big data and cognitive computing

MDPI

Article

Ensemble of Convolutional Neural Networks for COVID-19 Localization on Chest X-ray Images

Karem D. Marcomini

Institute of Mathematics and Computer Sciences, University of Sao Paulo (ICMC/USP), 400 Trabalhador Saocarlense Avenue, São Carlos 13566-590, SP, Brazil; karem.dm@alumni.usp.br

Abstract: Coronavirus disease (COVID-19) is caused by the SARS-CoV-2 virus and has been declared as a pandemic. The early detection of COVID-19 is necessary to interrupt the spread of the virus and prevent its transmission. X-rays and CT scans can assist radiologists in disease detection. However, detecting COVID-19 on chest radiographs is challenging due to similarities with other bacterial and viral pneumonias. Therefore, it is essential to develop a fast and accurate algorithm for detecting COVID-19. In this work, we applied pre-processing in order to increase the contrast in X-rays. We then use the ResNet-50 model to differentiate between normal and COVID-19 images. Images classified as COVID-19 were investigated with an ensemble detection model (deep learning models—You Only Look Once version 5 and X). The classification model achieved an accuracy of 0.864 and an AUC of 0.904 in 5-fold cross-validation. The overlap between the predicted bounding boxes and the ground truth reached, in the ensemble model, a mAP of 59.63% in 5-fold cross-validation. Thus, we consider that the result was significant in terms of the global classification of the images, as well as in the location of suspicious regions that require greater attention from the specialist, which makes the developed model a fast and promising way to aid the specialist in decision making.

Keywords: COVID-19; abnormality classification; localization; lung opacity detection; deep neural network; chest X-ray

check for updates

Citation: Marcomini, K.D. Ensemble of Convolutional Neural Networks for COVID-19 Localization on Chest X-ray Images. *Big Data Cogn. Comput.* **2024**, *8*, 84. https://doi.org/10.3390/bdcc8080084

Academic Editor: Moulay A. Akhloufi

Received: 11 May 2024
Revised: 25 June 2024
Accepted: 11 July 2024
Published: 1 August 2024

Copyright: © 2024 by the author. 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/).

1. Introduction

The first cases of the novel coronavirus disease (COVID-19) started with reports of pneumonia of unknown causes in Wuhan city, Hubei province of China, at the end of 2019. The infectious disease was officially named Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) by the World Health Organization (WHO) [1,2].

This viral infection can lead to hospitalization in intensive care with acute respiratory syndrome, severe pneumonia, multiple organ failure, and death. The main symptoms include fever (38 °C or higher), dry cough, difficulty breathing, sore throat, fatigue, pain, and diarrhea for some individuals [3]. A sudden loss of smell, no nasal obstruction, and complete loss of taste are also observed symptoms in patients. However, the period between contamination and the onset of COVID-19 symptoms can extend up to 15 days. Therefore, asymptomatic individuals carrying the virus can unknowingly transmit it to others, significantly facilitating its spread [4–6]. The outbreak has impacted over 200 countries, resulting in more than 775 million confirmed cases and a death toll exceeding 7 million as of 27 April 2024.

The early detection of COVID-19 is crucial for interrupting the virus spread and preventing its transmission through early patient isolation [7]. Several techniques have been employed to detect and diagnose COVID-19, including reverse-transcription polymerase chain reaction (RT-PCR), chest computed tomography (CT) scans, and chest X-rays (CXRs). RT-PCR detects SARS-CoV-2 nucleic acids present in nasopharyngeal fluids, and it is being conducted for the early detection of the disease, including asymptomatic infected people [8]. Wikramaratna [9] reported that the sensitivity of RT-PCR on nasopharyngeal specimens

ranges from 96.40% (95% CI: 90.98 to 98.6) near symptom onset to 75.47% (95% CI: 66.88 to 82.51) by day 10 since symptom onset. The RT-PCR test typically requires a few hours. However, the overall time from sample collection to obtaining results can span several days when samples are processed in bulk. Despite its widespread use, RT-PCR can yield both false negative and false positive results [8].

Chest imaging is essential for the early diagnosis and treatment of COVID-19 patients. CT scans offer more precise imaging and have a higher sensitivity than CXR [10]. However, the use CT for COVID-19 detection places a significant burden on radiology units and poses challenges for infection control in CT suites. In contrast, CXR is still the primary imaging test for pneumonia due to its short radiation exposure time and low cost. Almost all clinics, urgent care facilities, emergency departments, and hospitals are equipped with stationary and mobile radiography units. These units can minimize the risk of cross-infection, since they can be easily protected from exposure or disinfected after use and can be used directly in a contained clinical environment without moving patients [3,11].

Medical image interpretation is typically carried out by radiologists and physicians. Since medical data vary significantly between patients depending on the disease, interpretation relies heavily on clinicians. However, this process is constrained by experience, interobserver variability, speed, and fatigue. With the limited capacity of the healthcare system and the work overload of radiologists, one way to aid diagnosis is to provide medical image analysis through an automated, precise, and efficient computational system [12].

Machine learning, deep learning, and artificial intelligence-based approaches have been used for the detection and classification of several diseases. These systems can discover patterns that normally would not be recognized by radiologists [13,14], which can augment the capabilities of radiologists in better decision making and a more effective management of the situation [15,16].

Hence, the main aim of this paper is in applying and evaluating pre-trained deep learning architectures as an automated tool to classify CXR images as normal/negative for pneumonia or COVID-19. In the case of a positive result, the method identifies the affected area.

The major contributions of this paper are summarized as follows:

- This study evaluates the performance of two state-of-the-art object detection models, YoloX and Yolov5, both independently and in combination. The focus is on accurately identifying lung areas that require specialist attention, thereby enhancing diagnostic efficiency.
- The research aims to automate the screening process for patients, facilitating the identification of cases that may require urgent medical intervention. By leveraging advanced AI models, the system aids in early detection and intervention, potentially improving patient outcomes.
- By employing AI-powered tools, the study seeks to streamline the interpretation of patient X-rays. This efficiency improvement aims to reduce the time clinicians spend on reading X-rays, thereby enabling the screening of more patients within a limited timeframe. This approach not only enhances healthcare delivery but also optimizes resource utilization in healthcare settings.

To reduce the stress on physicians, we propose a simple but efficient pipeline to process a large number of X-rays, to look for clues indicating signs of pneumonia. It allows the physician to process a higher number of images in a shorter period. Our model not only classifies the input images but can also determine the possible location of the infection, so that it can be treated with a proper procedure.

The paper is organized as follows. Section 2 discusses the previous state-of-the-art models designed for the classification and localization of COVID-19 from image exams. In Section 3, we present the characteristics of the dataset. Section 4 describes digital processing, and in Section 5, we present the assessment metrics. In Section 6, we present the results obtained in our study and discuss them based on the data presented in the literature. Finally, in Section 7, we offer our conclusions.

2. Related Work

Starting in 2020, after the discovery of COVID-19, several artificial intelligence (AI) systems based on machine learning (ML) and deep learning (DL) have been employed for differentiating COVID-19 from non-COVID-19 or normal on digital chest X-ray and CT images. This interest was motivated by the rapid spread of the disease, shortages of essential equipment, similar symptoms to other diseases such as pneumonia, and the lack of qualified health professionals. Thus, research on ML and DL methodologies has grown immensely in this field. Furthermore, the assessment of images by a human expert is expensive, tedious, time-consuming, often impractical on a large scale, and prone to inter- and intra-observer variability. Therefore, an automatic image analysis tool is essential to assist the specialist in diagnosing the disease.

Several studies have been conducted using traditional machine learning (ML) algorithms, where input features are required to be fed as input to the algorithm. The hand-crafted features are based on the texture, geometry, and morphological characteristics of the lung [17]. The traditional classifiers are Bayesian networks [17], fuzzy c-means [18], support vector machines (SVMs) [19,20], artificial neural networks (ANNs) [21], k-nearest neighbors (kNNs) [17], logistic regression [21,22], and decision trees [17,21–23]. Other works developed models to identify radiomic biomarkers for COVID-19 [24,25] and prognostic models for survival prediction [26–28].

Machine learning models have advantages when the dataset is limited and there is a low computational cost for data training [19]. However, the major limitation faced by ML-based model is the dependence on the feature extraction step. The lack of an efficient feature extraction makes it challenging to find the most relevant feature set, which is needed to obtain the most effective result. The feature extraction issue can be solved by using sophisticated deep learning techniques [15,17].

Convolutional neural networks (CNNs) have automatic feature-learning capabilities and have shown promising results in diverse fields. Advancements in technology, particularly in handling large datasets and high-speed Graphical Processing Units (GPUs), have contributed to improving their performance.

Several studies have evaluated the use of pre-trained networks on the ImageNet database; the most common networks are DenseNet [29,30], MobileNet [31], ResNet [32], Inception [33], Xception [34], VGG [35], and EfficientNet [35,36].

However, ML and DL models only provide global finding labels for classification for all the images. There are relatively few papers in the CXR literature that specifically address localization methods, despite their importance in medical imaging and their potential feasibility compared to precise segmentation. This scarcity is likely due to the challenges associated with accurately annotating a projection image and the high cost of detailed segmentations [37].

Localization refers to identifying a specific region within the image, usually denoted by a bounding box or a point location. This approach is valuable for identifying anatomical regions, abnormalities, or foreign objects within medical imaging. A straightforward system employing bounding boxes to highlight abnormal regions can effectively guide radiologists' attention, potentially enhancing their sensitivity to subtle findings or complex areas with overlapping structures [37].

The main studies [38–45] in this field used the dataset provided by the Radiological Society of North America (RSNA), which consists of 8964 pneumonia-labeled CXR images, and every pneumonia area of each image is marked with a bounding box. In addition, the remaining 20,025 non-pneumonia CXR images are available.

Nayyar [38] evaluated YoloV3 and Mask R-CNN for pneumonia detection and localization, achieving better results with Mask R-CNN, with a mean average precision (mAP) metric of 0.155. On the other hand, rather than using a single detector, some works develop an ensemble of two or more CNNs in order to improve the results. Sirazitdinov [39], for example, evaluated the RetinaNet and Mask R-CNN models separately and obtained a mAP of 0.202 and 0.165, respectively. The combined model produced a mAP of 0.204. Ko [40]

combined two Mask R-CNN models with different types of residual network (ResNet) as a backbone (ResNet150 and ResNet162), and three RetinaNet (ResNet178, ResNet184 and ResNet201) models handled their experiment. They used different weights in each model, and the mAP was 0.217 in the majority voting method. However, Jaiswal [41] obtained practically the same result with the combination of only two networks. This study evaluated Mask R-CNN using ResNet50 and ResNet101, and achieved a mAP of 0.184 and 0.199, respectively. After combining the models, the mAP increased to 0.218. By merging RetinaNet (mAP of 0.2196) with Mask R-CNN (mAP of 0.2102). Mao [42] improved the mAP to 0.2283. Li [43] proposed an enhanced squeeze-and-excitation network (SENet) architecture, noted for its exceptional ability to suppress useless features. This method achieved a mAP of 0.262. Yao [44] presented a method based on DeepConv-DilatedNetfor identifying and localizing pneumonia CXR images. This study adopted a Faster R-CNN network, leveraging dilated convolution and an attention mechanism to enhance detection accuracy. Additionally, a double K-means approach was employed to generate anchor boxes, further improving localization precision. The algorithm obtained a mAP of 0.382. Yao [45] also applied this modification on YoloV3 and achieved a mAP of 0.468.

The study proposed by Frid-Adar [46] estimated the severity of pneumonia in COVID-19 patients and reported the findings of a longitudinal study of disease progression. It presented a DL model, trained on the RSNA Pneumonia Detection Challenge Dataset, for the simultaneous detection and localization of pneumonia in CXR images, which has been demonstrated to generalize well to COVID-19 pneumonia. This achieved a mAP of 0.27 and reported 0.86, 0.91, and 0.89 for sensitivity, specificity, and accuracy, respectively.

Al-antari [47] proposed a system based on the Yolo predictor to detect and diagnose COVID-19, differentiating it from eight other respiratory diseases. This study used CXR images from patients with COVID-19 collected by Chowdhury [29] and Cohen [48]. The information regarding the ground truth localization (i.e., bounding box) is not publicly available. The CAD predictor obtained an overall detection and classification accuracy of 96.31% for COVID-19, achieving an average intersection over union (IoU) greater than 90%.

Duraisamy [49] and Yang [50] used the YoloV5 network to detect lesion areas of COVID-19 in the dataset proposed by Lakhani [51]. Duraisamy [49] obtained a mAP of 0.446, calculated at an IoU threshold of 0.5. This evaluation was conducted on a dataset partitioned into a 60:30:10 ratio for training, validation, and testing, respectively. The model underwent training for 25 epochs, utilizing a batch size of 16 and a learning rate set at 0.0001, alongside the application of thermal image pre-processing functions. Yang [50] proposed a Focal Loss function to improve the category imbalance problem in the data. This approach enhances the network's focus on learning complex cases, effectively improving model accuracy and achieving a mAP of 62.23%.

3. Dataset Characteristics

This paper conducted its experiments based on two public datasets that contain chest X-ray images, which are the BIMCV-COVID19 Data [52] and MIDRC-RICORD Data [53]. SIIM-FISABIO-RSNA COVID-19 Detection [51] combined both datasets into a single one (https://www.kaggle.com/c/siim-covid19-detection/data, assessed on 2 May 2024).

The dataset comprises 6334 chest scans in the DICOM format, with the original high-resolution scale. We used only CXR images of anteroposterior (AP) and posteroanterior (PA) projections of adult patients ($\geq$18 years old), obtained from both computed radiography (CR) and digital radiography (DX) devices. In order to protect patient privacy, the images were de-identified.

A panel of experienced radiologists labeled, in four classes, all the images for the presence of opacities, as well as overall appearance. To perform this task, 22 annotators were trained via an online web platform, and they received all annotation instructions, reference materials, and multiple example cases for each category. They were required to score the radiographs at an exam level into one of the following four categories: (a) typical appearance of COVID-19 (bilateral, peripherical, multifocal predominant opacities; diffuse

bilateral opacities including central and peripherical; and diffuse bilateral opacities with fibrosis/reduced lung volumes); (b) indeterminate appearance of COVID-19 (upper lung zone predominant opacities; unilateral opacities; central opacities with relative peripherical sparing); (c) atypical appearance of COVID-19 (pneumothorax without features of pneumonia; pleural effusion without features of pneumonia; mass(es) or nodule(s); lobar pneumonia; scarring/fibrosis); and (d) negative for pneumonia (no lung opacities) [51,54].

The main objective of our work presented in this paper is to determine whether a person has been infected by SARS-CoV-2 or not. Therefore, we only considered COVID-19-suspicious cases (typical and indeterminate appearance) and normal appearance, removing the images related to other types of pneumonia (labeled as atypical appearance). Thus, we obtained a total of 5639 images, of which 2854 are typical, 1049 indeterminate, and 1736 negative for pneumonia.

Annotators placed bounding boxes on pulmonary airspace opacities, whether the exam was scored as typical or an indeterminate pattern for COVID-19. However, bounding boxes were not placed on pleural effusions, masses/nodules, or pneumothoraxes, nor on the negative for pneumonia category.

4. The Proposed Detector Model

Often, the solutions in machine learning competitions are based on large and diverse ensembles in order to achieve more accurate results. In this regard, we evaluated the accuracy in locating the area of opacity in two ways: (1) In Experiment 1, we used a network to locate the area of infection, if any. (2) In Experiment 2, we first applied a network to differentiate between normal and COVID-19 images, and, if classified as COVID-19, then it locates the area of infection. We also investigated whether two location models, when combined, can increase diagnostic accuracy. Figure 1 illustrates the flowchart of our study.

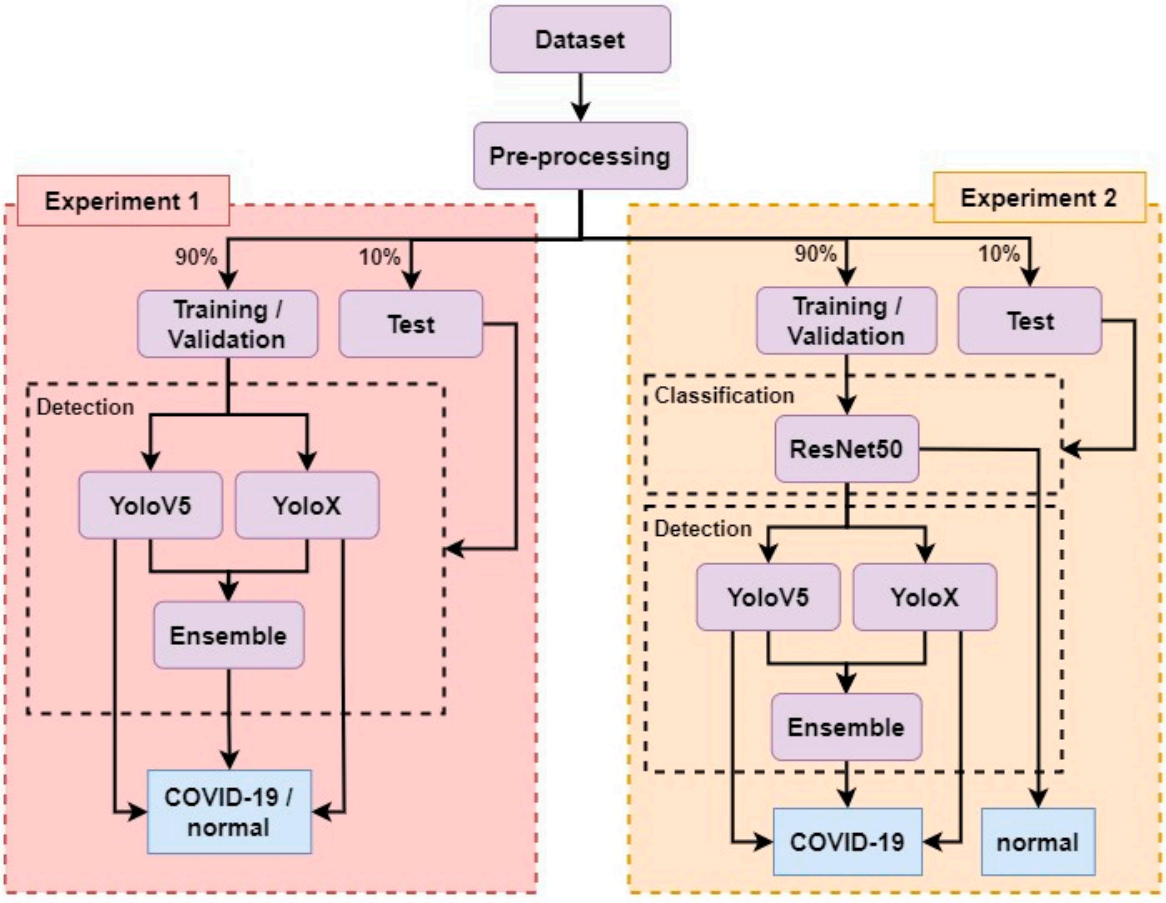

Figure 1. Flowchart of study developed.

4.1. Pre-processing

We applied a pre-processing step, which included different techniques, to the original images to reduce the large variability of these images.

Firstly, we converted all images to 8-bit resolution and resampled them to 640 × 640 pixels. Then, we calculated the normalized histogram of each image. To normalize, we divided the frequency of each bin by the total number of pixels in the image. As a result, the value 1 represents the maximum value of the cumulative histogram.

Next, we derived a lookup table which maps the pixel intensities to achieve an equalized histogram characteristic. For each discrete intensity level i, we calculated the mapped pixel value from the normalized cumulative histogram according to,

$$\text{MappedPixelValue(i)} = (\text{L} - 1) \times \text{NormalizedCumulativeHistogram(i)} \quad (1)$$

where L = 256 for a typical 8-bit unsigned integer representation of pixel intensity.

Once the lookup table was derived, we mapped the intensity of all pixels in the image to the new values. The result was an equalized image [55].

Figure 2 shows some examples of images after the abovementioned processing.

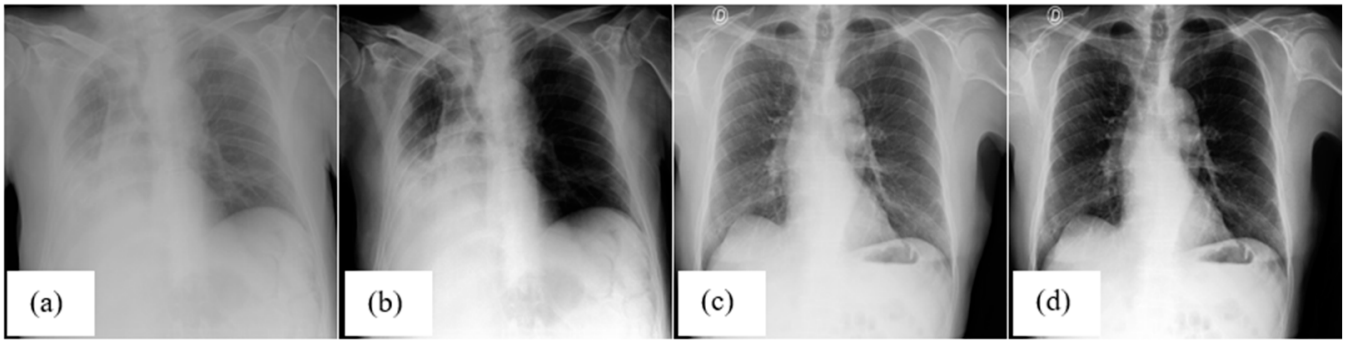

Figure 2. From the SIIM-FISABIO-RSNA COVID-19 Detection dataset [51], (**a**,**c**) represent the original images, (**b**,**d**) the images after pre-processing.

4.2. Classification

ResNet [56] is one of the well-known models that perform exceptionally well in various computer vision tasks. This model is trained on a dataset containing a wide range of image categories. We utilized the pre-trained ResNet-50 model, which was originally trained on the ImageNet dataset, to distinguish between classes (normal vs. COVID-19). Its architecture offers the advantage of being significantly deeper, while mitigating the issue of vanishing gradients. A vanishing gradient occurs when the gradient diminishes to a very small value as it propagates backward through the network layers, potentially reducing the model's accuracy and reaching a saturation point. ResNet-50 comprises 50 layers. To tailor it for our specific task, we replaced the original layers (Global Average Pooling, Flatten, and Dense) with a sequence that includes Global Average Pooling, Batch Normalization, a Dense layer (Fully Connected layer), and an activation function. We applied the Sigmoid function for the final classification. Figure 3 illustrates the modified architecture. The "+" symbol right after batch normalization in the shortcut (or skip connection) represents the addition operation. This means that the input tensor is added to the output tensor of the convolution and normalization process.

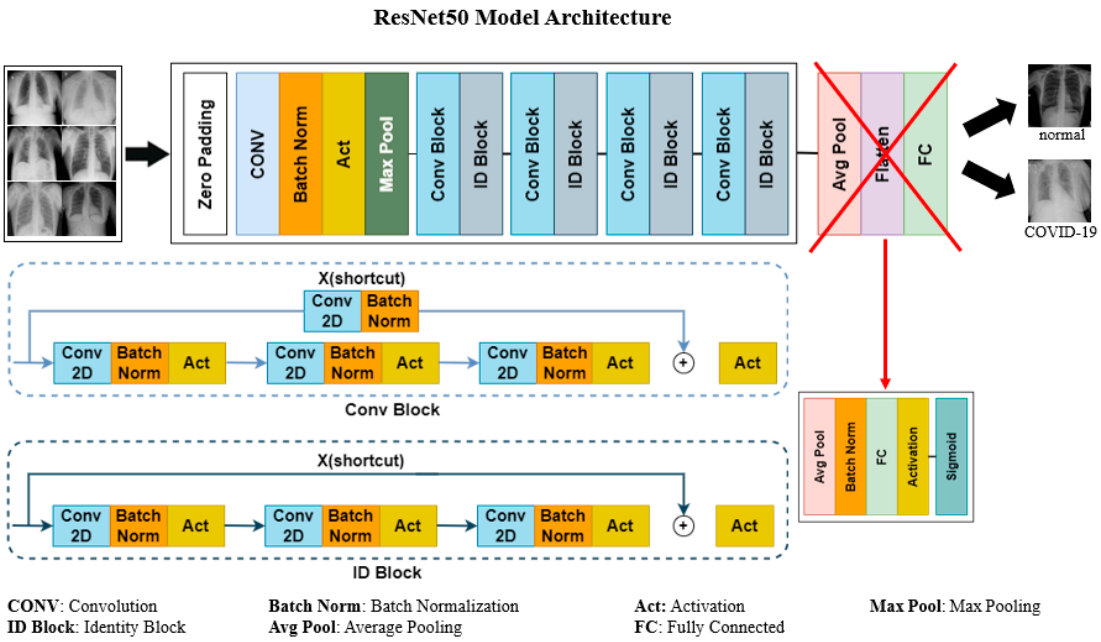

Figure 3. ResNet-50 architecture for CXR image classification. The feature map size is downsampled by a convolutional layer, which is followed by Batch normalization and Activation function layer. Each stage has Conv block and ID block. The identity block contains three sets of a convolutional layer, and the Convolutional block has one extra layer to match the input and output dimension.

For the classification step, we split the dataset into training and testing sets with a ratio of 90:10. We applied a 5-fold cross-validation on the training set, where the dataset is divided into five equally sized folds. The model is trained and evaluated five times. In each iteration, one fold is used as the validation set (20%), while the remaining four folds (80%) serve as the training set. We trained the network for 30 epochs, with a batch size of 8 images per step, employing the Adam optimizer with a learning rate of 0.0001.

To enhance learning performance, we applied data augmentation specifically to instances categorized as "normal" within the training dataset to achieve a better data balance. Data augmentation involves applying various distortions to the original dataset, creating multiple altered copies. The distortions randomly applied to each instance in our dataset include scaling, rotating, sharpening, shearing, and Gaussian blur.

4.3. Detection

The Yolo (You Only Look Once) series represents a significant advancement in the field of object recognition and detection. Its architecture generates an output vector for each detected object, encapsulating crucial information such as object class and confidence score, as well as the coordinates for the object's center (x, y), along with its width and height. These coordinates enable the generation of bounding boxes around detected objects. Yolo pursues an optimal speed and accuracy trade-off for real-time applications.

YoloV5 was originally published in May 2020 by Glenn Jocher [56] on GitHub (https://github.com/ultralytics/yolov5/releases, assessed on 5 May 2024). The YoloV5 model has shown a substantial performance increase from its predecessors, with the best trade-off performance on Common Objects in Context (COCO, https://cocodataset.org, assessed on 5 May 2024) [57].

YoloV5 is structured into three distinct parts: the backbone, neck, and detection head. The backbone employs the Cross-Stage Partial (CSP) networks technique to extract features from input images. The CSP module divides features into two parts and integrates them through a cross-stage hierarchy, enhancing accuracy while reducing the computational load. The neck structure utilizes a path aggregation network to generate a feature pyramid. This pyramid enables the model to detect objects at multiple scales, facilitating robust object

identification. Additionally, the neck module incorporates a PAN (Path Aggregation Network) structure based on the FPN (Feature Pyramid Network), facilitating the generation of the feature pyramid from bottom to top. The head of YoloV5 is responsible for predictions, including calculating the confidence scores and bounding boxes for detected objects. [56].

YoloV5 comes in various versions, each having its unique characteristic, these versions being -s, -m, -l, and -x, which differ in their number of layers and parameters (from a small to extra-large version).

While YoloV5 is recognized as a top performer, it remains an anchor-based detector. Using this detector necessitates optimizing a set of anchors prior to training, which adds complexity to predictions for each image and can slow down inference speeds.

In 2021, YoloX [58] was awarded First Place in the Streaming Perception Challenge at the Workshop on Autonomous Driving (WAD) in conjunction with the Conference on Computer Vision and Pattern Recognition (CVPR). YoloX is an anchor-free object detector using the Yolov3 [59] baseline with DarkNet53. The decoupled head uses separate convolution branches to predict each score for object detection, and it proves that the decoupled head has advantages in both the efficiency of model training and the accuracy of the predictions.

Ensembles of networks typically perform better than individual models [37], as they merge predictions from different models into a final prediction, which allows for an increasing overall accuracy.

Unlike non-maximum suppression (NMS) or Soft-NMS methods that simply remove part of the predictions, the Weighted Boxes Fusion (WBF) [60] method uses the confidence scores of all the proposed bounding boxes to construct the average boxes. This method will not discard any bounding box; instead, it combines them. This leads to an improvement in the quality of the combined predictions.

WBF updates a fused box at each step and uses it to check for overlap with subsequent predicted boxes. Additionally, WBF incorporates information on how many models predict a specific box within a cluster, enabling it to produce optimal results for model ensembles [60].

We also evaluated the ensemble model, whose YoloV5 and YoloX predictions were merged using the WBF method.

We used the base network YoloV5 and YoloX with an input resolution of 640 $\times$ 640, which was pre-trained on Common Objects in Context (COCO). The dataset contains 3903 images with 7316 labeled as COVID-19. We used the same split to the binary classification, with a proportional rate of 90:10 (3903:397 images and 7316:749 labels). We trained for 50 epochs, with a batch size of 8.

5. Evaluation

We used four parameters, true positive (TP), true negative (TN), false positive (FP) and false negative (FN), to evaluate the overall performance of the proposed classification model based on the performance metrics: accuracy, precision, recall, specificity, F1-score, and area under the receiver operating characteristics (AUROC).

Accuracy is the measure of all the correctly recognized cases. It is defined as the ratio between the number of correct predictions and the total number of predictions. This is commonly used, especially when both the classes are equally important.

$$\text{Accuracy} = (\text{TP} + \text{TN})/(\text{TP} + \text{TN} + \text{FP} + \text{FN}) \quad (2)$$

Precision is used to evaluate the exactness of the classifier. A low precision value indicates that the classifier suffers from a large number of FPs. The mathematical formulation is given by the ratio between the number of correct positive predictions to the total number of positive cases.

$$\text{Precision} = \text{TP}/(\text{TP} + \text{FP}) \quad (3)$$

Recall, also called sensitivity, is defined as a measure of a classifier's completeness, where a low recall value indicates that the classifier suffers from a large number of FNs.

It is defined as the ratio between the number of correct positive predictions to the total number of correct results.

$$\text{Recall} = \text{TP}/(\text{TP} + \text{FN}) \tag{4}$$

Specificity defines the proportion of negative cases that are correctly classified as such.

$$\text{Specificity} = \text{TN}/(\text{TN} + \text{FP}) \tag{5}$$

The F1-score considers how the data is distributed, testing the accuracy of the model. It is useful when the dataset contains imbalanced classes. It is introduced as the harmonic mean of precision and recall.

$$\text{F1} = (2 \times \text{Pr} \times \text{Rc})/(\text{Pr} + \text{Rc}) \tag{6}$$

The receiver operating characteristics (ROC) curve plots the true positive rate (sensitivity) vs. the false positive rate (1-specificity) across a wide range of values. The area under this curve (AUROC) represents the predictability of a binary classifier. The higher the value of the area under the curve, the better the model under analysis performs.

A predicted object is considered a TP if its intersection over union with a ground truth object is greater than 0.4; otherwise, it is an FP. False negative indicates that there is no predicted object associated with the ground truth object.

At each threshold value t, the average precision (AP) is measured based on all objects that are predicted in a single image. The AP evaluates the ability of the detector to find all relevant objects and the ability of the detector to detect objects correctly.

$$\text{AP(t)} = \text{TP(t)}/(\text{TP(t)} + \text{FP(t)} + \text{FN(t)}) \tag{7}$$

In order to evaluate the accuracy of the entire dataset, we used the mean average precision (mAP) metric. It is the most common performance metric in the field of object detection. mAP metrics do not count true negative (TN) image predictions, because a non-COVID-19 sample with zero predictions does not participate in averaging over all the images.

$$\text{mAP} = \frac{1}{|\text{images}|}\sum_{\text{img}} \frac{1}{|\text{thresholds}|}\sum_{\text{t}} \frac{\text{TP(t)}}{|\text{TP(t)} + \text{FP(t)} + \text{FN(t)}|} \tag{8}$$

6. Results and Discussion

This section presents the results achieved and detailed discussions regarding each experiment described earlier.

Table 1 presents the way in which the data are arranged in each set in each experiment.

Table 1. Distribution of data in training and test sets in each experiment.

		Images				**Labels**	
		Train		**Test**		**Train**	**Test**
		COVID-19	**Normal**	**COVID-19**	**Normal**	**COVID-19**	
Experiment 1		3514	1561	389	175	6576	740
Experiment 2	Classifier	3514	1561	389	175	-	-
	Locator	3514	-	389	-	6576	740

6.1. Experiment 1

We trained the YoloV5 and YoloX networks with images considered normal and with COVID-19. In this way, the network becomes more robust in generalizing the data, being able to recognize the normal cases in the test set.

Table 2 shows the mAP of the YoloV5, YoloX, and ensemble models in the test set.

Table 2. Evaluation of detectors in identifying areas with COVID-19.

	YoloX	YoloV5	Ensemble
Fold 1	49.88	53.34	53.51
Fold 2	52.34	52.17	53.66
Fold 3	52.01	52.39	55.01
Fold 4	51.85	55.91	55.00
Fold 5	50.43	51.68	53.52
Mean	51.30	53.1	54.16
Standard deviation	0.97	1.51	0.73

6.2. Experiment 2

We divided this experiment into two stages. In the first one, we performed a binary classification for normal and COVID-19 cases.

Table 3 presents the performance of three classification models with a 5-fold cross-validation, enabling the evaluating of generalization and comparison between different architectures. We assessed each model's performance using the confusion matrix, consisting of true positives (TPs), false negatives (FNs), true negatives (TNs), and false positives (FPs). Additionally, we computed the accuracy, precision, recall, specificity, F1-score, and AUROC to provide comprehensive performance metrics.

Table 3. Performance of multiple classifiers with 5-fold cross-validation in differentiating between normal- and COVID-19-appearance images.

Model	Fold	Confusion Matrix and Results									
		TP	FN	TN	FP	Accuracy	Precision	Recall	Specificity	F1	AUROC
InceptionV3	1	363	26	122	53	0.860	0.873	0.933	0.697	0.902	0.911
	2	335	54	142	33	0.846	0.910	0.861	0.811	0.885	0.924
	3	354	35	122	53	0.844	0.870	0.910	0.697	0.889	0.920
	4	312	77	153	22	0.824	0.934	0.802	0.874	0.863	0.917
	5	341	48	131	44	0.837	0.886	0.877	0.749	0.881	0.923
Mean		341	48	134	41	0.84	0.89	0.877	0.766	0.88	0.919
Standard deviation		17	17	12	12	0.01	0.02	0.045	0.067	0.01	0.005
ResNet50	1	354	35	123	52	0.846	0.872	0.910	0.703	0.891	0.910
	2	335	54	140	35	0.842	0.905	0.861	0.800	0.883	0.916
	3	362	27	129	46	0.871	0.887	0.931	0.737	0.908	0.925
	4	329	60	147	28	0.844	0.922	0.846	0.840	0.882	0.919
	5	340	49	138	37	0.848	0.902	0.874	0.789	0.888	0.927
Mean		344	45	135	40	0.85	0.90	0.884	0.774	0.89	0.919
Standard deviation		12	12	8.5	8	0.01	0.02	0.031	0.048	0.01	0.006
EfficientNetB4	1	323	66	143	32	0.826	0.910	0.830	0.817	0.868	0.917
	2	289	100	153	22	0.784	0.929	0.743	0.874	0.826	0.888
	3	327	62	145	30	0.837	0.916	0.841	0.829	0.877	0.904
	4	357	32	110	65	0.828	0.846	0.918	0.629	0.880	0.905
	5	326	63	147	28	0.839	0.921	0.838	0.840	0.878	0.893
Mean		324	65	140	35	0.82	0.90	0.834	0.798	0.87	0.902
Standard deviation		22	22	15	15	0.02	0.03	0.055	0.087	0.02	0.010

In the study proposed by Dhont [61], they applied five neural networks (VGG19, ResNet50, InceptionV3, DenseNet201, and COVID-Net) in order to classify normal and COVID-19 images. The authors used four datasets; BIMCV was one of these sets. The performance of these networks for this particular dataset had a precision of 0.65–0.76 and a recall of 0.59–0.71. Notice that our model's precision and recall values are always above those obtained in the models evaluated by Dhont [61].

In order to investigate radiomic features in the classification of COVID-19, Moura [23] used the PyRadiomics library to extract the first- and second-order statistical texture-based features for each lung mask. They also used ShapRFECV from the Probatus python library to perform the feature selection. Then, they applied XGBoost and Random Forest models to classify the lungs. The XGBoost classifier provided an accuracy, F1-score, and precision of 0.82. Random Forest presented an accuracy, F1-score, and precision of 0.77, 0.78 and 0.75, respectively.

In this work, the ResNet50 network provided better results for the measures evaluated, with an accuracy of 0.85 and precision of 0.9 for the 5-fold evaluation. Thus, the values obtained for this dataset were higher than those found in the literature.

It is important to highlight that many public datasets provide lower-quality images than those used by radiologists in clinical practice. This can be a cause for the decreased performance of deep learning systems, particularly for more subtle abnormalities. A decreasing in quality is usually related reducing image size or bit depth before making the images publicly available. This practice is commonly employed to reduce the overall download size of datasets. In clinical workflows, DICOM files are the industry standard for storing CXRs, typically utilizing 12 bits per pixel, with image dimensions ranging from approximately 2 to 4 thousand pixels on each axis (X and Y). However, there are public datasets sourced from screenshots of online literature, which may lead to an unquantifiable degradation in data quality. Additionally, it is important to note that if these data undergo pre-processing before being made available, it is desirable to provide a description of all the steps to enable reproducibility in other datasets [37].

There is another important factor to consider when we compared our study to the results found in the literature, which are studies that merge a dataset with healthy pediatric patients or with other pneumonias with adult patients diagnosed with COVID-19. These studies also aim to classify COVID-19, but there is a bias due to age-related characteristics [37]. In our study, we only used X-rays from adult patients.

After completing the differentiation between normal and COVID-19 cases, we trained the YoloV5 and YoloX networks only on COVID-19 images, removing the normal cases. In Table 4, we present the mAP values for each model.

Table 4. Evaluation of opacity detectors only on positive images for COVID-19.

	YoloX	**YoloV5**	**Ensemble**
Fold 1	59.92	57.61	61.71
Fold 2	60.47	57.61	59.71
Fold 3	62.3	56.5	61.71
Fold 4	61.53	57.98	63.05
Fold 5	60.29	61.05	64.73
Mean	60.90	58.15	62.18
Standard deviation	0.88	1.53	1.66

In order to include normal cases in the evaluation system, we chose the ResNet50 network to differentiate between normal and COVID-19 images. We selected the 'fold' with the best AUROC value and the least amount of FNs, since there were not many FP cases either (Table 3, fold 3). Data predicted as COVID-19 by the ResNet50 network in the test set were fed into the locator (weights resulting from training performed only with

COVID-19 positive images—Table 4). Table 5 presents the mAP of the location networks after removing the normal cases by the ResNet50 classifier.

Table 5. Evaluation of opacity detectors after differentiating between normal and COVID-19 cases by ResNet50 network.

	YoloX	YoloV5	Ensemble
Fold 1	57.39	55.87	58.73
Fold 2	57.59	55.87	58.77
Fold 3	59.26	54.24	59.28
Fold 4	59.33	55.66	60.29
Fold 5	57.61	58.31	61.1
Mean	58.24	55.99	59.63
Standard deviation	0.87	1.31	0.92

In the location model training stage (Table 4), YoloX presented a higher mAP for fold 3. However, when we performed the classification and location, there was a slight increase in fold 4 in relation to fold 3.

When we compare the mAP obtained by the locator acting in the locate and classify functions (Table 2) with the locator applied after the classifier (Table 5), we have a considerable increase in all 5-folds. In the ensemble, for example, the mAP was greater than 60% (fold 5) in the overlap rate between the predicted region and the GT.

We evaluated the present study and the results described in the literature. We noted the innovative applicability in chest X-ray images for COVID-19 patients. Furthermore, among the studies related to the location of pneumonia regions [5,38–45], the authors evaluated networks such as YoloV3, Mask R-CNN, and RetinaNet and ensemble models. In these works, the mAP ranged from 15.5 to 21.96% for individual models and from 20.4 to 46.8% for ensemble models. Ko [40] combined five architectures (two Mask R-CNNs with different types of backbone and three RetinaNets), but there was no significant increase in mAP compared to other works that combined only two architectures.

6.3. Final Results

In order to measure the performance of the image classification model based on its global appearance, we investigated the models developed in Experiments 1 and 2 in terms of the confusion matrix and the metrics already discussed.

In Experiment 1, we considered positive for COVID-19 every image in which the localizer determined at least one region as lung opacity. In cases where it did not provide any bounding box, we considered the image negative for COVID-19. We present the results of this experiment in Table 6.

When we used the model discussed in Experiment 1, we obtained a more expressive result in terms of accuracy by the YoloV5 network with 4-fold weights, whose value was 0.853. However, it resulted in 42 FP cases. If we consider AUROC, 3-fold was more expressive (0.903). In the ensemble model (combination of the Yolos) the measures are practically equal to the 5-fold.

In Experiment 2, we used ResNet50 to differentiate between normal and COVID-19 images. Due to the fact that the classifier distinguishes between both classes, we trained the location networks only with positive images for COVID-19; therefore, there was no need for the locator to have a prior knowledge of normal cases. The results are shown in Table 7.

Table 6. Performance of location models in differentiating between COVID-19-positive and normal cases.

Model	Fold	Confusion Matrix and Results									
		TP	FN	TN	FP	Accuracy	Precision	Recall	Specificity	F1	AUROC
YoloX	1	298	91	161	14	0.814	0.955	0.766	0.92	0.850	0.864
	2	295	94	161	14	0.809	0.955	0.758	0.92	0.845	0.863
	3	305	84	157	18	0.819	0.944	0.784	0.897	0.857	0.872
	4	305	84	159	16	0.823	0.950	0.784	0.909	0.859	0.873
	5	302	87	157	18	0.814	0.944	0.776	0.897	0.852	0.867
Mean		301	88	159	16	0.816	0.950	0.774	0.909	0.853	0.868
Standard deviation		3.9	3.9	1.8	1.8	0.005	0.005	0.010	0.010	0.005	0.004
YoloV5	1	342	47	130	45	0.837	0.884	0.879	0.743	0.881	0.892
	2	335	54	138	37	0.839	0.901	0.861	0.789	0.880	0.880
	3	354	35	124	51	0.848	0.874	0.910	0.709	0.892	0.903
	4	347	42	134	41	0.853	0.894	0.892	0.766	0.893	0.899
	5	329	60	145	30	0.840	0.916	0.846	0.829	0.880	0.887
Mean		341	48	134	41	0.843	0.894	0.878	0.767	0.885	0.892
Standard deviation		8.8	8.8	7.1	7.1	0.006	0.014	0.023	0.040	0.006	0.008
Ensemble	1	303	86	160	15	0.821	0.953	0.779	0.914	0.857	0.871
	2	300	89	158	17	0.812	0.946	0.771	0.903	0.850	0.867
	3	307	82	153	22	0.816	0.933	0.789	0.874	0.855	0.870
	4	308	81	155	20	0.821	0.939	0.792	0.886	0.859	0.874
	5	311	78	152	23	0.821	0.931	0.799	0.869	0.860	0.873
Mean		305	83	156	19	0.818	0.940	0.786	0.889	0.856	0.871
Standard deviation		3.8	3.9	3.0	3.0	0.004	0.008	0.010	0.017	0.004	0.002

Table 7. Performance of opacity locators preceded by ResNet50 classifier to predict COVID-19 and normal cases.

Model	Fold	Confusion Matrix and Results									
		TP	FN	TN	FP	Accuracy	Precision	Recall	Specificity	F1	AUROC
ResNet50 + YoloX	1	298	51	147	28	0.860	0.923	0.869	0.84	0.895	0.900
	2	295	61	148	27	0.844	0.924	0.843	0.846	0.882	0.888
	3	305	46	143	32	0.861	0.915	0.882	0.817	0.898	0.901
	4	305	54	144	31	0.849	0.915	0.861	0.823	0.887	0.897
	5	302	53	148	27	0.858	0.926	0.864	0.846	0.894	0.896
Mean		301	53	146	29	0.855	0.921	0.864	0.834	0.891	0.897
Standard deviation		3.9	4.9	2.1	2.1	0.007	0.005	0.012	0.012	0.006	0.005
ResNet50 + YoloV5	1	350	39	140	35	0.869	0.909	0.910	0.8	0.904	0.903
	2	350	39	140	35	0.869	0.909	0.910	0.8	0.904	0.903
	3	347	42	139	36	0.862	0.906	0.892	0.794	0.899	0.898
	4	338	51	139	36	0.846	0.904	0.869	0.794	0.886	0.893
	5	345	44	142	33	0.863	0.913	0.887	0.811	0.910	0.899
Mean		346	43	140	35	0.862	0.908	0.890	0.8	0.899	0.899

Table 7. *Cont.*

Model	Fold	Confusion Matrix and Results									
		TP	**FN**	**TN**	**FP**	**Accuracy**	**Precision**	**Recall**	**Specificity**	**F1**	**AUROC**
Standard deviation		4.4	4.4	1.1	1.1	0.008	0.003	0.011	0.006	0.007	0.003
ResNet50 + Ensemble	1	345	44	144	31	0.867	0.918	0.887	0.823	0.902	0.905
	2	335	54	144	31	0.849	0.915	0.861	0.823	0.887	0.889
	3	353	36	140	35	0.874	0.910	0.907	0.8	0.909	0.912
	4	348	41	138	37	0.862	0.904	0.895	0.789	0.899	0.908
	5	347	42	143	32	0.869	0.916	0.892	0.817	0.904	0.906
Mean		346	43	142	33	0.864	0.912	0.888	0.810	0.900	0.904
Standard deviation		5.9	5.9	2.4	2.4	0.008	0.005	0.015	0.014	0.007	0.008

We can notice that with the previous differentiation between the classes, the locator reached more expressive values than when used to perform the functions of locating and classifying the opacity regions. YoloX had a significant increase in accuracy; its value was 0.819 in Experiment 1 (fold 3), and in Experiment 2 (fold 3) it increased to 0.861. Consequently, when we combined the Yolo models (ensemble), the accuracy increased from 0.816 to 0.874 (fold 3), reaching a 0.912 AUROC.

One of the key ideas behind the ensemble model is to combine low-bias uncorrelated models. Although the models used in our study belong to the same family of deep convolutional networks, this double-checking of regions can help to eliminate the difficult negative examples, i.e., non-pneumonia or other areas of pulmonary opacity of pathology. As a result, a set of two networks presents higher indices than the individual networks.

Figure 4 presents the individual results of the detectors, as well as those obtained with the ensemble model. To enhance data visualization, we have overlaid the bounding boxes of the detectors on the ground truth.

We can observe three situations in Figure 4: (1) both detectors accurately locate all suspicious areas in the image; (2) one detector fails to locate a suspicious area, while the other does so accurately; and (3) one detector identifies an area that is not actually suspicious, resulting in a false positive (FP), while the other detects it correctly. In relation to this, we can assert that the ensemble model exhibits fewer false negative (FN) cases, as demonstrated in situation 2. However, if the confidence score exceeds the allowable threshold, the ensemble model may produce false positive cases, as depicted in situation 3.

A notable factor concerning the ground truth (GT) of this dataset is that in several images, the bounding boxes encompass the entire lung (Figure 5). This can potentially confuse the detector, as it might include healthy regions within these bounding boxes.

One of the biggest challenges in building a deep learning system for locating regions is collecting a large set of data with the respective annotations. Drawing such annotations is time-consuming and costly, and it is dependent on the observer's interpretation [37,62].

The 'Image-level prediction' term is used to refer to the task of predicting a label of a class (classification) or continuous value (regression), and it is implemented by analyzing the entire image. The literature is heavily focused on predictions of this type, with few publications focused on locating findings. Many studies provide an unvalidated view of the area of interest, using methods such as grad-cam, which produce heatmaps indicating which regions are important in the network output. Although heat maps can be useful for certain conditions or indicate signals, the lack of an assessment of their accuracy is problematic. In addition, many conditions can be difficult to explain with a heat map—for example, emphysema, which is identified by irregular radiolucency throughout the lung (among other features) [37].

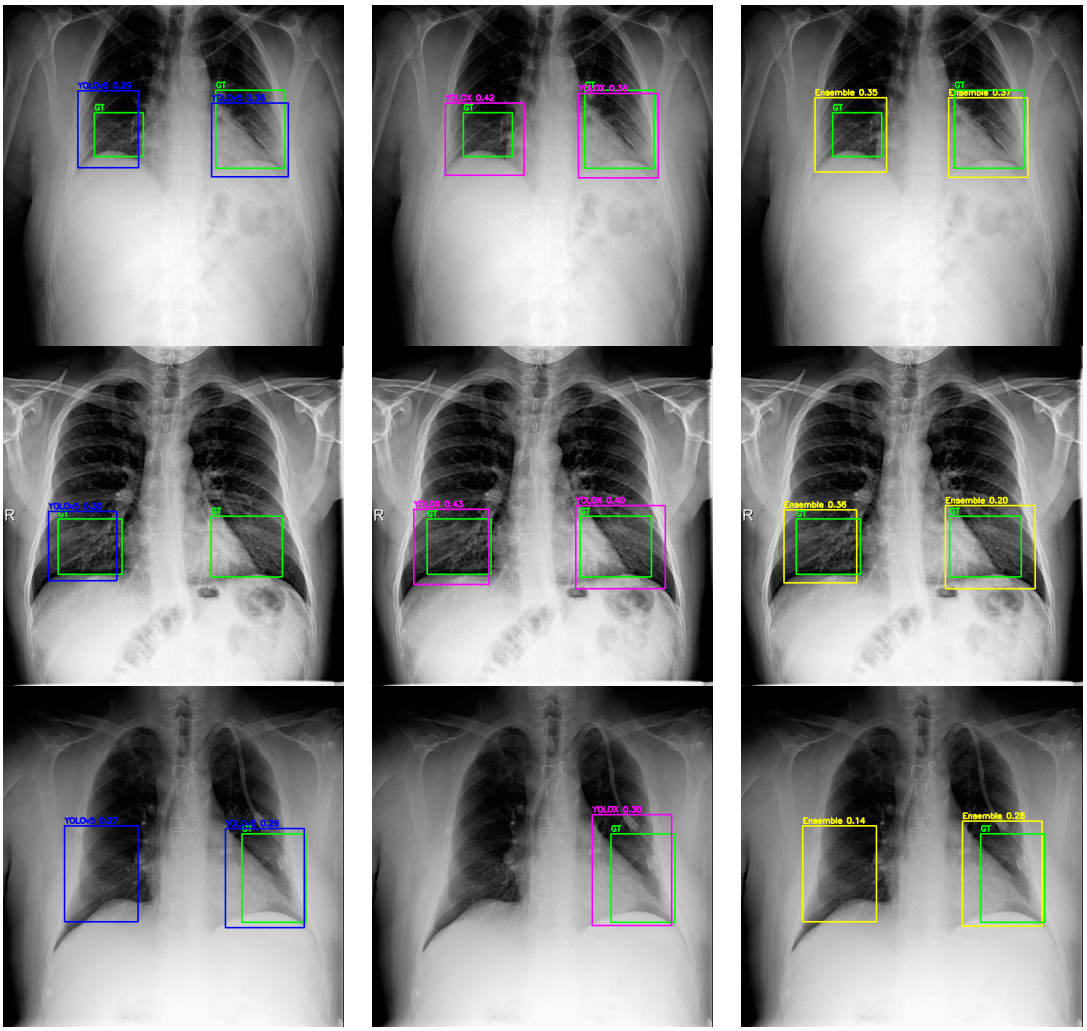

Figure 4. Examples of COVID-19 detection in CXR images. The green bounding boxes represent the ground truth (GT), while the blue, magenta, and yellow ones represent the results obtained by YoloV5, YoloX, and the ensemble model, respectively.

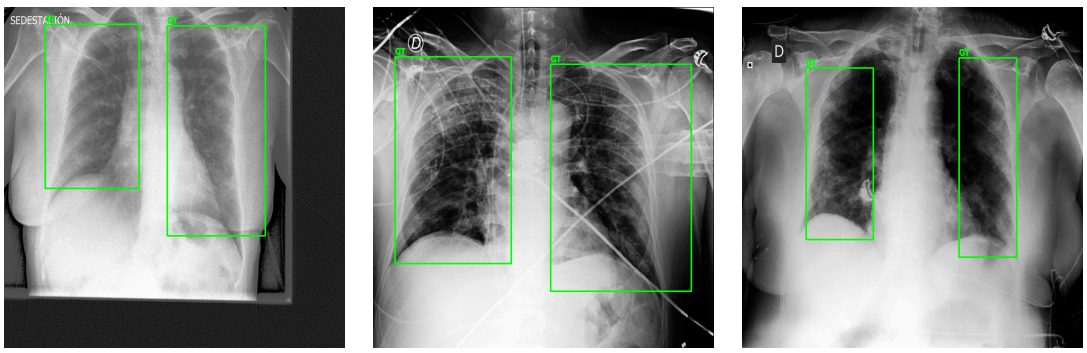

Figure 5. Examples of annotations where the ground truth (green bounding box) covers the entire pulmonary region.

The proposed system has the following important advantages: results superior to those found in the literature in the classification of the dataset; a promising predictive accuracy in locating regions of pulmonary opacity, indicating to the specialist the area that the specialist should direct greater attention to; and agility in predicting COVID-19.

7. Conclusions

In this study, we evaluated the accuracy of differentiating between normal and COVID-19 cases on CXR images. In addition, we also investigated two methods based on "guided attention", i.e., convolutional neural networks capable of highlighting suspicious areas that the specialist needs to pay more attention to. The approach using the ResNet50 classification network, followed by the ensemble location model, provided an expressive result compared to the works found in the literature involving this issue. The result was significant in terms of the global classification of the image, as well as in the location of suspicious regions that require greater attention from the specialist. The classification model achieved an accuracy of 0.864 and an AUC of 0.904 in 5-fold cross-validation, with the emphasis on fold 3, which reached an accuracy of 0.874 and an AUC of 0.912. The overlap between the predicted bounding boxes and the GT reached, in the ensemble model, a mAP of 59.63% in 5-fold cross-validation and 61.1% for the same model in fold 5.

Despite our successes, we encountered several challenges. The dataset, while comprising high-quality images with minimal compression, exhibited significant variability in the collection methods and image quality across clinical settings. Furthermore, the categorization of COVID-19 cases relied on the visual analysis of CXRs rather than RT-PCR results, introducing potential subjectivity and error in radiologist classification. Additionally, it is crucial to consider interobserver variability in the delineation of bounding boxes.

As part of our future work, our aim is to further develop a rapid and precise ensemble model for localizing suspicious lung regions. This involves exploring novel networks and architectures, as well as introducing a third detector to assist in resolving ambiguities and mitigating false positives and false negatives.

In conclusion, we consider that its good performance and fast diagnostic prediction make the model developed a promising way to aid the expert in decision making.

Funding: This research was supported by the Sao Paulo Research Foundation (FAPESP) under grant nos. 2020/14180-4, 2016/17078-0 and 2020/07200-9.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data included in this study are available upon request by contacting the corresponding author.

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

References

1. World Health Organization Coronavirus Disease 2019 (COVID-19): Situation Report 72. 2020. Available online: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200401-sitrep-72-covid-19.pdf (accessed on 1 May 2024).
2. Pham, T.D. Classification of COVID-19 Chest X-rays with Deep Learning: New Models or Fine Tuning? *Health Inf. Sci. Syst.* **2021**, *9*, 2. [CrossRef]
3. Maior, C.B.S.; Santana, J.M.M.; Lins, I.D.; Moura, M.J.C. Convolutional Neural Network Model Based on Radiological Images to Support COVID-19 Diagnosis: Evaluating Database Biases. *PLoS ONE* **2021**, *16*, e0247839. [CrossRef]
4. Wang, Z.; Xiao, Y.; Li, Y.; Zhang, J.; Lu, F.; Hou, M.; Liu, X. Automatically Discriminating and Localizing COVID-19 from Community-Acquired Pneumonia on Chest X-rays. *Pattern Recognit.* **2021**, *110*, 107613. [CrossRef]
5. Aishwarya, T.; Ravi Kumar, V. Machine Learning and Deep Learning Approaches to Analyze and Detect COVID-19: A Review. *SN Comput. Sci.* **2021**, *2*, 226. [CrossRef]
6. El Asnaoui, K.; Chawki, Y. Using X-ray Images and Deep Learning for Automated Detection of Coronavirus Disease. *J. Biomol. Struct. Dyn.* **2021**, *39*, 3615–3626. [CrossRef]
7. Kassania, S.H.; Kassanib, P.H.; Wesolowskic, M.J.; Schneidera, K.A.; Detersa, R. Automatic Detection of Coronavirus Disease (COVID-19) in X-ray and CT Images: A Machine Learning Based Approach. *Biocybern. Biomed. Eng.* **2021**, *41*, 867–879. [CrossRef]
8. Khatami, F.; Saatchi, M.; Zadeh, S.S.T.; Aghamir, Z.S.; Shabestari, A.N.; Reis, L.O.; Aghamir, S.M.K. A Meta-Analysis of Accuracy and Sensitivity of Chest CT and RT-PCR in COVID-19 Diagnosis. *Sci. Rep.* **2020**, *10*, 22402. [CrossRef] [PubMed]
9. Wikramaratna, P.S.; Paton, R.S.; Ghafari, M.; Lourenço, J. Estimating the False-Negative Test Probability of SARS-CoV-2 by RT-PCR. *Eurosurveillance* **2020**, *25*, 2000568. [CrossRef]

10. Jaiswal, A.; Gianchandani, N.; Singh, D.; Kumar, V.; Kaur, M. Classification of the COVID-19 Infected Patients Using DenseNet201 Based Deep Transfer Learning. *J. Biomol. Struct. Dyn.* **2021**, *39*, 5682–5689. [CrossRef]
11. Zhang, R.; Tie, X.; Qi, Z.; Bevins, N.B.; Zhang, C.; Griner, D.; Song, T.K.; Nadig, J.D.; Schiebler, M.L.; Garrett, J.W.; et al. Diagnosis of Coronavirus Disease 2019 Pneumonia by Using Chest Radiography: Value of Artificial Intelligence. *Radiology* **2021**, *298*, E88–E97. [CrossRef]
12. Kugunavar, S.; Prabhakar, C.J. Convolutional Neural Networks for the Diagnosis and Prognosis of the Coronavirus Disease Pandemic. *Vis. Comput. Ind. Biomed. Art.* **2021**, *4*, 12. [CrossRef]
13. Wong, H.Y.F.; Lam, H.Y.S.; Fong, A.H.-T.; Leung, S.T.; Chin, T.W.-Y.; Lo, C.S.Y.; Lui, M.M.-S.; Lee, J.C.Y.; Chiu, K.W.-H.; Chung, T.W.-H.; et al. Frequency and Distribution of Chest Radiographic Findings in Patients Positive for COVID-19. *Radiology* **2020**, *296*, E72–E78. [CrossRef]
14. Zech, J.R.; Badgeley, M.A.; Liu, M.; Costa, A.B.; Titano, J.J.; Oermann, E.K. Variable Generalization Performance of a Deep Learning Model to Detect Pneumonia in Chest Radiographs: A Cross-Sectional Study. *PLoS Med.* **2018**, *15*, e1002683. [CrossRef]
15. Mishra, N.K.; Singh, P.; Joshi, S.D. Automated Detection of COVID-19 from CT Scan Using Convolutional Neural Network. *Biocybern. Biomed. Eng.* **2021**, *41*, 572–588. [CrossRef]
16. Chatterjee, P.; Biswas, M.; Das, A.K. Specialized COVID-19 Detection Techniques with Machine Learning. *J. Phys. Conf. Ser.* **2021**, *1797*, 012033. [CrossRef]
17. Hussain, L.; Nguyen, T.; Li, H.; Abbasi, A.A.; Lone, K.J.; Zhao, Z.; Zaib, M.; Chen, A.; Duong, T.Q. Machine-Learning Classification of Texture Features of Portable Chest X-ray Accurately Classifies COVID-19 Lung Infection. *Biomed. Eng. Online* **2020**, *19*, 88. [CrossRef]
18. Parveen, N.R.S.; Sathik, M.M. Detection of Pneumonia in Chest X-ray Images. *J. Xray Sci. Technol.* **2011**, *19*, 423–428. [CrossRef]
19. Xie, C.; Ng, M.-Y.; Ding, J.; Leung, S.T.; Lo, C.S.Y.; Wong, H.Y.F.; Vardhanabhuti, V. Discrimination of Pulmonary Ground-Glass Opacity Changes in COVID-19 and Non-COVID-19 Patients Using CT Radiomics Analysis. *Eur. J. Radiol. Open* **2020**, *7*, 100271. [CrossRef]
20. Fang, M.; He, B.; Li, L.; Dong, D.; Yang, X.; Li, C.; Meng, L.; Zhong, L.; Li, H.; Li, H.; et al. CT Radiomics Can Help Screen the Coronavirus Disease 2019 (COVID-19): A Preliminary Study. *Sci. China Inf. Sci.* **2020**, *63*, 172103. [CrossRef]
21. Mondal, M.R.H.; Bharati, S.; Podder, P.; Podder, P. Data Analytics for Novel Coronavirus Disease. *Inform. Med. Unlocked* **2020**, *20*, 100374. [CrossRef]
22. Almalki, Y.E.; Qayyum, A.; Irfan, M.; Haider, N.; Glowacz, A.; Alshehri, F.M.; Alduraibi, S.K.; Alshamrani, K.; Alkhalik Basha, M.A.; Alduraibi, A.; et al. A Novel Method for COVID-19 Diagnosis Using Artificial Intelligence in Chest X-ray Images. *Healthcare* **2021**, *9*, 522. [CrossRef] [PubMed]
23. Moura, L.V.d.; Mattjie, C.; Dartora, C.M.; Barros, R.C.; Marques da Silva, A.M. Explainable Machine Learning for COVID-19 Pneumonia Classification With Texture-Based Features Extraction in Chest Radiography. *Front. Digit. Health* **2022**, *3*, 662343. [CrossRef] [PubMed]
24. Ferreira Junior, J.R.; Cardona Cardenas, D.A.; Moreno, R.A.; de Sá Rebelo, M.d.F.; Krieger, J.E.; Gutierrez, M.A. Novel Chest Radiographic Biomarkers for COVID-19 Using Radiomic Features Associated with Diagnostics and Outcomes. *J. Digit. Imaging* **2021**, *34*, 297–307. [CrossRef] [PubMed]
25. Fang, X.; Li, X.; Bian, Y.; Ji, X.; Lu, J. Radiomics Nomogram for the Prediction of 2019 Novel Coronavirus Pneumonia Caused by SARS-CoV-2. *Eur. Radiol.* **2020**, *30*, 6888–6901. [CrossRef] [PubMed]
26. Shiri, I.; Sorouri, M.; Geramifar, P.; Nazari, M.; Abdollahi, M.; Salimi, Y.; Khosravi, B.; Askari, D.; Aghaghazvini, L.; Hajianfar, G.; et al. Machine Learning-Based Prognostic Modeling Using Clinical Data and Quantitative Radiomic Features from Chest CT Images in COVID-19 Patients. *Comput. Biol. Med.* **2021**, *132*, 104304. [CrossRef] [PubMed]
27. Fu, L.; Li, Y.; Cheng, A.; Pang, P.; Shu, Z. A Novel Machine Learning-Derived Radiomic Signature of the Whole Lung Differentiates Stable From Progressive COVID-19 Infection. *J. Thorac. Imaging* **2020**, *35*, 361–368. [CrossRef] [PubMed]
28. Pulgar-Sánchez, M.; Chamorro, K.; Fors, M.; Mora, F.X.; Ramírez, H.; Fernandez-Moreira, E.; Ballaz, S.J. Biomarkers of Severe COVID-19 Pneumonia on Admission Using Data-Mining Powered by Common Laboratory Blood Tests-Datasets. *Comput. Biol. Med.* **2021**, *136*, 104738. [CrossRef] [PubMed]
29. Chowdhury, M.E.H.; Rahman, T.; Khandakar, A.; Mazhar, R.; Kadir, M.A.; Mahbub, Z.B.; Islam, K.R.; Khan, M.S.; Iqbal, A.; Al Emadi, N.; et al. Can AI Help in Screening Viral and COVID-19 Pneumonia? *IEEE Access* **2020**, *8*, 132665–132676. [CrossRef]
30. Wehbe, R.M.; Sheng, J.; Dutta, S.; Chai, S.; Dravid, A.; Barutcu, S.; Wu, Y.; Cantrell, D.R.; Xiao, N.; Allen, B.D.; et al. DeepCOVID-XR: An Artificial Intelligence Algorithm to Detect COVID-19 on Chest Radiographs Trained and Tested on a Large U.S. Clinical Data Set. *Radiology* **2021**, *299*, E167–E176. [CrossRef]
31. Apostolopoulos, I.D.; Mpesiana, T.A. COVID-19: Automatic Detection from X-ray Images Utilizing Transfer Learning with Convolutional Neural Networks. *Phys. Eng. Sci. Med.* **2020**, *43*, 635–640. [CrossRef]
32. Jain, G.; Mittal, D.; Thakur, D.; Mittal, M.K. A Deep Learning Approach to Detect COVID-19 Coronavirus with X-ray Images. *Biocybern. Biomed. Eng.* **2020**, *40*, 1391–1405. [CrossRef] [PubMed]
33. Asif, S.; Wenhui, Y.; Jin, H.; Jinhai, S. Classification of COVID-19 from Chest X-ray Images Using Deep Convolutional Neural Network. In Proceedings of the 2020 IEEE 6th International Conference on Computer and Communications (ICCC), Chengdu, China, 11–14 December 2020; pp. 426–433.

34. Jain, R.; Gupta, M.; Taneja, S.; Hemanth, D.J. Deep Learning Based Detection and Analysis of COVID-19 on Chest X-ray Images. *Appl. Intell.* **2021**, *51*, 1690–1700. [CrossRef] [PubMed]
35. Zebin, T.; Rezvy, S. COVID-19 Detection and Disease Progression Visualization: Deep Learning on Chest X-rays for Classification and Coarse Localization. *Appl. Intell.* **2021**, *51*, 1010–1021. [CrossRef] [PubMed]
36. Marques, G.; Agarwal, D.; Díez, I.T. Automated Medical Diagnosis of COVID-19 through EfficientNet Convolutional Neural Network. *Appl. Soft. Comput.* **2020**, *96*, 106691. [CrossRef] [PubMed]
37. Çallı, E.; Sogancioglu, E.; van Ginneken, B.; van Leeuwen, K.G.; Murphy, K. Deep Learning for Chest X-ray Analysis: A Survey. *Med. Image Anal.* **2021**, *72*, 102125. [CrossRef] [PubMed]
38. Nayyar, A.; Jain, R.; Upadhyay, Y. Object Detection Based Approach for Automatic Detection of Pneumonia. In Proceedings of the 2020 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD), Durban, South Africa, 6–7 August 2020; pp. 1–6.
39. Sirazitdinov, I.; Kholiavchenko, M.; Mustafaev, T.; Yixuan, Y.; Kuleev, R.; Ibragimov, B. Deep Neural Network Ensemble for Pneumonia Localization from a Large-Scale Chest X-ray Database. *Comput. Electr. Eng.* **2019**, *78*, 388–399. [CrossRef]
40. Ko, H.; Ha, H.; Cho, H.; Seo, K.; Lee, J. Pneumonia Detection with Weighted Voting Ensemble of CNN Models. In Proceedings of the 2019 2nd International Conference on Artificial Intelligence and Big Data (ICAIBD), Chengdu, China, 25–28 May 2019; pp. 306–310.
41. Jaiswal, A.K.; Tiwari, P.; Kumar, S.; Gupta, D.; Khanna, A.; Rodrigues, J.J.P.C. Identifying Pneumonia in Chest X-rays: A Deep Learning Approach. *Measurement* **2019**, *145*, 511–518. [CrossRef]
42. Mao, L.; Yumeng, T.; Lina, C. Pneumonia Detection in Chest X-rays: A Deep Learning Approach Based on Ensemble RetinaNet and Mask R-CNN. In Proceedings of the 2020 Eighth International Conference on Advanced Cloud and Big Data (CBD), Taiyuan, China, 5–6 December 2020; pp. 213–218.
43. Li, B.; Kang, G.; Cheng, K.; Zhang, N. Attention-Guided Convolutional Neural Network for Detecting Pneumonia on Chest X-rays. In Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 23–27 July 2019; pp. 4851–4854.
44. Yao, S.; Chen, Y.; Tian, X.; Jiang, R. Pneumonia Detection Using an Improved Algorithm Based on Faster R-CNN. *Comput. Math. Methods Med.* **2021**, *2021*, 8854892. [CrossRef] [PubMed]
45. Yao, S.; Chen, Y.; Tian, X.; Jiang, R.; Ma, S. An Improved Algorithm for Detecting Pneumonia Based on YOLOv3. *Appl. Sci.* **2020**, *10*, 1818. [CrossRef]
46. Frid-Adar, M.; Amer, R.; Gozes, O.; Nassar, J.; Greenspan, H. COVID-19 in CXR: From Detection and Severity Scoring to Patient Disease Monitoring. *IEEE J. Biomed. Health Inform.* **2021**, *25*, 1892–1903. [CrossRef]
47. Al-antari, M.A.; Hua, C.-H.; Bang, J.; Lee, S. Fast Deep Learning Computer-Aided Diagnosis of COVID-19 Based on Digital Chest X-ray Images. *Appl. Intell.* **2021**, *51*, 2890–2907. [CrossRef] [PubMed]
48. Cohen, J.P.; Morrison, P.; Dao, L. COVID-19 Image Data Collection. *arXiv* **2020**, arXiv:2006.11988.
49. Duraisamy, P.; D., A.; V., N. Enhancing COVID-19 Detection with YOLOv5 Object Detection Algorithm in Chest X-ray Images. In Proceedings of the 2024 3rd International Conference for Innovation in Technology (INOCON), Bangalore, India, 1–3 March 2024; pp. 1–6.
50. Yang, W.; Wang, R.; Zhang, J.; Liu, C.; Zeng, L.; Shen, X. COVID-19 Detection and Localization: A Novel Fused Attention Mechanism Network Based on YOLO V5. In *Electronic Engineering and Informatics*; IOS Press: Amsterdam, The Netherlands, 2024.
51. Lakhani, P.; Mongan, J.; Singhal, C.; Zhou, Q.; Andriole, K.P.; Auffermann, W.F.; Prasanna, P.; Pham, T.; Peterson, M.; Bergquist, P.J.; et al. The 2021 SIIM-FISABIO-RSNA Machine Learning COVID-19 Challenge:Annotation and Standard Exam Classification of COVID-19 ChestRadiographs. *J. Digit. Imaging* **2022**, *36*, 1–8. [CrossRef] [PubMed]
52. Vayá, M.I.; Saborit-Torres, M.d.l.I.; Jose, M.; Serrano, J.A.M.; Oliver-Garcia, E.; Pertusa, A.; Bustos, A.; Cazorla, M.; Galant, J.; Barber, X.; et al. BIMCV COVID-19+: A Large Annotated Dataset of RX and CT Images from COVID-19 Patients. *IEEE Dataport* **2021**, 1–22. [CrossRef]
53. Tsai, E.B.; Simpson, S.; Lungren, M.P.; Hershman, M.; Roshkovan, L.; Colak, E.; Erickson, B.J.; Shih, G.; Stein, A.; Kalpathy-Cramer, J.; et al. The RSNA International COVID-19 Open Radiology Database (RICORD). *Radiology* **2021**, *299*, E204–E213. [CrossRef]
54. Litmanovich, D.E.; Chung, M.; Kirkbride, R.R.; Kicska, G.; Kanne, J.P. Review of Chest Radiograph Findings of COVID-19 Pneumonia and Suggested Reporting Language. *J. Thorac. Imaging* **2020**, *35*, 354–360. [CrossRef] [PubMed]
55. Sherrier, R.H.; Johnson, G.A. Regionally Adaptive Histogram Equalization of the Chest. In *IEEE Transactions on Medical Imaging*; IEEE: Piscataway, NJ, USA, 1987; pp. 1–7.
56. Jocher, G.; Stoken, A.; Borovec, J.; NanoCode012; ChristopherSTAN; Changyu, L.; Laughing; Tkianai; Hogan, A.; Lorenzomammana; et al. Ultralytics/Yolov5: V3.1—Bug Fixes and Performance Improvements. *Zenodo* **2020**. [CrossRef]
57. Iyer, R.V.; Ringe, P.S.; Bhensdadiya, K.P. Comparison of YOLOv3, YOLOv5s and MobileNet-SSD V2 for Real-Time Mask Detection. *Int. Res. J. Eng. Technol. (IRJET)* **2021**, *8*, 1156–1160.
58. Ge, Z.; Liu, S.; Wang, F.; Li, Z.; Sun, J. YOLOX: Exceeding YOLO Series in 2021. *arXiv* **2021**, arXiv:2107.08430.
59. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767.
60. Solovyev, R.; Wang, W.; Gabruseva, T. Weighted Boxes Fusion: Ensembling Boxes from Different Object Detection Models. *Image Vis. Comput.* **2021**, *107*, 104117. [CrossRef]

61. Dhont, J.; Wolfs, C.; Verhaegen, F. Automatic Coronavirus Disease 2019 Diagnosis Based on Chest Radiography and Deep Learning—Success Story or Dataset Bias? *Med. Phys.* **2022**, *49*, 978–987. [CrossRef] [PubMed]
62. Wu, J.; Gur, Y.; Karargyris, A.; Syed, A.B.; Boyko, O.; Moradi, M.; Syeda-Mahmood, T. Automatic Bounding Box Annotation of Chest X-ray Data for Localization of Abnormalities. In Proceedings of the 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), Iowa City, IA, USA, 3–7 April 2020; pp. 799–803.

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.