ELSEVIER

Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

Radiography 29 (2023) 109–118

Contents lists available at ScienceDirect

Radiography

journal homepage: www.elsevier.com/locate/radi

Strong semantic segmentation for Covid-19 detection: Evaluating the use of deep learning models as a performant tool in radiography

H. Allioui [a,*], Y. Mourdi [b], M. Sadgal [a]

[a] Computer Sciences Department, Faculty of Sciences Semlalia, Cadi Ayyad University, Morocco
[b] Polydisciplinary Faculty Safi, Cadi Ayyad University, Morocco

ARTICLE INFO

Article history:
Received 30 June 2022
Received in revised form
12 October 2022
Accepted 17 October 2022
Available online 24 October 2022

Keywords:
Covid-19
CT images
Deep learning
Image processing
Semantic segmentation

ABSTRACT

Introduction: With the increasing number of Covid-19 cases as well as care costs, chest diseases have gained increasing interest in several communities, particularly in medical and computer vision. Clinical and analytical exams are widely recognized techniques for diagnosing and handling Covid-19 cases. However, strong detection tools can help avoid damage to chest tissues. The proposed method provides an important way to enhance the semantic segmentation process using combined potential deep learning (DL) modules to increase consistency. Based on Covid-19 CT images, this work hypothesized that a novel model for semantic segmentation might be able to extract definite graphical features of Covid-19 and afford an accurate clinical diagnosis while optimizing the classical test and saving time.
Methods: CT images were collected considering different cases (normal chest CT, pneumonia, typical viral causes, and Covid-19 cases). The study presents an advanced DL method to deal with chest semantic segmentation issues. The approach employs a modified version of the U-net to enable and support Covid-19 detection from the studied images.
Results: The validation tests demonstrated competitive results with important performance rates: Precision (90.96% ± 2.5) with an F-score of (91.08% ± 3.2), an accuracy of (93.37% ± 1.2), a sensitivity of (96.88% ± 2.8) and a specificity of (96.91% ± 2.3). In addition, the visual segmentation results are very close to the Ground truth.
Conclusion: The findings of this study reveal the proof-of-principle for using cooperative components to strengthen the semantic segmentation modules for effective and truthful Covid-19 diagnosis.
Implications for practice: This paper has highlighted that DL based approach, with several modules, may be contributing to provide strong support for radiographers and physicians, and that further use of DL is required to design and implement performant automated vision systems to detect chest diseases.
© 2022 The College of Radiographers. Published by Elsevier Ltd. All rights reserved.

Introduction

The understanding of 3D images is highlighted, nowadays, as an important concern in the field of computer vision, considering the increasing number of applications that learn and develop from inferring knowledge from images. Semantic segmentation is a high-level task that guarantees a complete understanding of the 3D image.[1,2] It consists of understanding the image universe by assigning each pixel to a class of objects, which is essential for a complete understanding of a whole 3D image.[1–3] The semantic segmentation aims to classify each pixel belonging to a specific label. It does not differentiate the several occurrences of the same object. For example, if there are two nodules in a chest image, semantic segmentation provides the same label to all the pixels of both nodules.[4,5]

Medical images come from various imaging techniques such as ultrasound, X-ray, computed tomography (CT), and magnetic resonance imaging (MRI).[6] The goal of using semantic segmentation is to gather pixels in a meaningful way so that pixels belonging to a specific object are grouped separately. Hence, numerous research studies are seeking to define and provide new methods for more speediness, efficient interactions, and accurate processes to enhance traditional image segmentation. Yet, no universal or standard method could be employed for segmentation. Several factors could contribute to reaching high-quality results, including accurate preparation of the needed input data. To overcome the related data

* Corresponding author. Computer Sciences Department, Faculty of Sciences Semlalia, Bd Prince Moulay Abdellah, Marrakech 40000, Morocco.
E-mail addresses: hananeallioui@gmail.com (H. Allioui), mourdiyoussef@gmail.com (Y. Mourdi), sadgal@hotmail.com (M. Sadgal).

https://doi.org/10.1016/j.radi.2022.10.010
1078-8174/© 2022 The College of Radiographers. Published by Elsevier Ltd. All rights reserved.

issues, preparing learning and test sets is usually a time-overwhelming and costly task, but at the same time crucial to obtain accurate results.[7]

A key factor in the segmentation, of medical images, is the higher cost of manual class extraction. Manual segmentation is performed by experts, who are often radiologists or specialized clinicians. This process is usually performed in a slice-by-slice way, with the expert encircling the region of interest (ROI) or labeling the voxels of interest. Manual segmentation, which is considered the gold standard, utilizes expert knowledge and experience. Yet, it is very time-consuming and susceptible to a large variability, due to increasing workload as well as exhaustion factors. Automatic methods do not rely on experts' interaction. The advantage of these methods is that once the approach has been built, the image segmentations can be accomplished relatively much faster than manual processes. Consequently, decreasing human interaction can optimize segmentation costs, regarding time and efforts, to facilitate Covid-19 detection. Therefore, DL and semantic segmentation, as automatic techniques, can provide highly relevant results.

Although various methods can automatically segment CT images, when an unforeseen and unprecedented illness arises, as in the case of Covid-19, the inexistence of huge training datasets with annotations remains the main challenge. Thus, the DL revolution has introduced several advanced approaches that are valuable and applicable to almost all medical imaging modalities.[8,9] The wide adoption of DL techniques for solving segmentation problems is based on the use of deep neural networks that further surpass the traditional segmentation approaches in terms of accuracy and efficiency.[10,11] For that reason, many cutting-edge scientific approaches related to DL rely on these capabilities, especially in medical imaging.

Mainly, current clinical research studies are focused on the Covid-19 diagnosis by:

- PCR (Polymerase Chain Reaction) which is the key test for the diagnosis of Covid infection since the beginning of the epidemic. SARS-CoV-2 PCR or RT-PCR is a laboratory technique for detecting the genetic material of the virus. It requires a sample to be taken from the most accessible place where the concentration of the virus is the most important such as the nasopharynx, behind the nostril ducts. The Covid PCR test can be used to determine, at the time of sampling, whether the person is a carrier of the Covid-19 virus.
- The rapid diagnostic test, which is an antigenic test, is recommended for patients suspected of having SARS-CoV-2 coronavirus. These tests provide information on the presence of antigens, not antibodies, and therefore serve the same purpose as a PCR test for Covid-19.
- Chest imaging still plays a major role in the detection of Covid-19. However, its indications need to be reviewed by clinical experts considering the experience gained and the progress made in access to virological tests.

Unlike the first wave of Covid-19, where patients arrived at the emergency room without a diagnosis, the majority now arrive at the emergency room with a positive PCR test. For these patients, a CT scan is not a systematic test, but it is indicated when there are signs of poor respiratory tolerance (dyspnea, desaturation, or severe hypoxemia). The CT scan allows grading of the severity of lung damage, which provides prognostic information. It can also help in the orientation of patients (hospitalization vs ambulatory management) even if this decision is based primarily on clinical exams. In the same context, a variety of segmentation methodologies have been approved by the scientific community considering their effectiveness in detecting chest diseases were examined.[12]

For this purpose, automated detection systems can enhance the sensitivity of radiologists to accurately analyze chest irregularities.[3,6,7] However, automatic diagnosis is still an open challenge since feature quantification changes remain complicated to predict. Accordingly, numerous semantic segmentation variants have been developed based on the use of different components, which provides self-improvement of the whole vision system.[12] However, the previous approaches have rather low scalability and remain limited to low-dimensional subjects concerning the increased complexity of the medical imaging field. Yet, the use of advanced techniques has offered new solutions to solve these problems and provide a powerful approximation to the deep feature representations and functions previously used.

The objective of this study is to present an advanced approach using DL techniques to automatically detect Covid-19. Therefore, the current study presents a fully deep convoluted architecture.

Methods

In this study, feature learning and prediction are performed using DL to accomplish different tasks of CT semantic segmentation.

The CT semantic segmentation process

The presented automatic process provides a CT processing unit, which can segment inhomogeneous object boundaries, as is often the case with biomedical images. The semantic segmentation of chest CT images intends to associate each pixel with a group of a predefined set. The presented process is based on a developed version of classical convolutional neural networks (CNN).[13] The adopted architecture has a combined structure that samples the input CT image to produce a high dimensional characteristics result.

This approach plays an important role in the separation of tissues that belong to different parts of the chest in CT images. The general structure, illustrated in Fig. 1, can be used for semantic segmentation using some new features for the fulfillment of a robust and efficient segmentation process.

The adopted architecture is composed of two main stages:

- Training: The model uses the labeled data, so it can be trained to identify and learn ROI and segment the input images. At this step, the training images are assessed to form a consistent foundation for network learning. The training data are processed utilizing the initial weights. Then, the obtained results are compared, then classified to minimize the training loss.[14,15] Afterward, the network adapts the weights. This process will be repetitive, and the weights will be constantly adjusted. The key goal of training the model is to discover and learn all the features of CT images and use this knowledge to make an accurate diagnosis later.
- Test and analysis: the fully end-to-end architecture locate object to ensure proper processing of Chest CT images. The proposed model is evaluated constantly by computing the learning accuracy and loss. This constant analysis allows ensuring that the model performs truthful analysis to detect and locate objects within the CT image.

Semantic segmentation is a vital task for a complete understanding of 3D and CT images. For this purpose, the use of DL allows the building of a model capable of learning samples from the visual inputs to predict the classes of objects constituting the CT image.

Several segmentation models were tested among the vast amount of DL versions: U-Net[16] as well as other 3D volumetric networks (such as 3D V-net,[17] 3D U-Net,[18] U-Net++,[19] Dense-UNet,[20] and Attention-UNet[21]). They were selected for a particular

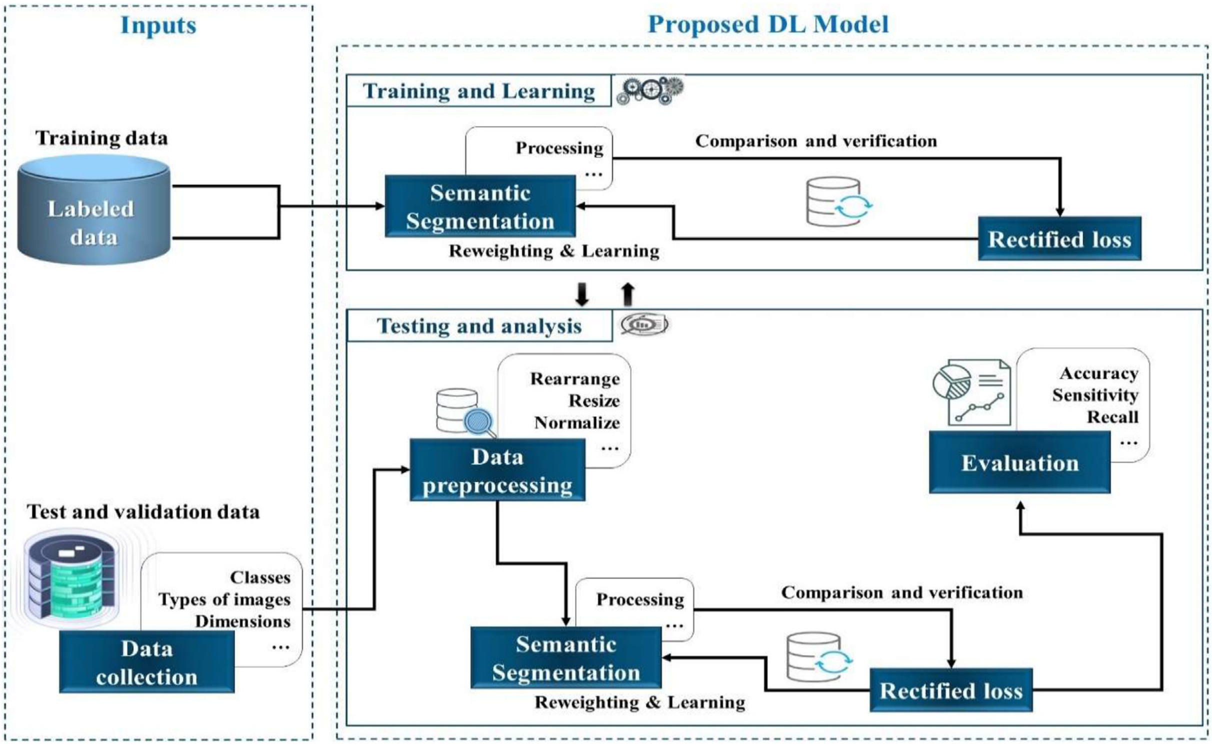

Figure 1. The adopted processing architecture.

motive related to our objective. As U-Net[14] was built for real-time segmentation of biomedical images, the other models offer optimized versions to allow better learning of the global context of an image. To obtain an accurate comparison of these models, they were trained and tested using the same datasets. The next task was to build an appropriate and optimized network for semantic segmentation. The presented network in this study combines U-Net,[16] and Long short-term memory (LSTM).[22]

Clinical data is a vital resource for effective medical and health research. Clinical data can be collected either during medical examinations of patients or as part of a clinical study program. In the case of the present study, the choice to use publicly available datasets, rather than those directly from clinical practice, is supported by the fact that it can provide a solid baseline for future work because they were used in important research studies led by a huge number of scientists. Testing different approaches using publicly available clinical datasets will decrease the time of ethics procedures, thus providing a solid basis for future work aiming at the realization of a monitoring environment of a narrow range of key data for certain thoracic diseases to be able to provide essential information for the management of the patient's condition in real-time.

As an early diagnosis of Covid-19 is important to raise patients' survival rate, DL-based methods are widely used for diagnosis and detection. One of the benefits of the proposed method is that it reduces human error in Covid-19 diagnosis. Chest images may contain interferences such as nodules or small nodes. For this purpose, noise elimination is required. The noise removal in CT images is very significant for the correct semantic segmentation.

Using DL architecture for both 2D and 3D data

For training, the study uses a training dataset of Covid-19 viral pneumonia as well as other interstitial lung diseases to analyze the impact of Covid-19 areas. The proposed architecture aims to process both a 3D and 2D semantic segmentation approach and target feature learning of high-resolution 2D images, by taking three axial slices as input to the network. The segmentation model is trained first by collecting the annotated data prepared by the label block.

- For 2D images: The input images are resampled to have a resolution of 0.6×0.6 mm. Then, the geometric center is calculated to crop the images with a bounding box of size 256×256. Subsequently, the network is trained with the Adam optimizer with decoupled weight decay regularization.[13] A dice loss is applied to the output prediction to penalize the difference from the Covid-19 truth plane annotation during training.[17]
- For 3D images: The input 3D CT volumes are preprocessed resampled to 1×1x3mm resolution and cropped based on the lung segmentation on a $224 \times 224 \times 224$ size box. The 3D network is trained using the AdaBound optimizer which adaptively combines the Adam optimizer with The Saccharomyces Genome Database (SGD) for faster convergence.[23] The Jaccard index[24] is used as a training loss function to have stable behavior for unbalanced labels. Among the advantages of using a 3D architecture is the ability to use the 3D context to deal with partial volume effects in the 2D plane as well as the global lung context.

Ethical approval

Ethical approvals were checked and related arrangements and data collection permissions were achieved respecting Cadi Ayyad University procedure and the Moroccan law n°09-08.[25] This law provides effective protection to people against abuses of the various types of data use that could violate their privacy. In addition, respecting this law, no ethical approvals are mandatory in this study where ethical standards were always respected, involving the use of online accessible data.

Evaluation baselines

For the Covid-19 infection region detection experiments, the proposed model is compared with previous segmentation models in the medical field. As the emergence of Covid-19 is very recent, none of the large repositories contains a large collection of Covid-19 labeled data, which requires relying on different image sources of normal, pneumonia, and Covid-19 cases. As a first step, different 3D images were collected to perform semantic segmentation, which can facilitate the detection and classification of Covid-19 infections from four open datasets:

- Cov19-A: for subjects without Cov-19, The Cancer Imaging Archive (TCIA) dataset[26] contains a total of 60 3D CT lung scans, to train the network for lung delineation learning. These data can be viewed in the publicly available imaging archive.[27]
- Cov19-B: The public data set of Ma et al.[28] consists of 20 annotated volumes of Covid-19 chest scans for training the tag system. The scans were collected from the Coronacases initiative and Radiopaedia and were licensed under CC BY-NC-SA.
- Cov19- C: consists of 9 volumetric Covid-19 CTs in DICOM format containing a total of 829 axial slices, Each CT axial slice was labeled by a radiologist for ground truth, consolidation, and pleural effusion. These data are publicly available.[29]
- Cov19- D: consists of 20 volumes of Covid-19 CT images posted online. Lungs and infection areas were labeled and verified by three experienced radiologists.[30]

Statistics for the four data sets are presented in Table 1.

The main objective of the data selection was to make the data public so that it could be accessed and extended by researchers. The use of these datasets in future studies may also allow for a more efficient diagnosis of patients with Covid-19.

Training details

Training, validation, and testing of each experiment were performed on a machine equipped with an NVidia TitanX GPU with 12 GB of memory. The methods were implemented using the Python 3.6 library, 1.1.0.[31] The training parameters for each method were initialized using the initialization of He et al. when training from scratch[32] and were optimized using stochastic gradient descent with a momentum of 0.9, and the initial learning rate set to $10e^{-6}$. The initial models were trained using CT scans from the Cov19-A data set. Therefore, these models may not be familiar with the visual patterns of the Covid-19 scans. For effective training on the new visual models, all models were trained with Cov19-A and Cov19-B using a combined loss between the generalized dice loss[33] (used to train the initial models) and the top-K cross-entropy loss (K = 30% of all voxels in the input). The top-K cross-entropy loss was applied simply as the voxel-level cross-entropy loss but selected only the K voxels with the largest cross-entropy to back-propagate.

Pre- and post-processing

All training and test scans were normalized by setting the intensity values to the range [−1200–400] before resizing them to [0–1]. Then, all scans were down-sampled by trilinear interpolation to achieve 256 × 256 in-plane resolution, while the z-spacing is adjusted to make the scan isotropic. The input size of the second network for the proposed method consisted of two 3D patches of size 116 × 116 × 116. The preprocessed scan was downsampled by a factor of 2 using trilinear interpolation as input for the first step (padding from zero is required if the z-axis size is not divisible by 16). The SoftMax probability outputs of all 3D patches in the second step were overlaid by sliding over the entire scan without overlap to produce a scan-level probability map, which is used to generate the final prediction by assigning each voxel to the label with the highest probability. In post-processing, the predictions were then resampled by nearest-neighbor interpolation to match the original scan resolution. All evaluations are performed using reference predictions and segmentations at the original resolution.

Table 1
Statistical description of evaluation data sets.

Datasets	3D Volumes	Covid-19 subjects	% Slice with infection
Cov19-A	60	_	_
Cov19-B	20	20	100%
Cov19-C	9	9	44.9%
Cov19-D	20	20	52.3%

Subjective segmentation quality evaluation: Likert scale

Accurate segmentation of 3D lung images was used as the basis for qualitative and quantitative analysis. Thus, qualitative results derived from the DL approach were visually evaluated by radiology experts (the first had 14 years of cumulative experience in lung imaging, the second had 7 years of experience in cardiopulmonary imaging, and the third had 2 years of experience in general radiology), who independently visualized the segmentation. The three radiologists did not consider the clinical status of the patients or the data sources. The procedure for scoring results was as follows: The three radiology experts reviewed the segmentation results displayed as detected regions. They used a scoring criterion based on the adequacy of the segmentation task with the actual lung opacification. Specifically, the degree of adequacy was qualified using a Likert[34] score ranging from 0 to 5.

Widely used in surveys and questionnaires, the Likert scale consists of evaluating a respondent's opinion on a specific subject. This measurement scale offers several choices of answers allowing each respondent to express his or her degree of agreement or disagreement with a given question. Responders were able to specify their level of agreement according to five points: (1) Strongly oppose, (2) Disagree, (3) Not sure about the concept or results, (4) Agree, and (5) Strongly agree. To reduce the subjectivity of the radiologist's assessment, the final score was the average of the three scores for each image. The experienced radiologists individually checked the segmented CT images and classified the segmentation quality of each succeeding object and the complete result in comparison to the corresponding chest CT images.

The radiologists evaluated the semantic segmentation results as 5, 5, and 4, respectively, on a five-point Likert scale. The assessment result on a five-point Likert scale is presented in Fig. 6. The parameters to evaluate for the experimentation were proposed by the authors:

1) Covid-19 areas are well perceived.
2) Covid-19 is well observed.
3) Develop knowledge of the disease.
4) Could help to decrease errors.
5) Improve understanding.
6) Provide a realistic view of the Covid-19 case.

Evaluation measures

Evaluating a semantic segmentation model and providing a fair comparison remains a complex task. This is due to the numerous existing evaluation measurement techniques, and the wide use of different distinct data sets. Once input images are processed, qualitative and quantitative outcomes can be acquired. In this study, the proposed model, as well as the tested models, were tested using the same public datasets and evaluated with the same measures to afford an accurate comparison regarding the performance of the study proposal. The adopted evaluation metrics (Precision, F1-score, accuracy, sensitivity, and specificity)[35–37] are presented in Table 2.where:

- TP denotes the value of the true positive (the number of subjects having Covid-19)
- TN represents the true negative (the number of normal images)
- FP is the false positive (the number of control subjects with Covid-19)
- FN represents the false-negative value (the number of subjects with Covid-19 considered as controls).

Results

To prove the effectiveness of the presented model, various experiments were designed to evaluate the proposed method qualitatively and quantitatively on public datasets. The choice for these datasets was driven by their public access to the scientific community, their reputation, and the quality of the provided images.

After training the proposed model, numerous sequences of tests were achieved. These assessments allowed to measure the operational functioning of the presented semantic segmentation model, then contribute to the enhancement of the learning of the proposed model. Firstly, the proposed model was run to visualize all the possible differences between the normal case, and variant Covid-19 infections stage. Fig. 2 illustrates the difference between all the stages.

The second test concentrates on the segmentation results of the presented model compared to the ground truth. Fig. 3 shows an example of the differences between the segmented images and the segmentation ground truth. By analyzing the results of the detected areas, the areas related to Covid-19 can be correctly located on the images even with variations of Covid-19.

The performance of the proposed model, dedicated to the semantic segmentation of Covid-19 CT images, is assessed and compared to other tested methods as identified in Fig. 4. To make a fair comparison in the third test, a more detailed analysis was performed by comparing the results of the different tested approaches including the proposed one. This comparison is done by training the models using the same datasets as well as details and using the same validation protocols. The qualitative results for lung infection segmentation, shown in Fig. 3, indicate that the proposed method provides competitive results compared to the tested state-of-the-art methods.

Furthermore, the advantage of the 3D semantic segmentation strategy is also confirmed in Fig. 4. As can be observed, compared to other methods, the current approach gives segmentation results with more precise boundaries. Each volume of the used datasets (described in Table 1) contains hundreds of slices. 25% of these data were used to test the proposed approach, among the numerous tests, the best match has been selected and a bias is therefore introduced to present each area detected in comparison with the ground truth. In addition, the values in Table 3 aim to support these results.

The U-Net forms the basis of the tested models, so compared with its results, which remain modest, the other approaches remarkably offer better results at different progressive stages. The proposed method outperformed the other methods because it offered a tighter margin and could detect approximately the same objects compared to the ground truth. This illustrates the power of combining two important networks to form a strong semantic segmentation model.

In the previous tests, examples of the qualitative results were shown and analyzed. In this way, the evaluation of the proposed method proved effective results. The implementation of the semantic segmentation model reached 93,46% precision rate, 94,28% F1 score, 94,57% accuracy, 99,68% sensitivity, and 99,21% specificity. The performance results of the evaluated methods are presented in Table 3. Each of the models was tested using the same databases mentioned above. To ensure a fair comparison. The tested models prove competitive results, Attention-UNet and Dense-UNet. However, the proposed method was able to demonstrate its superiority in terms of performance. Results analysis has shown that a combination of U-Net and LSTM in the proposed model has significant effects on the performance of Covid-19 detection. The proposed system could distinguish Covid-19-related infections even in the early stages with high accuracy.

The tested methods were evaluated to describe how well they perform Covid-19 detection. Numerous loss variations were gained over the same periods. In fact, the variation in the results presented in Fig. 5 explains that most methods swiftly converge toward the minimum. Yet, the proposed model uses the total variation loss.[14,15] This strategy mimics the way real clinicians segment lung infection regions from CT slices, and thus achieves promising performance.

The final score obtained confirmed that the proposed method is considered appropriate to meet clinical requirements. This describes how the current semantic segmentation model can differentiate between lung abnormalities and detect Covid-19 while giving a clear definition of the patient stage to provide perceived information for clinicians to choose the right treatment process.

As the computational proposed segmentation method presented an excellent overlap with the other tested methods, it was necessary to have the opinion of human readers, the experienced radiologists. The proposed model successfully segmented the chest tissues while conserving the quantitative semantic segmentation close to the reference evaluation. The segmentation quality was rated by the three experts (Fig. 6). The overlaps between the proposed and the manual segmentations were not significantly different. The subjective evaluation by the experienced radiologists also resulted as good to excellent for Covid-19 detection.

Discussion

A successful and accurate diagnosis of Covid-19 remains vital for early and effective treatment. With this aim, numerous research studies have promoted the progress of DL use for early disease diagnosis.[14–19] Most of the tested techniques have proven a valuable overall accuracy above 90%. Nevertheless, continuous morphological changes by Covid-19 in the lung were not easy to distinguish. Deep neural models were experimented with using several segmentation techniques. The tested models demonstrated the appropriateness of DL techniques to provide strong tools for Covid-19 detection. Outcomes indicate that using DL models, to segment chest CT images, can provide potential support for radiologists and clinicians.

Table 2
The used evaluation metrics.

Metrics	Formulas
Precision	$\frac{TP}{TP + FP}$
F1- score	$2x\frac{\text{Precision x Recall}}{\text{Precision} + \text{Recall}}$
Accuracy	$\frac{TP + TN}{TP + 2TN + FP}$
Sensitivity	$\frac{TP}{TP + FN}$
Specificity	$\frac{TN}{TN + FP}$

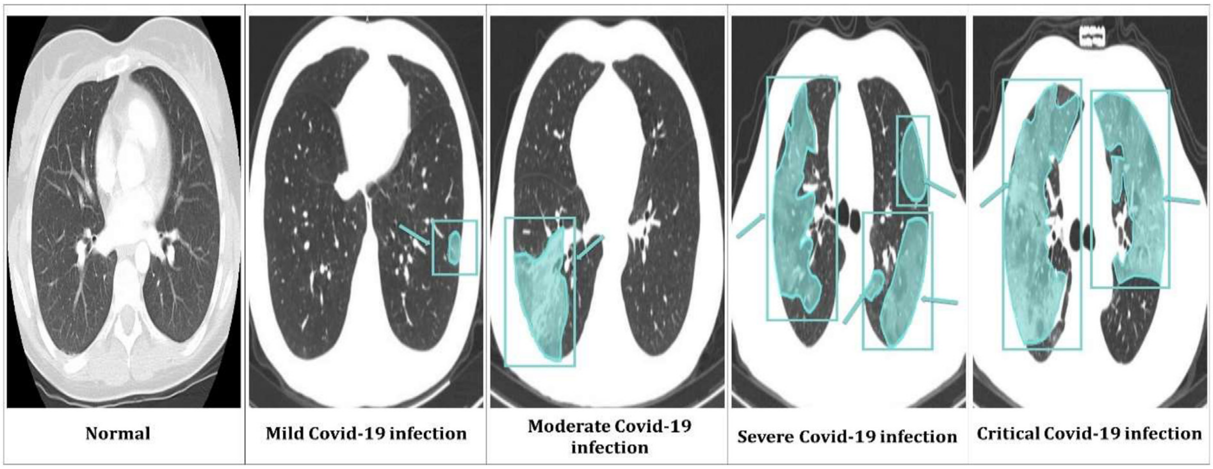

Figure 2. Examples of variations, caused by the Covid-19 infection, observed by the proposed approach.

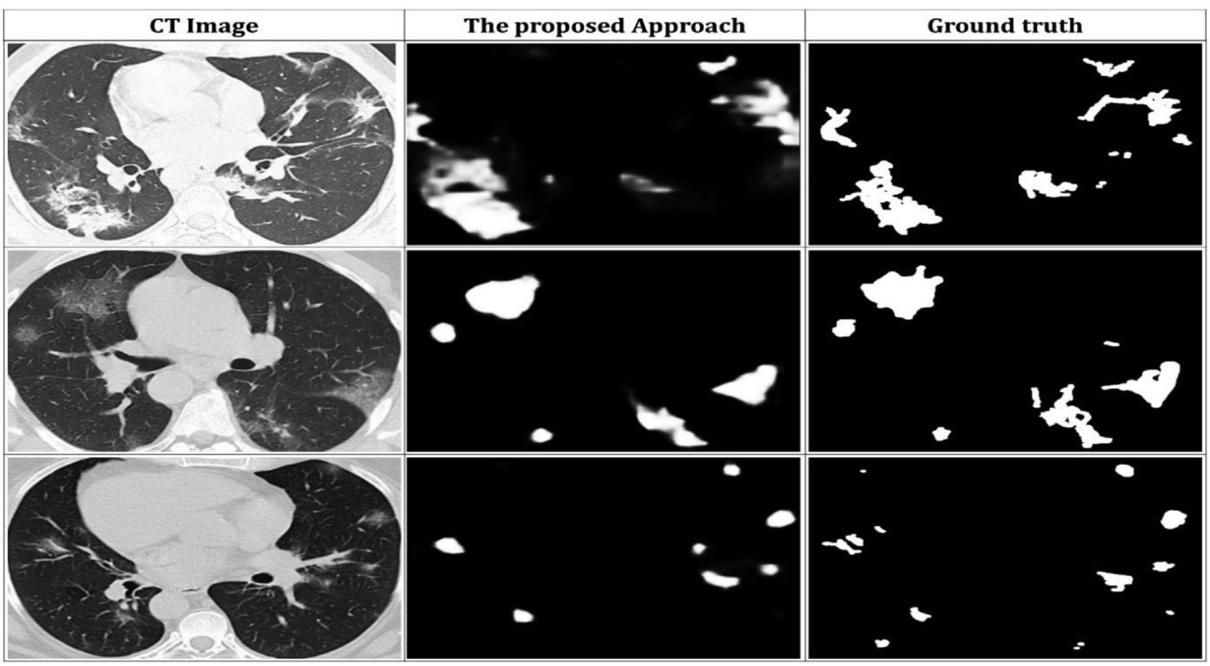

Figure 3. Qualitative comparison of Covid-19 infection segmentation results.

Various studies have aimed to employ DL in the radiology context to expand healthcare quality and simplify data retrieval. Thus, there are promising opportunities for using DL to improve Covid-19 detection and develop healthcare management.

In this paper, the presented results demonstrate the effort made to improve the chest semantic segmentation. During this study, the model architecture was trained by the CT dataset. Firstly, different segmentation architectures were tested to detect the best combination of base networks. The training, of all tested models, was performed with the same datasets, and the best results were obtained by combining the U-Net architecture with LSTM as a base network for semantic segmentation. The proposed approach (U-Net + LSTM) reached an average precision of 90,96% (±2,5%) while U-Net 69,23% (±10,6). Nevertheless, 3D V-net, 3D U-Net, U-Net++, Dense-UNet, and Attention-UNet performed 71,73% (±10,1), 77,19% (±6,7), 85,94% (±4,2), 87,65% (±4,3) and 89,28% (±4,4) respectively. Specifically, it produces segmentation results that are extremely close to the ground truth with fewer miss-segmented tissue. Similarly, the performance rates prove that the proposed method outperformed all the other models in the metrics used, except the F1 score for Dense U-net, and this was relatively comparable.

The response frame rate of the real-time segmentation was satisfactory regarding the hardware used for testing (104s). With better hardware abilities the proposed approach can reach a response frame rate of approximately 40s, which is a decent value for a segmentation, where the restrictions on real-time are not so severe.

Since the current study aims to present an advanced approach using DL techniques to automatically detect Covid-19, the choice of combining U-Net and LSTM networks ensures a good compromise between speed and precision. As illustrated in Fig. 5, the proposed model was trained until the loss converges, which means that the model can perform better on a new validation dataset. All the tested models start converging toward the minimum in an average time of the 20 s. However, the use of the rectified loss allowed the

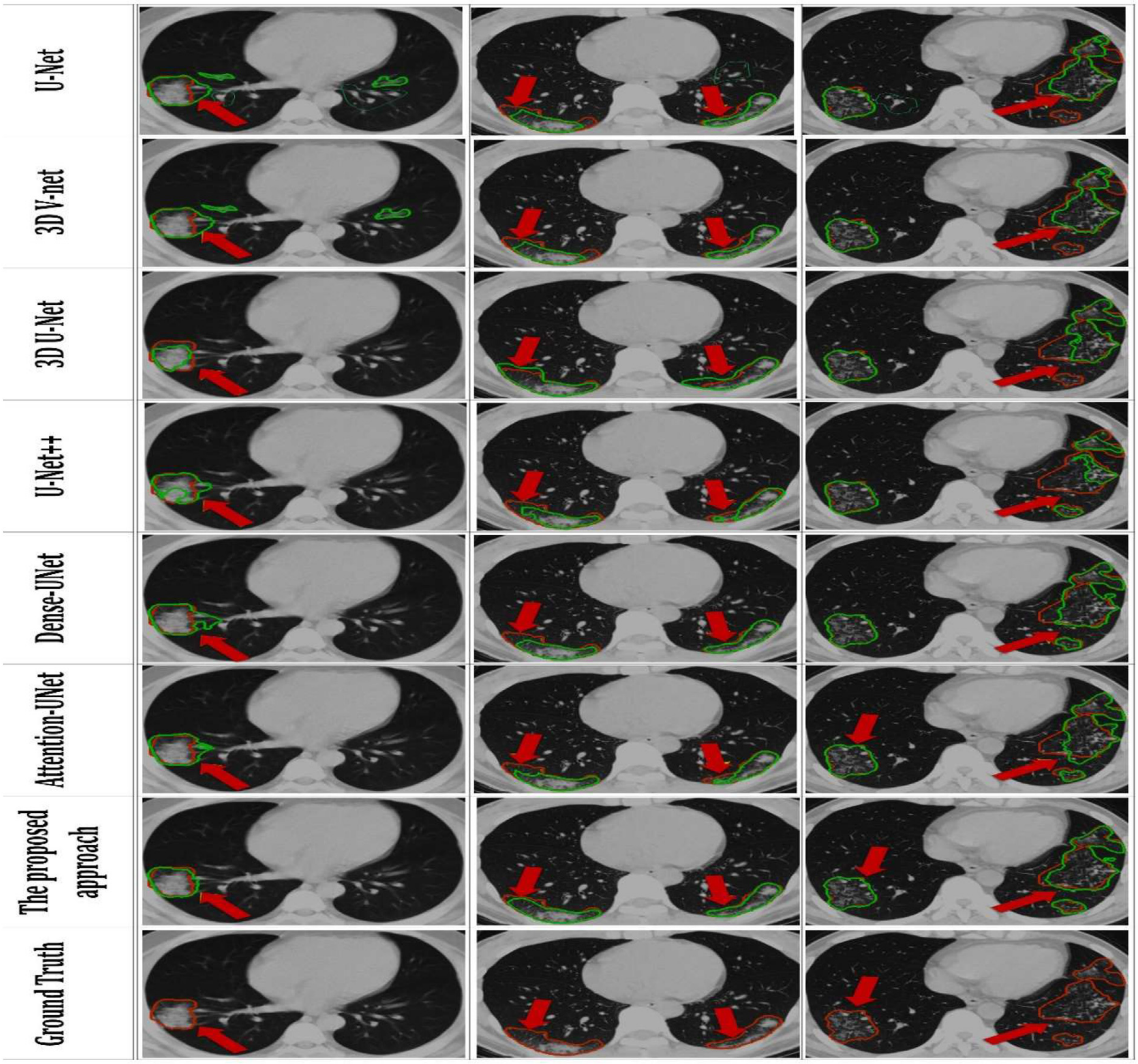

Figure 4. Visual comparison of the segmentation performance of the multi-class Covid-19 infections of different networks tested (red indicates true infection areas, green indicates detected areas).

Table 3
Quantitative result of the performance measures of the tested methods.

Architecture	Precision %	F1- score%	Accuracy%	Sensitivity%	Specificity%
U-Net[16]	69.23 ± 10.6	73.04 ± 8.47	76.85 ± 10.45	75.29 ± 7.89	67.60 ± 12.83
3D V-net[17]	71.73 ± 10.1	81.64 ± 7.5	90.68 ± 3	95.59 ± 3.5	89.29 ± 3.3
3D U-Net[18]	77.19 ± 6.7	61.23 ± 3.4	89.27 ± 1.6	51.21 ± 5.3	83.84 ± 2.1
U-Net++[19]	85.94 ± 4.2	85.48 ± 1.8	92.84 ± 1	85.29 ± 4.4	85.92 ± 5
Dense-UNet[20]	87.65 ± 4.3	91.96 ± 2.2	91.08 ± 3.2	86.89 ± 5.4	86.31 ± 2.4
Attention-UNet[21]	89.28 ± 4.4	86.40 ± 2.1	90.32 ± 2.9	92.81 ± 4.2	91.42 ± 1.3
Proposed approach	90.96 ± 2.5	91.08 ± 3.2	93.37 ± 1.2	96.88 ± 2.8	96.91 ± 2.3

proposed model to learn quickly and produce semantic segmentation results.

When testing CT data, comparing the tested models as well as the proposed approach demonstrated a significant increase in segmentation performance combining U-Net with another network, especially LSTM. In the very last line of Table 3, it is obvious that the proposed method reached the best results. These results surpassed most of the tested approaches.

Even with the strengths of the proposed method, it still has certain limits. Although the comparison with radiology experts was based on three different datasets, it remains essential to survey the variations of the same subject. Another drawback, the study used only online databases during the experiments, and the measured conclusions are likely to be beyond general. Lastly, even though the proposed method is aligned with a DL model, it primarily intends to enhance the neural network learning to process Chest CT images.

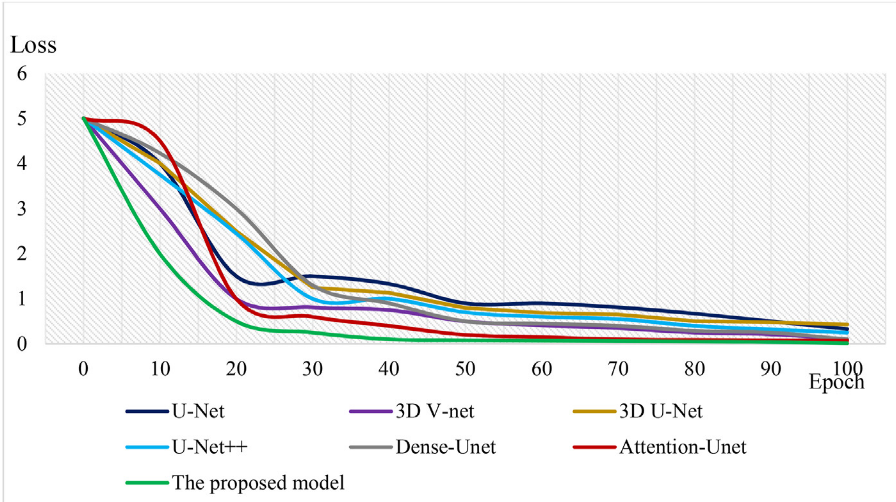

Figure 5. Comparison of losses between the three methods.

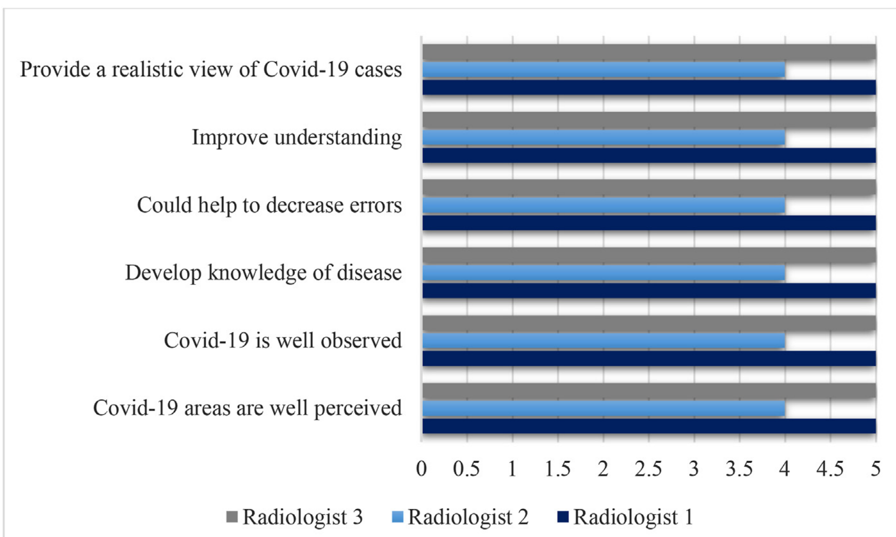

Figure 6. Subjective assessment results in a five-point Likert scale.

Therefore, it is significant to highlight the need to train the adopted network on much larger datasets, this will potentially reinforce the processing of the proposed model and decrease the uncertainty rate considerably while warranting a good exchange between sensitivity and specificity.

To better ensure clinical applicability, further prospective and retrospective studies are required. In future works, the DL model would be adopted to explore different chest features to compute the prevalence of Covid 19. It is also essential to be able to employ the proposed model to segment other medical imaging (X-ray, PET, and MRI) to smooth the discovery and management of abnormal tissue.

The qualities of semantic segmentations were evaluated by three experienced radiologists on a five-point Likert scale. The result of the proposed model of semantic segmentation was compared to manual segmentations. Likert results showed that the DL-based proposed model Offred promising results in most of the cases. In clinical routine, semantic segmentation can be used as a higher quality control measure to warn physicians about possible related Covid-19 problems.

The visual investigation results suggest that the model error was low compared to the manual segmentation, but in areas with a mismatch, it is suggested to examine more carefully into object segmentation to avoid the underestimation of tumors and the overestimation of healthy tissue.

The presented method performance was evaluated using the commonly used assessment scores such as precision, F1-score, accuracy, sensitivity, and specificity. Strengths of this work also include the potential to dectect the Covid-19, visualize the affected area, and rapidly track the disease variations. Moreover, the proposed model has the ability to segment chest CT cimages even with low-intensity contrast, and to diffirenciate between abnormal and healthy tissues. Even though the achieved promising results, it is worth mentioning that there are some limitations.

The semantic segmentation quality assesment require reliable ground truth (GT) of the objects and mask structures. Therefore, an experienced medical expert remains the only one that can provide the most precise manual segmentation that serves as a reference. Second, Covid-19 have imaging features similar to the pneumonia caused by other types of viruses. As the study used online available datssets, it was not possible to detect other viral pneumonia, for comparison purposes, due to the lack of successive laboratory and clinical evidence for each of these cases. As future work, Authors consider extending the validation of the proposed model by collecting the chest CT images from different severity types of viral pneumonia through numerous institutions and countries.

Furthermore, the presented DL model should be optimized to separately pinpoint segment patterns that are classified as ground-glass opacity, crazy paving, and consolidation. Also it would be more intersting to combine imaging data with clinical indications and laboratory analysis results to provide a differential detection, and diagnosis. Since Covid-19 continues to spread across the world following an unpredicted strategy, the wide adoption of advanced DL-based tools will be helpful to health systems facing similar challenges, including abnormalities caused by other viruses and diseases.

Conclusion

The current study presented a DL method based on U-Net and LSTM architectures to provide an improved semantic segmentation of chest images to detect Covid-19. The results of the proposed method revealed its robustness compared to the other tested methods. From the experiences, it should be reported that the incorporation of various networks can be utilized to process medical images without further computational cost. The proposed method permits the combination of U-Net and LSTM information to provide a performant segmentation pipeline. Moreover, the proposed approach can automatically segment chest CT images and offer an accurate analysis of chest structures to detect Covid-19. The use network for semantic segmentation, for chest CT images, has been trained from scratch, so it is expected to be pertinent for many chest analyses.

The DL network has been capable to reach competitive results in distinguishing Covid-19 regions. In conclusion, the study hypothesizes that a larger training set would further improve performance, especially for cases with gross pathological changes that are not yet well represented in current training scans. Nevertheless, the results presented are sufficient for further analysis and would be useful in other research topics in the future.

Author contribution

Hanane Allioui: Methodology, Investigation, Writing - Review & Editing, Supervision; **Youssef Mourdi**: Investigation, Writing - Review & Editing; **Mohamed Sadgal**: Methodology, Writing - Review & Editing.

Conflict of interest statement

None.

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

1. Li F, Tran L, Thung KH, Ji S, Shen D, Li J. A robust deep model for improved classification of AD/MCI patients. *IEEE J Biomed Health Inform* 2015;**19**(5):1610–6.
2. Zeng N, Zhang H, Song B, Liu W, Li Y, Dobaie AM. Facial expression recognition via learning deep sparse autoencoders. *Neurocomputing* 2018;**273**:643–9. https://doi.org/10.1016/j.neucom.2017.08.043.
3. Lin W, Tong T, Gao Q, Guo D, Du X, Yang Y, et al. Convolutional neural networks-based MRI image analysis for the Alzheimer's disease prediction from mild cognitive impairment. *Front Neurosci* 2018;**12**:777. https://doi.org/10.3389/fnins.2018.00777.
4. Sundaram SG, Aloyuni SA, Alharbi RA, Alqahtani T, Sikkandar MY, Subbiah C, et al. Deep transfer learning based unified framework for COVID19 classification and infection detection from chest X-ray images. *Arabian J Sci Eng* 2022;**47**:1675–92. https://doi.org/10.1007/s13369-021-05958-0.
5. Chavan M, Varadarajan V, Gite S, Kotecha K. Deep neural network for lung image segmentation on chest X-ray. *Technologies* 2022;**10**:105. https://doi.org/10.3390/technologies10050105.
6. Allioui H, Sadgal M, El Fazziki A. Deep MRI segmentation: a convolutional method applied to Alzheimer disease detection. *Int J Adv Comput Sci Appl* 2019;(11):10.
7. Litjens G, kooi T, Bejnordi BE, Setio AA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. *Med Image Anal* 2017;**42**:60–88.
8. Greenspan H, Ginneken BV, Summers RM. Guest editorial deep learning in medical imaging: overview and future promise of an exciting new technique. *IEEE Trans Med Imag* 2016;**35**(5):1153–9.
9. Eigen D, Fergus R. Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture. In: *IEEE international conference on computer vision*; 2015. p. 2650–8.
10. Nie D, Wang L, Gao Y, Shen D. Fully convolutional networks for multi-modality isointense infant brain image segmentation. In: *Proc IEEE 13th international symposium on biomedical imaging (ISBI)*; 2016. p. 1342–5. https://doi.org/10.1109/ISBI.2016.7493515.
11. Shin HC, Roth HR, Gao M, Lu L, Xu Z, Nogues I, et al. Deep convolutional neural networks for computer-Aided detection: CNN architectures, dataset characteristics, and transfer learning. *IEEE Trans Med Imag* 2016;**35**(5):1285–98.
12. Salehi M, Ardekani MA, Taramsari AB, Ghaffari H, Haghparast M. Automated deep learning-based segmentation of COVID-19 lesions from chest computed tomography images. *Pol J Radiol* 2022;**87**(1):478–86. https://doi.org/10.5114/pjr.2022.119027.
13. Kamnitsas K, Ledig C, Newcombe VFJ, Simpson JP, Kane AD, Menon DK, et al. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med Image Anal* 2017;**36**:61–78.
14. Brosch T, Yoo Y, Tang LY, Li DK, Traboulsee A, Tam R. Deep convolutional encoder networks for multiple sclerosis lesion segmentation. In: *MICCAI 2015*; 2015. p. 3–11.
15. Christoffersen P, Jacobs K. The importance of the loss function in option valuation. *J Financ Econ* 2004;**72**(2):291–318.
16. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: *International conference on medical image computing and computer-assisted intervention*. Springer; 2015. p. 234–41.
17. Milletari F, Navab N, Ahmadi SA. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: *4th Int. Conf. 3D Vis. (3DV)*; 2016. p. 565–71.
18. Chang J, Zhang X, Chang J, Ye M, Huang D, Wang P, et al. *Brain tumor segmentation based on 3D Unet with multi-class focal loss, 11th international congress on image and signal processing BioMedical engineering and informatics (CISP-BMEI)*. 2018. https://doi.org/10.1109/CISP-BMEI.2018.8633056.
19. Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J. *UNet++: a nested U-net architecture for medical image segmentation*. 2018. arXiv:1807.10165v1.
20. Li X, Chen H, Qi X, Dou Q, Fu CW, Heng PA. H-DenseUNet: Hybrid densely connected UNet for liver and tumor segmentation from CT volumes. *IEEE Trans Med Imag* 2018;**37**(12):2663–74.
21. Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K. Attention U-net: learning where to look for the pancreas. In: *International conference on medical imaging with deep learning*; 2018.
22. Demir F. DeepCoroNet: a deep LSTM approach for automated detection of COVID-19 cases from chest X-ray images. *Appl Soft Comput* 2021;**103**:107160. https://doi.org/10.1016/j.asoc.2021.107160. Epub 2021 Feb 8. PMID: 33584157; PMCID: PMC7868740.
23. Loshchilov I, Hutter F. *Decoupled weight decay regularization arXiv preprint*. 2017. arXiv:1711.05101.
24. Luciano da F. Costa. Further Generalizations of the Jaccard Index. arXiv: 2110.09619. https://doi.org/10.48550/arXiv.2110.09619.
25. The Moroccan law n°09-08. https://www.dgssi.gov.ma/sites/default/files/attached_files/loi_09-08protection_donnees_personnelles.pdf. Accessed September 2022.
26. Yang J, Veeraraghavan H, Armato SG, Farahani K, Kirby JS, Kalpathy-Kramer J, et al. Autosegmentation for thoracic radiation treatment planning: a grand challenge at AAPM 2017. *Med Phys* 2018;**45**(10):4568–81.
27. *TCIA*. http://doi.org/10.7937/K9/TCIA.2017.3r3fvz08; 2021.
28. Ma J, Wang Y, An X, Ge C, Yu Z, Chen J, et al. *Towards efficient COVID-19 CT annotation: a benchmark for lung and infection segmentation*. 2020. arXiv:2004.12537v1.
29. MedSeg. *Covid-19 ct segmentation dataset*. 2021.
30. Jun M, Cheng G, Yixin W, Xingle A, Jiantao G, Ziqi Y, et al. *Covid-19 ct lung and infection segmentation dataset (version verson 1.0)*. http://doi.org/10.5281/zenodo.375747; 2021.
31. Paszke A, Gross S, Chintala S, Chanan G, Yang E, DeVito Z, et al. *Automatic differentiation in PyTorch*. NIPS Autodiff Workshop; 2017.
32. He K, Zhang X, Ren S, Sun J. Delving deep into rectifiers: surpassing human-level performance on imagenet classification. In: *IEEE International Conference on Computer Vision (ICCV), ICCV '15*; 2015. p. 1026–34 [Washington].
33. Luo L, Xiong Y, Liu Y, Sun X. *Adaptive gradient methods with dynamic bound of learning rate*. arXiv e-prints. 2019. arXiv:1902.09843.

34. South Laura, Saffo David, Vitek Olga, Dunne Cody, Borkin Michelle A. Effective use of Likert scales in visualization evaluations: a systematic Review. *Comput Graph Forum* 2022. https://doi.org/10.1111/cgf.14521.
35. Haubro M, Stougaard C, Torfing T, Overgaard S. Sensitivity and specificity of CT- and MRI-scanning in evaluation of occult fracture of the proximal femur. *Injury* 2015;**46**(8):1557–61.
36. Hashim U, Bosaily AE-S, Brown LC, Gabe R, Kaplan R, Parmar MK, et al. Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study. *Lancet* 2017;**389**(10071):815–22.
37. Bautista-Gomez L, Benoit A, Cavelan A, Raina SK, Robert Y, Sun H, et al. Coping with recall and precision of soft error detectors. *J Parallel Distr Comput* 2016;**98**: 8–24.