43.84	86.16	87.11	82.78	84.89
	Densenet3D121 [39]	43.06	87.02	88.97	82.78	85.76
	COVID-AL [15]	-	86.60	-	-	-
	VGG-Ensemble [21]	-	88.12	84.04	89.19	86.54
	MNas3DNet [14]	22.91	87.14	84.44	86.09	87.25
	CovidNet3D [16]	53.26	88.69	90.48	88.08	89.26
	EMARS-APS [17]	3.38	89.61	91.48	89.97	90.72
	RACNet	8.60	**93.94**	**93.69**	**94.30**	**94.00**
MosMedData	MC3_18 [18]	43.84	80.04	79.43	98.43	87.92
	Densenet3D121 [39]	43.06	79.55	84.23	92.16	88.01
	DeCoVNet [2]	-	82.43	-	-	-
	CovidNet3D [16]	60.39	82.29	79.50	98.82	88.11
	EMARS-APS [17]	10.69	88.09	93.52	90.59	92.03
	RACNet	8.60	**89.87**	**94.69**	**92.85**	**93.76**
CT-image DB	ResNet50 [40]	98.0	76.32	-	-	-
	VGG19 [40]	549.0	84.80	-	-	-
	RACNet	8.60	**95.27**	**93.15**	**97.14**	**95.10**

5.6. Anchor Set Creation

In the following we implemented the procedure of latent variable extraction and anchor set generation when training RACNet with the COV19-CT-DB database. This resulted in a set of 11 anchors, each represented by a vector in

the 128-dimensional space. 7 of them corresponded to COVID-19 cases, with the rest corresponding to non COVID-19 cases.

Table 6 provides the number of CT scans, belonging to RACNet training data, assigned to every generated cluster and their COVID-19, or non-COVID-19 category. It also provides a ranking of the severity of COVID-19, as classified by our medical experts, in the range from 1 to 4, with 4 denoting the critical status. Table 7 describes each of these four categories [11]. The centers of the above 11 clusters formed the anchor set on COV19-CT-DB.

Table 6: Number of CT Scans per cluster, cluster category & Severity category in COV19-CT-DB

Cluster ID	Number of CT Scans	Category	Severity Category
0	231	COVID-19	3
1	360	COVID-19	2
2	344	COVID-19	4
3	106	COVID-19	1
4	195	COVID-19	4
5	156	COVID-19	3
6	242	COVID-19	4
7	107	non COVID-19	1
8	586	non COVID-19	1
9	557	non COVID-19	1
10	322	non COVID-19	1

Table 7: Description of the Severity Categories

Category	Severity	Description
1	Mild	Few or no Ground glass opacities. Pulmonary parenchymal involvement $\leq 25\%$ or absence
2	Moderate	Ground glass opacities. Pulmonary parenchymal involvement $25-50\%$
3	Severe	Ground glass opacities. Pulmonary parenchymal involvement $50-75\%$
4	Critical	Ground glass opacities. Pulmonary parenchymal involvement $\geq 75\%$

For validation, we used this anchor set to classify the COV19-CT-DB test set. In particular, we fed each 3-D CT scan in the test set of the RACNet architecture; we extracted the corresponding dense layer neuron outputs; we computed their euclidean distance from each anchor. Then they were classified according to the label of their nearest cluster center. The obtained classification performance over the test dataset was similar to the original RACNet's classification performance.

Table 8: Description of findings in each cluster center in COV19-CT-DB

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; mild category
9	Normal CT scan; mild category
10	Normal CT scan; mild category

Moreover, our medical experts examined the 3-D scan inputs corresponding to the 11 cluster centres and produced justification for the respective diagnosis. Table 8 presents the findings detected in each cluster center.

Some examples of CT slices from the cluster centers are given below. Figure 7 shows 10 consecutive slices from COVID-19 cluster center 0. Medical experts have annotated it as 'bilateral ground glass regions that appear, especially in lower lung lobes'. Figure 8 shows 10 slices from COVID-19 cluster center 2. According to medical experts' annotation, this is consistent with 'COVID-19 pneumonia bilateral thickening filtrates'. Figure 9, on the contrary, shows 10 slices from non COVID-19 cluster center 9.

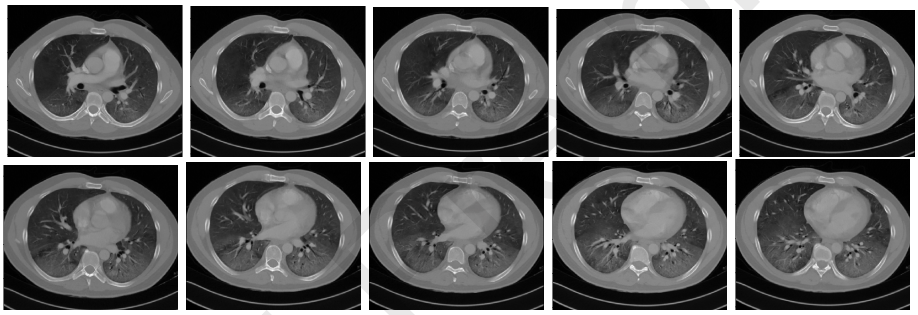


Figure 7: Slices from cluster center 0 of COVID-19 category in COV19-CT-DB. Bilateral ground glass regions are seen especially in lower lung lobes.

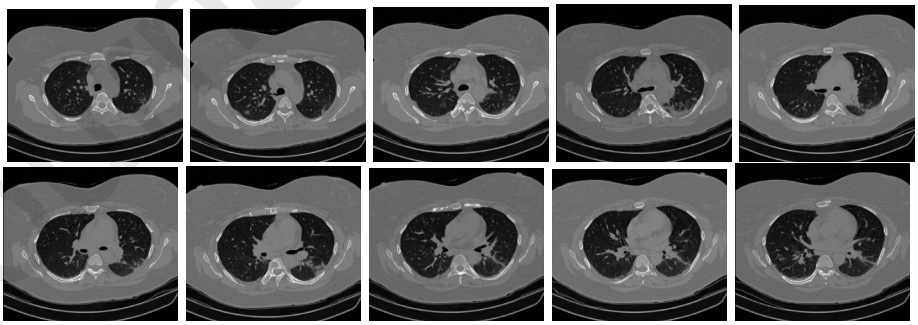


Figure 8: Slices from COVID-19 cluster center 2 in COV19-CT-DB, which is consistent with COVID-19 pneumonia bilateral thickening filtrates.

The major advantage of the anchor set model is the insight that it introduces

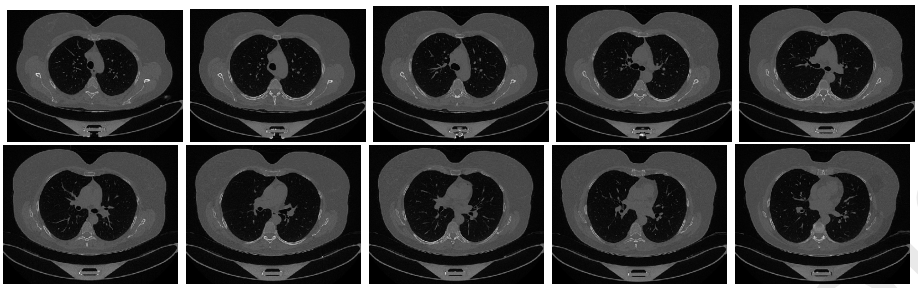


Figure 9: Slices from non COVID-19 cluster center 9 in COV19-CT-DB

in the diagnosis process. In each new test case, the generated decision is accompanied by the information about the anchor to which this case was assigned through the above nearest neighbor classification procedure. As a result, the patient, or the doctor, can see which part of RACNet data-driven knowledge was used to make the specific diagnosis.

5.7. Anchor Set Unification across Databases

In the following we used the RACNet trained on COV19-CT-DB and the set of 11 anchors for unification with similar data-driven knowledge generated from other databases, i.e., the CC-CCII and CT-image Database. We developed an efficient unification procedure based on the generated anchor set, which, on the one hand, alleviates the problem of 'catastrophic forgetting' when transfer learning is used and, on the other hand, reduces the high computational cost needed for retraining the deep learning model.

Utilization of CC-CCII database. At first, we computed the 128-dimensional features for each CT scan of the CC-CCII database using the RACNet's model which had been trained with the COV19-CT-DB training set. Then, these 128-dimensional features formed the input to train a neural network, say $\mathrm{NN}^{(1)}$, consisting of 3 fully connected layers, so as to predict the Covid/non-COVID status of the CC-CCII data. The three layers included 64, 128 and 2 (output) neurons respectively. In a similar way, as we extracted the 11 clusters from RACNet in 4.4, we extracted a set of representations from the layer with 128

neurons; then, through clustering, we generated another set of cluster centres. In this case, the number of cluster centers that produced the best performance over the CC-CCII test partition set was 13. Figure 10, shows 10 slices from one of the extracted COVID-19 cluster centers.

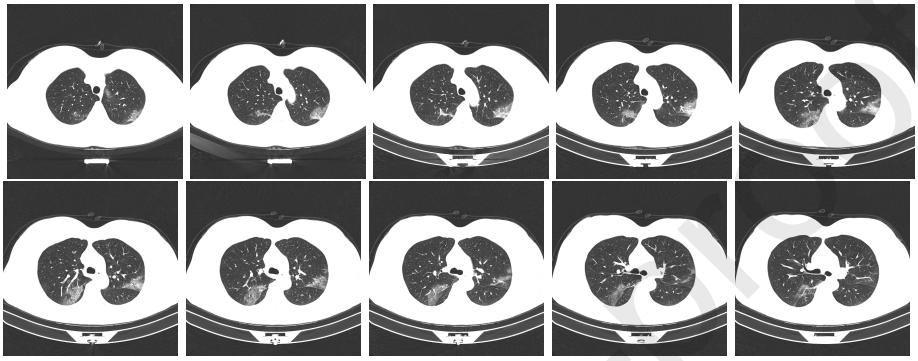


Figure 10: Slices of a new COVID-19 anchor in CC-CCII database, with ground glass regions in the lungs.

As a result, we had created: a) a set of 11 clusters & respective cluster centers, using COV19-CT-DB and RACNet, b) a set of 13 clusters & respective cluster centers, using CC-CCII, RACNet and $NN^{(1)}$.

In the following, we generated a unified prediction model, by: a) merging the 11 cluster centers from COV19-CT-DB and the 13 cluster centers from CC-CCII, b) using the RACNet-$NN^{(1)}$ as the combined test model. In particular, we could classify any CT scan in the COV19-CT-DB and CC-CCII databases, by passing it through RACNet-$NN^{(1)}$ and computing which one out of the 24 cluster centers was nearest to the extracted representation in the 128-dimensional space. The obtained performance was almost identical to the one obtained when processing each database independently. This result was achieved without exchanging any data between the holders of the two databases. It was only assumed that the RACNet and $NN^{(1)}$ networks and the cluster center representations in the 128-dimensional space were made available to each other.

Further Utilization of CT-image Database. To further illustrate the effectiveness of this procedure, we extended the unified representation so as to include the CT-image Database [29]. In a similar way, we created a set of 5 more clusters and merged them with the 24 (11 and 13) already derived cluster centres over COV19-CT-DB and CC-CCII, together with RACNet, $NN^{(1)}$ and $NN^{(2)}$ networks, where $NN^{(2)}$ had the same structure as $NN^{(1)}$. The resulting representation was able to provide a performance over all databases that was similar to the one obtained when processing each one of them independently. Figure 11 shows the derived 5 cluster centers from the CT-image Database, two of which belong to the COVID-19 category and three to the non-COVID-19 category.

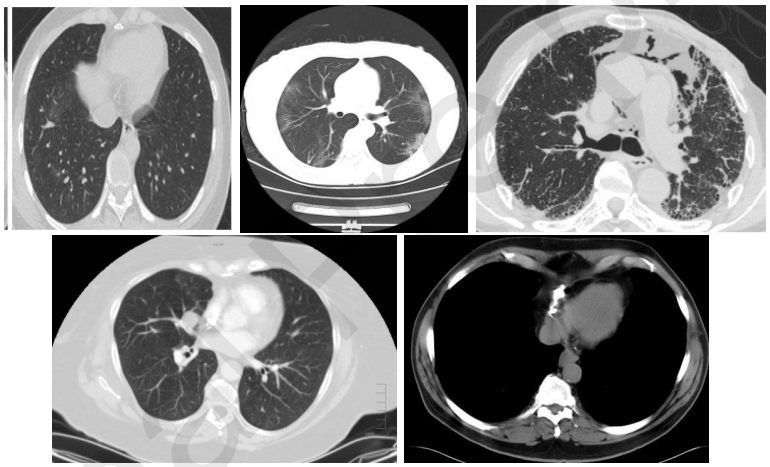


Figure 11: The 5 derived cluster centers in the CT-image Database; the 3 top ones correspond to the non-COVID-19 category and the 2 bottom ones correspond to the COVID-19 category

6. Conclusions and Future work

In this paper we presented COV19-CT-DB, a large database with Chest 3-D CT scans, annotated for COVID-19 diagnosis. We also presented RACNet, a new method which: a) harmonizes analysis of 3-D image volumes consisting of different number of slices and annotated per volume, b) unifies decisions made over different datasets, thus enriching data-driven knowledge and improving its

trusted use. RACNet was successfully used for obtaining high performance in COVID-19 diagnosis based on chest 3-D CT scans, over six different datasets, whilst permitting continual learning and avoiding catastrophic forgetting.

Future work includes extension of the RACNet model and of the presented approach, so as to include uncertainty estimation and domain adaptation to a large variety of related applications, referenced in the Introduction of the paper.

Acknowledgment

We would like to thank GRNET, National Infrastructures for Research & Technology, for supporting us through Project "Machaon - Advanced networking and computational services to hospital units in public cloud environment".

References

[1] A. Amyar, R. Modzelewski, H. Li, S. Ruan, Multi-task deep learning based ct imaging analysis for covid-19 pneumonia: Classification and segmentation, Computers in Biology and Medicine 126 (2020) 104037.

[2] X. Wang, X. Deng, Q. Fu, Q. Zhou, J. Feng, H. Ma, W. Liu, C. Zheng, A weakly-supervised framework for covid-19 classification and lesion localization from chest ct, IEEE transactions on medical imaging 39 (8) (2020) 2615–2625.

[3] L. Zhang, M. Wang, M. Liu, D. Zhang, A survey on deep learning for neuroimaging-based brain disorder analysis, Frontiers in neuroscience (2020) 779.

[4] A. Tagaris, D. Kollias, A. Stafylopatis, G. Tagaris, S. Kollias, Machine learning for neurodegenerative disorder diagnosis—survey of practices and launch of benchmark dataset, International Journal on Artificial Intelligence Tools 27 (03) (2018) 1850011.

[5] F. Gu, M.-H. Chung, M. Chignell, S. Valaee, B. Zhou, X. Liu, A survey on deep learning for human activity recognition, ACM Computing Surveys (CSUR) 54 (8) (2021) 1–34.

[6] T. Zhang, Z. Tan, Deep emotion recognition using facial, speech and textual cues: A survey.

[7] F. Caliva, F. S. De Ribeiro, A. Mylonakis, C. Demazi'ere, P. Vinai, G. Leontidis, S. Kollias, A deep learning approach to anomaly detection in nuclear reactors, in: 2018 International joint conference on neural networks (IJCNN), IEEE, 2018, pp. 1–8.

[8] J. Wang, C. Lan, C. Liu, Y. Ouyang, W. Zeng, T. Qin, Generalizing to unseen domains: A survey on domain generalization, arXiv preprint arXiv:2103.03097.

[9] S. Wang, B. Kang, J. Ma, X. Zeng, M. Xiao, J. Guo, M. Cai, J. Yang, Y. Li, X. Meng, et al., A deep learning algorithm using ct images to screen for corona virus disease (covid-19), European radiology (2021) 1–9.

[10] K. Zhang, X. Liu, J. Shen, Z. Li, Y. Sang, X. Wu, Y. Zha, W. Liang, C. Wang, K. Wang, L. Ye, M. Gao, Z. Zhou, L. Li, J. Wang, Z. Yang, H. Cai, J. Xu, L. Yang, W. Cai, W. Xu, S. Wu, W. Zhang, S. Jiang, L. Zheng, X. Zhang, L. Wang, L. Lu, J. Li, H. Yin, W. Wang, O. Li, C. Zhang, L. Liang, T. Wu, R. Deng, K. Wei, Y. Zhou, T. Chen, J. Yiu-Nam Lau, M. Fok, J. He, T. Lin, W. Li, G. Wang, Clinically applicable AI system for accurate diagnosis, quantitative measurements, and prognosis of COVID-19 pneumonia using computed tomography, Cell 182 (5) (2020) 1360.

[11] S. P. Morozov, A. E. Andreychenko, I. A. Blokhin, P. B. Gelezhe, A. P. Gonchar, A. E. Nikolaev, N. A. Pavlov, V. Y. Chernina, V. A. Gombolevskiy, Mosmeddata: data set of 1110 chest ct scans performed during the covid-19 epidemic, Digital Diagnostics 1 (1) (2020) 49–59.

[12] X. Yang, X. He, J. Zhao, Y. Zhang, S. Zhang, P. Xie, Covid-ct-dataset: a ct image dataset about covid-19, arXiv preprint arXiv:2003.13865.

[13] A. Jaiswal, N. Gianchandani, D. Singh, V. Kumar, M. Kaur, Classification of the covid-19 infected patients using densenet201 based deep transfer learning, Journal of Biomolecular Structure and Dynamics (2020) 1–8.

[14] X. He, S. Wang, S. Shi, X. Chu, J. Tang, X. Liu, C. Yan, J. Zhang, G. Ding, Benchmarking deep learning models and automated model design for covid-19 detection with chest ct scans, medRxivarXiv:https://www.medrxiv.org/content/early/2021/11/04/2020.06.08.20125963.full.pdf, doi:10.1101/2020.06.08.20125963.
URL https://www.medrxiv.org/content/early/2021/11/04/2020.06.08.20125963

[15] X. Wu, C. Chen, M. Zhong, J. Wang, J. Shi, COVID-AL: The diagnosis of COVID-19 with deep active learning, Med Image Anal 68 (2020) 101913.

[16] X. He, S. Wang, X. Chu, S. Shi, J. Tang, X. Liu, C. Yan, J. Zhang, G. Ding, Automated model design and benchmarking of deep learning models for covid-19 detection with chest ct scans, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 35, 2021, pp. 4821–4829.

[17] X. He, G. Ying, J. Zhang, X. Chu, Evolutionary multi-objective architecture search framework: Application to covid-19 3d ct classification, in: L. Wang, Q. Dou, P. T. Fletcher, S. Speidel, S. Li (Eds.), Medical Image Computing and Computer Assisted Intervention – MICCAI 2022, Springer Nature Switzerland, Cham, 2022, pp. 560–570.

[18] D. Tran, H. Wang, L. Torresani, J. Ray, Y. LeCun, M. Paluri, A closer look at spatiotemporal convolutions for action recognition (2017). doi:10.48550/ARXIV.1711.11248.
URL https://arxiv.org/abs/1711.11248

[19] A. Khadidos, A. O. Khadidos, S. Kannan, Y. Natarajan, S. N. Mohanty, G. Tsaramirsis, Analysis of covid-19 infections on a ct image using deepsense model, Frontiers in Public Health 8.

[20] S. A. Ali Ahmed, M. C. Yavuz, M. U. Şen, F. Gülşen, O. Tutar, B. Korkmazer, C. Samancı, S. Şirolu, R. Hamid, A. E. Eryürekli, T. Mammadov, B. Yanikoglu, Comparison and ensemble of 2d and 3d approaches for covid-19 detection in ct images, Neurocomputing 488 (2022) 457–469. doi:https://doi.org/10.1016/j.neucom.2022.02.018.
URL https://www.sciencedirect.com/science/article/pii/S092523122200162X

[21] X. Li, W. Tan, P. Liu, Q. Zhou, J. Yang, Classification of covid-19 chest ct images based on ensemble deep learning, Journal of Healthcare Engineering 2021 (2021) 5528441. doi:10.1155/2021/5528441.
URL https://doi.org/10.1155/2021/5528441

[22] D. Arthur, S. Vassilvitskii, k-means++: The advantages of careful seeding, Tech. rep., Stanford (2006).

[23] D. Kollias, A. Arsenos, L. Soukissian, S. Kollias, Mia-cov19d: Covid-19 detection through 3-d chest ct image analysis, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 537–544.

[24] D. Kollias, N. Bouas, Y. Vlaxos, V. Brillakis, M. Seferis, I. Kollia, L. Sukissian, J. Wingate, S. Kollias, Deep transparent prediction through latent representation analysis, arXiv preprint arXiv:2009.07044.

[25] D. Kollias, Y. Vlaxos, M. Seferis, I. Kollia, L. Sukissian, J. Wingate, S. D. Kollias, Transparent adaptation in deep medical image diagnosis., in: TAILOR, 2020, pp. 251–267.

[26] D. Kollias, A. Tagaris, A. Stafylopatis, S. Kollias, G. Tagaris, Deep neural

architectures for prediction in healthcare, Complex & Intelligent Systems 4 (2) (2018) 119–131.

[27] D. Kollias, A. Arsenos, S. 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, Springer, 2023, pp. 677–690.

[28] A. Arsenos, D. Kollias, S. 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), IEEE, 2022, p. 1–5.

[29] J. Zhao, Y. Zhang, X. He, P. Xie, Covid-ct-dataset: a ct scan dataset about covid-19, arXiv preprint arXiv:2003.13865.

[30] K. Hara, H. Kataoka, Y. Satoh, Can spatiotemporal 3d cnns retrace the history of 2d cnns and imagenet?, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 6546–6555.

[31] Z. Han, B. Wei, Y. Hong, T. Li, J. Cong, X. Zhu, H. Wei, W. Zhang, Accurate screening of covid-19 using attention-based deep 3d multiple instance learning, IEEE transactions on medical imaging 39 (8) (2020) 2584–2594.

[32] S. Chen, K. Ma, Y. Zheng, Med3d: Transfer learning for 3d medical image analysis, arXiv preprint arXiv:1904.00625.

[33] C.-C. Hsu, G.-L. Chen, M.-H. Wu, Visual transformer with statistical test for covid-19 classification (2021). doi:10.48550/ARXIV.2107.05334.
URL https://arxiv.org/abs/2107.05334

[34] R. Miron, C. Moisii, S. Dinu, M. Breaban, Covid detection in chest cts: Improving the baseline on cov19-ct-db (2021). doi:10.48550/ARXIV.2107.04808.
URL https://arxiv.org/abs/2107.04808

[35] J. Hou, J. Xu, R. Feng, Y. Zhang, F. Shan, W. Shi, Cmc-cov19d: Contrastive mixup classification for covid-19 diagnosis, in: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops, 2021, pp. 454–461.

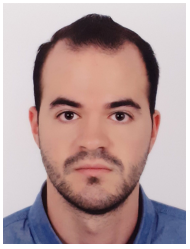
[36] R. Turnbull, Cov3d: Detection of the presence and severity of COVID-19 from CT scans using 3D ResNets [Preliminary Preprint] doi:10.26188/20226087.v1.
URL https://melbourne.figshare.com/articles/preprint/Cov3d_Detection_of_the_presence_and_severity_of_COVID-19_from_CT_scans_using_3D_ResNets_Preliminary_Preprint_/20226087

[37] J. Hou, J. Xu, R. Feng, Y. Zhang, Fdvts's solution for 2nd cov19d competition on covid-19 detection and severity analysis (2022). doi:10.48550/ARXIV.2207.01758.
URL https://arxiv.org/abs/2207.01758

[38] C.-C. Hsu, C.-H. Tsai, G.-L. Chen, S.-D. Ma, S.-C. Tai, Spatiotemporal feature learning based on two-step lstm and transformer for ct scans (2022). doi:10.48550/ARXIV.2207.01579.
URL https://arxiv.org/abs/2207.01579

[39] A. Diba, M. Fayyaz, V. Sharma, A. H. Karami, M. M. Arzani, R. Yousefzadeh, L. Van Gool, Temporal 3d convnets: New architecture and transfer learning for video classification (2017). doi:10.48550/ARXIV.1711.08200.
URL https://arxiv.org/abs/1711.08200

[40] W. Zouch, D. Sagga, A. Echtioui, R. Khemakhem, M. Ghorbel, C. Mhiri, A. B. Hamida, Detection of covid-19 from ct and chest x-ray images using deep learning models, Annals of Biomedical Engineering 50 (7) (2022) 825–835. doi:10.1007/s10439-022-02958-5.
URL https://doi.org/10.1007/s10439-022-02958-5



Dimitrios Kollias , Fellow of the Higher Education Academy and member of the IEEE, is currently a Lecturer (Assistant Professor) in Artificial Intelligence with the School of Electronic Engineering and Computer Science, Queen Mary University of London. He is a member of Multimedia and Vision (MMV) research group and of the Queen Mary Computer Vision Group; an associate member of the Centre for Advanced Robotics (ARQ); an academic fellow of the Digital Environment Research Institute (DERI). He has been the recipient of the prestigious Teaching Fellowship of Imperial College London. He has obtained the Ph.D. from Imperial College London, where he was a member of the iBUG group. He has been a Senior Lecturer in Artificial Intelligence at the University of Greenwich. He has published his research in the top journals and conferences on machine learning, perception and computer vision such as IJCV, CVPR, ICCV, ECCV, BMVC, IJCNN and ECAI. He is reviewer of top journals and conferences, such as CVPR, IEEE TPAMI, IEEE TIP, ECCV, ICCV, AAAI, TNNL, TAC, Pattern Recognition and Neural Networks. He is the Chair of the ABAW series of Competitions and Workshops in IEEE CVPR 2023, IEEE CVPR 2022, ECCV 2022, ICCV 2021 and IEEE FG 2020; he has also been a Chair of the MIA-COV19D Competition and Workshop in IEEE ICASSP 2023, ECCV 2022, ICCV 2021. He has won many grants and awards, such as from the City and Guilds College Association, the Imperial College Trust and the Complex & Intelligent Systems Journal. He has h-index 26 and i10-index 30. His research interests span the areas of machine and deep learning, deep neural networks, computer vision, affective computing and medical imaging.

Anastasios Arsenos is a Ph.D. student in the Artificial Intelligence and Learning Systems Laboratory of the National Technical University of Athens (NTUA), performing research on Transparent and Explainable Deep Learning and Artificial Intelligence. He has an M.Sc. from the School of Electrical and Computer Engineering of the NTUA in 2020, having presented his work on propaganda detection in news articles in the 2020 International Workshop on Semantic Evaluation. His research interests include Artificial Intelligence, Machine & Deep Learning and Pattern Recognition. He is currently collaborating with the Greek National Infrastructures for Research and Technology in the HARMONI project on analysis of medical image data for COVID-19 diagnosis.

Stefanos Kollias, Fellow of the IEEE, Fellow of HEA, has been Professor in the School of Electrical and Computer Engineering of the National Technical University of Athens, since 1997. He has been Professor of Machine Learning in the School of Computer Science of the University of Lincoln, UK, since 2016. He has been Chairman of the Greek National Infrastructures for Research and Technology, since 2019. He has been member of the Executive Committee of the European Neural Network Society, 2007-2016. He has been member of the Member States Expert Group on Digitization and Digital Preservation, 2007-2016. He is member of the European Commission Expert Group on Cultural Heritage (responsible for AI), since 2019. His research covers machine and deep learning and artificial intelligence technologies and systems, multimedia analysis, search, retrieval and recognition, and their application in vision, healthcare, cultural heritage, human computer interaction, agri-food, industrial monitoring and anomaly prediction. He has published 110 journal papers and 310 conference papers. He has supervised 43 Ph.D. students. He has led his Group participation in more than 100 European research projects.