Multi-task deep learning based CT imaging analysis for COVID-19 pneumonia: Classification and segmentation

Amine Amyar [a,b,*], Romain Modzelewski [b,c], Hua Li [d], Su Ruan [b]

[a] *General Electric Healthcare, Buc, France*
[b] *LITIS - EA4108 - Quantif, University of Rouen, Rouen, France*
[c] *Nuclear Medicine Department, Henri Becquerel Center, Rouen, France*
[d] *Department of Bioengineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA*

A B S T R A C T

This paper presents an automatic classification segmentation tool for helping screening COVID-19 pneumonia using chest CT imaging. The segmented lesions can help to assess the severity of pneumonia and follow-up the patients. In this work, we propose a new multitask deep learning model to jointly identify COVID-19 patient and segment COVID-19 lesion from chest CT images. Three learning tasks: segmentation, classification and reconstruction are jointly performed with different datasets. Our motivation is on the one hand to leverage useful information contained in multiple related tasks to improve both segmentation and classification performances, and on the other hand to deal with the problems of small data because each task can have a relatively small dataset. Our architecture is composed of a common encoder for disentangled feature representation with three tasks, and two decoders and a multi-layer perceptron for reconstruction, segmentation and classification respectively. The proposed model is evaluated and compared with other image segmentation techniques using a dataset of 1369 patients including 449 patients with COVID-19, 425 normal ones, 98 with lung cancer and 397 of different kinds of pathology. The obtained results show very encouraging performance of our method with a dice coefficient higher than 0.88 for the segmentation and an area under the ROC curve higher than 97% for the classification.

1. Introduction

The novel coronavirus disease (COVID-19) spread rapidly around the world, changing the daily lives of billions of people. The infection can lead to severe pneumonia that can causes death. Also, COVID-19 is highly contagious, which is why it must be detected quickly, in order to isolate the infected person very fast to limit the spread of the disease. Today, the gold standard for detecting COVID-19 is the Reverse Transcription Polymerase Chain Reaction (RT-PCR) [42], which consists of detecting viral RNA from sputum or nasopharyngeal swab. The limitation with the RT-PCR test is due to the time needed to get the results, the availability of the material which remains very limited in hospitals [42] and its relatively low sensitivity, which does not meet the major interest of rapidly detecting positive cases as soon as possible in order to isolate them [39]. An alternative solution for rapid screening could be the use of medical imaging such as chest X-ray images or computed tomography

(CT) scanners [39].

Identifying COVID-19 at an early stage through imaging would indeed allow the isolation of the patient and therefore limit the spread of the disease [39]. However, physicians are very busy fighting this disease, hence the need to create decision support tools based on artificial intelligence to not only detect but also segment the infection at the lung level in the image [42]. Artificial intelligence has seen a major and rapid growth in recent years with deep neural networks [43] as a first tool to solve different problems such as object detection [29,37]speech recognition [37], drug interaction [27] and image classification [10]. More specifically, convolutional neural networks [22] showed astonishing results for image processing [21]. For image segmentation, several works have shown the power and robustness of these methods [5]. CNNs architectures for medical imaging also have been used with very good results [14], for both image classification [3,6] or image segmentation [19].

1.1. Related work

For the detection of COVID-19 and the segmentation of the infection at the lung level, several deep learning works on chest X-ray images and CT scans have emerged and reported in Ref. [34]. In Ref. [25] Ali Narin et al. created a deep convolutional neural network to automatically detect COVID-19 on X-ray images. To that end, they used transfer learning based approach with a very deep architectures such as ResNet50, InceptionV3 and Inception-ResNetV2. The algorithms were trained on the basis of 100 images (50 COVID vs 50 non-COVID) in 5 cross-validation. Authors claimed 97% of accuracy using InceptionV3 and 87% using Inception-ResNetV2. However, due to the very limited size of the dataset and the very deep models, overfitting would rise and could not be ruled-out, hence the need to validate those results in a larger database is necessary. Also in Ref. [15], Hemdan et al. created several deep learning models to classify X-ray images into COVID vs non-COVID classes reporting best results with an accuracy of 90% using VGG16. Again, the database was very limited with only 50 cases (25 COVID vs 25 non-COVID). A resembling study was conducted by wang and wang [38] where they trained a CNN on the ImageNET database [11] then fine-tuned on X-ray images to classify cases into one of four classes: normal, bacterial, non-COVID-19 viral and COVID-19 viral infection, with an overall performance of 83.5%. For CT images, Jinyu Zhao et al. [45] created a container for CT scans initially with 275 CT COVID-19 on which they also applied a transfer learning algorithm using chest-x-ray14 [40] with 169-layer DenseNet [17]. The performance of the model is 84.7% with an area under the ROC curve of 82.4%. As of today, the database contains 347 CT images for COVID-19 patients and 397 for non-COVID patients.

Instead of using CNNs, other works have used network capsules which were first proposed in Refs. [16] to solve the problems of CNNs architectures which need a large amount of data and many parameters. In Ref. [2] the authors opted for this method. They created a capsule network to identify COVID-19 cases in X-ray images. The results were encouraging with an accuracy of 95.7%, sensitivity at 90% and specificity at 95.8%. They compared their results with Sethy et al. [33] where they created a model based on resnets50 with SVMs and obtained a performance of 95.38%, a sensitivity of 97.29% and a specificity of 93.47%.

In [18], Jin et al. created and deployed an AI tool to analyze CT images of COVID-19 in 4 weeks. To do this, a multidisciplinary team of 30 people collaborated together using a database of 1136 images including 723 positive COVID-19 images from five hospitals, to achieve a sensitivity of 0.974 and a specificity of 0.922. The system was deployed in 16 hospitals and performed over 1300 screenings per day. They proposed a combined model for classification and segmentation showing lesion regions in addition to the screening results. The pipeline is divided into 2 steps: segmentation and classification. They used several models including 3D U-NET++, V-NET, FCN-8S for segmentation and InceptionV3, ResNet50 and others for classification. They were able to achieve a dice coefficient of 0.754 using 3D U-NET ++ trained on 732 cases. The combination of 3D U-NET ++ and ResNet50 resulted in an area under the OCR curve of 0.991 with a sensitivity of 0.974 and a specificity of 0.922. In practice, the model continued to improve by re-training. The model proved to be very useful to physicians by highlighting regions of lesions which improved the diagnosis. What should be noted here is that the two models are independent and therefore they cannot help each other to improve both classification and segmentation performances.

Other works have also emerged recently with interesting results. In Ref. [26] T. Ozturk et al. proposed a deep neural network called Dark-CovidNet to detect automatically COVID-19 cases on X-ray images. The model inspired from Darknet-19, is a classifier model that forms the basis of a real-time object detection system named YOLO (You only look once) [30]. They implemented a 17 convolutional layers network achieving a results of 98.08% for binary classes and 87.02% for multi-class cases. In Ref. [27], Y. Pathak et al. used a transfer learning approach to classify COVID-19 infected patients. They introduced a top-2 smooth loss function with cost-sensitive attributes to handle noise and imbalanced dataset. The model was trained on a public dataset of chest CT images, and then used to classify COVID-19 infected patients. The model achieved an accuracy of 0.93, a sensitivity of 0.91 and a specificity of 0.94. Other works such as in S. Dilbag et al. [35] tried a multi-objective differential evolution–based convolutional neural networks to classify COVID-19 patients from chest CT images.

1.2. Motivation

Multi-task learning (MTL) [8] is a type of learning algorithm whose goal is to combine several pieces of information from different tasks to improve the performance of the model and its ability to better generalize [44]. The basic idea of MTL is that different tasks can share a common features representation [44], and therefore, training them jointly. The use of different datasets of different tasks makes it possible to learn an effective feature representation that is common to all tasks, because all datasets are used to obtain it, even if each task has a small dataset, which leads to improve the performance of each task. Different approaches have been proposed in MTL such as hard parameter sharing [8] or soft parameter sharing [32]. Hard parameter sharing is the most commonly used approach to MTL in neural networks and greatly reduces the risk of overfitting [32], see Fig. 4. It is generally applied by sharing the hidden layers between all tasks, while keeping several task-specific output layers. Soft parameter sharing defines a model for each task with its own parameters, and the distance between the parameters of the model is regularized in order to encourage the parameters to be similar.

In this work, we propose a novel multi-task deep learning model for jointly detecting COVID-19 image and segmenting lesions. The main challenges of this work are: 1) the lack of data and annotated data, because the databases were collected from multiple sources with a huge variation in images and most of the images are noisy (see Fig. 3). 2) instead of expensive models like ResNet 50 or DenseNet, developing a multitasking approach to reduce overfitting and improve results. In Fig. 1 we can see an example of different image formats (png (B C E F), Nifti (D), DICOM (A), different scales of visualisation window (strong centering on lung (B E C), low (A,D, F), different image sizes, with/without annotations (with (C,D E), without (A B F). Facing these challenges, we proposed to train our neural network with three tasks: reconstruction, classification and segmentation, in order to both classify COVID/Non-COVID images and to segment the lesion regions. We add the reconstruction task often used in unsupervised learning, to better learn the disentangled feature representation with a single common encoder. Based on this features representation, three neural networks are designed to finally accomplish the three tasks.

The paper is organized as follows. In Section 2, we describe our multi-task model, which is mainly based on classification and segmentation tasks. Section 3 presents the experimental studies. In section 4, we describe the validation methodology used in this study. Section 5 is showing the results of our work. Section 6 and 7 are for discussion and conclusion respectively.

1.3. Data

In this study, three datasets from different hospitals including one thousand three hundred and ninety-six CT images are used. The first one is a public available data set coming from Refs. [45] which includes 347 COVID-19 images and 397 non-COVID images with different types of pathologies. The database was pre-processed and stored in png format. The dimension varies from 153 to 1853 with an average of 491 for the height, while the width varies from 124 to 383 with an average of 1485. The second dataset coming from [http://medicalsegmentation.com/covid19/] in which 100 COVID-19 CT scan with ground truths lesions segmentation are available. The ground truth was defined by the
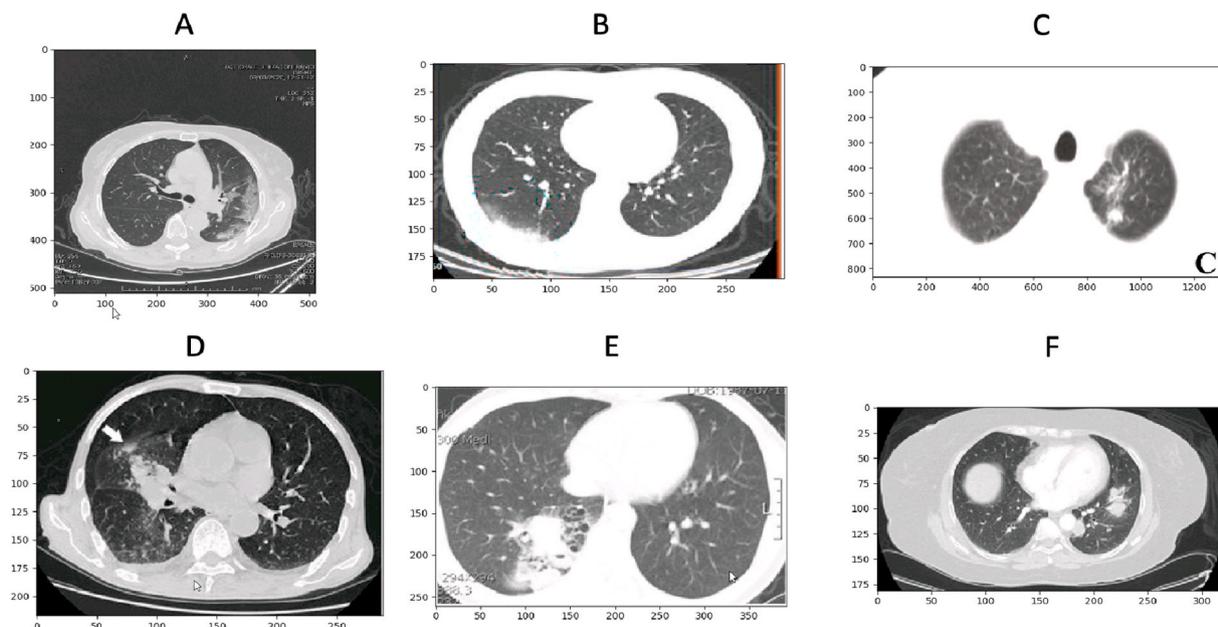
Fig. 1. An example of exams heterogeneities between different CT images for COVID (upper) and non-COVID (bottom). Patient images do not have the same resolution. Also, images show different image formats (Dicom(A), png (B C E F), Nifti (D).
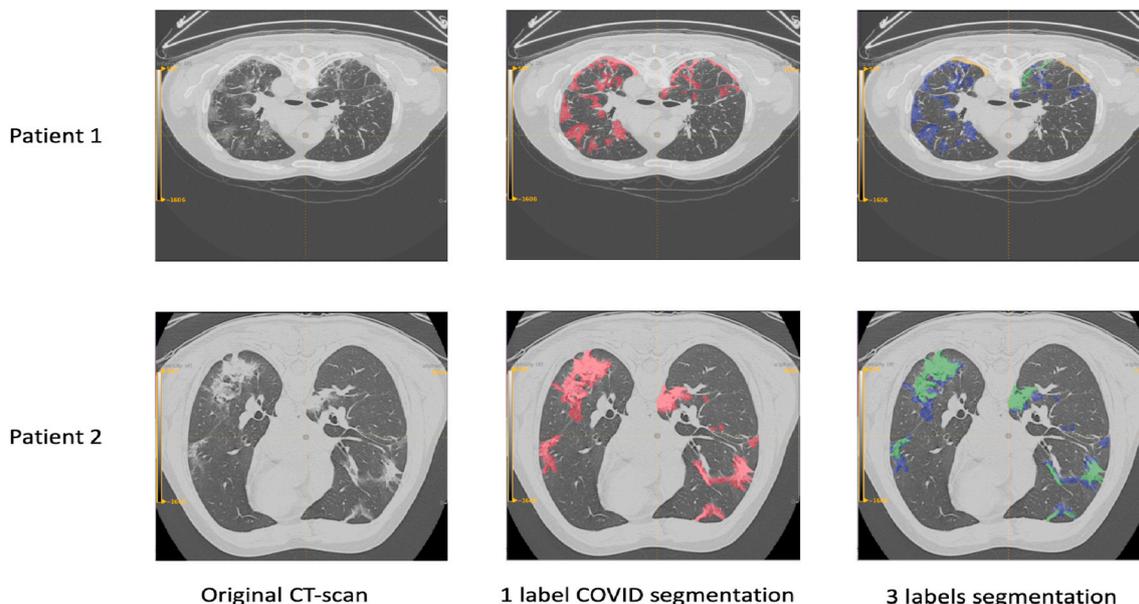

Fig. 2. Example of COVID-19 segmentation on CT scan for 2 patients: first column - original CT scanner; Second column - one label segmentation, Third column - 3 labels segmentation: ground glass (green), consolidation (blue) and pleural effusion (yellow).

Fig. 3. The different databases used in this study.

physicians of different hospitals. Three lesion labels are provided: ground glass, consolidation and pleural effusion. As all lesion labels are not given in all images, for the purpose of this study, we merged the three labels into one lesion label (See Fig. 2). The third dataset coming from the Henri Becquerel Cancer Center (HBCC) in Rouen city of France includes 425 CT scans of normal patients and 98 of lung cancer. All the three image datasets were resized to have the same size 256×256 and the intensity normalized between 0 and 1 prior to analysis. Table 1 summarizes how to split the datasets for training, validation and test.
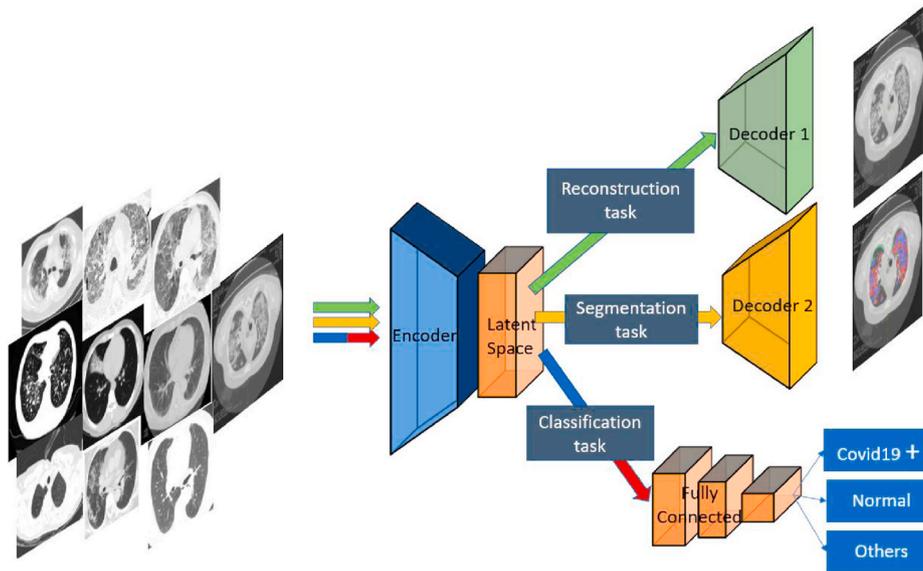
Fig. 4. Hard parameter sharing for multi-task learning in deep neural networks used in our proposed architecture.

Table 1
Data split.

	Train	Validation	Test	Total
Normal	325	50	50	425
COVID-19	349	50	50	449
Other Infections	395	50	50	495
Total	1069	150	150	1369

2. Method

2.1. Model description

We propose a new MTL architecture based on 3 tasks: 1) COVID vs Normal vs Other Infections classification, 2) COVID lesion segmentation, 3) image reconstruction. The two first tasks are essential ones, while the third task is added to enhance the feature representation extracted. In this work, we choose to use a hard parameter sharing to share parameters between the different tasks (see Fig. 4). We create a common encoder for the three tasks which takes a CT scan as input, and its output is then used to the reconstruction of the image via a first decoder, to the segmentation via a second decoder, and to the classification of COVID vs Normal vs Other Infections classification image via a multi-layer perceptron.

Each convolutional layer, denoted $C^{(m)}$, consists of $F^{(m)}$ feature maps, where m is the layer number. For the first layer, $C^{(1)}$, each feature map is obtained by convolving the volume of interest with a weight matrix $W_i^{(1)}$ to which a bias term $b_i^{(1)}$ is added, where i is the feature map number. Then, the output goes through a non linear function f(x), where x is the input to a neuron, such as:

$$c_i^{(1)} = f\left(b_i^{(1)} + W_i^{(1)} * x\right) \quad \text{with} \quad i = 1, \ldots, F^{(1)}. \tag{1}$$

Each element of a feature map, $c_i^{(1)}$, is obtained by convolving the input x with a kernel. The $F^{(1)}$ weight matrices (one matrix per feature map) are learned by looking at different position of the input, leading to the extraction of the description of features. Thus, the weight parameters are shared for all lesion or infection input sites, so that the layer has an equivariance property and is invariant to the input lesion transformations (such as translation and rotation). It also results in a sparse weight, which means that the kernel can detect small, but meaningful features.

In order to extract high-level features from the low-level ones ob-

tained in the initial layer, other layers are added. Each feature map in the other layers are obtained as follow:

$$c_i^{(n)} = f\left(b_i^{(n)} + \sum_{s=1}^{P^{(n-1)}} W_i^{(n)} * P_s^{(n-1)}\right) \quad \text{with} \quad i = 1, \ldots, F^{(n)}. \tag{2}$$

2.1.1. The encoder-decoder

The encoder-decoder is based on the U-NET architecture [31] for both reconstruction and segmentation tasks (Fig. 5). The encoder is used to obtain the disentangled feature representation. It includes a convolutional block followed by skip connection. In order to maintain the spatial information, we use a convolution with stride = 2 to replace pooling operation. It's likely to require different receptive fields when segmenting different regions in an image. All convolutions are 3×3 and the number of filter is increased from 64 to 1024. Each decoder level begins with upsampling layer followed by a convolution to reduce the number of features by a factor of 2. Then the upsampled features are combined with the features from the corresponding level of the encoder part using concatenation.

2.1.2. The reconstruction task T1

We trained the model with a linear activation for the output and a mean squared error for the loss function (Lrecon) and used accuracy as the metric:

$$Lrecon = \frac{1}{n} \sum_{t=1}^{n} \left(y\_true - y\_predict\right)^2 \tag{3}$$

where y_true is the true label and y_predict is the predicted label.

2.1.3. The segmentation task T2

We used the same architecture as the reconstruction except for the activation function for the output, which is a sigmoid. The loss function is based on the dice coefficient loss (Lseg) which is considered as the metric:

$$dice\_coef = \frac{2*|X \cap Y| + \varepsilon}{|X| + |Y| + \varepsilon} \tag{4}$$

$$Lseg = -dice\_coef \tag{5}$$

where ε is the smoothing factor and used to avoid a division by zero.
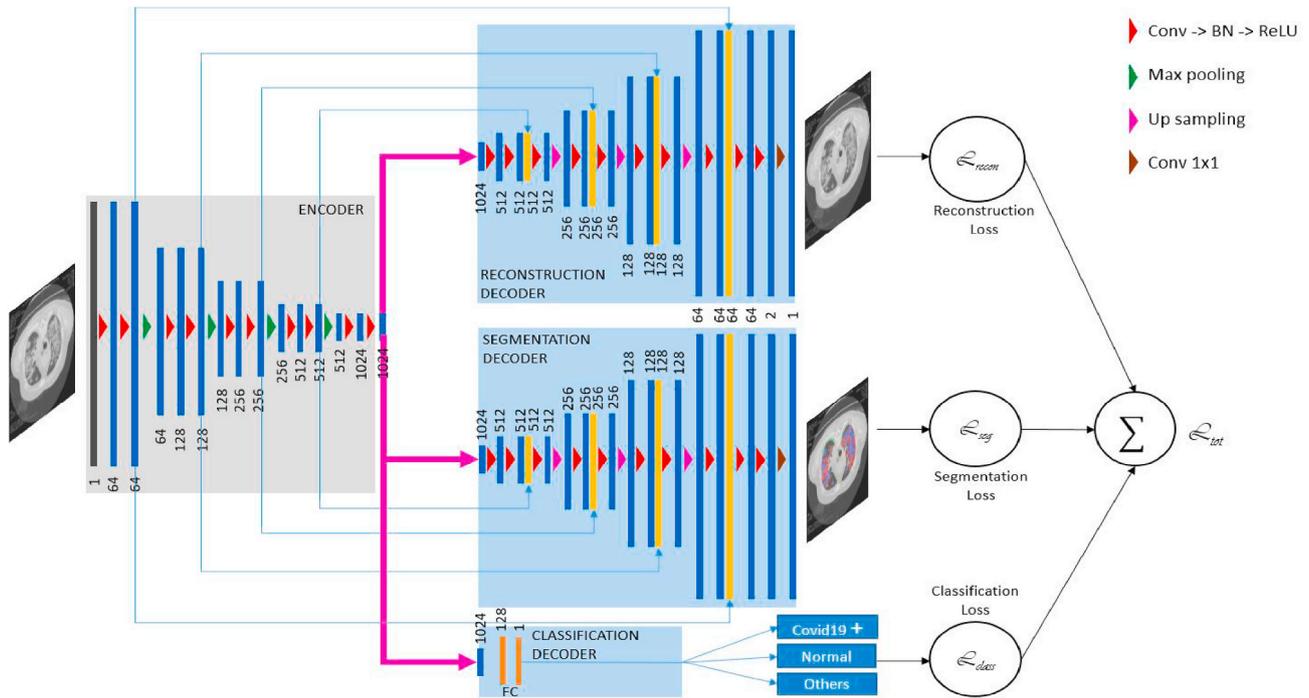
Fig. 5. Our proposed architecture, composed of an encoder and two decoders for image reconstruction and infection segmentation. A fully connected layers are added for classification (COVID vs Normal vs Other Infections classification).

2.1.4. The classification task T3

The resulting set of feature maps, encloses the entire spatial local information, as well as the hierarchical representation of the input. Then, each feature map is flattened out, and all the elements are collected into a single vector V of dimension K, providing the input for a fully connected hidden layer, called h, consisting of H units. The activation of the i$^{(th)}$ unit of the h hidden layer is given by:

$$h_i = f(b_i + Wh_i * V) \quad \text{with} \quad i = 1, \ldots, H. \tag{6}$$

In details, the output of the encoder is a tensor of mini_batch x 32 × 32 x 1024 to which we add a convolutional layer followed by a max-pooling, and then a flatten operation to convert the data to a mono-dimensional tensor to perform the classification. The multilayer perceptron consist of a two Dense layer with 128 and 64 neurons respectively, with a dropout of 0.5 and the activation function *elu*. The last layer is a Dense layer with three neurons for image classification using a sigmoid activation and a binary cross entropy is used as the loss function (Lclass):

$$Lclass = -\frac{1}{n} \sum_{i=1}^{n} \left[y_i \log\left(\widehat{y}_i\right) + (1 - y_i)\log\left(1 - \widehat{y}_i\right) \right] \tag{7}$$

which is a special case of the multinomial cross-entropy loss function for m = 2:

$$L(\theta) = -\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m} y_{ij} \log\left(\widehat{y}_{ij}\right) \tag{8}$$

where n is the number of patients, y is the class label. The output layer consists of 3 neurons, each one output a binary value such as: ŷij∈ (0,1): $\sum$ j ŷij = 1 ∀i,j is the prediction of a COVID presence for the first neuron. The second neuron returns 1 if the patient is normal, 0 otherwise, and for the third neuron 1 if the patient has another infection and 0 otherwise. The patient is positive if he has the COVID-19. In our experiments, the Adam optimizer [20] algorithm was used with a mini batches of 4 and a learning rate of 0.0001. The global loss function (loss glob) for the 3 tasks is defined by:

$$loss\_glob = Lrecon + Lseg + Lclass \tag{9}$$

Our model was trained for 500 epochs with an early stopping of 10.

2.2. Implementation

The implementation of our method was done using the keras library with tensorflow in backend. The model was performed on an nvidia p6000 quadro gpu with 24 GB, and 128 RAM.

3. Experiments

Three experiments are conducted to evaluate our model.

Experiment 1: The first experiment consisted of tuning the hyper-parameters and add/remove a task to find the best model using only the training dataset. Several models were developed by combining the tasks 2 by 2 and the 3 tasks with different resolutions of images (512 × 512 and 256 × 256). The combination of the first task and the second one is only to evaluate segmentation results, since it is for image reconstruction and infection segmentation, while the peer T1 and T3 is for classification.

Experiment 2: The second experiment is to compare our model with state of the art method U-NET in order to evaluate the performance on the segmentation task. Two U-NET with different resolutions were trained: 512 × 512 and 256 × 256.

Experiment 3: Different state of the art models were compared to ours on the classification task. We use: Alexnet, VGG-16, VGG-19, ResNet50, 169-layer DenseNet, InceptionV3, Inception-ResNet v2 and Efficient-Net. We have also added an 8 layer deep neural network with 6 convolutional layers, where each one is followed by a Maxpooling and a Dropout regularization of 25% to prevent the model from overfitting. The feature maps go from 8 to 256 by a factor of 2 between each two layers. We used 3 × 3 filter for convolution and 2 × 2 for Maxpooling. Then a Flatten followed by two Dense layers with 128 neurons and 3 neurons respectively. A Dropout of 50% is also applied to the first layer to reduce and prevent overfitting. The activation function is *elu* for all layers except the last one which is a *sigmoid*. The loss function is the

binary cross-entropy and the metric is the accuracy, with the Adam optimizer. The CNN was optimized in order to ensure a fair comparison with our proposed model. The model was trained for 500 epochs with an early stopping of 10, in the same condition as our model.

To find the best hyperparameters, the influence of F, the number of feature maps (8–64), the number of neurons (128–4096) were evaluated, as well as different receptive field sizes ($3 \times 3, 5 \times 5$) and different sizes of mini-batch (2–16). Several expressions of f(x), the activation function, were also evaluated (relu, elu, selu and tanh), to choose finally *relu*.

4. Validation methodology

For the validating methodology, we split the data for training, validation and test as shown in Table 1. Among the 349 COVID cases in the training, the ground truth for the infection label (segmentation task) was available for 50 CT scans. Twenty others were in the validation and thirty in the test set. For normal patients, 50 were used in validation, 50 in test and 325 in training. For other infections, different kinds of pathology such as lung cancer or cases were selected randomly in training, validation and test. For a fair comparison, the other methods were trained, validated and tested on the same group of data. The performance of the models were evaluated using the dice coefficient for the segmentation task, area under the ROC curve (AUC) for the classification, and the accuracy (Acc), sensitivity (Sens), specificity (Spec) for both [1], such as:

$$Sens = \frac{TP}{TP + FN} \qquad (10)$$

where TP is the true positives, FN is the false negatives and TP + FN is the number of patients classified positively, or the segmented lesion infection.

$$Spec = \frac{TN}{TN + FP} \qquad (11)$$

where TN is the true negatives, FP is the false positives and TN + FP is the number of patients classified negatively, or the non segmented region.

$$ACC = \frac{TP + TN}{TP + FN + TN + FP} \qquad (12)$$

For each curve, the definition of the thresholds was determined using the method proposed by Fawcett [13], and the optimal cut-off point was defined using Youden's index.

5. Results

The main results of the three experiments are shown in Tables 2 and 3. Metrics include: dice coefficient, accuracy, sensibility, specificity and the area under the ROC curve. The neural network was trained for 500

Table 2
Segmentation results. Experiment 1: results with different tasks (T1: classification, T2: Segmentation, T3: reconstruction). Experiment 2: segmentation results comparing with U-NET.

Method	Dice_coef	ACC	Sens	Spec
Experiment 1				
T1 & T2	80.3%	88.34%	76.8%	98.8%
T2 & T3	79.34%	85.67%	74.7%	98.3%
Ours (T1&T2&T3 512 x 512)	84.8%	91.03%	84.10%	99.41%
Ours 256 x 256	88.0%	95.23%	90.2%	99.7%
Experiment 2				
U-NET 512×512	76.03%	81.14%	69.51	97.34%
U-NET 256×256	77.69%	83.40%	73.10%	98.24%
Ours 256 x 256	88.0%	95.23%	90.2%	99.7%

Table 3
Classification results: Experiment 1 for optimizing hyperparameters and choosing the best combination of tasks. Experiment 3 for classification.

	Method	ACC	Sens	Spec	AUC
Experiment 1	T1 & T3	78.66	0.88	0.79	0.83
	T2 & T3	79.33	0.81	0.75	0.81
	Ours (T1 & T2 & T3 512 × 512)	91.13	0.94	0.85	0.94
	Ours(T1 & T2 & T3 256 x 256)	**94.67**	**0.96**	**0.92**	**0.97**
Experiment 3	CNN 8-layers	74.67	0.8	0.70	0.78
	Encoder-Dense	70.04	0.75	0.61	0.72
	AlexNET	56.67	0.67	0.64	0.66
	VGG-16	62.67	0.67	0.65	0.66
	VGG-19	66.14	0.77	0.61	0.69
	ResNet50	86.67	0.9	0.83	0.88
	169-layer DenseNet	83.33	0.91	0.83	0.88
	InceptionV3	82.67	0.88	0.78	0.82
	Inception-ResNet V2	85.33	0.84	0.88	0.90
	Efficient-Net	90.67	0.91	0.85	0.93
	Ours	**94.67**	**0.96**	**0.92**	**0.97**

epochs with an early stopping of 10. In Fig. 6, the learning curve and the loss curve obtained from the train and validation sets respectively show the stability of our model. From Fig. 6 we can observe that the training and validation losses decrease to a point of stability, and have a small gap between them. Early stopping of the training provides a robust mechanism to avoid overfitting, like the behavior of our model.

Experiment 1: As shown in Table 2 for the segmentation and Table 3 for the classification, the best dice coefficient = 88%, accuracy (ACC = 0.94.7) and area under the curve (AUC = 0.97) were obtained with the combination of the three tasks of image reconstruction, infection segmentation and image classification, with all images resized to 256 × 256. The results of 4 other experiments were also shown with multi-task learning but with a higher resolution of 512 × 512, and the combination 2 by 2 of the other tasks. The major differences between our best model and the model with higher resolution are in sensitivity (0.96 vs 0.94) and specificity (0.92 vs 0.85). Compared to the peer combination of T1 and T3 for segmentation our model proved to be more performant with an improvement of +9% of dice, and a higher AUC and accuracy compared to the peer T1 and T3 for classification only. On the ROC curve in Fig. 8, the advantage of using all three tasks combined can be seen by obtaining a significantly better area under the curve than when two tasks are used. The same results were observed for the peer segmentation and classification without reconstruction. Those results confirm the usefulness of the reconstruction task to extract meaningful features and to help improve the results of the other two tasks. For multi-class classification, as we can observe from the confusion matrix of our model in Fig. 10, 47 of 50 COVID cases were classified correctly, while only 3 were miss-classified as other infections. The same observation for normal patients, only 2 were misclassified as other infections, while no normal patient was misclassified as COVID.

Experiment 2: In Table 2, the best results for image segmentation was obtained using our method with a dice_coef of 88% versus 77.69% and 76.69% using U-NET with 256 × 256 and 512 × 512 resolutions respectively. The combination of the reconstruction, segmentation and classification provide a higher accuracy to detect infection regions, compared to the use of the U-NET model alone. Fig. 7 shows a comparison between our model and U-NET for infection segmentation.

Experiment 3: The results of experiment 3 are also given in Table 3. We compared our multi-task deep learning model with multiple deep convolutional neural networks. The obtained results show that our model outperformed the CNN in both accuracy and AUC. The ROC curve for experiment 3 is shown in Fig. 9.

Finally, we compared our model with the state of the art methods on COVID-19 for classification and segmentation. Table 4 shows the results on the classification task. Results vary from an accuracy of 66.67% in Ref. [24] to 92.6% [38] in X-ray images, and from 84.7% to 90.8% for CT scans. Our model outperformed state of the art methods with an
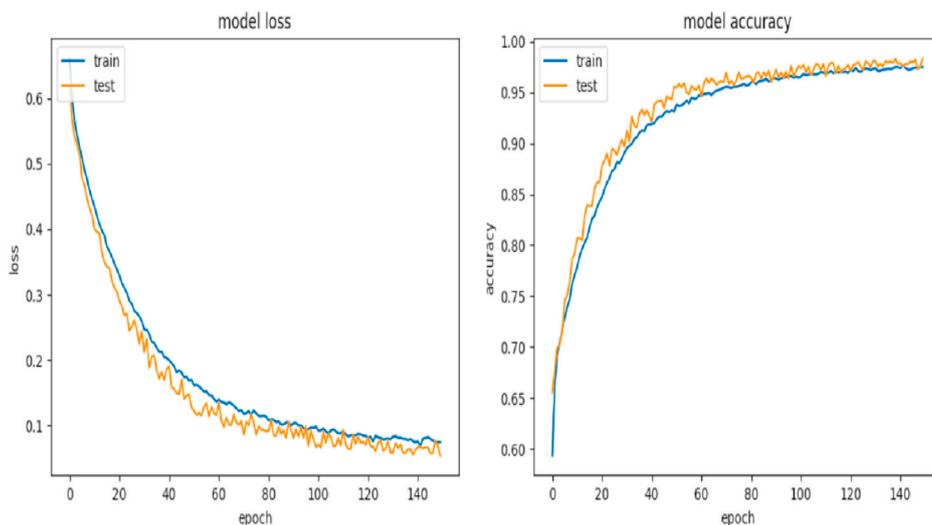
Fig. 6. Learning curve of our proposed model. Left is the model loss and right is the model accuracy per epoch.
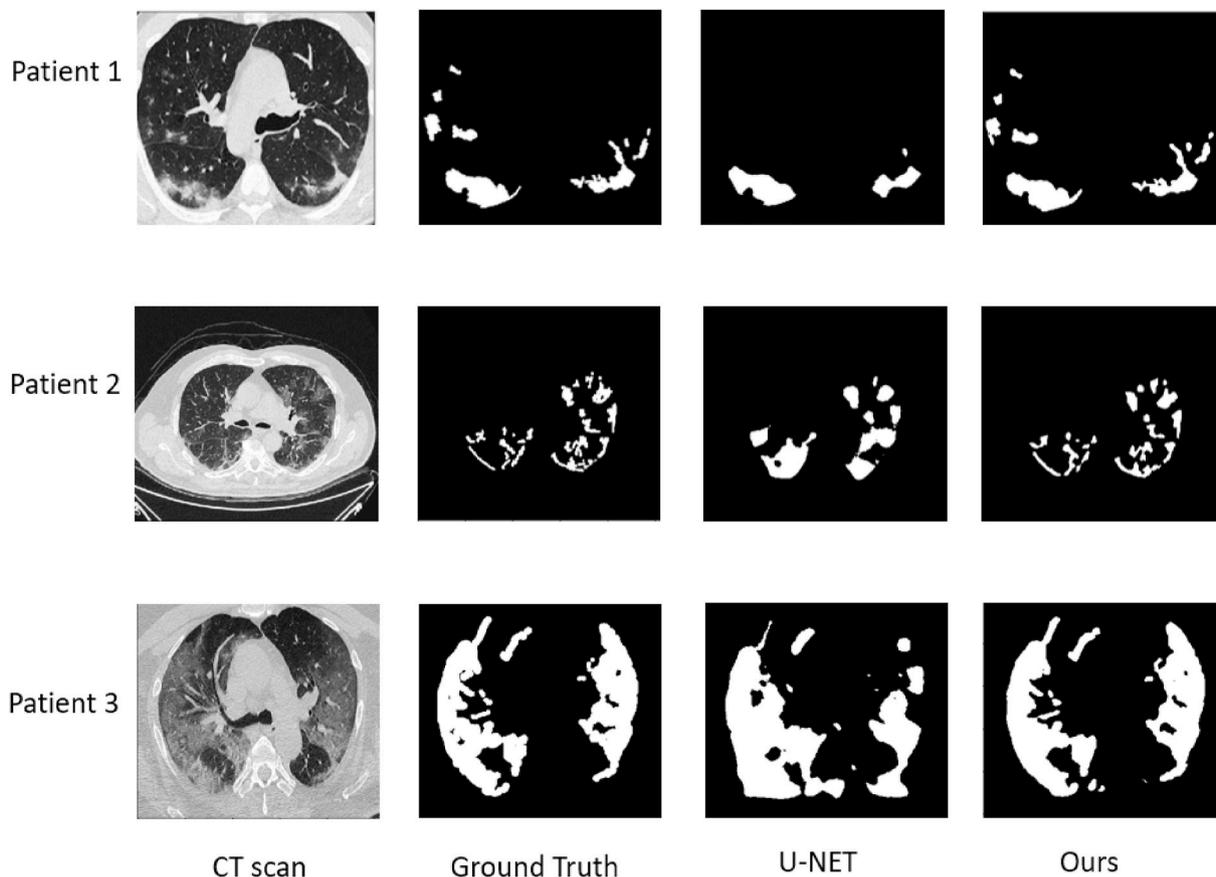

Fig. 7. A comparison between our model and U-NET for infection segmentation. From left to right columns: input images, ground truth, results of U-NET, results of our method.

accuracy of 94.67%. Zhou et al. [47], who performed only the segmentation task achieved 61.0% using a modified U-NET and 69.1% using an attention mechanism. Other results reported in Ref. [9] reached a dice coefficient of 85.0%, which is less than our model with a dice coefficient of 88.0%. The results for the segmentation are shown in Table 5.

6. Discussion

We have developed a new deep learning multi-task model to jointly detect COVID-19 CT images and segment the regions of infection. Our architecture is general, which means that it can be used for other segmentation-classification applications. We have also compared with several state of the art algorithms such as U-NET and CNNs. To show the performance of our method, we tested the different combinations of
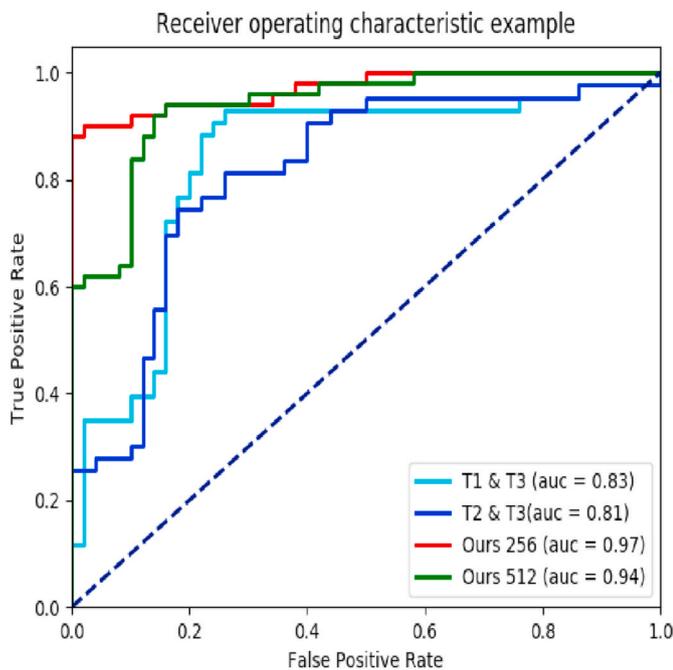
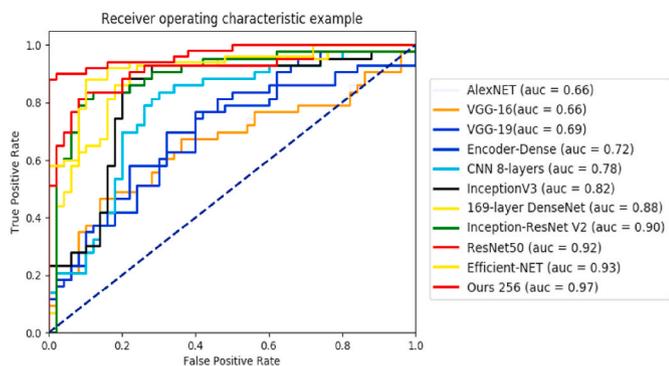**Fig. 8.** ROC curve of Experiment 1 for COVID-19 classification.

Fig. 9. ROC curve of Experiment 3 for COVID-19 classification.

tasks 2 by 2 and all the 3 tasks simultaneously with different images resolutions. Our motivation was to leverage useful information contained in multiple related tasks to improve both segmentation and classification performances.

Multitasking can handle small data problems well, although each task can have a relatively small data set. In our study, we were able to increase the size of the database, in total 1044 images by the combination of several databases, to learn the disentangled representation. Although we have a database of 100 images for segmentation, thanks to the learned latent representation, we have obtained good segmentation results.

The state of the art U-NET has shown impressive results to deal with image segmentation in recent years, just like the classification with Alexnet, VGG-16, VGG-19, ResNet50, DensNet. However, these segmentation or classification methods usually require a large amount of annotated datasets to work efficiently. Due to the lack of annotated data in the medical imaging field, other mechanisms can be included to improve its generalisation ability. In this work, we propose to use a multitask learning approach that can jointly improve the U-NET model and classification models by enhancing its encoder. Indeed, by using a shared encoder for the classification and segmentation tasks, it is able to extract more meaningful information from the CT scan relating to the COVID-19 characteristics, which improves both tasks simultaneously
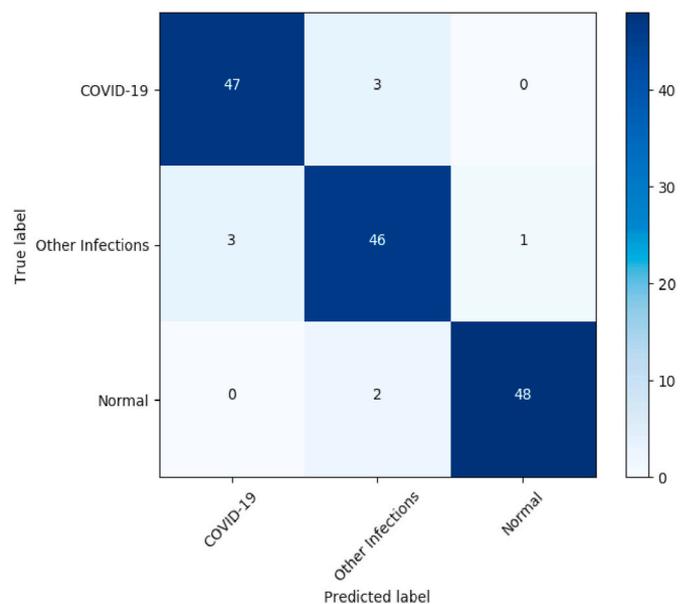

Fig. 10. Confusion matrix of our model.

Table 4
A quantitative comparison between our model and state of the art for the classification task.

Method	Modality	ACC	Sens	Spec
Alexnet Loey et al. (2020) [24]	X-ray	66.67	66.67	–
Resnet18 Loey et al. (2020) [24]	X-ray	69.47	66.67	–
ShuffleNet + SVM Sethy and Behera [33]	X-ray	70.66	65.26	–
Googlenet Loey et al. (2020) [24]	X-ray	80.56	80.56	–
CNN Zhao et al. (2020) [45]	CT	84.7	76.2	–
Ying et al. [36]	CT	86	–	–
Xu et al. [7]	CT	86.7	–	–
Ozturk et al. (multiclass) [26]	X-ray	87.02	–	–
Li and Zhu [23]	X-ray	88.9	–	–
Hemdan et al. [15]	X-ray	90	–	–
Zheng et al. [46]	CT	90.8	–	–
Wang et al. [38]	X-ray	92.6	–	–
Ours	**CT**	**94.67**	**96**	**92**

Table 5
A quantitative comparison between our proposed model and state of the art for the segmentation task.

Method	Dice_coef
U-Net + DL Zhou et al. (2020) [47]	61.0%
U-Net + FTL Zhou et al. (2020) [47]	66.7%
AU-Net + DL Zhou et al. (2020) [47]	68.5%
AU-Net + FTL Zhou et al. (2020) [47]	69.1%
Backbone + PPD + RA + EA Fan et al. (2020) [12]	73.9%
JCS Wu et al. (2020b) [41]	77.5%
JCS′ Wu et al. (2020b) [41]	78.3%
U-net Chen et al. (2020b) [9]	82.0%
M − A Chen et al. (2020b) [9]	85.0%
M − R Chen et al. (2020b) [9]	84.0%
Ours	**88.0%**

with less annotated datasets. Furthermore, adding a third task for image reconstruction allows the encoder to refine the image characteristics to make a further improvement for both the classification and segmentation tasks. Thus, multitask learning can be used to improve U-NET and other classification models especially in the case where annotated data are limited.

In addition to the many advantages of using CT images to spot early COVID-19 patients and isolate them, deep learning methods using CT

images can be used as a tool to assist physicians fighting this new spreading disease, as they can be used also to not only classify and segment images in the medical field, but also to predict the outcome of treatment [4,28].

7. Conclusion

In this paper, we proposed a multi-task learning approach to detect COVID-19 from CT images and segment the regions of interest simultaneously. Our method can improve segmentation results even if we don't have many segmentation ground truths. This thanks to the ground truths of classification which are relatively easy to obtain compared to those of segmentation. Our method shows very promising results. It outperformed the state of the art methods when used alone for image segmentation such as U-NET or image classification such as CNNs. We have shown that by combining jointly these two tasks, the method improves for both segmentation and classification performances. Moreover, adding a third task such as image reconstruction, the encoder can extract meaningful feature representation which help the other tasks (classification and segmentation) to improve even more their performances.

From experiment 2, we observe a neat improvement when using the multitask approach with a dice coefficient of 88% for segmentation, 10% higher than when using the state of the U-net alone. With a specificity of 99.7% and a sensitivity of 90.2%, the segmentation results outperformed other models without multitask learning approach and when combining a peer of tasks. For the classification results, with an AUC = 0.97 and an accuracy higher than 94%, our model shows a big improvement compared to other models with results ranging from 56% to 90%. Our method uses only CT images. Other information, like patient information, is not included in our architecture. In addition, the performance of our method was performed from a dateset of 150 patients. In future work, we will study new types of networks to take into account other useful information and test our method on a larger database to confirm its good performance.

Declaration of competing interest

None Declared.

References

[1] A. Abbasian Ardakani, A. Bitarafan-Rajabi, A. Mohammadzadeh, A. Mohammadi, R. Riazi, J. Abolghasemi, A. Homayoun Jafari, M. Bagher Shiran, A hybrid multilayer filtering approach for thyroid nodule segmentation on ultrasound images, J. Ultrasound Med. 38 (2019) 629–640.

[2] P. Afshar, S. Heidarian, F. Naderkhani, A. Oikonomou, K.N. Plataniotis, A. Mohammadi, Covid-caps: A Capsule Network-Based Framework for Identification of Covid-19 Cases from X-Ray Images, 2020 arXiv preprint arXiv: 2004.02696.

[3] A. Amyar, S. Ruan, I. Gardin, C. Chatelain, P. Decazes, R. Modzelewski, 3-d rpet-net: development of a 3-d pet imaging convolutional neural network for radiomics analysis and outcome prediction, IEEE Transactions on Radiation and Plasma Medical Sciences 3 (2019) 225–231.

[4] A. Amyar, S. Ruan, I. Gardin, R. Herault, C. Clement, P. Decazes, R. Modzelewski, Radiomics-net: convolutional neural networks on fdg pet images for predicting cancer treatment response, J. Nucl. Med. 59 (2018) 324–324.

[5] V. Badrinarayanan, A. Kendall, R. Cipolla, Segnet: a deep convolutional encoder-decoder architecture for image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 39 (2017) 2481–2495.

[6] H.S. Basavegowda, G. Dagnew, Deep learning approach for microarray cancer data classification, CAAI Trans. Intell. Technol. 5 (2020) 22–33.

[7] C. Butt, J. Gill, D. Chun, B.A. Babu, Deep Learning System to Screen Coronavirus Disease 2019 Pneumonia, vol. 1, Applied Intelligence, 2020.

[8] R. Caruana, Multitask learning, Mach. Learn. 28 (1997) 41–75.

[9] X. Chen, L. Yao, Y. Zhang, Residual Attention U-Net for Automated Multi-Class Segmentation of Covid-19 Chest Ct Images, 2020 arXiv preprint arXiv:2004.05645.

[10] D. Ciregan, U. Meier, J. Schmidhuber, Multi-column deep neural networks for image classification, in: 2012 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2012, pp. 3642–3649.

[11] J. Deng, W. Dong, R. Socher, L.J. Li, K. Li, L. Fei-Fei, Imagenet: a large-scale hierarchical image database, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, Ieee, 2009, pp. 248–255.

[12] D.P. Fan, T. Zhou, G.P. Ji, Y. Zhou, G. Chen, H. Fu, J. Shen, L. Shao, Inf-net: automatic covid-19 lung infection segmentation from ct images, IEEE Trans. Med. Imag. 39 (8) (2020).

[13] T. Fawcett, An introduction to roc analysis, Pattern Recogn. Lett. 27 (2006) 861–874.

[14] H. Greenspan, B. Van Ginneken, R.M. Summers, Guest editorial deep learning in medical imaging: overview and future promise of an exciting new technique, IEEE Trans. Med. Imag. 35 (2016) 1153–1159.

[15] E.E.D. Hemdan, M.A. Shouman, M.E. Karar, Covidx-net: A Framework of Deep Learning Classifiers to Diagnose Covid-19 in X-Ray Images, 2020 arXiv preprint arXiv:2003.11055.

[16] G.E. Hinton, S. Sabour, N. Frosst, Matrix Capsules with Em Routing, 2018.

[17] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4700–4708.

[18] S. Jin, B. Wang, H. Xu, C. Luo, L. Wei, W. Zhao, X. Hou, W. Ma, Z. Xu, Z. Zheng, et al., Ai-assisted Ct Imaging Analysis for Covid-19 Screening: Building and Deploying a Medical Ai System in Four Weeks, 2020 medRxiv.

[19] B. Kayalibay, G. Jensen, P. van der Smagt, Cnn-based Segmentation of Medical Imaging Data, 2017 arXiv preprint arXiv:1701.03056.

[20] D.P. Kingma, J. Ba, Adam: A Method for Stochastic Optimization, 2014 arXiv preprint arXiv:1412.6980.

[21] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, in: Advances in Neural Information Processing Systems, 2012, pp. 1097–1105.

[22] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE 86 (1998) 2278–2324.

[23] X. Li, D. Zhu, Covid-xpert: an Ai Powered Population Screening of Covid-19 Cases Using Chest Radiography Images, 2020 arXiv preprint arXiv:2004.03042.

[24] M. Loey, F. Smarandache, N.E. M Khalifa, Within the lack of chest covid-19 x-ray dataset: a novel detection model based on gan and deep transfer learning, Symmetry 12 (2020) 651.

[25] A. Narin, C. Kaya, Z. Pamuk, Automatic Detection of Coronavirus Disease (Covid-19) Using X-Ray Images and Deep Convolutional Neural Networks, 2020 arXiv preprint arXiv:2003.10849.

[26] T. Ozturk, M. Talo, E.A. Yildirim, U.B. Baloglu, O. Yildirim, U.R. Acharya, Automated detection of covid-19 cases using deep neural networks with x-ray images, Comput. Biol. Med. 121 (2020), 103792.

[27] Y. Pathak, P.K. Shukla, A. Tiwari, S. Stalin, S. Singh, P.K. Shukla, Deep Transfer Learning Based Classification Model for Covid-19 Disease, IRBM, 2020.

[28] D. Paul, R. Su, M. Romain, V. Sébastien, V. Pierre, G. Isabelle, Feature selection for outcome prediction in oesophageal cancer using genetic algorithm and random forest classifier, Comput. Med. Imag. Graph. 60 (2017) 42–49.

[29] G. Qi, H. Wang, M. Haner, C. Weng, S. Chen, Z. Zhu, Convolutional neural network based detection and judgement of environmental obstacle in vehicle operation, CAAI Trans. Intell. Technol. 4 (2019) 80–91.

[30] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: unified, real-time object detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 779–788.

[31] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 234–241.

[32] S. Ruder, An Overview of Multi-Task Learning in Deep Neural Networks, 2017 arXiv preprint arXiv:1706.05098.

[33] P.K. Sethy, S.K. Behera, Detection of Coronavirus Disease (Covid-19) Based on Deep Features, 2020.

[34] F. Shi, J. Wang, J. Shi, Z. Wu, Q. Wang, Z. Tang, K. He, Y. Shi, D. Shen, Review of Artificial Intelligence Techniques in Imaging Data Acquisition, Segmentation and Diagnosis for Covid-19, 2020 arXiv preprint arXiv:2004.02731.

[35] D. Singh, V. Kumar, M. Kaur, Classification of covid-19 patients from chest ct images using multi-objective differential evolution–based convolutional neural networks, Eur. J. Clin. Microbiol. Infect. Dis. (2020) 1–11.

[36] Y. Song, S. Zheng, L. Li, X. Zhang, X. Zhang, Z. Huang, J. Chen, H. Zhao, Y. Jie, R. Wang, et al., Deep Learning Enables Accurate Diagnosis of Novel Coronavirus (Covid-19) with Ct Images, medRxiv, 2020.

[37] C. Szegedy, A. Toshev, D. Erhan, Deep neural networks for object detection, in: Advances in Neural Information Processing Systems, 2013, pp. 2553–2561.

[38] L. Wang, A. Wong, Covid-net: A Tailored Deep Convolutional Neural Network Design for Detection of Covid-19 Cases from Chest Radiography Images, 2020 arXiv preprint arXiv:2003.09871.

[39] 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), medRxiv, 2020.

[40] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, R.M. Summers, Chestx-ray8: hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2097–2106.

[41] Y.H. Wu, S.H. Gao, J. Mei, J. Xu, D.P. Fan, C.W. Zhao, M.M. Cheng, Jcs: an Explainable Covid-19 Diagnosis System by Joint Classification and Segmentation, 2020 arXiv preprint arXiv:2004.07054.

[42] X. Xu, X. Jiang, C. Ma, P. Du, X. Li, S. Lv, L. Yu, Y. Chen, J. Su, G. Lang, et al., Deep Learning System to Screen Coronavirus Disease 2019 Pneumonia, 2020 arXiv preprint arXiv:2002.09334.

[43] J. Yosinski, J. Clune, Y. Bengio, H. Lipson, How transferable are features in deep neural networks?, in: Advances in Neural Information Processing Systems, 2014, pp. 3320–3328.

[44] Y. Zhang, Q. Yang, A Survey on Multi-Task Learning, 2017 arXiv preprint arXiv: 1707.08114.

[45] J. Zhao, Y. Zhang, X. He, P. Xie, Covid-ct-dataset: A Ct Scan Dataset about Covid-19, 2020 arXiv preprint arXiv:2003.13865.

[46] C. Zheng, X. Deng, Q. Fu, Q. Zhou, J. Feng, H. Ma, W. Liu, X. Wang, Deep Learning-Based Detection for Covid-19 from Chest Ct Using Weak Label, medRxiv, 2020.

[47] T. Zhou, S. Canu, S. Ruan, An Automatic Covid-19 Ct Segmentation Based on U-Net with Attention Mechanism, 2020 arXiv preprint arXiv:2004.06673.

Hua Li, Publications:

1- Use of image registration and fusion algorithms and techniques in radiotherapy: Report of the AAPM Radiation Therapy Committee Task Group No. 132. KK Brock, S Mutic, TR McNutt, H Li, ML Kessler Medical physics 44 (7), e43-e76

2- Vessels as 4-D curves: Global minimal 4-D paths to extract 3-D tubular surfaces and centerlines H Li, A Yezzi IEEE transactions on medical imaging 26 (9), 1213-1223

3- Current status of Radiomics for cancer management: Challenges versus opportunities for clinical practice H Li, I El Naqa, Y Rong Journal of applied clinical medical physics 21 (7), 7

Amine Amyar, Publications:

1- Amyar, A. et al. "

3-D RPET-NET: development of a 3-D pet imaging convolutional neural network for radiomics analysis and outcome prediction." *IEEE Transactions on Radiation and Plasma Medical Sciences* 3.2 (2019): 225-231.

2-Amyar, Amine, et al. "Radiomics-net: Convolutional neural networks on FDG PET images for predicting cancer treatment response." *Journal of Nuclear Medicine* 59.supplement 1 (2018): 324-324

3- Amyar, Amine, et al. "Weakly supervised pet tumor detection usingclass response." *arXiv preprint arXiv:2003.08337* (2020).

Su Ruan, Publications:

1- Nie, Dong, et al. "Medical image synthesis with context-aware generative adversarial networks." *International Conference on Medical Image Computing and Computer-Assisted Intervention.* Springer, Cham, 2017.

2- Joint tumor segmentation in PET-CT images using co-clustering and fusion based on belief functions C Lian, S Ruan, T Denoeux, H Li, P Vera IEEE Transactions on Image Processing 28 (2), 755-766

3- Segmenting multi-source images using hidden markov fields with copula-based multivariate statistical distributions J Lapuyade-Lahorgue, JH Xue, S Ruan IEEE Transactions on Image Processing 26 (7), 3187-3195

Romain Modzelewski, Publications:

1- Belharbi, Soufiane, et al. "Spotting L3 slice in CT scans using deep convolutional network and transfer learning." *Computers in biology and medicine* 87 (2017): 95-103.

2- Paul, Desbordes, et al. "Feature selection for outcome prediction in oesophageal cancer using genetic algorithm and random forest classifier." *Computerized Medical Imaging and Graphics* 60 (2017): 42-49.

3- Lanic, Hélène, et al. "Sarcopenia is an independent prognostic factor in elderly patients with diffuse large B-cell lymphoma treated with immunochemotherapy." *Leukemia & lymphoma* 55.4 (2014): 817–823.