

System for Quantitative Diagnosis of COVID-19-Associated Pneumonia Based on Superpixels With Deep Learning and Chest CT

Everton Tetila (■ evertontetila@ufgd.edu.br)

Universidade Federal da Grande Dourados

Keno Bressem

Charité, Department of Radiology

Gilberto Astolfi

Instituto Federal de Mato Grosso do Sul

Diego André Sant'Ana

Instituto Federal de Mato Grosso do Sul

Marcio Carneiro Pache

Instituto Federal de Mato Grosso do Sul

Hemerson Pistori

Universidade Católica Dom Bosco

Research Article

Keywords: COVID-19, RT-PCR, CT

Posted Date: December 17th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-123158/v1

License: © 1) This work is licensed under a Creative Commons Attribution 4.0 International License.

Read Full License

System for quantitative diagnosis of COVID-19associated pneumonia based on superpixels with deep learning and chest CT

Everton Castelão Tetila^{1,*}, Keno K. Bressem^{2,3}, Gilberto Astolfi^{4,5}, Diego André Sant'Ana^{6,7}, Marcio Carneiro Brito Pache^{6,7}, and Hemerson Pistori^{5,6}

¹Universidade Federal da Grande Dourados, FACET, Dourados, 79825-070, Brasil

ABSTRACT

COVID-19 is a disease caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) that can lead to complications such as acute respiratory distress syndrome, acute heart injury and secondary infections in a relatively high proportion of patients and, consequently, significant mortality. The definitive diagnosis of COVID-19 is performed by real-time Polymerase Chain Reaction (RT-PCR). However, as the result of RT-PCR, at least for now, has been made available within a longer period of time than the computed tomography (CT) report, this has taken on an important role in the detection of patients infected with COVID-19. A rough estimate of the extent of lung involvement by the disease is also important and considered an additional criterion for deciding on discharge or hospitalization. Recent research has adopted deep neural networks and other machine learning approaches to detect the presence of lung infection caused by COVID-19. However, the extent of lung involvement (volume) caused by the disease has been little investigated. In this work, we created an end-to-end computer vision system to automatically quantify the Percentage Of Infection (POI) in chest CT images of COVID-19 cases confirmed by the laboratory. Trying to obtain high accuracy, we evaluated the performance of three deep neural networks well known in the literature, trained with three different training strategies: (1) no transfer learning with randomly initialized weights; (2) transfer learning without fine-tuning with ImageNet weights; and (3) transfer learning with 100% fine-tuning. Data augmentation and dropout were used during networks training to reduce overfitting and increase the generalization capacity of the models. Our approach consists of segmenting a chest CT with the SLIC Superpixels method and classifying each segment (superpixel) into a specific class (COVID or non-COVID). We used the weights of the deep neural network best evaluated for accuracy in our computer vision system in order to classify the superpixels in the image and quantify the regions of COVID-19 infection, thus calculating the POI on chest CT. The results indicate that deep learning models can be successfully used to support radiologists in the quantitative diagnosis of lung infection caused by COVID-19, reaching an accuracy of up to 98.4% with Inception-Resnet-v2 architecture.

Introduction

COVID-19 is a disease caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) that can lead to complications such as acute respiratory distress syndrome, acute heart injury and secondary infections in a relatively high proportion of patients and, consequently, significant mortality¹. The disease was first identified in December 2019, when the zoonotic virus SARS-CoV-2, originating from the bat *Rhinolophus*, was transmitted to man. Huanan Seafood Wholesale Market in Wuhan City, Hubei province, China, was the epicenter of the outbreak of the disease that spread rapidly around the world and was declared a pandemic by the World Health Organization (WHO) on March 11, 2020².

Two alternatives are believed to be valid by the medical community for diagnosing COVID-19: laboratory tests and computed tomography of the chest³. The definitive diagnosis of COVID-19 is performed by the Real-Time Polymerase Chain Reaction (RT-PCR) - a test that verifies the presence of genetic material of the virus - as a normal (negative) chest computed tomography (CT) does not exclude it. However, as the result of RT-PCR, at least for now, has been made available within a longer time frame than the CT report, it has assumed an essential role in the overall evaluation of patients, also because it has

²Charité, Department of Radiology, Hindenburgdamm 30, 12203 Berlin, Germany

³Berlin Institute of Health (BIH), Anna-Louisa-Karsch-Str. 2, 10178 Berlin, Germany

⁴Instituto Federal de Mato Grosso do Sul, Campus Campo Grande, Campo Grande, 79100-510, Brasil

⁵Universidade Federal de Mato Grosso do Sul, FACOM, Campo Grande, 79070-900, Brasil

⁶Universidade Católica Dom Bosco, Inovisão, Campo Grande, 79117-010, Brasil

⁷Instituto Federal de Mato Grosso do Sul, Campus Aquidauana, Aquidauana, 79200-000, Brasil

^{*}evertontetila@ufgd.edu.br

already proved to be very sensitive, although not very specific, in the detection of the most frequent pulmonary findings in the disease⁴.

In this context of high demand and urgency to make the results available, the radiological report's contents should be quite direct and as straightforward as possible to the requesting physicians. The most important information to be communicated relates to the existence (or not) of pulmonary involvement, whether the appearance of the results is consistent with the infectious process and, in positive cases, whether the changes imply viral etiology, particularly of COVID-19, even if there is a viral etiology of COVID-19⁴. It is also important to have a rough estimate of the extent of lung involvement generally called Percentage Of Infection (POI)^{5,6}, caused by the disease COVID-19, which has been considered useful in the management of patients, together with other clinical data and physical examination⁴. Zhao et al.⁷ proposed a link between chest CT findings and the clinical conditions of coronavirus pneumonia (COVID-19) as an additional criterion for deciding on hospitalization.

COVID-19 CT images manifestations display individual characteristics of their own that differ from other forms of viral pneumonia, such as influenza-A viral pneumonia¹. Given this, a chest CT can be used in deep learning (DL)-based strategies to predict COVID-19 outcomes. However, the quantification of the POI based on superpixel with deep learning and chest CT was not investigated. Clinically, there is no automated analysis system capable of quantifying the regions of infection to assist the radiologist in the hospitalization decision. Therefore, there is a need for new contributions in this research domain, that is, optical diagnostic methods capable of estimating the extent of pulmonary involvement based on DL and CT images. In addition, the medical community calls for rapid and accurate diagnosis of the disease to reduce mortality.

In this sense of urgency for data delivery, this is important, since a large number of chest CT reports need to be made available and automatically analyzed. Besides, as an alternative criterion for determining whether to discharge or hospitalize, the outcome of the relationship between COVID-19 chest CT findings can be used. In this respect, the DL technology with chest CT can serve as an efficient way of diagnosing quantitative lung infection to differentiate against community-acquired pneumonia in COVID-19 patients.

This is essential in this context of urgency to provide results, since a large number of chest CT reports need to be made available and analyzed automatically. In addition, the result of the relationship between the findings of the chest CT scan by COVID-19 can be used as an additional criterion to decide on discharge or hospitalization. In this regard, DL technology with chest CT can serve as an effective form of quantitative diagnosis of lung infection to differentiate community-acquired pneumonia from patients with COVID-19.

In this work, an end-to-end computer vision system was developed to automatically measure the percentage of lung infection in chest CT images of COVID-19 laboratory-confirmed cases. First, we considered an image segmentation phase with the SLIC Superpixels technique to segment the infection regions and the entire lung from CT. A collection of 20,000 superpixels (image segments) produced from 2,481 chest CT of 120 patients was used to test the proposed approach. Precision, recall, F-measure and training time were evaluated for the deep learning models' performance.

Our methodology tests the efficiency of three well-known deep neural networks from the literature, trained with three distinct training strategies: (1) no transfer learning with randomly initialized weights; (2) transfer learning without fine-tuning with ImageNet weights; and (3) transfer learning with 100% fine-tuning, gradually extracting from the input images high-level features and building the classification model. To identify the superpixels in the image and calculate the regions infected with COVID-19, we use the deep neural network weights best tested for accuracy in our computer vision system, thus measuring the POI in the chest CT.

Related Works

A variety of methods to assist radiologists in screening suspected cases of COVID-19 using images collected from chest radiographs have been suggested within the Artificial Intelligence framework and, more specifically, in the field of computer vision. In two publicly available datasets, CheXpert and COVID-19 Image Data Collection, Bressem et al.⁸ compared sixteen deep learning models for chest radiograph classification. With shorter training times, the shallower networks achieved results comparable to their deeper and more complex counterparts, enabling classification efficiency in medical imaging information close to state-of-the-art techniques, even when using minimal hardware. In^{9–11} the authors tested various models based on DL and chest radiography images to diagnose patients infected with COVID-19.

Rahimzadeh & Attar¹² modified a deep neural network based on the concatenation of Xception and ResNet50V2 to classify X-ray images into three classes: normal, pneumonia and COVID-19, based on two merged open source datasets. As a result, the modified neural network achieved the best accuracy when using many resources derived from two robust networks. A double transfer learning strategy was implemented by Bassi & Attux¹³, using the NIH ChestX-Ray14 dataset as an intermediate stage. The layered-sensitive propagation of interest was implemented after training the deep neural networks (DenseNet121, DenseNet201), generating X-ray heat maps, along with the probabilities of COVID-19, viral pneumonia and healthy lungs.

A generative adversary network (GAN) and three deep learning models (Alexnet, Googlenet and Restnet18) have been used by Loey et al.¹⁴ for coronavirus detection in chest X-ray images. This work's key motivation was the lack of datasets

for COVID-19, particularly chest X-ray images. Up until the moment of study, the authors gathered all possible images for COVID-19 and used the GAN to create more radiographic images and improve the model's accuracy. The performance of an DL model (CAPE: Covid-19 AI Predictive Engine) for predicting severe patient outcomes (ICU admission and mortality) of COVID-19 from chest radiographs was assessed by Liew et al.¹⁵.

Other works in the literature used images acquired from chest computed tomography (CT). Qingcheng et al. ¹⁶ used the DL-based Unet architecture in chest CT for COVID-19 patient discharge management. In order to standardize discharge requirements in a "square cabin" hospital in China, they correlated CT images with consecutive negative RT-PCR test results.

Two three-dimensional classification models of convolution neural networks were tested by Butt et al. ¹⁷ - original ResNet and another model designed to concatenate the localization attention function in the ResNet network - and achieved an overall accuracy of 86.7 percent using ResNet with the Coronavirus vs. non-coronavirus cases localization attention model through chest CT studies. A deep learning-based CT diagnostic framework (DeepPneumonia) has been developed by Song et al. ¹⁸ to classify COVID-19 patients. The system could discriminate against patients with COVID-19 infection and bacterial pneumonia and classify the key lesion characteristics.

Multiple pulmonary and cardiovascular metrics derived from the initial chest CT of COVID-19 patients from deep neural networks were used by Weikertet et al. ¹⁹ to predict the severity of clinical care in three groups: group 1 (outpatient), group 2 (general ward) and group 3 (ICU). Among other things, with the severity of clinical treatment, the average percentage of lung volume affected by ground-glass opacities and other results increased significantly (from group 1 to 3). A deep learning supervised framework (COVNet) was developed by Li et al. ²⁰ to detect COVID-19 and community-acquired pneumonia on chest computed tomography.

Yan et al.²¹ used a blood sample database of 485 infected patients in the Wuhan region of China to classify critical predictive biomarkers of disease mortality to help decision-making and logistical preparation in health systems. For this purpose, machine learning tools selected three biomarkers that predict 10-day advance mortality and 90 percent accuracy of individual patients: Lymphocytes and high-sensitivity C-reactive protein (PCR-as), lactate dehydrogenase (DHL). Uddin et al.²² have developed an intelligent deep learning-based surveillance system that can be built at various locations to direct people to protect themselves from symptoms of high exposure to the virus (SARS-CoV-2) or symptoms of COVID-19 (such as fever and cough) and to follow protection recommendations to prevent the disease.

For the identification of cases of COVID-19, conventional supervised learning methods used to classify images, including support vector machine (SVM), decision tree and k-NN, were also used. Artificial intelligence algorithms (SVM, random forest, multi-layer perceptron) were used by Mei et al.²³ to correlate chest CT findings with clinical symptoms, history of exposure and laboratory testing to rapidly diagnose patients positive for COVID-19.

Yu²⁴ et al. explored four pre-trained DL models (Inception-V3, ResNet-50, ResNet-101 and DenseNet-201) to extract the high-level resources of CT. These resources were then provided to various classifiers (linear discriminant, linear SVM, cubic SVM, k-NN and Adaboost with decision trees) to identify serious and non-serious cases of COVID-19. A new computational vision tool, also based on shallow algorithms, has been developed by Tuncer et al.²⁵ to classify Covid-19 in X-ray images. The proposed method consists of steps for preprocessing, feature extraction and feature selection. Lastly, decision tree, linear discriminant, SVM and k-NN methods are used in the classification stage. Farhat et al.²⁶ presented a literature review on state-of-the-art deep learning approaches in medical image analysis, produced between February 2017 and May 2020, highlighting several tasks such as classification, segmentation and identification, as well as various lung pathologies such as airway diseases, lung cancer, COVID-19 and other infections, was presented. Hussain et al.²⁷ presented an overview of various approaches and methods of artificial intelligence, including neural networks, classical SVM and major state-of-the-art learning that can be applied to different forms of pandemics.

In all the works cited, deep neural networks and other machine learning approaches were trained to detect the presence of lung infection caused by COVID-19 or to locate the main characteristics of the lesion in the image. Despite this, the authors did not investigate the extent of lung involvement (volume) caused by the disease. In this work, we developed an end-to-end computer vision system based on superpixel with deep learning to automatically calculate the POI in chest CT images.

Proposed Approach

This section presents a computer vision approach to automatically quantify the percentage of lung infection in chest CT images of confirmed cases in the COVID-19 laboratory. The proposed approach adopts the SLIC Superpixels method to segment the infection regions in the image, as well as the entire lung. The SLIC method employs the k-means²⁸ algorithm to generate similar regions, called superpixels. The algorithm's k parameter refers to the number of superpixels in the image and allows to control the size of superpixels. The parameter m corresponds to the compression control of the generated regions. We set k = 200 and m = 30 to segment the main characteristics of the lesions in the image, defined by adherence to the size and compression limits of the SLIC algorithm.

SLIC groups regions of pixels in the 5-D space defined by l, a, b (values of the CIELAB color scale) and the coordinates x and y of the pixels. An input image is segmented into regular regions, defining the number k of superpixels with approximately $\frac{N}{k}$ pixels, where N is the number of pixels in the image. Each region composes an initial superpixel of dimensions $S \times S$, where $S = \sqrt{\frac{N}{k}}$. The centers of the superpixel clusters $C_k = [l_k, a_k, b_k, x_k, y_k]$ with k = [1, k] are sampled on a regular grid spaced S pixels apart. In a 3×3 neighborhood, the centers are moved to the lowest gradient position, preventing the allocation of centroids on an edge and to reduce the chance of seeding a superpixel with a noisy pixel. Each pixel is associated to the nearest cluster center, an update step adjusts the cluster centers to be the mean labxy vector of all the pixels belonging to the cluster²⁹.

A schematic diagram of the proposed approach is shown in Figure 1. It illustrates the proposed system that consists of 4 steps: (a) Image acquisition, (b) SLIC segmentation, (c) Image annotation, and finally (e) Classification of COVID-19 infected regions and other lung regions. First, we used a large publicly available dataset, called SARS-COV-2 Ct-Scan Dataset³⁰, containing 1,252 CT scans positive for SARS-CoV-2 infection (COVID-19) and 1230 CT scans for patients not infected with SARS-CoV-2, see step (a) of Figure 1. Using the SLIC Superpixels method, these 2,482 CT scans were segmented one by one, see step (b) of Figure 1. After the image segmentation, each superpixel was annotated by a senior radiologist to generate a set of superpixel data for training and system testing, see step (c) of Figure 1. The annotated images were divided into two classes: (1) COVID, (2) non-COVID. The COVID class describes the three-dimensional anomalies associated with COVID-19: ground-glass opacities, crazy paving and consolidations. Finally, a deep neural network was trained to learn the visual features of the superpixels and classify the lung regions, see step (d) of Figure 1.

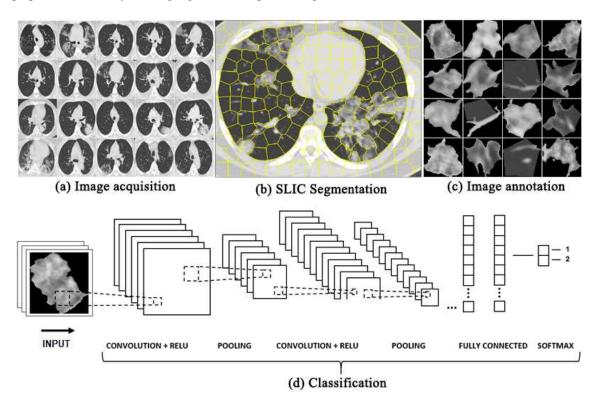


Figure 1. Proposal of a system for quantitative diagnosis of COVID-19-associated pneumonia based on superpixels with deep learning and chest CT. (a) Image acquisition. (b) SLIC segmentation. (c) Image annotation. (d) Classification.

The post-processing stage consists of segmenting a chest CT using the SLIC method and classifying each segment (superpixel) in a specific class. The system scans the image from left to right, top to bottom, classifying each superpixel individually using the deep neural net trained, while providing the color of the class simultaneously. Thus, a colored map is generated by providing one class per segment. The POI on chest CT is obtained by adding the superpixels corresponding to the regions infected with COVID-19, in relation to the number of superpixels identified by our computer vision system. The computational complexity of the segmentation process, based on the SLIC algorithm, limits the search space to a region proportional to the superpixel size. This reduces the linear complexity to the number of superpixels k, instead of the number of pixels n.

Materials and Methods

Dataset Description

We use a large publicly available TC SARS-COV-2 Ct-Scan dataset³⁰. The dataset consists of 2,482 CT images, divided between 1,252 CT scans for patients infected with SARS-CoV-2 and 1,230 CT scans for patients not infected with SARS-CoV-2, but who had other lung diseases. These data were collected from 120 patients in São Paulo hospitals, Brazil. The number of patients considered to compose the dataset is 60 patients infected with SARS-CoV-2, of which 32 are men and 28 are women; and 60 patients not infected with SARS-CoV-2, of which 30 are men and 30 are women. Figure 2 illustrates some examples of CT for patients infected and not infected with SARS-CoV-2 that make up the dataset. Detailed characteristics of each patient were omitted by hospitals for ethical reasons.

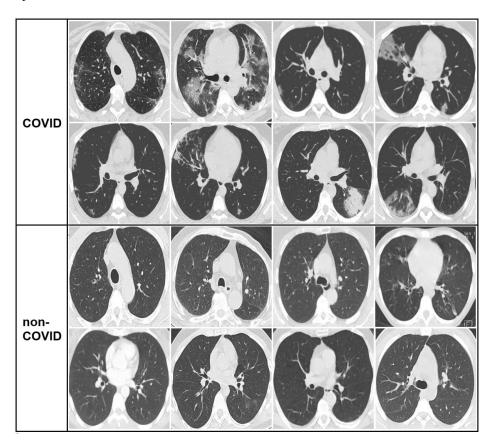


Figure 2. Examples of CT scans for infected and non-infected patients with SARS-CoV-2 (COVID-19) that make up the SARS-COV-2 Ct-Scan Dataset.

This dataset aims to promote the research and development of artificial intelligence methods capable of identifying whether an individual is infected with SARS-CoV-2 by analyzing their CT scans. As a result of the baseline for this dataset, Soares et al.³⁰ used the eXplainable deep learning (xDNN) approach which achieved a F-measure of 97.31%, which is very promising. The CT-Scan Dataset for the SARS-COV-2 is available at: www.kaggle.com/plameneduardo/sarscov2-ctscan-dataset and the xDNN code is available at: https://github.com/Plamen-Eduardo/xDNN-SARS-COV-2-CT-Scan.

Then, each acquired CT image was segmented by the SLIC Superpixels method and segments chosen at random (10,000 superpixel for each class COVID and non-COVID) were labeled by a senior radiologist and doctor of Medicine in the Radiology Department of the Charité Universitätsmedizin Berlin, Germany, thus generating a set of superpixel references for the training and test set (see Figure 3), called COVID20K2C Superpixels Dataset and available to the public at³¹.

Experimental Design

We adopted five-fold stratified cross-validation with validation and test sets on the COVID20K2C. Five-fold stratified cross-validation on the dataset was performed to determine the generalization capability of the models. In each fold, we kept aside the 20% allocated for testing. Then, we split the 80% allocated to training into two subsets, 60% to be the actual train set and

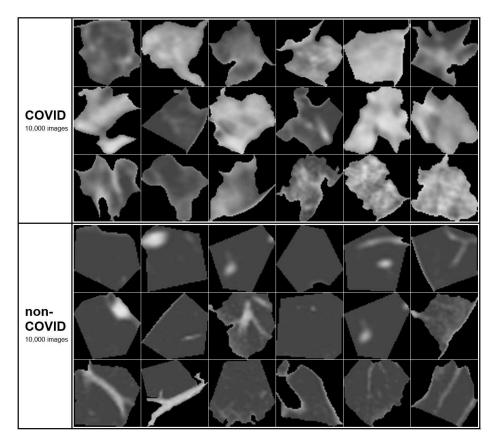


Figure 3. Examples of superpixels from the COVID20K2C Superpixels Dataset, divided into COVID, non-COVID and image numbers by class.

20% to be the validation set. Thus, in each fold the dataset was split into 60% for training, 20% for validation and 20% for the test, allowing that the deep learning models be iteratively trained and validated on different sets. Finally, the classifier's output is given by the average of the five-fold on the test set. We applied the Scikit-learn package³² to construct this five-fold cross-validation model.

The performance of the deep learning models was assessed using five metrics: precision, accuracy, recall, F-measure and training time. We calculate the average results of the five assessment criteria for each model studied. We are using the ANOVA hypothesis test to determine whether the models vary statistically in their results. We report the p-value found for each metric and the level of significance has been set at 5%.

Data Increase and Training Analysis

Deep learning models are trained with labeled images for image classification to learn how to recognise and classify them according to visual patterns. In our experiments, we use open source implementations of three ImageNet competition-recognized deep learning models: DenseNet-201³³, Inception-Resnet-v2³⁴ and ResNet-152³⁵. In the experiments, we use the following input parameters for the trained models. The input image's width and height are both set to 256 pixels and a batch size of 32 samples and training with 50 epochs. We also used the SGD optimizer³⁶ with a learning rate set at 0.0001 and momentum at 0.9.

We also perform data augmentation in the training set to increase data by applying rotation, rescaling, scrolling and zoom operations. This procedure aims to reinforce the rotation invariance and scale invariance in the classification task. Data augmentation includes: horizontal inversion, random rotation at $+30^{\circ}/-30^{\circ}$, resize factor set to 3.92×10^{-3} , zoom with image magnification factor between 70% and 130% and shift range 0.3 for changing the horizontal and vertical image size by 30%.

We adopted three distinct training strategy to statistically evaluate the ability of the deep learning models to recognize lung infections induced by COVID-19. In the first training approach, which is called transfer learning (TL), we instantiated a base model and loaded pre-trained weights from ImageNet into it. After, we froze all layers in the base model and created new layers on top of it. The new layers on top consist of one Dense layer with 1024 units, followed by a Dropout layer with the rate set at 0.5 and two other Dense layers with 1024 and 2 units, respectively. In the second strategy, which is called transfer learning and

fine-tuning (TL+FT), we configured the networks like at the TL strategy. However, we kept the model base in inference mode during the training. Then, we unfroze the base model and retrained it using a lower learning rate (0.00001). Finally, in the third strategy, which is called no transfer learning (No TL), we configured the networks like at the TL strategy and trained them with the weights randomly initialized.

We used a workstation with an AMD Ryzen 1800X 3.6 GHz (4.0 GHZ MAX TURBO) 20 MB (8 N, 16 T) processor, Nvidia Titan Xp 12 GB 3840 Cuda color graphics card, Ballistix DDR4 2400 MHz 16 GB RAM, Kingston SSD storage A1000 240 GB M.2.

Results and Discussion

Table 1 presents the results of the five metrics used to evaluate the performance of the deep learning models. Figure 4 shows the confusion matrix of each model evaluated for accuracy. The percentage values represent the average of the five iterations in the test set; each matrix follows the same order as the models in Table 1. Figure 5 shows training and validation accuracy during 50 epochs. The values plotted in the graph refer to the average of five-fold in the training and validation set. Using randomly initialized weights (No TL) with Inception-Resnet-v2 (98.4%), our approach obtained the best result for accuracy. The F-measure (98.4%) of this model applied to our superpixel dataset surpassed the baseline result of the Soares et al. dataset, which reached an F-measure of 97.31%. The table also shows the total training time, in hours, to build the classification model. The time performance in Table 1 refers to the hardware specifications presented in the Materials and Methods section.

Model	Training strategy	Time (h:m:s)	Accuracy (%)	Precision (%)	Recall (%)	F-measure (%)
DenseNet-201	TL + FT	03:48:51	$96.5 \pm .0030$	$96.5 \pm .0028$	$96.5 \pm .0030$	$96.5 \pm .0030$
	TL	02:04:07	$95.7 \pm .0052$	$95.8 \pm .0045$	$95.7 \pm .0052$	$95.7 \pm .0052$
	No TL	02:11:27	$91.3 \pm .0773$	$93.1 \pm .0503$	$91.3 \pm .0773$	$91.0 \pm .0822$
Inception-Resnet-v2	TL + FT	03:53:26	$96.0 \pm .0028$	$96.1 \pm .0027$	$96.0 \pm .0028$	$96.0 \pm .0028$
	TL	02:06:12	$95.8 \pm .0014$	$95.8 \pm .0014$	$95.8 \pm .0014$	$95.8 \pm .0014$
	No TL	02:22:53	$98.4 \pm .0062$	$98.4 \pm .0060$	$98.4 \pm .0062$	$98.4 \pm .0062$
ResNet-152	TL + FT	03:58:20	$89.5 \pm .0029$	$90.1 \pm .0025$	$89.5 \pm .0029$	$89.5 \pm .0030$
	TL	02:01:47	$88.1 \pm .0309$	$89.2 \pm .0189$	$88.1 \pm .0309$	$88.0 \pm .0325$
	No TL	03:00:01	$97.8 \pm .0052$	$97.9 \pm .0049$	$97.8 \pm .0052$	$97.8 \pm .0052$

Table 1. Performance metrics used to evaluate deep learning models.

ANOVA test results indicate that there is no evidence to determine if there is a significant difference in the average performance of the models tested, at a 5% significance level, using accuracy (p-value = .404), precision (p-value = .352), recall (p-value = .404), F-measure (p-value = .416) and training time (p-value = .362) as metrics.

Comparison with Shallow Learning Methods

In this experiment, the deep learning proposed approach is compared with shallow learning methods (SVM, Random Forest, J48, Naive Bayes, k-NN and Adaboost). For this purpose, we used the same implementation as the authors and provided the same set of CT presented in section Materials and Methods with five-fold stratified cross-validation in the test set. Table 2 shows five metrics proposed to evaluate the performance of the deep learning models and compares them with shallow learning methods. The training time for shallow learning methods considers the time for the extraction of attributes (00:07:82), added to the time to create the classification model for each method.

Model	Training strategy	Time (h:m:s)	Accuracy (%)	Precision (%)	Recall (%)	F-measure (%)
DenseNet-201	TL + FT	03:48:51	96.5	96.5	96.5	96.5
Inception-Resnet-v2	No TL	02:22:53	98.4	98.4	98.4	98.4
ResNet-152	No TL	03:00:01	97.8	97.9	97.8	97.8
Random Forest	Combined feature	00:08:43	96.4	96.4	96.4	96.4
SVM	extraction based	00:08:93	95.7	95.7	95.7	95.7
J48	on color ³⁷ , gradient ³⁸ ,	00:08:20	94.1	94.1	94.1	94.1
AdaBoost	texture ^{39,40} and shape ⁴¹	00:08:32	92.2	92.2	92.2	92.2
k-NN		00:07:82	91.1	91.4	91.1	91.1
Naive Bayes		00:07:84	92.2	92.2	92.2	92.2

Table 2. Comparison of deep learning models with shallow learning methods.

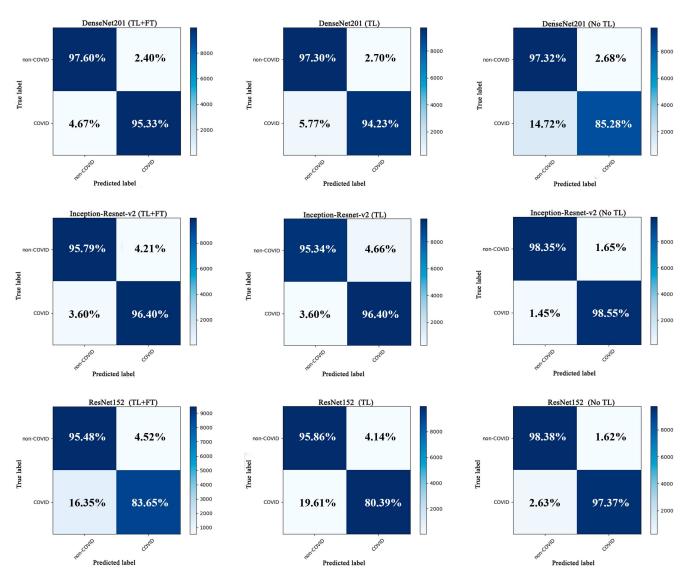


Figure 4. Confusion matrix of each deep learning model. The values in percent represent the average of the 5 iterations in the test set.

As shown in Table 2, deep learning models outperformed all shallow learning methods. The Random Forest and SVM classification algorithms provided 96.4% and 95.7%, respectively. While the deep learning models varied between 96.5% and 98.4% with the best training strategy, showing a good ability to distinguish typical features of COVID-19 from other viral pneumonias or infection-free regions.

Percentage of lung infection caused by COVID-19

In this section, we have implemented the approach proposed in our computer vision system called Pynovisão. In order to classify superpixels in the image and quantify regions infected with COVID-19, we employ the weights of the deep neural network best evaluated for accuracy - Inception-ResNet-v2 No TL - in the Pynovisão.

First, the system employs the SLIC Superpixels method to segment the lung regions from CT, as shown in Figure 6a. Here, the parameter k=200 was adjusted to segment the pulmonary regions. The system then runs the image from top to bottom from left to right, individually classifying each superpixel using the Inception-Resnet-v2 No TL model, while simultaneously supplying the class color. Thus, regions infected with COVID-19 can be easily identified by the corresponding segment color in an image, see Figure 6b. First, the system uses the SLIC Superpixels method to segment the lung regions from CT, as shown in Figure 6a. Here, the parameter k=200 was adjusted to segment the pulmonary regions. The system scans the image from left to right from top to bottom, classifying each superpixel individually using the Inception-Resnet-v2 No TL model, while

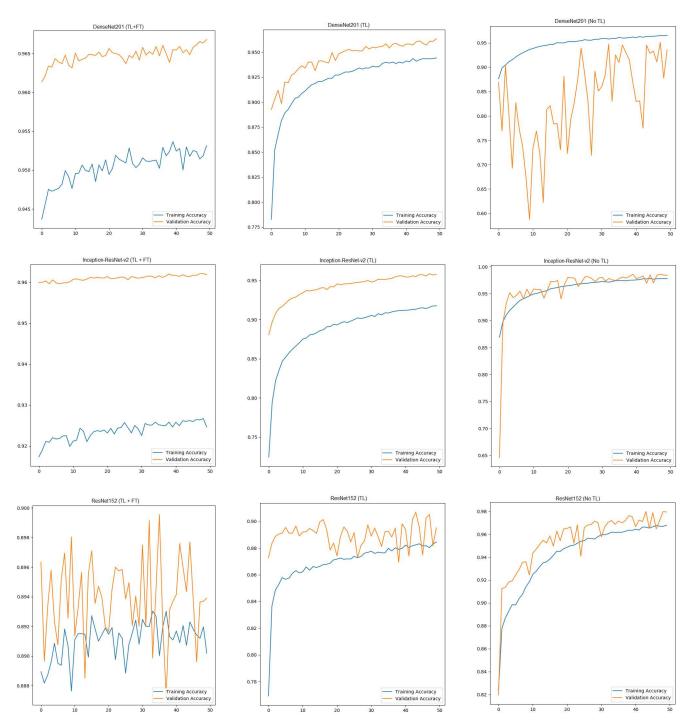
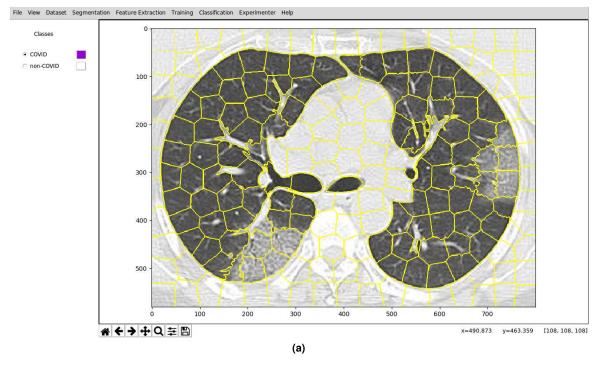


Figure 5. Training and validation accuracy of each deep learning model during 50 epochs.

providing the color of the class simultaneously. Thus, the regions infected with COVID-19 can be easily located in the image by the color of the corresponding segment, as well as the quantitative result of each class, see Figure 6b.

Pynovisão recognizes for diagnostic purposes whether pulmonary findings are compatible with an infectious process of viral etiology, particularly COVID-19, including ground-glass opacities, ground-glass opacities associated with interlobular septal thickening, characterizing the pattern of crazy paving and ground-glass opacities associated with consolidations. Our system is a deep neural network for quantitative diagnosis of COVID-19 infection based on chest CT superpixels with 98.4% accuracy and F-measure, but should not be used for first-line diagnostic testing for COVID-19 or as a replacement for RT-PCR.



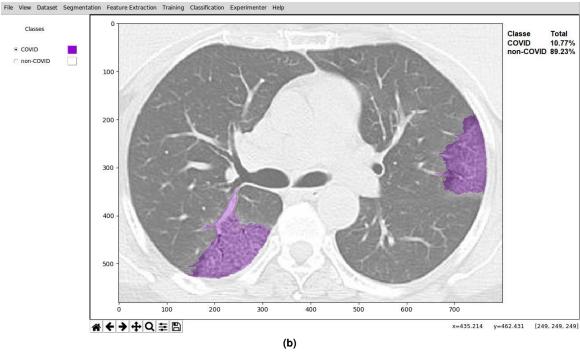


Figure 6. Screen capture of our computer vision system evaluating a chest CT. In Figure (a), the system called Pynovisão presents the segmentation step of the pulmonary regions. In Figure (b), the color labels represent the categories of our problem. In this case, the system uses the Inception-Resnet-v2 No TL model for the classification of superpixels. Pynovisão was registered by INPI with the number BR 51 2019 000427 2 and is available at:

http://git.inovisao.ucdb.br/inovisao/pynovisao.

Conclusion

In this work, we evaluated different state-of-the-art deep learning models and different training strategies to automatically quantify the POI on chest CT from COVID-19 cases confirmed by RT-PCR laboratory. First, we consider an image segmentation step with the SLIC Superpixels method to segment the regions of infection, as well as the entire lung from chest CT. In the classification stage, we compared three deep learning models recognized by the competition on ImageNet: DenseNet-201, Inception-Resnet-v2 and ResNet-152, trained with three different training strategies: transfer learning and fine-tuning (TL+FT), transfer learning (TL) and trained network with the weights randomly initialized (No TL). Inception-Resnet-v2 using randomly initialized weights was the most appropriate deep learning model for classifying COVID-19 superpixels.

Experimental results also showed that deep learning models lead to higher ratings compared to other shallow learning methods such as SVM, Random Forest, J48, Naive Bayes, k-NN and Adaboost, with the Inception-ResNet-v2 model achieving an accuracy of up to 98.4%. The 98.4% F-measure surpassed the same dataset's baseline score, which reached a 97.31% F-measure.

It has also been shown how a deep learning model can be implemented in an end-to-end computer vision system to automatically quantify the percentage of infection in chest CT. In our Pynovisão computer vision system, we used the deep neural network weights best evaluated for accuracy - Inception-ResNet-v2 No TL - to classify the superpixels in the image and quantify the regions infected with COVID-19, thus calculating the percentage of infection in chest CT. The results indicate that deep learning models can effectively help radiologists in detecting the most frequent pulmonary disease findings, thus reducing the time to control COVID-19.

As part of the future work, we intend to evaluate the performance of the proposed approach to distinguish the various tomographic findings observed in COVID-19 patients, including ground-glass opacities, crazy paving and consolidations.

References

- **1.** Xu, X. *et al.* A deep learning system to screen novel coronavirus disease 2019 pneumonia. *Engineering* DOI: https://doi.org/10.1016/j.eng.2020.04.010 (2020).
- 2. WHO. Coronavirus disease 2019 (covid-19) situation report 51. Tech. Rep., World health organization (2020).
- **3.** Qureshi, S. A. & Rehman, A. u. Optical techniques, computed tomography and deep learning role in the diagnosis of covid-19 pandemic towards increasing the survival rate of vulnerable populations. *Photodiagnosis Photodyn. Ther.* **31**, DOI: https://doi.org/10.1016/j.pdpdt.2020.101880 (2020).
- **4.** Shoji, H. *et al.* Structured thoracic computed tomography report for covid-19 pandemic. *Einstein (Sao Paulo)* **18**, DOI: 10.31744/einstein journal/2020ed5720 (2020).
- 5. Shan, F. et al. Lung infection quantification of covid-19 in ct images with deep learning (2020). 2003.04655.
- **6.** Zhang, H.-t. *et al.* Automated detection and quantification of covid-19 pneumonia: Ct imaging analysis by a deep learning-based software. *Eur. J. Nucl. Medicine Mol. Imaging* **47**, DOI: https://doi.org/10.1007/s00259-020-04953-1 (2020).
- 7. Zhao, W., Zhong, Z., Xie, X., Yu, Q. & Liu, J. Relation between chest ct findings and clinical conditions of coronavirus disease (covid-19) pneumonia: A multicenter study. *Am. J. Roentgenol.* 214, 1–6, DOI: https://doi.org/10.2214/AJR.20. 22976 (2020).
- **8.** Bressem, K. *et al.* Comparing different deep learning architectures for classification of chest radiographs. *Sci. Reports* **10**, DOI: https://doi.org/10.1038/s41598-020-70479-z (2020).
- **9.** Asif, S., Wenhui, Y., Jin, H., Tao, Y. & Jinhai, S. Classification of covid-19 from chest x-ray images using deep convolutional neural networks. *medRxiv* DOI: https://doi.org/10.1101/2020.05.01.20088211 (2020). https://www.medrxiv.org/content/early/2020/06/18/2020.05.01.20088211.full.pdf.
- **10.** Sarker, L., Islam, M., Hannan, T. & Ahmed, Z. Covid-densenet: A deep learning architecture to detect covid-19 from chest radiology images. *Preprints* DOI: https://doi.org/10.20944/preprints202005.0151.v1 (2020).
- 11. Hassantabar, S., Ahmadi, M. & Sharifi, A. Diagnosis and detection of infected tissue of covid-19 patients based on lung x-ray image using convolutional neural network approaches. *Chaos, Solitons & Fractals* 140, 110170, DOI: https://doi.org/10.1016/j.chaos.2020.110170 (2020).
- 12. Rahimzadeh, M. & Attar, A. A modified deep convolutional neural network for detecting covid-19 and pneumonia from chest x-ray images based on the concatenation of xception and resnet50v2. *Informatics Medicine Unlocked* 19, 100360, DOI: https://doi.org/10.1016/j.imu.2020.100360 (2020).
- 13. Bassi, P. & Attux, R. A deep convolutional neural network for covid-19 detection using chest x-rays. CoRR (2020).

- **14.** Loey, M., Smarandache, F. & Khalifa, N. E. Within the lack of chest covid-19 x-ray dataset: A novel detection model based on gan and deep transfer learning. *Symmetry* **12**, 651, DOI: https://doi.org/10.3390/sym12040651 (2020).
- **15.** Liew, C., Quah, J., Goh, H. & Venkataraman, N. A chest radiography-based artificial intelligence deep-learning model to predict severe covid-19 patient outcomes: the cape (covid-19 ai predictive engine) model. *Medrxiv* DOI: https://doi.org/10.1101/2020.05.25.20113084 (2020).
- **16.** Meng, Q. *et al.* Role of novel deep-learning-based ct used in management and discharge of covid-19 patients at a "square cabin" hospital in china. *Res. Sq.* DOI: https://doi.org/10.21203/rs.3.rs-28201/v1 (2020).
- 17. Butt, C., Gill, J., Chun, D. & Babu, B. Deep learning system to screen coronavirus disease 2019 pneumonia. *Appl. Intell.* DOI: https://doi.org/10.1007/s10489-020-01714-3 (2020).
- **18.** Song, Y. *et al.* Deep learning enables accurate diagnosis of novel coronavirus (covid-19) with ct images. *Medrxiv* DOI: https://doi.org/10.1101/2020.02.23.20026930 (2020).
- **19.** Weikert, T. *et al.* Treatment intensity stratification in covid-19 by fully automated analysis of pulmonary and cardiovascular metrics on initial chest ct using deep learning. *Res. Sq.* DOI: https://doi.org/10.21203/rs.3.rs-35878/v1 (2020).
- **20.** Li, L. *et al.* Artificial intelligence distinguishes covid-19 from community acquired pneumonia on chest ct. *Radiology* **296**, 200905, DOI: https://doi.org/10.1148/radiol.2020200905 (2020).
- 21. Yan, L. et al. An interpretable mortality prediction model for covid-19 patients. Nat. Mach. Intell. 1–6, DOI: https://doi.org/10.1038/s42256-020-0180-7 (2020).
- 22. Uddin, M. I., Ali Shah, S. & Al-Khasawneh, M. A novel deep convolutional neural network model to monitor people following guidelines to avoid covid-19. *J. Sensors* 2020, 1–15, DOI: https://doi.org/10.1155/2020/8856801 (2020).
- 23. Mei, X. *et al.* Artificial intelligence–enabled rapid diagnosis of patients with covid-19. *Nat. Medicine* 26, 1–5, DOI: https://doi.org/10.1038/s41591-020-0931-3 (2020).
- **24.** yu, Z. *et al.* Rapid identification of covid-19 severity in ct scans through classification of deep features. *Biomed Eng Online* DOI: https://doi.org/10.21203/rs.3.rs-30802/v1 (2020).
- **25.** Tuncer, T., Dogan, S. & Özyurt, F. An automated residual exemplar local binary pattern and iterative relieff based corona detection method using lung x-ray image. *Chemom. Intell. Lab. Syst.* **203**, 104054, DOI: https://doi.org/10.1016/j.chemolab. 2020.104054 (2020).
- **26.** Farhat, H., Sakr, G. E. & Kilany, R. Deep learning applications in pulmonary medical imaging: recent updates and insights on covid-19. *Mach. vision applications* **31**, 53–53, DOI: https://doi.org/10.1007/s00138-020-01101-5 (2020). 32834523[pmid].
- 27. Hussain, A., Bouachir, O., Al-Turjman, F. & Aloqaily, M. Ai techniques for covid-19. *IEEE Access* **PP**, 1–1, DOI: https://doi.org/10.1109/ACCESS.2020.3007939 (2020).
- **28.** Hartigan, J. A. & Wong, M. A. Algorithm as 136: A k-means clustering algorithm. *J. Royal Stat. Soc. Ser. C (Applied Stat.* **28**, 100–108 (1979).
- **29.** Achanta, R. *et al.* Slic superpixels compared to state-of-the-art superpixel methods. *IEEE transactions on pattern analysis machine intelligence* **34**, DOI: https://doi.org/10.1109/TPAMI.2012.120 (2012).
- **30.** Soares, E., Angelov, P., Biaso, S., Higa Froes, M. & Kanda Abe, D. Sars-cov-2 ct-scan dataset: A large dataset of real patients ct scans for sars-cov-2 identification. *medRxiv* DOI: https://doi.org/10.1101/2020.04.24.20078584 (2020). https://www.medrxiv.org/content/early/2020/05/14/2020.04.24.20078584.full.pdf.
- 31. Tetila, E. C. Covid20k2c superpixels dataset: Banco de imagens de superpixels criado a partir de tomografias computadorizadas para infeccao por sars-cov-2 (covid-19). github https://github.com/EvertonTetila/COVID20K2C-Superpixels-Dataset (2020).
- 32. Pedregosa, F. et al. Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830 (2011).
- **33.** Huang, G., Liu, Z., Van Der Maaten, L. & Weinberger, K. Q. Densely connected convolutional networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2261–2269 (2017).
- **34.** Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. & Wojna, Z. Rethinking the inception architecture for computer vision. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), DOI: https://doi.org/10.1109/CVPR.2016.308 (2016).
- **35.** He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770–778 (2016).

- **36.** Bottou, L. & Bousquet, O. The tradeoffs of large scale learning. In Platt, J., Koller, D., Singer, Y. & Roweis, S. (eds.) *Advances in Neural Information Processing Systems*, vol. 20, 161–168 (Curran Associates, Inc., 2008).
- 37. Swain, M. & Ballard, D. Color indexing. Int. J. Comput. Vis. 7, 11–32 (2004).
- **38.** Dalal, N. & Triggs, B. Histograms of oriented gradients for human detection. In 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol. 1, 886–893 vol. 1 (2005).
- **39.** Haralick, R. Haralic rm.statistical and structural approaches to texture. proc ieee 67:786-804. *Proc. IEEE* **67**, 786 804, DOI: https://doi.org/10.1109/PROC.1979.11328 (1979).
- **40.** Ojala, T., Pietikainen, M. & Maenpaa, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis Mach. Intell.* **24**, 971–987 (2002).
- 41. Ming-Kuei Hu. Visual pattern recognition by moment invariants. IRE Transactions on Inf. Theory 8, 179–187 (1962).

Acknowledgments

We thank the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), NVIDIA Corporation for the donation of the graphics card and the Fundação de Apoio ao Desenvolvimento do Ensino, Ciência e Tecnologia do estado de Mato Grosso do Sul (FUNDECT).

Author contributions statement

E.C.T. wrote and reviewed the manuscript text, registered raw data and prepared all figures. K.K.B. annotated a set of superpixel data for training and system testing and reviewed the manuscript text. G.A. and D.A.S. wrote all the programs and run the experiments. M.C.B.P. translated, formatted and reviewed the manuscript text. H.P. also reviewed the manuscript text.

Competing interests

The author(s) declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to E.C.T.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

©The Author(s) 2020

Figures

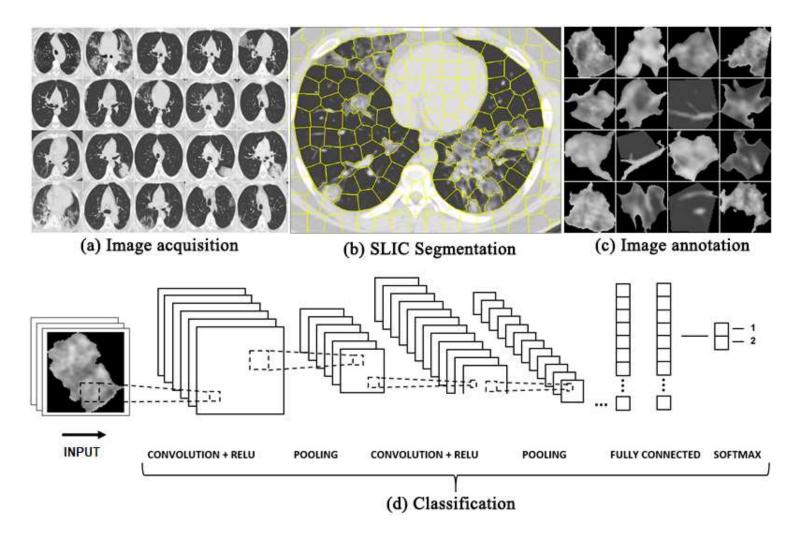


Figure 1

Proposal of a system for quantitative diagnosis of COVID-19-associated pneumonia based on superpixels with deep learning and chest CT. (a) Image acquisition. (b) SLIC segmentation. (c) Image annotation. (d) Classification.

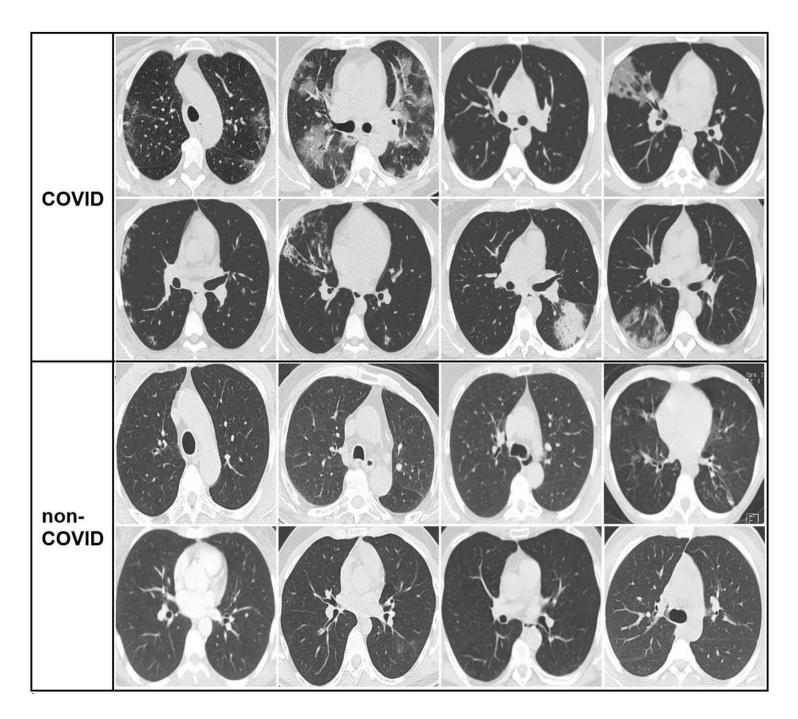


Figure 2

Examples of CT scans for infected and non-infected patients with SARS-CoV-2 (COVID-19) that make up the SARS-COV-2 Ct-Scan Dataset.

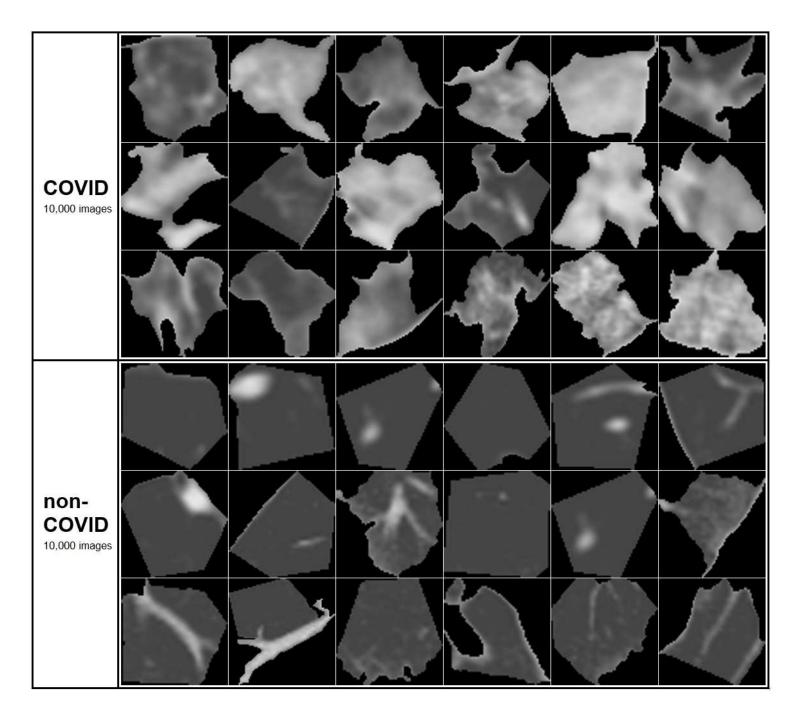


Figure 3

Examples of superpixels from the COVID20K2C Superpixels Dataset, divided into COVID, non-COVID and image numbers by class.

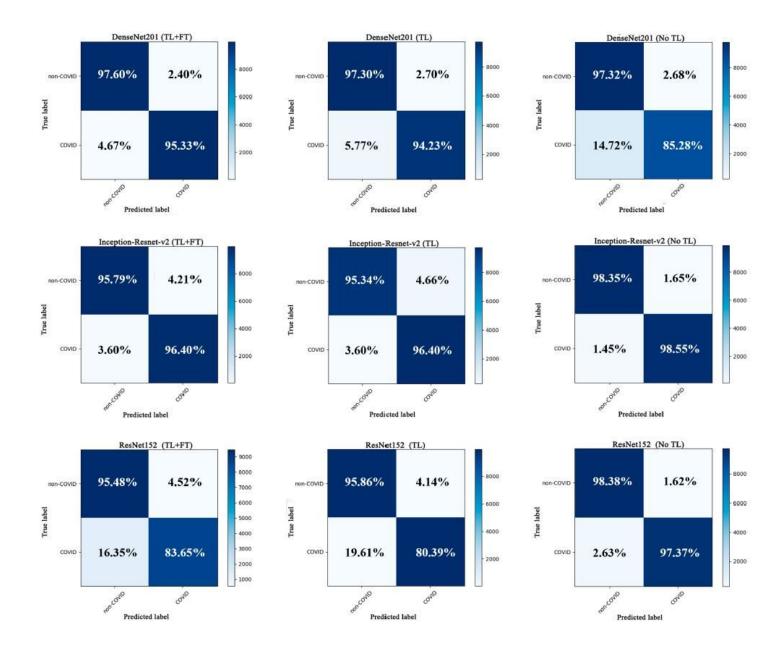


Figure 4

Confusion matrix of each deep learning model. The values in percent represent the average of the 5 iterations in the test set.

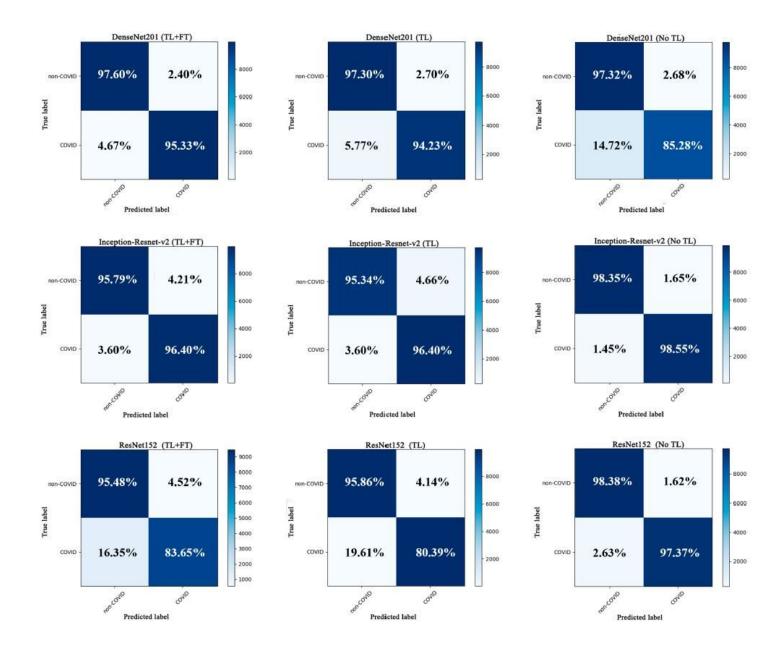
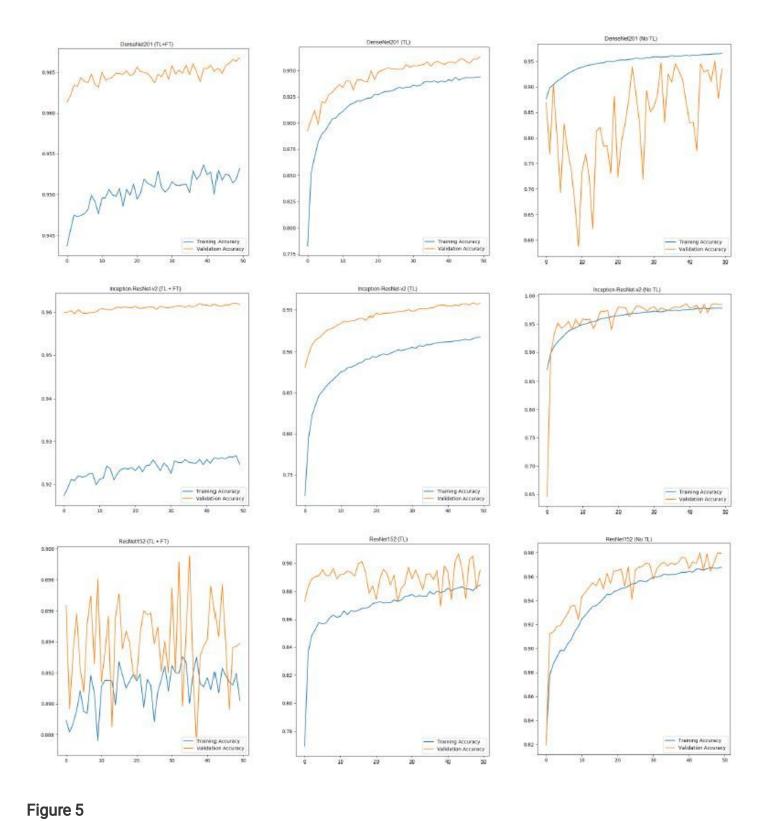


Figure 4

Confusion matrix of each deep learning model. The values in percent represent the average of the 5 iterations in the test set.



Training and validation accuracy of each deep learning model during 50 epochs.

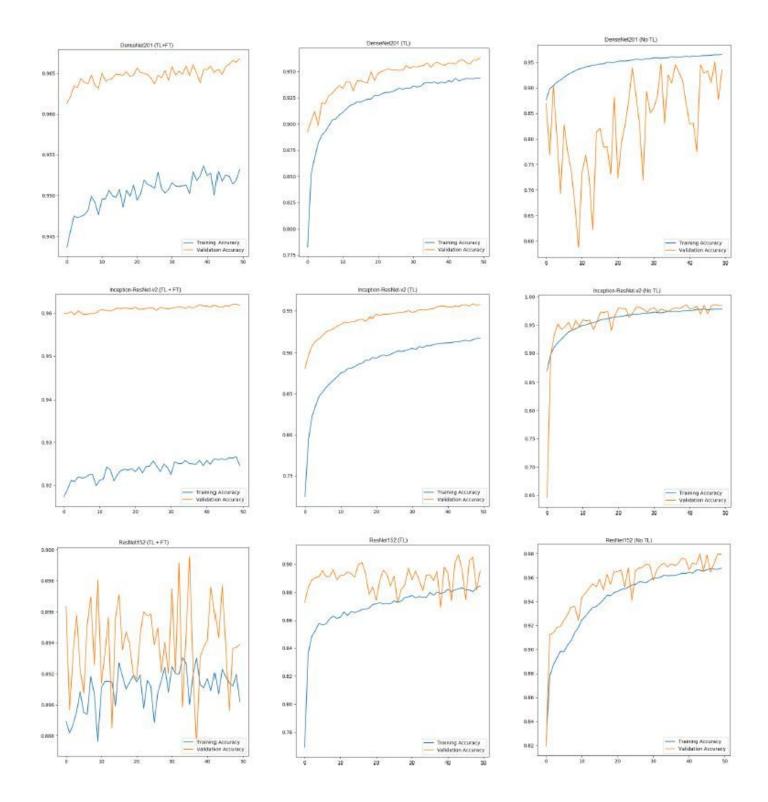


Figure 6

Training and validation accuracy of each deep learning model during 50 epochs.



Figure 6

Screen capture of our computer vision system evaluating a chest CT. In Figure (a), the system called Pynovisão presents the segmentation step of the pulmonary regions. In Figure (b), the color labels represent the categories of our problem. In this case, the system uses the Inception-Resnet-v2 No TL model for the classification of superpixels. Pynovisão was registered by INPI with the number BR 51 2019 000427 2 and is available at: http://git.inovisao.ucdb.br/inovisao/pynovisao.

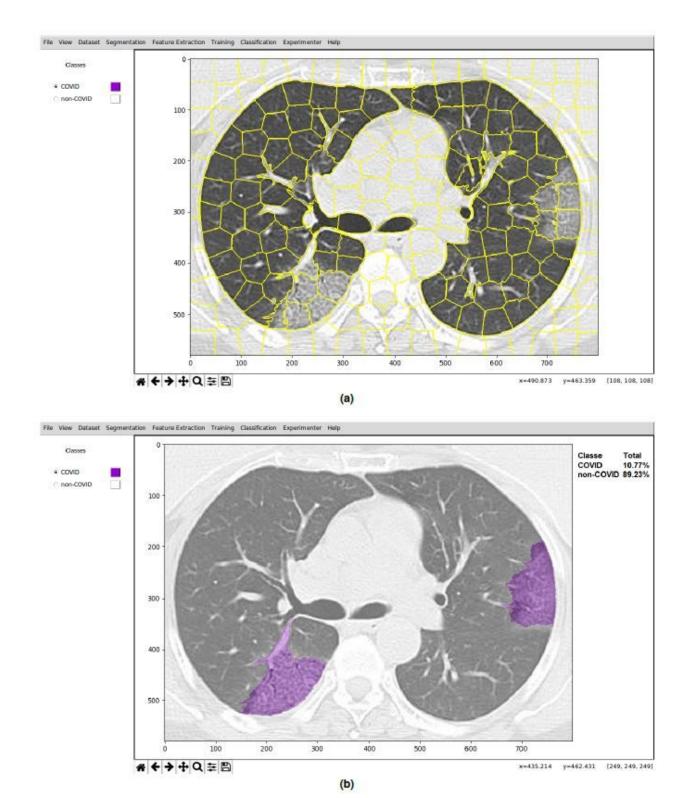


Figure 7

Screen capture of our computer vision system evaluating a chest CT. In Figure (a), the system called Pynovisão presents the segmentation step of the pulmonary regions. In Figure (b), the color labels represent the categories of our problem. In this case, the system uses the Inception-Resnet-v2 No TL model for the classification of superpixels. Pynovisão was registered by INPI with the number BR 51 2019 000427 2 and is available at: http://git.inovisao.ucdb.br/inovisao/pynovisao.