

Article

Diagnosis and Proposed Treatment for COVID-19 Patients Based on Deep Learning Analysis of Computed Tomography Images

Zofia Knapieńska ¹, Jan Mulawka ^{1,*} and Maciej Kierzkiewicz ²¹ Institute of Computer Science, Warsaw University of Technology, Nowowiejska 15/19, 00-665 Warsaw, Poland² Międzyleski Specialist Hospital in Warsaw, Bursztynowa 2, 04-749 Warsaw, Poland; kierzkiewicz@op.pl

* Correspondence: jan.mulawka@pw.edu.pl

Abstract: In this contribution, we consider computed tomography (CT) as a diagnostic tool for identifying coronavirus disease 2019 (COVID-19) pneumonia. However, interpreting CT scans can be subjective, leading to interobserver variability and potential misdiagnosis. To address these challenges, a deep learning-based chest approach was developed to create a precise diagnostic tool for COVID-19 pneumonia and a personalized therapeutic strategy for individual patients. The study collected chest CT images from patients with different lung conditions, creating a diverse convolutional neural network (CNN) training material. Three different CNN-based models were tested for diagnostic purposes, with the output stating whether the patient was healthy or infected. The models facilitated selecting regions of interest (ROIs) and extracting the radiomic features from the input data, resulting in satisfactory results with core classification quality measures above the 50% threshold. For therapeutic purposes, a custom U-Net-based model was used to extract lung and infection masks from a provided CT slice. The percentage of the pathologically altered tissue was calculated, and the COVID-19 severity score was computed and then matched with an optimal therapeutic strategy. Overall, the models delivered high-quality results, representing a functioning deep learning-based application that could be advantageous as a doctor-friendly support tool. The use of deep learning techniques in medical imaging shows promising results, improving the accuracy and speed of diagnosis and treatment of not only COVID-19 but also many different diseases.



Citation: Knapieńska, Z.; Mulawka, J.; Kierzkiewicz, M. Diagnosis and Proposed Treatment for COVID-19 Patients Based on Deep Learning Analysis of Computed Tomography Images. *Appl. Sci.* **2023**, *13*, 7565. <https://doi.org/10.3390/app13137565>

Academic Editor: Andrea Prati

Received: 29 April 2023

Revised: 21 June 2023

Accepted: 25 June 2023

Published: 27 June 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: radiomics; COVID-19; deep learning; chest computed tomography scan; convolutional neural networks; U-Net; pathologically changed lung tissue; treatment planning; diagnostic tool; clinical decision support; personalized therapeutic strategy

1. Introduction

Accurate detection and diagnosis of coronavirus disease 2019 (COVID-19) are pivotal in curbing its spread and managing affected individuals. Currently, a combination of molecular tests, serology tests, and chest computed tomography (CT) scans forms the backbone of diagnostic strategies. However, these established methods, while essential, have limitations. Therefore, there is an ongoing pursuit of alternative approaches that offer improved sensitivity and specificity and faster detection times [1].

Among the commonly used molecular tests, the reverse transcription polymerase chain reaction (RT-PCR) technique stands out as a widely employed method endorsed by the World Health Organization (WHO). Despite this endorsement, PCR-based methods have drawbacks; particularly, limited sensitivity, especially during the early stages of infection, which can result in false-negative results. Moreover, RT-PCR tests focus solely on detecting the presence of the virus, making them unable to track individuals who have silently recovered from asymptomatic infection [1].

In contrast, chest CT scans have emerged as a promising tool for identifying abnormal findings associated with COVID-19 and other types of pneumonia. By providing valuable

support in cases with strong clinical suspicion, CT scans offer several advantages, including non-invasiveness and swiftness. However, the use of chest CT scans is constrained by the requirement for expensive equipment and trained technicians. Furthermore, the overlapping features shared between COVID-19 and other types of viral pneumonia pose challenges in terms of specificity. Therefore, careful consideration is necessary before routinely implementing CT scans for screening purposes, especially due to potential radiation exposure risks [2,3].

Comparative studies have shed light on the limitations of PCR-based methods and highlighted the superior sensitivity of chest CT scans in detecting COVID-19. These studies have consistently demonstrated the ability of chest CT scans to identify abnormalities in nearly all COVID-19 patients, including those who are asymptomatic or initially test negative with RT-PCR tests, underlining the superior effectiveness of CT scans as a primary diagnostic method, particularly in epidemic areas [1,4].

The continuous search for improved approaches to COVID-19 diagnosis and treatment has led to the exploration of machine learning-based techniques. In this article, we aim to investigate the potential of these techniques in enhancing the sensitivity, specificity, and efficiency of COVID-19 diagnosis through the analysis of chest CT scans. Additionally, we explore how these techniques may offer valuable insights for therapeutic suggestions regarding the effective management of the disease. While the field of machine learning holds promise, it is important to approach these techniques with careful consideration and evaluate their feasibility and practical applications in the context of COVID-19.

Typical findings related to patients' chest CT scans are depicted in Figure 1. In particular, they include:

1. Ground glass opacities (GGOs): The most common finding in COVID-19-infected patients' CT scans. They are commonly present as multifocal, bilateral, peripheral, and asymmetrical lesions. In the less advanced disease, they may appear as a unifocal pathology in the inferior lobe of the right lung [5];
2. Crazy paving: An alveolar filling associated with thickened interlobular and intralobular lines. This pathology is recognized in the later stages of the disease [5];
3. Vascular dilatation: An asymmetrical internal enlargement of pulmonary arterial branches [6];
4. Architectural distortion within the lung tissue that usually leads to the formation of subpleural bands [7];
5. Traction bronchiectasis: An asymmetrical varicose dilatation of the segmental bronchi [8].

However, the primary objective of this contribution was to test various radiological tools for detection, classification, and treatment planning for COVID-19 pneumonia. The study also aimed to demonstrate the effectiveness of radiomics methods in clinical decision making for COVID-19 patients. To achieve this, lung CT scans were selected and analyzed in consultation with experienced radiologists. Their feedback was used to identify specific features of COVID-19 infection and choose optimal deep learning methods for data processing and classification.

A program utilizing a convolutional neural network model was subsequently developed. This program extracts characteristic features from the regions of interest and analyzes them to deliver a final diagnosis to clinicians. This program indicates whether the patient is healthy or infected with COVID-19, estimates the percentage of pathologically changed tissue, and determines the severity of the disease. These findings are then used to propose the most appropriate therapeutic approach. The effectiveness of this approach highlights the potential of radiomics methods in aiding clinical decision making for patients with COVID-19 infection.

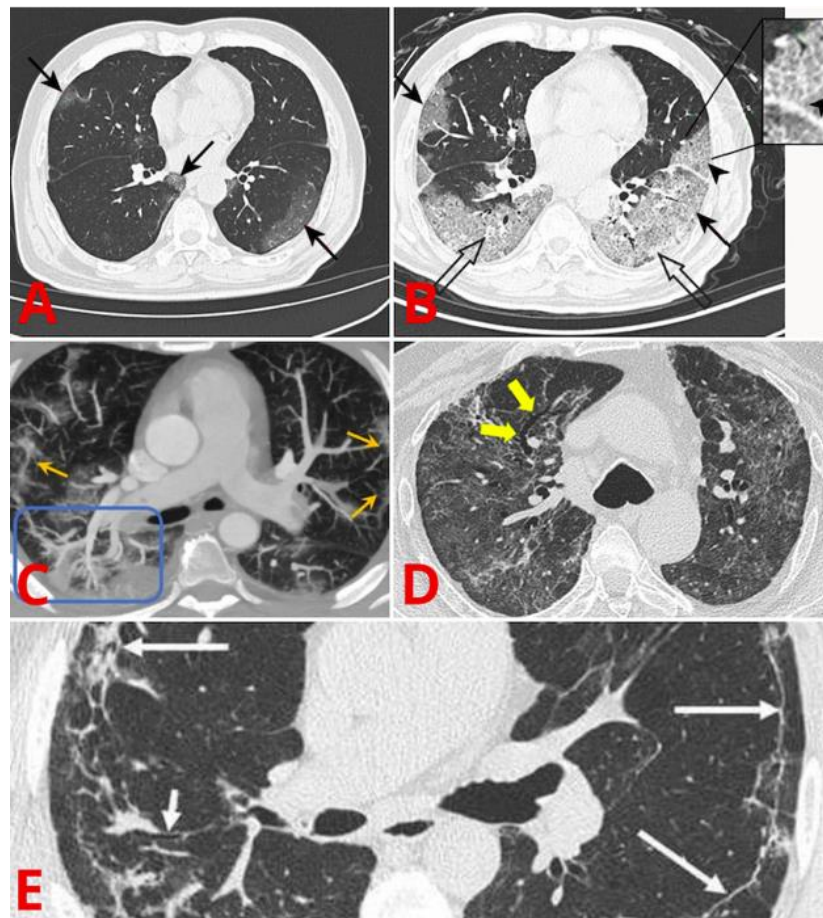


Figure 1. (A) GGOs (black arrows): left lower and right middle lobes. (B) Crazy paving (arrow-heads), confluent tissue concentrations (clear arrows), and extensive bilateral GGOs (black arrows). (C) Subpleural peripheral GGOs (orange arrows), pulmonary vascular enlargement (PVE), and asymmetrically dilated right lower lobar pulmonary arterial branches (blue square). (D) Traction bronchiectasis visible in anterior segmental bronchus and right upper lobe (yellow arrows). (E) Secondary bronchial dilation (short arrows), architectural distortion of the pulmonary tissue, and subpleural parenchymal bands (long arrows).

Recent Progress in Convolutional Neural Networks

Neural networks imitate the behavior and properties of neurons in the human brain. They can recognize patterns in data and are the essential basis of artificial intelligence. An artificial neural network (ANN) consists of interconnected nodes organized into layers. The first layer is the input layer and the last is the output layer. There can be multiple hidden layers in between, with the number depending on the complexity of the task. The input layer accepts and transfers the input data matrix, which includes relevant features. Hidden layers apply mathematical transformations to the input matrices to improve accuracy and the output layer provides the final results. For classification tasks, the number of neurons in the output layer corresponds to the number of classes [9]. They consist of interconnected nodes, and the connections between them have varying weights to optimize the learning process.

Neural networks are advantageous due to their learning ability, whereby they process many examples to analyze associations within and between data groups. In this regard, supervised learning is mainly discussed, where the output from the model is compared with the default output during the training phase.

Generally, there are three common types of neural networks: ANNs, recurrent neural networks (RNNs), and convolutional neural networks (CNNs). CNNs are the primary

research focus in deep learning and are useful for image recognition, classification, and processing tasks. A CNN has multiple layers and each layer detects different image features through the application of new filters as the input image passes through each layer. On the whole, CNNs have several advantages, such as the ability to automatically capture an image's spatial and temporal dependencies without the need for manual pre-processing of data. The architecture includes three main types of layers: convolution, pooling, and fully connected (FC) layers. Furthermore, CNNs are flexible, and pre-trained models can be used for new image recognition tasks [10].

The convolution layer is a crucial component of the CNN architecture responsible for feature extraction through a combination of linear and nonlinear operations. The layer requires an RGB image input with three dimensions (height, width, and depth), a filter, and a feature map. The feature detector or filter moves across the image and conducts mathematical operations to check for the presence of specific features. The filter is typically a 3×3 matrix that is shifted by a stride across the image to extract distinct feature maps representing different characteristics of the input tensor. The feature detector's weights are shared as it moves across the image, allowing for fixed local feature patterns and spatial hierarchies of feature patterns and reducing the number of parameters to learn. The convolution layer's hyperparameters, including the number of filters, stride, and zero-padding, need to be adjusted before training to affect the output's volume size. A rectified linear unit (ReLU) transformation is applied to the feature map to introduce nonlinearity to the model. If one convolution layer follows another, the CNN structure becomes hierarchical [11].

The pooling layer is an essential component of CNNs responsible for reducing the input's dimensionality by decreasing its spatial size. Unlike the convolutional layer, the pooling operation does not carry any weights but applies an aggregation function to the values within the receptive field. The two most commonly used types of pooling are max pooling, which outputs the maximum value from a selected set of pixels, and average pooling, which returns the average value of the given pixel portion as an output. Max pooling is more commonly used nowadays and can be considered a denoising operation applied to the input. These layers limit the risk of overfitting and improve classification efficiency.

The final crucial element of a CNN is the FC layer. This layer effectively processes and learns from the output of the previous layer, which typically represents complex, nonlinear combinations of high-level features. In an FC layer, each output node is directly connected to the previous layer's node, making it particularly effective for classification tasks. To produce a probability distribution as the final output, FC layers often use a softmax activation function.

It should be mentioned that there are a wide range of pre-trained CNN architectures available for download, each specialized for different classification tasks. When applied appropriately to the input data, these architectures have demonstrated versatility and efficiency [9,12–14].

2. Materials and Methods

The entire program script and the models were created using the Python programming language. As it is developed under an open-source license, it gives the user access to many libraries, including the ones used for working with deep learning, neural networks, and artificial intelligence (AI).

2.1. Python Libraries Used and Data Selection

The libraries used included TensorFlow and Keras, which provide an interface for solving classification problems and essential building blocks for developing deep learning models, and Matplotlib, NumPy, Pandas, Plotly, SciPy, Scikit-learn, OpenCV, Pydicom, Pyradiomics, NiBabel, Imageio, Pillow, and Tkinter.

The datasets used for training the deep learning models were downloaded from Kaggle. They are summarized in Table 1. Below, they are presented together with their descriptions:

1. COVID-19 CT scans [15]: This image collection encompasses a set of 20 CT scans obtained from patients diagnosed with COVID-19. The dataset also includes the results of lung segmentation in the form of annotated lung and infection masks produced by experienced radiologists;
2. Large COVID-19 CT scan slice dataset [16]: A curated COVID-19 CT scan collection with a compilation of images drawn from six publicly available datasets. The scans have been meticulously divided into three distinct categories: “COVID-19 Positive”, “Normal”, and “Community-Acquired Pneumonia (CAP)”. This dataset encompasses a total of 7593 images of COVID-19-positive cases obtained from 466 infected individuals, 6893 images of healthy individuals categorized as “Normal” obtained from 604 patients, and 2618 images categorized as “CAP” obtained from 60 additional individuals;
3. COVID-19 Lung CT Scans [17]: This dataset comprises a total of 349 CT scans, of which 216 were obtained from individuals diagnosed with COVID-19. In compliance with the ethical guidelines, all patient information was anonymized prior to inclusion in the set. The CT scans were sourced from a range of peer-reviewed journals, including medRxiv, bioRxiv, NEJM, JAMA, and The Lancet, and were carefully annotated by a team of radiologists.

Table 1. Summary of the contents of the databases used for model fitting and subsequent training.

Dataset Used	Number of Patients		
	COVID-19-Infected	Healthy	CAP
COVID-19 Lung CT Scans	216	133	
Large COVID-19 CT scan slice dataset	466	604	60
COVID-19 CT scans	20		

2.2. CT Image Segmentation

Lung segmentation is a critical step in CT scan analysis, focusing on identifying and analyzing the specific region of interest within the lungs. By eliminating surrounding tissue, this technique enhances analysis accuracy and efficiency, as illustrated in Figure 2. However, it is crucial to avoid losing any part of the ROI, such as lung nodules near the thoracic wall and structures such as bronchi within the lung area, during segmentation [18]. In CT scans, the lung area typically appears darker than neighboring sections, while brighter regions within the lungs may indicate air, blood vessels, or pathologically altered tissue. The following steps are involved in lung segmentation [19]:

- Conversion of the image with RGB channels into a grayscale image;
- Binarization of the CT slice using a user-defined threshold value;
- Removal of objects outside the ROI but still connected to the image border;
- Labeling of regions within the image;
- Measurement of the properties of labeled image regions and elimination of everything except for the two largest entities (right and left lung);
- Conducting binary erosion with a disk of a chosen radius to separate the lung nodules attached to the blood vessels from surrounding tissue;
- Removal of blood vessels from the image;
- Filling in the holes created during the previous operation while preserving the nodules attached to the pleura;
- Detection of lung edges using the Roberts operator to estimate and highlight the spatial gradient zones in the image;
- Filling in any remaining holes within the ROI;
- Superimposing both binary and infected lung masks on the original chest CT scan.

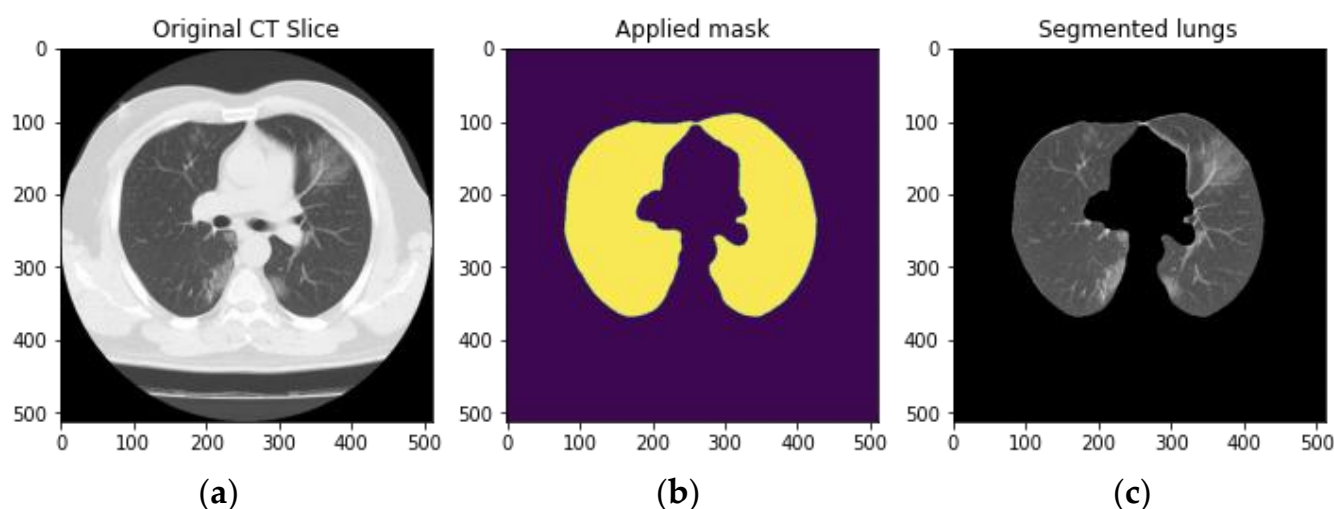


Figure 2. Lung segmentation—example: (a) pathological chest CT scan; (b) applied segmented lung mask; (c) final segmentation outcome.

2.3. Preliminary Data Processing and Analysis

To prevent overfitting or underfitting in a deep learning model, it is necessary to split the available CT images into separate sets for training and validation, as outlined in Table 2. This ensures that the model does not use the same information twice and helps to mitigate potential errors.

Table 2. Dataset division characteristics.

Main Dataset Division into Three Smaller Sets		
Training	Validation	Testing
Fits the parameters of each proposed model	Helps evaluate the model's predictive performance and adjust the hyperparameters	Enables estimation of the generalization error for one chosen model

Data augmentation is a technique used to artificially increase the training dataset when access to a comprehensive set is not possible. It involves generating multiple versions of a single image, which can prevent overfitting and improve the model's ability to generalize.

When working with small datasets, it is important to ensure adequately high accuracy for the model's output. By modifying the images in the training set with translations, distortions, rotations, flips, shifts, brightness adjustments, sheerness adjustments, and scale changes, the model is exposed to a variety of slightly different images in each epoch. This method helps ensure that the model trains with a more diverse set of images and ultimately leads to improved accuracy [20,21].

2.4. Neural Networks Characteristics Utilized

Transfer learning is conducted prior to model training when a pre-trained model is used for feature detection/classification. A user can freeze all or some layers, add new ones on top of the model, and then retrain it with a fine-tuning process after an initial training session. The Keras CNN library [22] offers pre-trained VGG19 and InceptionResNetV2 models for image multiclassification [23,24].

The U-Net-based model was primarily intended to extract the infection masks from the CT lung slices. The infection mask is the entire infected area within the lung tissue border. Extracting it can also be considered a further segmentation step.

These data allowed us to calculate the percentage of the lung area covered by pathological changes in a given CT cross-section. This percentage helped to assess the COVID-19

severity score, which can be used as a basis for appropriate therapeutic strategy implementation. The patient will be treated differently depending on how advanced their condition is.

Additionally, the extracted infection mask can be graphically converted into a heatmap, which helps the viewer visualize which lung areas have been affected by the pathological changes the most.

For all the models summarized in Table 3, accuracy and loss functions were chosen as quality measures for classification. If the loss function value increased or stopped improving, the learning rate was decreased. The best-performing epoch was selected at the end of training and the associated weights or entire model were saved.

Table 3. Summary of the proposed models' characteristics.

Type of Model	Basic Architectural Attributes		
Standard VGG19 model	19 layers	Input characteristics	
		RGB images of size 224×224	
		Input matrix: (224, 224, 3)	
		Output characteristics	
New VGG19-inspired model	Eight new layers on top of a frozen, pre-trained VGG19 model Input matrix: (128, 128, 3) The new model has 75,074 trainable parameters	Softmax activation function	
		Input characteristics	
		RGB images of size 299×299	
		Input matrix: (299, 299, 3)	
Standard InceptionResNetV2 model	164 layers	Output characteristics	
		Softmax activation function	
		Input characteristics	
		RGB images of size 299×299	
New InceptionResNetV2-inspired model	Four new layers on top of a frozen, pre-trained model Input matrix: (128, 128, 3) The model has 54,673,507 trainable parameters	Input matrix: (299, 299, 3)	
		Output characteristics	
		Softmax activation function	
		Input characteristics	
New custom CNN model	54,673,507 trainable parameters 960 non-trainable parameters 1,898,202 parameters in total	Greyscale images of size 128×128	
		Input matrix: (128, 128, 1)	
		Output characteristics	
		Softmax activation function	
Custom U-Net	The contracting path has a structure where many convolution blocks are followed by maxpool downsampling and ReLU. This helps encode the input image into varied complexity- and depth-feature representations. The expansive path semantically links the input features to the pixel space. As a result, a dense classification is obtained. It is achieved thanks to the upsampling, concatenations, and a sequence of upconvolutions with high-resolution features 2,158,417 trainable parameters 2,158,417 parameters in total		

2.5. Neural Network Training and Model-Specific Overview

It is essential to adjust the weights during model training to achieve the best classification results. The weights are gradually modified to increase the accuracy of the outcomes. Log loss/binary cross-entropy is a commonly used loss function for binary classification tasks.

Various gradient-based optimization algorithms are available to optimize the learning process and thus lower the value of the loss function, including Adam and RMSprop. The learning rate needs to be manually set. Monitoring the loss function during training and adjusting the learning rate accordingly are essential.

Callbacks such as ModelCheckpoint and ReduceLROnPlateau can automate tasks and optimize the training process. ModelCheckpoint saves the model at a specific frequency at the end of every epoch. ReduceLROnPlateau reduces the learning rate if the monitored classification quality metric is not improving. Implementing these callbacks can help achieve better accuracy and faster training times. In summary, adjusting weights and loss function, selecting the appropriate learning rate, monitoring the training process, and using callbacks are essential components in optimizing a model's training process [25–27].

To ensure high-quality classification in all models, the accuracy and loss function were carefully monitored during the training. If the loss function value ceased to improve, the learning rate was automatically decreased in all cases. Once the training was complete, the best-performing epoch was identified, and the associated weights and the entire model were saved. The training parameters of different models are summarized in Table 4.

Table 4. Overview of the different models' training and parameter settings.

Type of Model	Epochs	Steps per Epoch	Tolerance	Additional Remarks
VGG19	100	16	Five epochs	The learning rate was lowered if the loss function value remained unchanged over five epochs
InceptionResNetV2	40	Dependent on the batch size	One epoch	Training stopped after the 20th epoch, as the loss function value stopped decreasing and remained unchanged over five subsequent epochs
Custom CNN	20	363	Five epochs	The learning rate was lowered if the loss function value remained unchanged over five epochs
U-Net	315	315	Five epochs	The learning rate was lowered if the loss function value remained unchanged over five epochs

During the binary classification training task, the performance of the first three distinct models was evaluated based on their validation accuracy (see Figures 3–5). The custom CNN model achieved the highest value for accuracy of 98.21%, followed by the InceptionResNetV2-based model with a score of 97.20%. This particular model performed the classification task by sorting the data into three distinct classes instead of two as in the other models. This is also why the appearance of the graphs (see Figure 4) was different. The VGG-19-based model had the lowest performance, with a validation accuracy of 83%. Based on the overall performance in the binary classification task, the custom CNN model was determined to be the most appropriate architecture for the investigated problem.

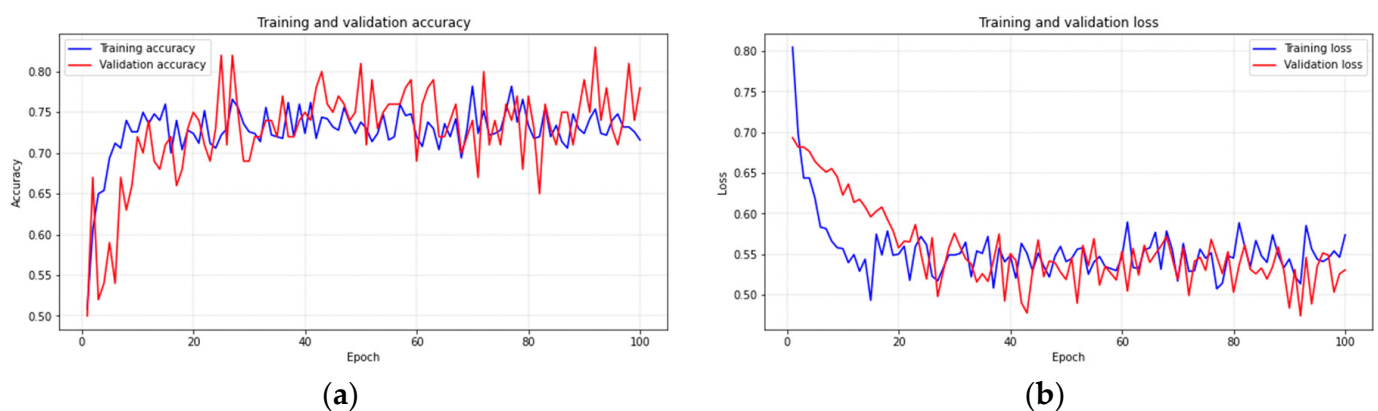


Figure 3. Custom VGG-19-based model—assessment of the training quality: (a) accuracy; (b) loss function.

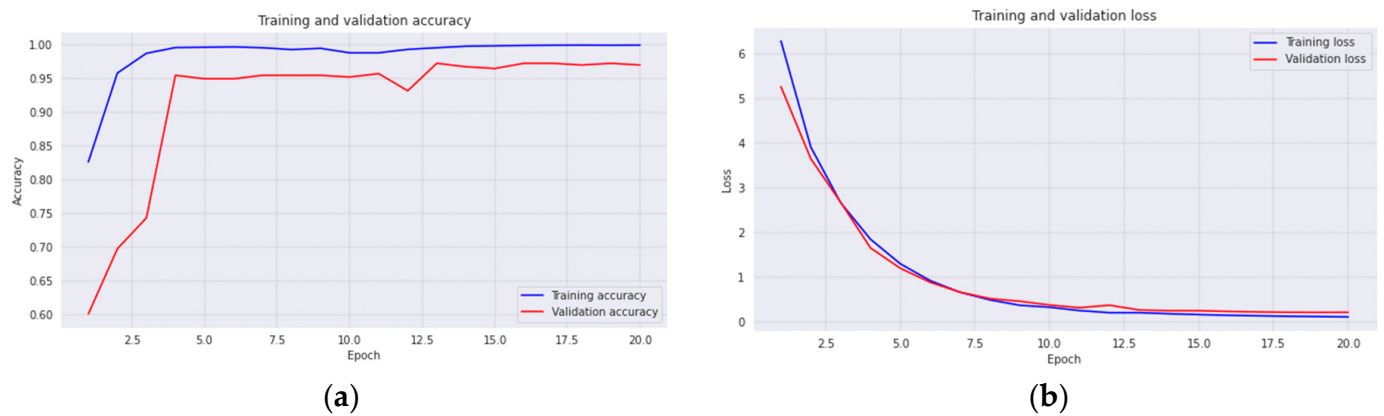


Figure 4. Custom InceptionResNetV2-based model—assessment of the training quality (in contrast to the other models, this one sorted the data into three distinct classes): (a) accuracy; (b) loss function.

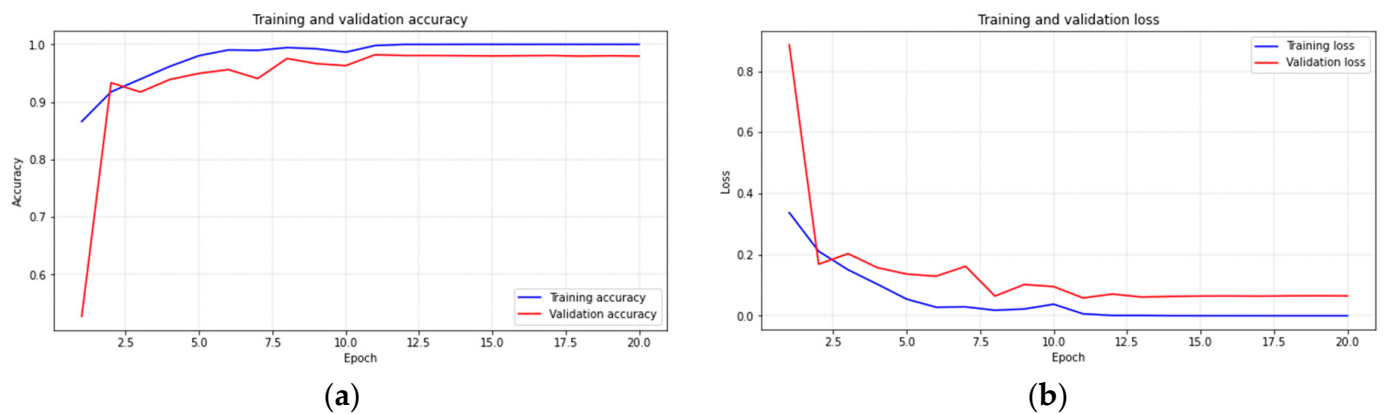


Figure 5. Custom CNN model—assessment of the training quality: (a) accuracy; (b) loss function.

The fourth model, the U-Net model, was trained on numerous lung CT slices obtained from various patients, each previously annotated by a radiologist (see Figure 6). Each CT slice was associated with a corresponding lung mask, infection mask, and superimposed lung with an infection mask. The U-Net model was trained to accurately extract both the lung and infection masks from each CT slice. To assess the quality of the extraction results, the accuracy and loss function were computed and compared against the original annotations.

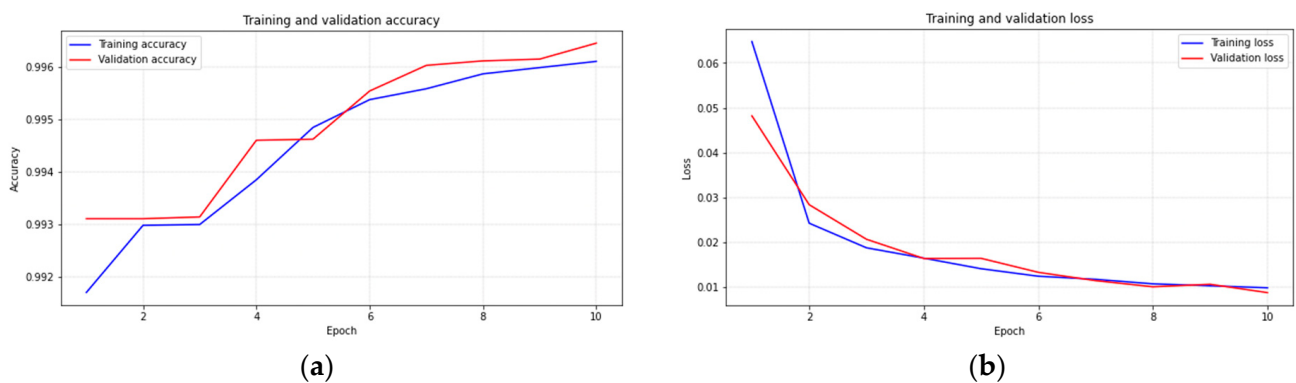


Figure 6. Custom U-Net-based model—assessment of the training quality: (a) accuracy; (b) loss function.

2.6. COVID-19 Severity Score

The manifestation of COVID-19 in the lungs can take various forms, underscoring the importance of an accurate numerical assessment of tissue damage. A COVID-19 severity score is critical for patient triage, predicting disease progression, and developing effective treatment plans. While RT-PCR is currently the most common diagnostic tool, CT scans may be necessary for patients suspected of COVID-19 infection and those who have negative RT-PCR test results.

The U-Net-based model can generate masks that facilitate the calculation of the percentage of lung tissue affected by the pathological process. The following steps can be taken to obtain this information:

1. Determine the number of pixels corresponding to the total lung area;
2. Calculate the number of pixels corresponding to the infected area within the segmented lung borders;
3. Utilize the values from the previous steps to calculate the percentage of lung tissue affected by the pathological process.

The resulting percentage can be used to derive a COVID-19 severity score, which provides valuable insights for optimal patient hospital management and therapeutic approaches.

3. Results

Machine learning models are frequently used for binary classification problems. Statistical analysis in the form of a confusion matrix can help understand their performance, with each row representing the actual class and each column representing the predicted class. A color-coded matrix can make it easy to determine whether the model can accurately differentiate between the classes.

The receiver operating characteristic (ROC) curve is another statistical measure that can be used for binary classification problems. The ROC curve is a plot that visualizes the diagnostic ability of a binary classification model. It presents sensitivity as a function of specificity, with each point on the curve representing a different threshold for the model's prediction. A perfect ROC curve is obtained when the model's performance is plotted in the upper-left corner, representing that both sensitivity and specificity are approximately 100%. The ROC curve is an efficient tool that can be used for optimal model selection.

The diagonal line in the ROC space splits it into two parts, with points in the upper part representing good classification. The area under the ROC curve (AUC) should also be considered when comparing different models. The AUC ranges from 0 to 1, with a score of 1 representing perfect model performance and 0 representing random categorization to a given category.

3.1. Summary of the Models' Performance

1. VGG-19-based model

A model was used to accurately differentiate between lung images with COVID-19-induced pathological tissue changes and those images without such changes. The number of true-negative and true-positive cases was much higher than that of false-positive and false-negative cases, indicating that the model did not randomly assign objects to either of the two available classes.

The model's performance was assessed (Figure 7) using a confusion matrix and ROC curve, which showed that 75% of the processed images were correctly categorized into one of the two classes. The final accuracy achieved by the model was considered appropriate for research settings. Therefore, the model accurately distinguished between pathological changes induced by COVID-19 and healthy lungs.

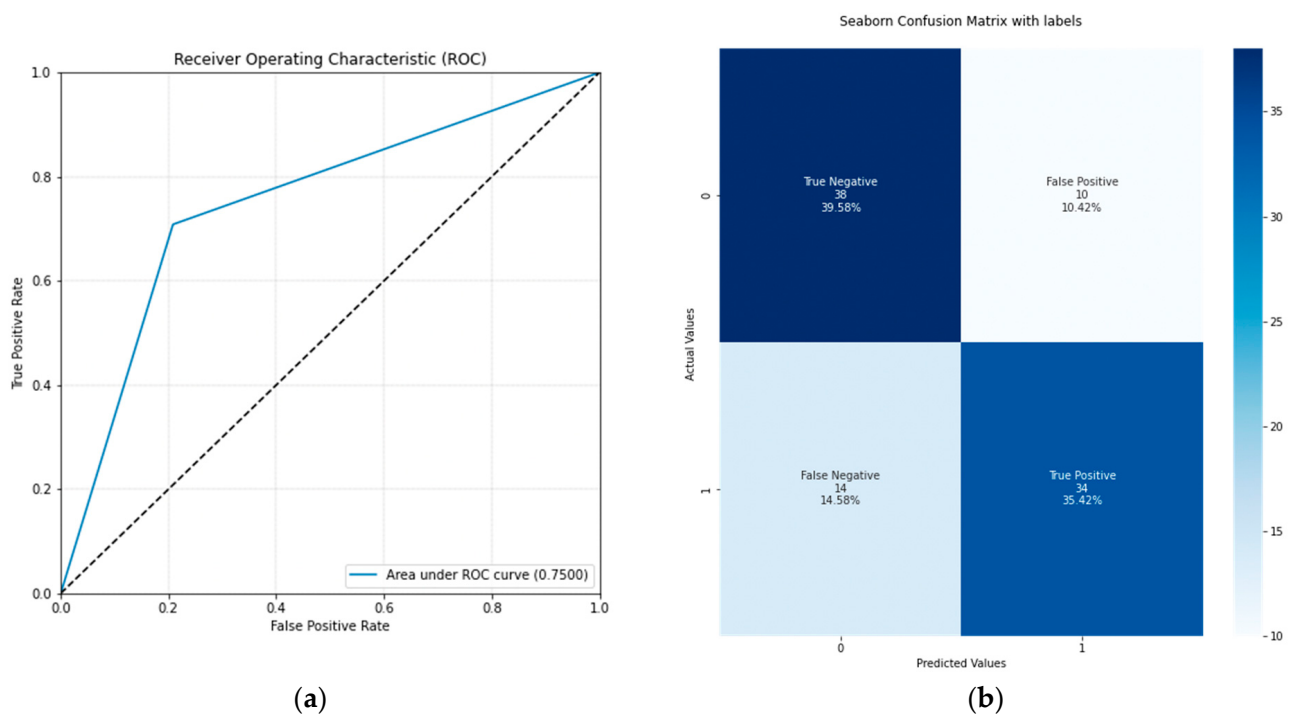


Figure 7. VGG-19-based model Testing results and performance evaluation: (a) ROC curve; (b) confusion matrix. 0—healthy patient, 1—COVID-19-infected patient.

2. InceptionResNetV2-based model

This model was trained to categorize objects into three classes, as shown in the confusion matrix. The ROC curve, however, only assessed the performance in differentiating COVID-19 cases from two other diagnostic possibilities. These results are shown in Figure 8.

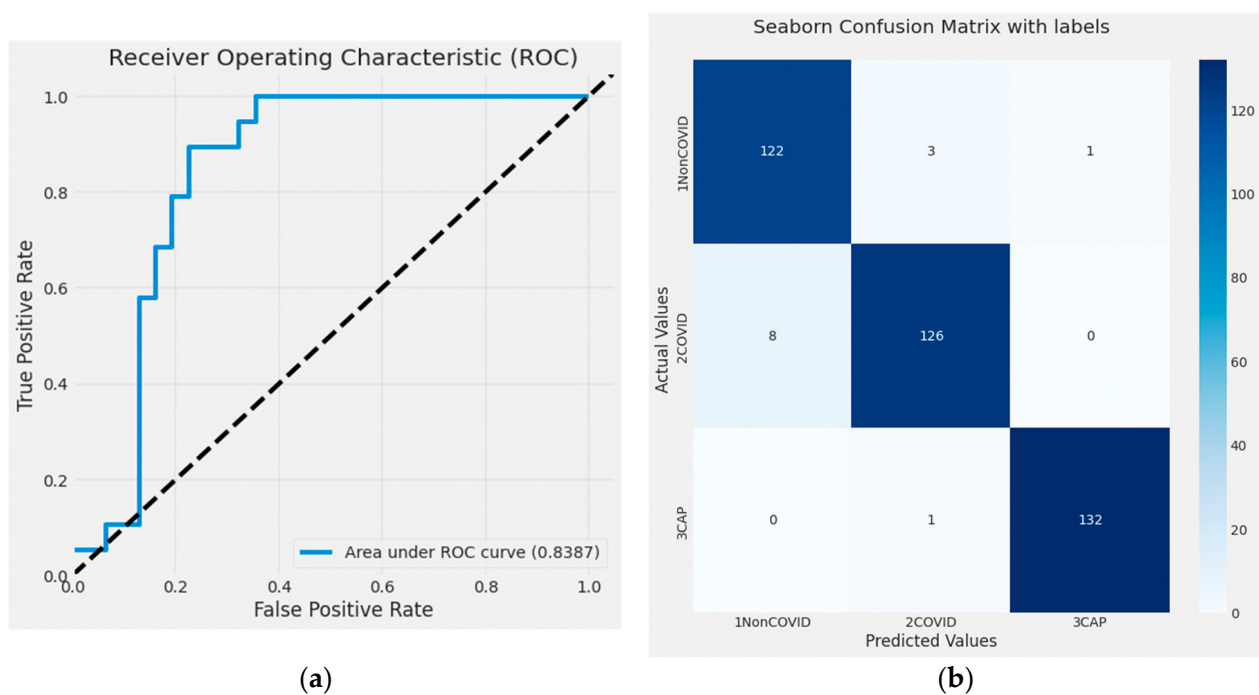


Figure 8. InceptionResNetV2-based model Testing results and performance evaluation: (a) ROC curve; (b) confusion matrix. 3CAP—community-acquired pneumonia-infected patient, 2COVID—COVID-19-infected patient, 1Non-COVID—healthy patient.

The confusion matrix indicated that the model effectively sorted images into categories without randomly assigning them. This was confirmed by the ROC curve, which showed a high number of true-positive and true-negative cases for all three conditions. The model performed best in detecting community-acquired pneumonia from lung CT slices, while it struggled the most with categorizing COVID-19-infected patients. Nonetheless, the confusion matrix and ROC curve demonstrated that the model correctly identified 84% of COVID-19 cases. Therefore, the model's final accuracy is suitable for research use.

3. Custom CNN model

The confusion matrix and ROC curve demonstrated that the model accurately performed binary classification without randomly assigning images to either class. The ROC curve closely aligned with the left corner of the plot, indicating a high number of true-positive and true-negative cases (see Figure 9).

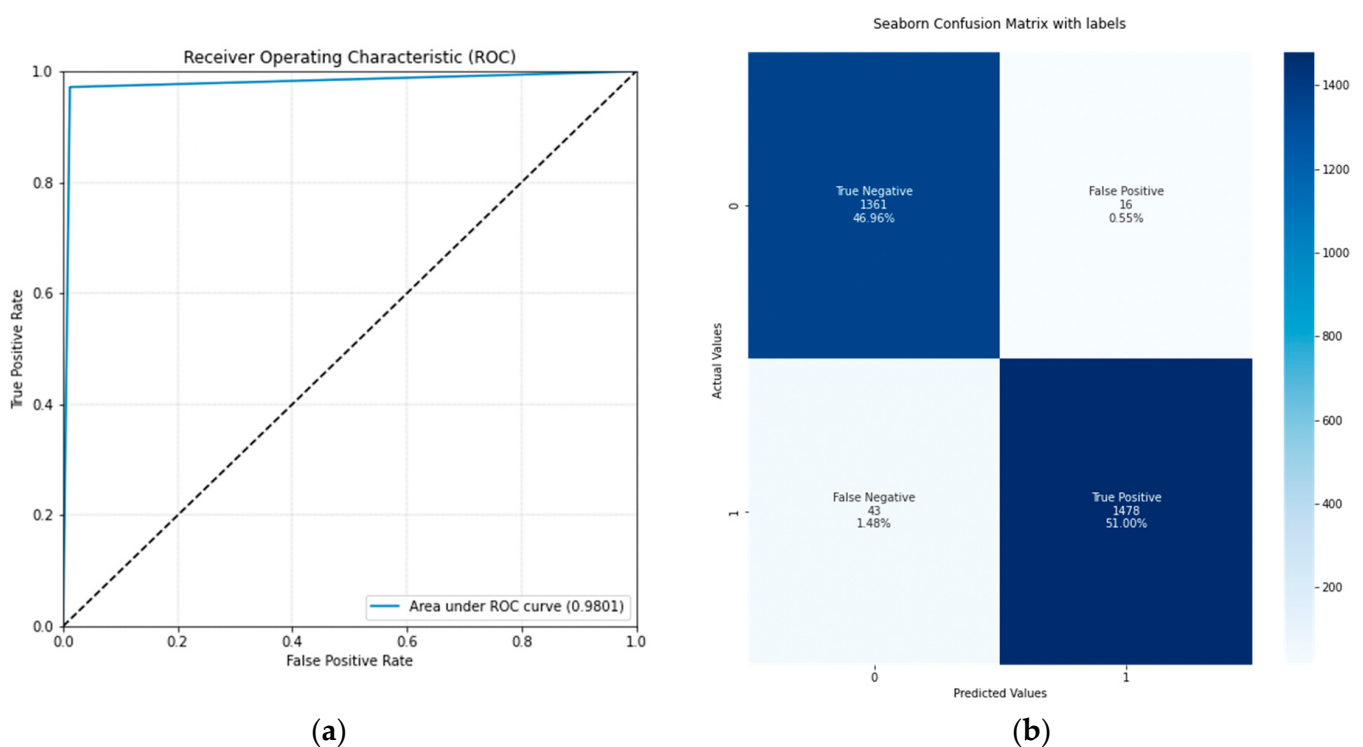


Figure 9. Custom CNN model Testing results and performance evaluation: (a) ROC curve; (b) confusion matrix. 0—healthy patient, 1—COVID-19-infected patient.

The model achieved a diagnostic accuracy of 98.01%, as shown by both the confusion matrix and ROC curve. This result was considered satisfactory and demonstrates that the model can be applied in clinical research settings.

Table 5 shows the model testing results. Organizing the results into a table facilitated a comprehensive and in-depth analysis of the statistical measures. This approach enabled a more straightforward comparison of the outcomes of the three models evaluated in this study.

When reviewed, it was found that the custom CNN model achieved the highest diagnostic quality. Most of its cells corresponding to specific statistical measures were shaded in green, indicating superior performance compared to the other two models. It needs to be highlighted that the custom CNN model also demonstrated the highest accuracy, precision, specificity, and sensitivity, which are considered the most critical statistical measures.

The results for the two other models investigated in this study can be considered satisfactory, but the custom CNN model clearly outperformed them. The VGG-19-based

model demonstrated the lowest diagnostic quality, while the InceptionResNetV2-based model produced better outcomes, with accuracy, precision, specificity, and sensitivity scores above 50% in both cases.

The testing process for all three models can be considered successful and their diagnostic ability confirmed, and they can be efficiently utilized for research and scientific purposes. The primary goal was to develop a tool that supports radiologists and improves their efficiency. For this purpose, the custom CNN-based model was the best-suited model from the three discussed models.

Table 5. Comparison of the statistical measures associated with the quality and efficiency of the classification task undertaken by the three tested models. Color coding: green—best outcome, yellow—medium outcome, red—worst outcome.

Classification Quality Measures	Models		
	VGG-19-Based	InceptionResNetV2-Based	Custom CNN Model
Accuracy	0.75	0.97	0.98
Sensitivity	0.71	0.94	0.97
Specificity	0.79	0.98	0.99
Precision	0.77	0.97	0.99
Negative predictive value	0.73	0.97	0.97
False-negative rate	0.29	0.97	0.03
False-positive rate	0.21	0.06	0.01
False discovery rate	0.23	0.02	0.01
False omission rate	0.27	0.03	0.03
Threat score	0.59	0.03	0.96
F1 score	0.74	0.91	0.98
Positive likelihood ratio	2.43	0.97	34.37
Negative likelihood ratio	0.37	15.75	0.03
Balanced accuracy	0.75	0.06	0.98
Prevalence threshold	0.35	0.11	0.10
Log loss score	0.20	0.01	0.01

The U-Net-based model underwent testing with a lung CT slice that was not included in its training data. To ensure compatibility with the model input, the image dimensions were adjusted to fit the default matrix. Furthermore, the image was converted into grayscale. The model output was an infection mask that was presented as a heatmap and overlaid on a segmented lung image. As illustrated in Figure 10, the heatmap captured all the regions impacted by lung tissue degeneration effectively. The intensity of the color in the heatmap correlated with the severity of the observed damage. Thus, the model provided accurate information about the location and intensity of the detected pathologies.

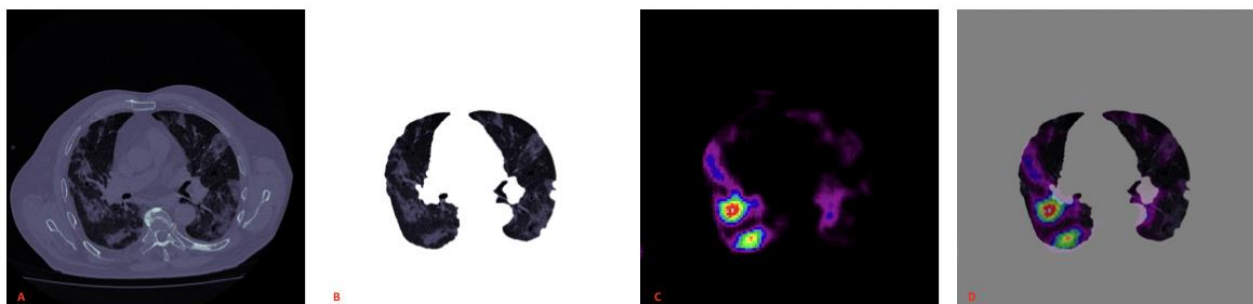


Figure 10. The testing outcome for the U-Net-based model in the form of an infection heatmap. It was produced using (A) a CT slice with visible lung tissue degeneration caused by COVID-19 infection. The next steps were (B) lung segmentation and (C) infection mask extraction. (D) The extracted mask in the form of a heatmap was superimposed on the segmented lung area to visualize where the infection focus points were localized.

3.2. Guidelines concerning the Optimal Therapeutic Approach for Patients with COVID-19 Infection

Various therapies have been developed to target severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Some of these therapies are effective for patients with mild lung involvement, while others are more suited for those with severe lung damage. CT scans are only performed on COVID-19 positive patients when they are hospitalized. Hence, a severity score is needed to determine the optimal therapeutic approach and patient management framework for these patients.

The COVID-19 lung involvement score, which can be objectively determined using the masks yielded by the U-Net-based model, plays a crucial role in establishing patient treatment plans.

Here are the updated guidelines for the optimal therapeutic approach for COVID-19 patients with different severity levels [28]:

1. Severity scores of 1 and 2—patients in stable condition who do not require oxygen supplementation:
 - Do not use dexamethasone or other corticosteroids;
 - Use remdesivir for patients who are at high risk of progressing to severe COVID-19;
 - Administer a prophylactic dose of heparin;
2. Severity score of 3—patients who require conventional oxygen supplementation:
 - Use remdesivir for patients who require minimal oxygen supplementation;
 - For most patients, administer dexamethasone and remdesivir or just dexamethasone (if remdesivir is unavailable);
 - Add either baricitinib or tocilizumab if the oxygen is needed and inflammation markers are rapidly increasing;
 - Administer a therapeutic dose of heparin for nonpregnant patients with D-dimer levels above the upper limit of normal without increased bleeding risk;
3. Severity score of 4—patients who require high-flow nasal cannula oxygen supplementation or noninvasive ventilation:
 - Use just dexamethasone, dexamethasone with baricitinib, or dexamethasone with tocilizumab;
 - Administer a prophylactic dose of heparin;
4. Severity score of 5—patients who require mechanical ventilation or extracorporeal membrane oxygenation:
 - Use just dexamethasone, dexamethasone with baricitinib, or dexamethasone with tocilizumab;
 - Administer a prophylactic dose of heparin.

Figure 11 summarizes the guidelines and highlights the most important information based on lung damage severity [28].

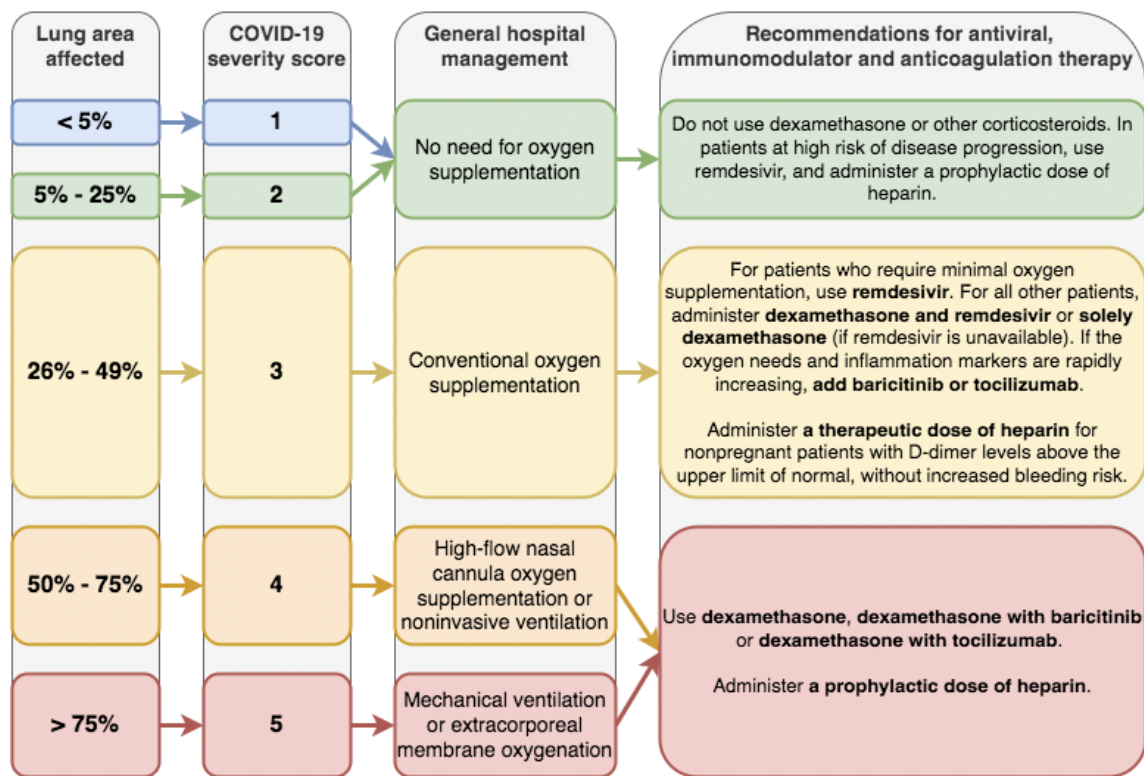


Figure 11. Proposed therapeutic management of hospitalized adults with COVID-19 based on the COVID-19 severity score.

4. Discussion

This study aimed to examine the potential of deep learning as a precise COVID-19 diagnostic tool and for suggesting a patient-tailored therapeutic strategy. The results of this scientific investigation imply that these objectives have been achieved.

The first step involved the collection and preparation of CT medical images from various patients with different conditions. A considerable number of diverse scans were obtained to ensure that the tested models would reach a high level of accuracy.

Three different models based on convolutional neural network architectures were constructed and tested for diagnostic purposes. They provided satisfactory results, with accuracy, sensitivity, and specificity that were well above the 50% threshold. The overall performance of the models was compared, and the third custom-made model delivered the best outcomes. This model was then used in the development of a prototype window application.

The next phase involved predicting the condition's severity and suggesting a therapeutic strategy. A U-Net-based model was trained on lung and infection masks, which enabled it to extract these masks from a CT slice and then calculate the percentage of pathologically altered tissue. This measure was then matched with the COVID-19 severity score, and a patient-tailored therapeutic approach was proposed accordingly. This functionality was also incorporated into the prototype window application.

The study results provide a functioning deep learning-based approach that may prove advantageous in experimental medical and/or scientific settings. The quality of the results ensures reliability, and the approach constitutes a simple clinical decision-support tool.

5. Conclusions

In conclusion, the results obtained from the study were generally satisfactory, highlighting the potential for further development and refinement of the application. However, several areas require attention and improvement.

Firstly, it is essential to acknowledge that the study utilized a limited dataset of medical images. Future research should consider incorporating different and more extensive databases to enhance the robustness and generalizability of the findings by providing a broader representation of cases.

Moreover, the segmentation of the lungs may have unintentionally excluded relevant surrounding tissues and structures, such as tracheal tubes. Modifying the segmentation process and algorithm is crucial to ensure the inclusion of these structures and thus enhance the range of features used in the diagnostic model training.

Additionally, the study only tested three neural network architectures for medical image classification in the diagnostic task. However, exploring a more comprehensive range of architectures may help identify potential superior alternatives. Different architectures offer distinct strengths and capabilities, potentially improving performance in medical image classification tasks.

In terms of therapeutic suggestions, the current U-Net model provided outcomes in terms of the percentage involvement of pathological tissue within the total lung area on each CT slice. However, modifications of the model are required for a more objective and accurate analysis. Analyzing 3D reconstructions of the lungs derived from all CT slices of an individual patient's scan would allow for a more precise evaluation of the extent of the disease. Assessing the percentage involvement of pathologically altered tissue within the total lung volume would allow for the optimization of therapeutic suggestions, mitigating the risk of misdiagnosis associated with relying solely on a limited number of CT slices.

Furthermore, constructing and training multiple models specifically for this task would facilitate a thorough performance comparison, enabling the identification of potential disadvantages associated with each model. This approach would make it possible to select the optimal model for accurate and personalized therapeutic suggestions.

To further enhance patient care and management, it is crucial to consider the integration of the developed models with other diagnostic modalities and medical procedures beyond lung-related CT imaging. This integration would broaden the scope of clinical decisions for COVID-19 patients, enabling a comprehensive approach to diagnosis and treatment. Moreover, it is worth noting that the potential applications of machine learning extend beyond the scope of this study, encompassing various medical conditions and imaging modalities. As medical imaging devices continue to advance, providing improved visualization of previously undetectable features, the utilization of machine learning holds promise for further enhancing the quality of results and expanding its impact in the field of healthcare.

AI integration in radiology holds promise for improving the early detection and diagnosis of malignancies. One significant advancement is the development of a deep learning model that can predict an individual's lung cancer risk up to six years in advance using a single low-dose computed tomography (LDCT) scan. This proactive approach enables personalized treatment plans, leading to improved patient outcomes [29].

The broader adoption of AI in radiology has the potential to revolutionize clinical practices by streamlining workflows and enhancing interpretative speed and accuracy. AI algorithms empower radiologists to deliver more precise and timely diagnoses, optimizing patient care [30,31].

In conclusion, AI and deep learning solutions are transformative in radiology. They enhance early detection, improve diagnostic accuracy, and optimize patient care across diverse medical conditions and imaging modalities. Continued research and development in this field will unlock further potential, benefiting medical professionals, patients, and engineers alike.

Author Contributions: Methodology, J.M.; Software, Z.K.; Validation, Z.K.; Formal analysis, Z.K. and M.K.; Investigation, M.K.; Data curation, M.K.; Writing—original draft, Z.K. and J.M.; Supervision, J.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable as processed material was anonymous.

Data Availability Statement: The datasets used for training the deep learning models were downloaded from Kaggle: Ma Jun et al., COVID-19 CT Lung and Infection Segmentation Dataset. Version 1.0. Zenodo, Apr. 2020. DOI: 10.5281/zenodo.3757476. URL: <https://doi.org/10.5281/zenodo.3757476> [15] (accessed on 19 March 2023). Maede Maftouni, Large COVID-19 CT scan slice dataset. 2021. DOI: 10.34740/KAGGLE/DSV/2321803. URL: <https://www.kaggle.com/dsv/2321803> [16] (accessed on 19 March 2023). Jinyu Zhao et al., COVID-19 Lung CT Scans. 2020. DOI: 10.34740/KAGGLE/DS/584020. URL: <https://www.kaggle.com/ds/584020> [17] (accessed on 19 March 2023).

Acknowledgments: The scope of the paper included researching the official radiological databases and cooperating with the Międzyzleski Specialist Hospital in Warsaw, Poland.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Taleghani, N.; Taghipour, F. Diagnosis of COVID-19 for controlling the pandemic: A review of the state-of-the-art. *Biosens. Bioelectron.* **2020**, *174*, 112830. [CrossRef] [PubMed]
2. Afshar, P.; Rafiee, M.J.; Naderkhani, F.; Heidarian, S.; Enshaei, N.; Oikonomou, A.; Fard, F.B.; Anconina, R.; Farahani, K.; Plataniotis, K.N.; et al. Human-level COVID-19 diagnosis from low-dose CT scans using a two-stage time-distributed capsule network. *Sci. Rep.* **2022**, *12*, 4827. [CrossRef] [PubMed]
3. Wikipedia Contributors. COVID-19—Wikipedia, The Free Encyclopedia. Available online: <https://en.wikipedia.org/w/index.php?title=COVID-19&oldid=1094307612> (accessed on 19 March 2023).
4. Maia, R.; Carvalho, V.; Faria, B.; Miranda, I.; Catarino, S.; Teixeira, S.; Lima, R.; Minas, G.; Ribeiro, J. Diagnosis Methods for COVID-19: A Systematic Review. *Micromachines* **2022**, *13*, 1349. [CrossRef] [PubMed]
5. Ma, H.; Zhang, Y. Computed tomography of COVID-19 pneumonia. *BMJ* **2020**, *370*, m1807. [CrossRef] [PubMed]
6. Khalifa, M.H.; Samir, A.; Baess, A.I.; Hendawi, S.S. COVID-19-induced vascular angiopathy: CTPA signs in critically ill patients other than acute pulmonary embolism and high-lung opacity scores. *Egypt. J. Radiol. Nucl. Med.* **2021**, *52*, 112. [CrossRef]
7. Alarcón-Rodríguez, J.; Fernández-Velilla, M.; Ureña-Vacas, A.; Martín-Pinacho, J.; Rigual-Bobillo, J.; Jaureguizar-Oriol, A.; Gorospe-Sarasúa, L. Radiological management and follow-up of post-COVID-19 patients. *Radiol. Engl. Ed.* **2021**, *63*, 258–269. [CrossRef]
8. George, P.M.; Barratt, S.L.; Condliffe, R.; Desai, S.R.; Devaraj, A.; Forrest, I.; Gibbons, M.A.; Hart, N.; Jenkins, R.G.; McAuley, D.F.; et al. Respiratory follow-up of patients with COVID-19 pneumonia. *Thorax* **2020**, *75*, 1009–1016. Available online: <https://thorax.bmj.com/content/75/11/1009> (accessed on 19 March 2023). [CrossRef] [PubMed]
9. IBM. What Are Neural Networks? Available online: <https://www.ibm.com/topics/neural-networks#toc-types-of-n-YgdII-Kt> (accessed on 19 March 2023).
10. Aravindpai Pai. ANN vs CNN vs RNN: Types of Neural Networks. 2023. Available online: <https://www.analyticsvidhya.com/blog/2020/02/cnn-vs-rnn-vs-mlp-analyzing-3-types-of-neural-networks-in-deep-learning/> (accessed on 19 March 2023).
11. Yamashita, R.; Nishio, M.; Do, R.K.G.; Togashi, K. Convolutional neural networks: An overview and application in radiology. *Insights Imaging* **2018**, *9*, 611–629. [CrossRef] [PubMed]
12. Wikipedia Contributors. Convolutional Neural Network—Wikipedia, The Free Encyclopedia. Available online: https://en.wikipedia.org/w/index.php?title=Convolutional_neural_network&oldid=1103715044 (accessed on 19 March 2023).
13. MathWorks. What Is a Convolutional Neural Network?: 3 Things you Need to Know. Available online: <https://www.mathworks.com/discovery/convolutional-neural-network-matlab.html> (accessed on 19 March 2023).
14. Sumit Saha. A Comprehensive Guide to Convolutional Neural Networks-the ELI5 Way. 2022. Available online: <https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53> (accessed on 19 March 2023).
15. Ma, J.; Ge, C.; Wang, Y.; An, X.; Gao, J.; Yu, Z.; Zhang, M.; Liu, X.; Deng, X.; Cao, S.; et al. COVID-19 CT Lung and Infection Segmentation Dataset. Version 1.0. Zenodo **2020**. [CrossRef]
16. Maede Maftouni. Large COVID-19 CT Scan Slice Dataset. Available online: <https://www.kaggle.com/dsv/2321803> (accessed on 19 March 2023).
17. Zhao, J.; Zhang, Y.; He, X.; Xie, P. COVID-19 Lung CT Scans. Available online: <https://www.kaggle.com/ds/584020> (accessed on 19 March 2023).
18. Liu, C.; Zhao, R.; Xie, W.; Pang, M. Pathological lung segmentation based on random forest combined with deep model and multi-scale superpixels. *Neural Process. Lett.* **2020**, *52*, 1631–1649. [CrossRef]
19. Kmader. DSB Lung Segmentation Algorithm. 2017. Available online: <https://www.kaggle.com/code/kmader/dsb-lung-segmentation-algorithm/notebook> (accessed on 19 March 2023).

20. Robinet, L. Data Augmentation and Handling Huge Datasets with Keras: A Simple Way. 2020. Available online: <https://towardsdatascience.com/data-augmentation-and-handling-huge-datasets-with-keras-a-simple-way-240481069376> (accessed on 19 March 2023).
21. MLrg-Cefet-Rj. ML-class/ppcicmlmodelselection.ipynb. Available online: https://github.com/MLRG-CEFET-RJ/ml-class/blob/master/ppcic_ml_modelselection.ipynb (accessed on 19 March 2023).
22. Team Keras. Keras Documentation: About Keras. Available online: <https://keras.io/about/> (accessed on 19 March 2023).
23. Aakash Kaushik. Understanding the VGG19 Architecture. 2020. Available online: <https://iq.opengenus.org/vgg19-architecture/> (accessed on 19 March 2023).
24. Elhamraoui, Z. InceptionResNetV2 Simple Introduction. 2020. Available online: <https://medium.com/@zahraelhamraoui1997/inceptionresnetv2-simple-introduction-9a2000edc6b6> (accessed on 19 March 2023).
25. Daniel Godoy. Understanding Binary Cross-Entropy/Log Loss: A Visual Explanation. 2022. Available online: <https://towardsdatascience.com/understanding-binary-cross-entropy-log-loss-a-visual-explanation-a3ac6025181a> (accessed on 19 March 2023).
26. Bushaev, V. Adam-Latest Trends in Deep Learning Optimization. 2018. Available online: <https://towardsdatascience.com/adam-latest-trends-in-deep-learning-optimization-6be9a291375c> (accessed on 19 March 2023).
27. Duong, A. Keras Callbacks Explained in Three Minutes. Available online: <https://www.kdnuggets.com/2019/08/keras-callbacks-explained-three-minutes.html> (accessed on 19 March 2023).
28. National Institutes of Health. Hospitalized Adults: Therapeutic Management. 2022. Available online: <https://www.covid19treatmentguidelines.nih.gov/management/clinical-management-of-adults/hospitalized-adults-therapeutic-management/> (accessed on 19 March 2023).
29. Mikhael, P.G.; Wohlgend, J.; Yala, A.; Karstens, L.; Xiang, J.; Takigami, A.K.; Bourgouin, P.P.; Chan, P.; Mrah, S.; Amayri, W.; et al. Sybil: A Validated Deep Learning Model to Predict Future Lung Cancer Risk From a Single Low-Dose Chest Computed Tomography. *J. Clin. Oncol.* **2023**, *41*, 2191–2200. [CrossRef] [PubMed]
30. Lin, A.; Kolossváry, M.; Motwani, M.; Išgum, I.; Maurovich-Horvat, P.; Slomka, P.J.; Dey, D. Artificial intelligence in cardiovascular CT: Current status and future implications. *J. Cardiovasc. Comput. Tomogr.* **2021**, *15*, 462–469. [CrossRef] [PubMed]
31. McLeavy, C.; Chunara, M.; Gravell, R.; Rauf, A.; Cushnie, A.; Talbot, C.S.; Hawkins, R. The future of CT: Deep learning reconstruction. *Clin. Radiol.* **2021**, *76*, 407–415. [CrossRef] [PubMed]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.