ORIGINAL ARTICLE

Detection and classification of COVID-19 by using faster R-CNN and mask R-CNN on CT images

M. Emin Sahin[1] · Hasan Ulutas[1] · Esra Yuce[1] · Mustafa Fatih Erkoc[2]

Abstract

The coronavirus (COVID-19) pandemic has a devastating impact on people's daily lives and healthcare systems. The rapid spread of this virus should be stopped by early detection of infected patients through efficient screening. Artificial intelligence techniques are used for accurate disease detection in computed tomography (CT) images. This article aims to develop a process that can accurately diagnose COVID-19 using deep learning techniques on CT images. Using CT images collected from Yozgat Bozok University, the presented method begins with the creation of an original dataset, which includes 4000 CT images. The faster R-CNN and mask R-CNN methods are presented for this purpose in order to train and test the dataset to categorize patients with COVID-19 and pneumonia infections. In this study, the results are compared using VGG-16 for faster R-CNN model and ResNet-50 and ResNet-101 backbones for mask R-CNN. The faster R-CNN model used in the study has an accuracy rate of 93.86%, and the ROI (region of interest) classification loss is 0.061 per ROI. At the conclusion of the final training, the mask R-CNN model generates mAP (mean average precision) values for ResNet-50 and ResNet-101, respectively, of 97.72% and 95.65%. The results for five folds are obtained by applying the cross-validation to the methods used. With training, our model performs better than the industry standard baselines and can help with automated COVID-19 severity quantification in CT images.

Keywords Classification · COVID-19 · CT images · Faster R-CNN · Mask R-CNN

1 Introduction

The infectious disease caused by severe acute respiratory syndrome is COVID-19 disease. In December 2019, this disease appeared for the first time in the Chinese city of Wuhan, Hubei Province [1]. The virus has also spread rapidly outside of China.

Chest radiography is an important imaging technique that has been shown in earlier research to be useful in predicting COVID-19 infection in patients [2]. Deep learning is one of the most effective tools in medical science. These approaches can provide a quick and effective outcome with a high accuracy rate for the diagnosis and prognosis of many diseases [3]. Specially trained models are available to categorize inputs into the many categories that programmers demand. Deep learning models are frequently employed in the medical industry for detecting cardiac abnormalities, tumours utilizing image analysis, cancer diagnosis, and many more applications [4, 5]. A number of review studies examine and evaluate several models in order to identify which is the most accurate and COVID-19 susceptible. According to these publications, feature-extracting models for chest CT scan data with various network designs provide the greatest overall accuracy across a range of datasets. It is frequently used, for example, to classify CT scan images of COVID-19-infected individuals as positive, negative, or non-infected. There are several studies in the literature that use artificial intelligence approaches, particularly deep learning models, to diagnose COVID-19 using X-ray and CT scans [6–11].

Artificial neural network-based techniques, in which effective properties of COVID-19 are learned directly from

---

✉ M. Emin Sahin
emin.sahin@bozok.edu.tr

[1] Department of Computer Engineering, Faculty of Engineering and Architecture, Yozgat Bozok University, Yozgat, Turkey

[2] Department of Radiology, Faculty of Medicine, Yozgat Bozok University, Yozgat, Turkey

imaging data: proving their satisfactory performance in many medical image processing and computer vision applications, such as the rapid growth and classification of deep neural networks [12–14]. Among them, researchers such as detection [15, 16] and segmentation [2, 17] have adopted deep learning methods, providing more accurate results for COVID-19 diagnosis and classification. When the studies for the detection of COVID-19 in the literature are examined, Zheng et al. (2020) used 3D CT for the diagnosis of COVID-19 disease [18]. They developed an application called DeCoVNet using the DR architecture. Four hundred and ninety-nine CT images were used for training, and 131 CT images were used for testing. As a result of the study, they predicted the COVID-19 disease with an accuracy rate of 0.90. Ucar and Korkmaz (2020) made an application that they defined as COVIDiagnosis-Net by using AI-based SqueezeNet Bayesian optimization algorithms to diagnose COVID-19 [19]. They used the chest X-ray data set obtained from the open database. In their study, they reached a test accuracy of 0.983. They reported that this study to detect COVID-19 outperformed other studies in this field and achieved a higher accuracy rate.

Jaiswal et al. (2020) used DenseNet201-based deep transfer learning to classify CT images of patients infected with the COVID-19 virus [14]. As a result of the study, they reported that CNN showed good success and reached an accuracy rate of 97%. They stated that they detected COVID-19 with this model.

Ozyurt et al. (2021) conducted a study by applying ANN and CNN algorithms to CT images [20]. In this study, they obtained an accuracy rate of 94.10% with the CNN algorithm and 95.84% with the ANN algorithm. As a result of the study, they obtained a higher accuracy rate with the CNN algorithm.

Mask R-CNN [21], an improved object detection model based on Faster R-CNN, outperforms other object detection and segmentation models [22]. Mask R-CNN is also an essential AI-based approach that has been employed in a variety of applications, including the identification and segmentation of lung nodules [23], liver segmentation [24], and multiorgan segmentation [25]. This approach has also been used to identify and segment faces [22], hand segmentation [26], early gastric cancer segmentation [27, 28], and breast tumour detection and classification [29]. There are very few studies in the literature using mask R-CNN and faster R-CNN methods in detecting COVID-19. When the studies using these methods are examined; Ter-Sarkisov presents the COVID-CT-Mask-Net model, which predicts the presence of COVID-19 on chest CT scans [17]. Three thousand images are used to train the model. On test data containing 21,192 images, they achieve a COVID-19 sensitivity of 90.80% and F1-score of 91.50%. Importantly,

they establish that the regional predictions detected by mask R-CNN can be used to classify entire images. Podder et al. in their article propose an application that can diagnose COVID-19 using X-ray images and deep learning techniques [2]. In order to classify COVID-19-infected versus uninfected patients, they used the mask R-CNN method to train and test on the dataset. Using 668 chest X-ray images, the proposed model achieved a sensitivity of 97.36% and an accuracy of 96.98%.

This study is based on the benefits of mask R-CNN and automated image segmentation in establishing the COVID-19 region bounding box and drawing the contour of the COVID-19. The goal of this paper is to develop a model for automatically detecting, segmenting, and classifying diseases using COVID-19 CT scans. The findings of this study are compared to those of radiologists who specialize in COVID-19 diagnosis. Automatic image segmentation and disease classification using CT scans can be a useful tool in the diagnosis and treatment of COVID-19. The mask R-CNN algorithm is a popular choice for image segmentation tasks, as it is able to accurately identify and locate objects within an image. Using this algorithm to identify the region bounding box and draw the contour of COVID-19 in CT scans could potentially improve the efficiency and accuracy of COVID-19 diagnosis, particularly if the findings of the automated approach are compared to those of radiologists who specialize in COVID-19 diagnosis. It is important to note that while automated approaches have the potential to improve diagnostic accuracy and efficiency, they should not be used as a replacement for the expertise and judgement of trained medical professionals.

The article's structure is given below. In Sect. 2, a basic overview of the dataset, data pre-processing and object detection is provided, as well as a detailed explanation of the faster R-CNN and mask R-CNN models. Furthermore, details on the evaluation parameters for the used deep learning models are provided. The work and outcomes of faster R-CNN and mask R-CNN models are given in Sect. 3, along with evaluation findings such as accuracy, recall, precision, and F1 scores, as well as cross-validation results. The discussion and conclusion sections are submitted in Chapters 4 and 5, respectively. An overview of the study is shown in Fig. 1.

2 Material and methods

2.1 Dataset

There are numerous data sets developed by data scientists and practitioners of machine learning for the diagnosis of COVID-19 disease. In this study, however, a unique dataset was created by obtaining retrospective CT images
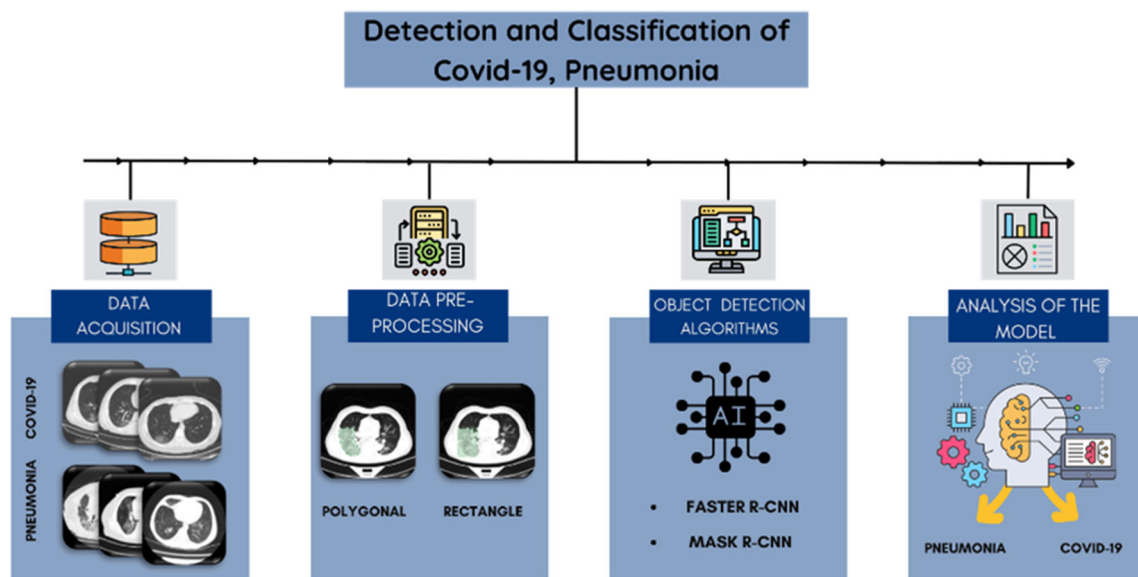
Fig. 1 An overview of the study

of patients with COVID-19 disease from the Yozgat Bozok University Faculty of Medicine after obtaining the necessary ethical approvals. All image labelling is overseen by a radiologist who is unaware of the clinical status of the patients (Table 1).

2.2 Data pre-processing

In computer vision applications, pre-processing is a common occurrence. Pre-processing techniques can aid in the training phase of deep learning by reducing unwanted noise and highlighting image regions that can aid in the recognition challenge, among other tasks.

This procedure requires the application of the Make-Sense data labelling tool. MakeSense is a free online application for image labelling. It draws polygons to completely outline images' objects. For this study, labels are saved as ".json" files for mask R-CNN, and the ".vgg" format for faster R-CNN. In this study, a total of 4000 images, including 3200 training images and 800 validation images, are labelled.

Table 1 Cases and image number information of the dataset

| No	Case name	Total cases	Total images	Used images
1	COVID-19	1058	15,650	2000
2	Pneumonia	1032	13,162	2000
3	Normal	1012	13,530	0

2.3 Object detection

Inspired by the success of deep learning algorithms (regional-based ESAs) for object detection on images, numerous deep learning-based object detection methods for remote sensing data have been developed. As a result of their superior capacity for learning high-level semantic representations, these methods are superior to conventional ones [30]. General object detection is the process of specifying an object's density in an image using a rectangular bounding box and labelling the object. Regional-based detectors and regression/classification-based detectors are the two categories of object detectors. The zone-based sensor employs the conventional method of object detection. First, the proposed object of the region is determined, and then the regions are classified. Regional-based systems include R-CNN [31], SPP-net [32], fast R-CNN [33], and mask R-CNN [21].

2.3.1 Faster R-CNN

Girshick proposed faster R-CNN, and what makes it more successful and appealing than its predecessors is that it introduces a mechanism (region proposal network) for estimating the region in the images where the object is believed to be located [33]. Similar to a fully convolutional network is a region bidding network (RPN). RPN's primary function is to generate object identification and detection suggestions. An image serves as the input for the RPN. It produces classification scores for object suggestions and flipping rectangles. A sliding window is applied to the feature map to generate suggested regions. At each sliding

Fig. 2 a COVID-19 original image **b** COVID -19 image marked polygonal **c** COVID-19 image marked rectangle

| Original image | Image marked polygonal | Image marked rectangle |
| a) | b) | c) |
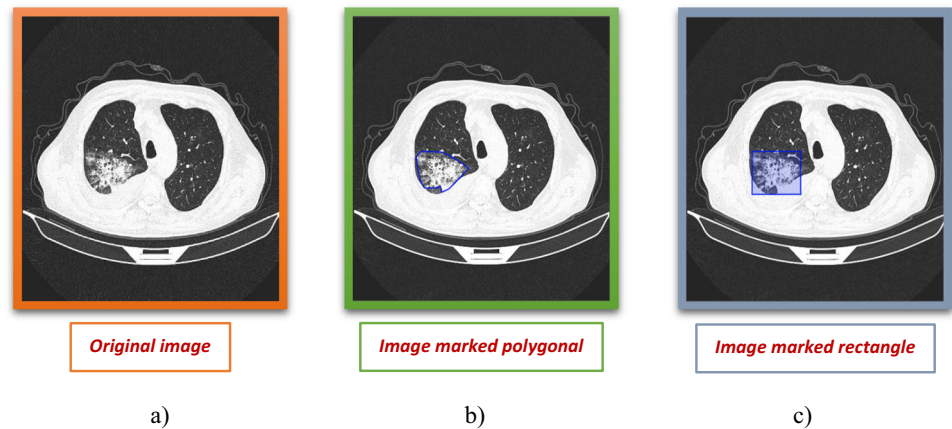
window position, several zone proposals are evaluated. Using intersection over union, the accuracy of the suggestions is checked (IoU). Due to the region proposal network mechanism, the cost of estimating the region where the object is believed to be located is significantly less than its predecessors. This feature has given the algorithm an edge over its competitors in terms of faster and more efficient work. Some sample marked dataset images are shown in Fig. 2.

2.3.2 Mask R-CNN

Mask region-based convolutional neural network (mask R-CNN) is a neural network model that was derived from the faster R-CNN model [21]. Faster R-CNN extracts a feature map from the input images. Regional proposal network (RPN) is utilized to identify image regions that may contain an object. Then, the regional proposals are pooled, and an estimate is made among the predetermined class or classes by finally presenting them to the fully connected MLP as input data for object estimation. While

these processes are performed more slowly in fast R-CNN, faster R-CNN yields faster results [34].

With the creation of faster R-CNN, mask R-CNN was born. During object detection, faster R-CNN determines the bounding box and class label of each object. Mask on the contrary, R-CNN creates the mask for each object. In this instance, mask R-CNN employs a two-step method. In the first stage, the regional bidding network is used to determine the bounding boxes of the objects (RPN). Using ROI Align, the mask of each object is determined in the second stage. Mask R-CNN architecture is shown in Fig. 3.

In mask R-CNN, minimizing the function of loss value:

$$L = L_{\text{class}} + L_{\text{box}} + L_{\text{mask}} \tag{1}$$

where $L_{\text{class}} + L_{\text{box}}$ are recognized the similar terms in faster R-CNN, and definition of $L_{\text{class}} + L_{\text{box}}$ is,
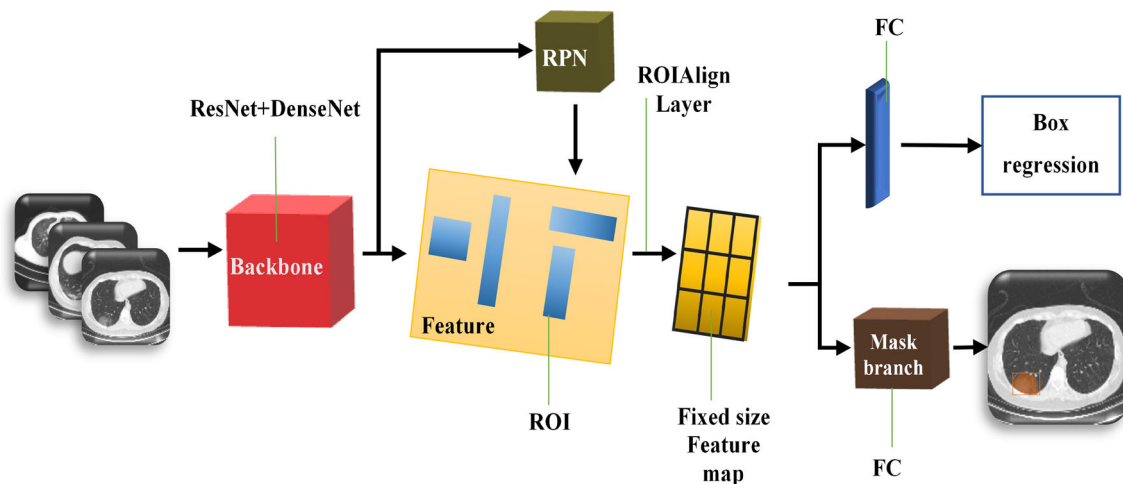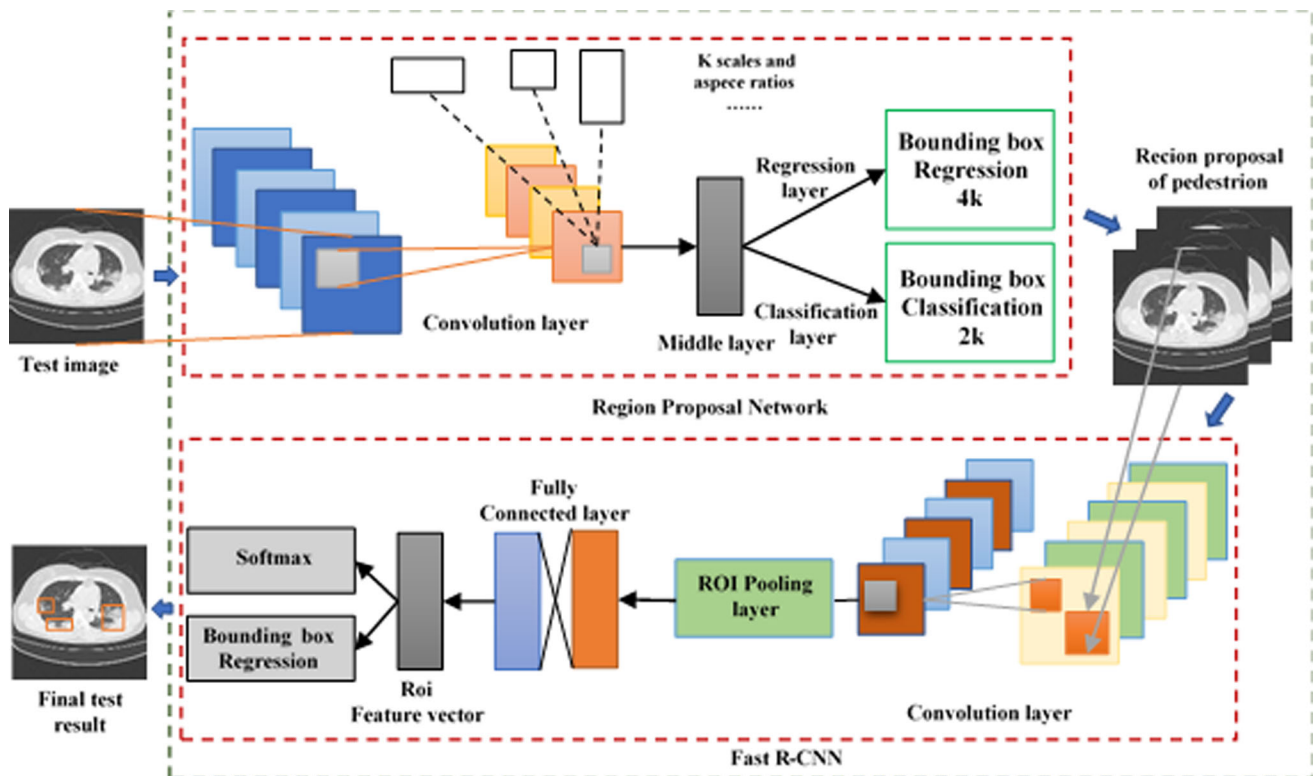

Fig. 3 Mask R-CNN architecture

Fig. 4 Procedure of the proposed method

$$L_{\text{class}} + L_{\text{box}} = \frac{1}{N_{\text{cls}}} \sum_i L_{\text{cls}}\left(p_i, p_i^*\right)$$
$$+ \frac{1}{N_{\text{box}}} \sum_i p_i^* L_i^{\text{smooth}}\left(t_i - t_i^*\right) \quad (2)$$

$$L_{\text{cls}}\left(\left\{\left(p_i, p_i^*\right)\right\}\right) = -p_i^* \log p_i^* - \left(1 - p_i^* \log\left(1 - p_i^*\right)\right) \quad (3)$$

Loss of average binary cross-entropy is L_{mask} defined as:

$$L_{\text{mask}} = \frac{1}{m^2} \sum_{0 \le i,j \le m} \left[y_{ij} \log o y_{ij}^k + \left(1 - y_{ij}\right) \log\left(1 - y_{ij}^k\right) \right] \quad (4)$$

In this study, the first stage is the backbone network (ResNet-50 and ResNet-101), where feature maps are generated from the input images. ROIs are generated by the RPN, which receives feature maps from the backbone network. In the third stage, the shared feature maps are mapped to extract the target features that correspond to the generated ROIs. Then, these characteristics are sent to the fully connected layer (FC) and the fully convolutional network (FCN). Examples include segmentation and classification of objects. In the final, fourth step, the positive regions of the ROI classifier are extracted, and classifier masks are created for them.

2.4 Evaluation methods

During the testing stage, the results of the algorithms are compared to the actual results in the marked data, and the success and performance value of the model are measured. These performance criteria are computed utilizing the confusion matrix.

Among the actual and approximate data contained in this data set:

Table 2 Faster R-CNN training configuration

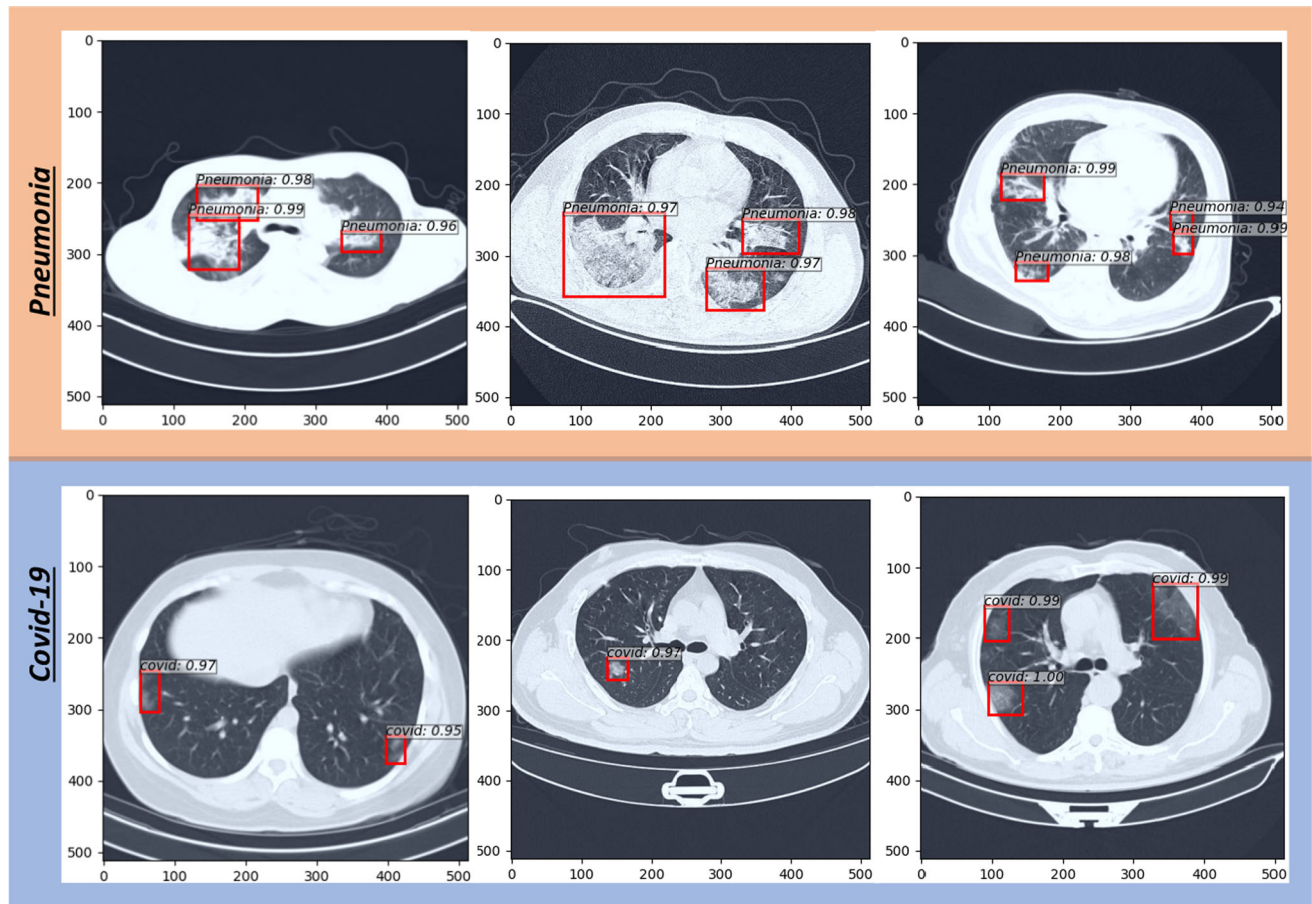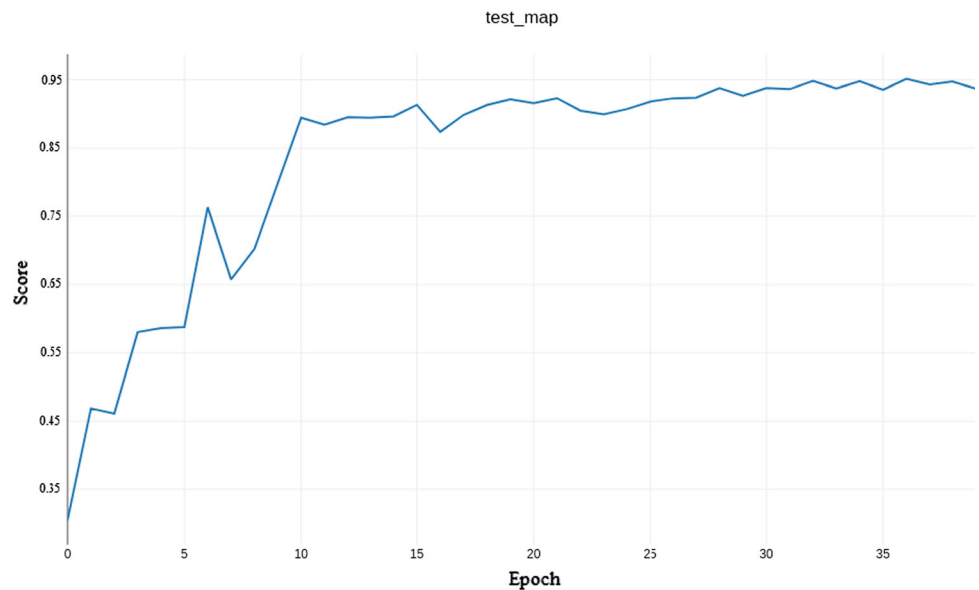| Parameter	Value
Num workers	3
Test num workers	3
RPN sigma	3
ROI sigma	3
Weight decay	0.0005
Learning decay	0.01
Learning rate	1e-4
Epoch	40
Optimizer	Adam
Dropout	True
Backbone	VGG-16
min_size, max_size	512, 512
Num_classes	2 + 1

Fig. 5 Faster R-CNN test results

**Fig. 6** Test mAP value for 40 epoch

True Positive (TP) The true positive value is the positive prediction of positive data.
True Negative (TN) A true negative value is an actual negative estimate of data.

False Positive (FP) A false positive value is the incorrect prediction of positive data.
False Negative (FN) False negative value is a negative estimate of actual positive data.

Fig. 7 ROI loss graph of faster R-CNN
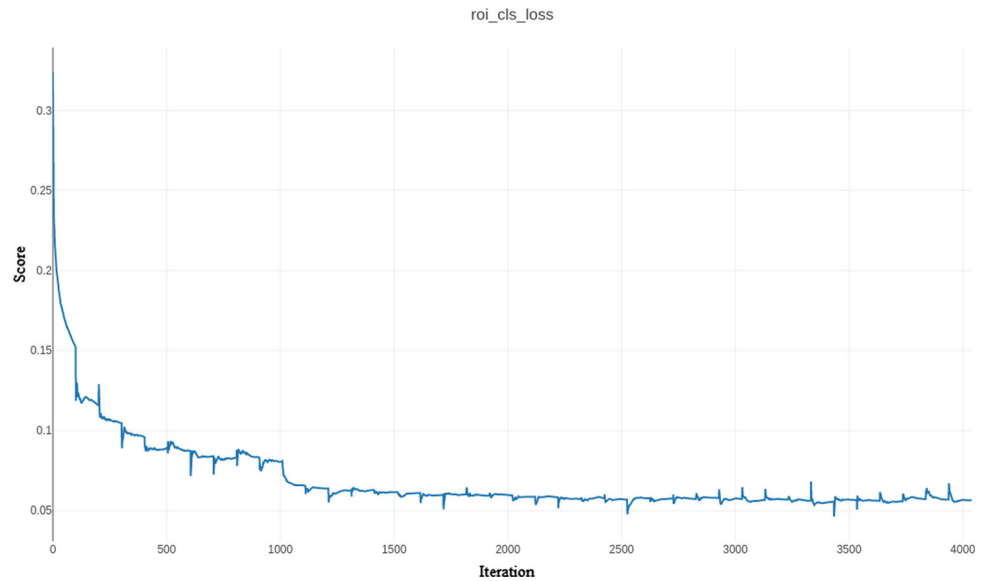
Fig. 8 RPN classification loss graph of faster R-CNN
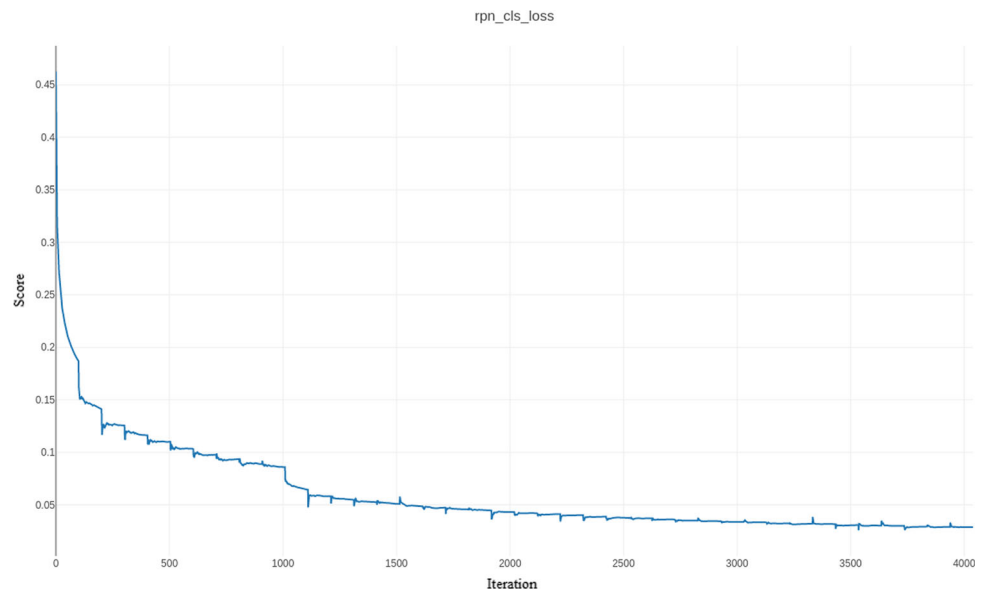
Using algorithms for deep learning, the success of the model is determined by the results of the confusion matrix. The diagnostic success rates of each algorithm for COVID-19 disease have been determined. The content and calculation formulas for these criteria are provided below.

The region bounded by the coordinate axis and the curve is known as average precision (AP). The AP values of all categories are averaged to produce mAP as well. The markers resemble Eqs. (5) through (8):

$$\text{Accuracy} = \frac{T_N + T_p}{T_N + T_p + F_N + F_p} \qquad (5)$$

$$\text{Precision} = \frac{T_p}{T_p + F_p} \times 100\% \qquad (6)$$

$$\text{Recall} = \frac{T_p}{T_p + F_N} \times 100\% \qquad (7)$$

$$F1 - \text{score} = \frac{2 \times P \times R}{P + R} \qquad (8)$$

The AP (average precision) value and the calculation of the area under the precision-sensitivity graph for each class separately are presented in Eq. (9).

$$\text{AP} = \int_{0}^{1} P(R)\mathrm{d}R \qquad (9)$$

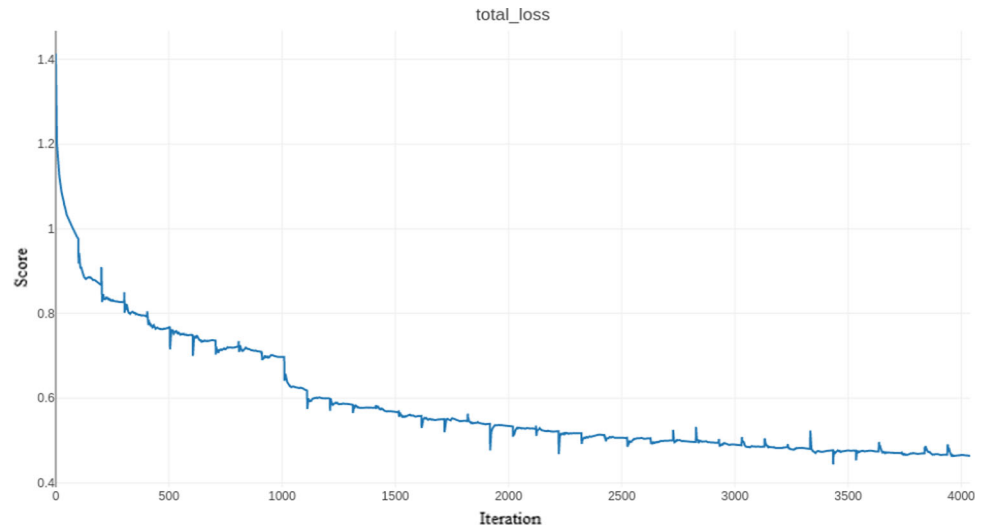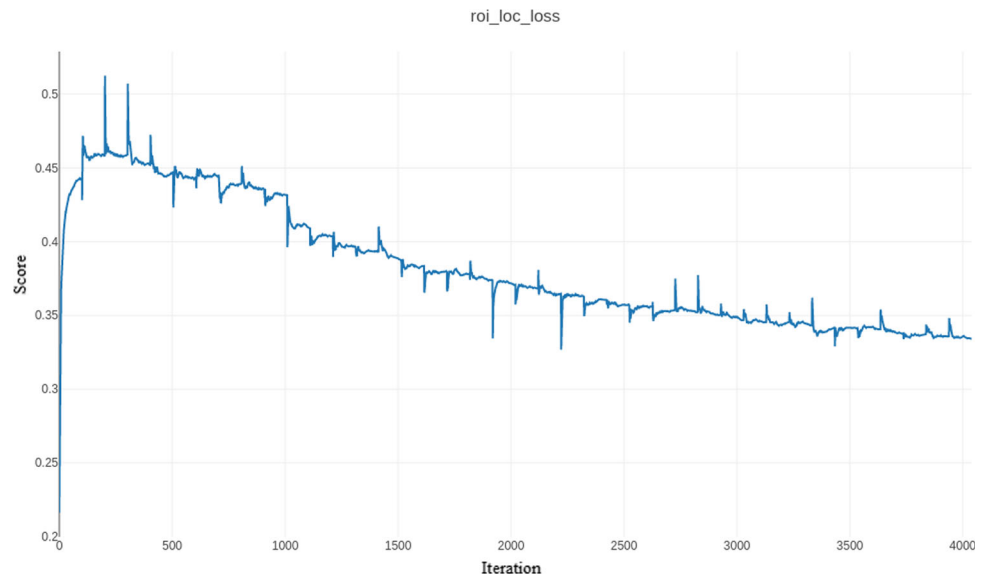**Fig. 9** Total loss graph of faster R-CNN

total_loss

Fig. 10 ROI localization loss

roi_loc_loss

The mean average precision (mAP) value is presented in Eq. (10) by taking the arithmetic average of the AP values calculated for each class.

$$mAP = \frac{A \cap B}{A \cup B} = \frac{1}{N_T} \sum_{i=1}^{N_T} \frac{N_i^{DR}}{N_i^D} \tag{10}$$

where segmentation result of the model is A and the radiologist experienced well will define consistent tumour contour as B, images number are represented by N_T; the area overlapped between the detected lesion model and the regions of lesion in true clinical value is given by N_i^{DR} i; and true lineal lesion size is N_i^D.

3 Experimental results

In this study, a unique data set consisting of CT scan images was prepared to classify COVID-19 patients, and faster R-CNN and mask R-CNN methods, whose data set is given in this section, are used. Procedure of the proposed method is given in Fig. 4.

3.1 Faster R-CNN

In the study, the VGG-16 network, which uses the residual and subsampling blocks in its layers as the backbone network, was preferred as the feature extractor of faster R-CNN. In the dataset, 3200 images of 512 × 512 size are used in model training, and 800 images of 512 × 512 size are used in training and validation.

Fig. 11 RPN localization loss

rpn_loc_loss
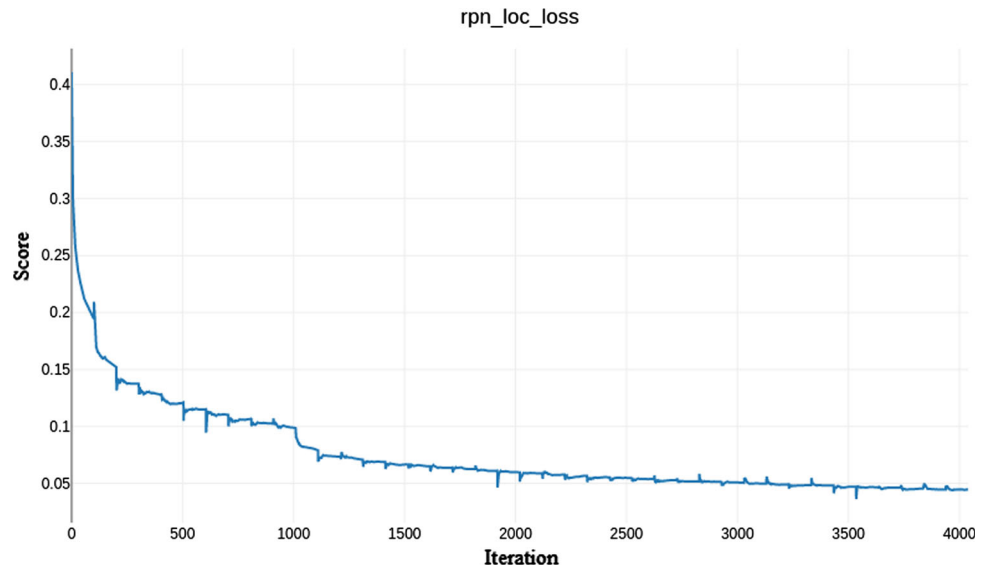
Table 3 Mask R-CNN parameter configuration

Parameters	Value
Images_Per_Gpu	1
Steps_Per_Epoch	100
Validation_Steps	250
Backbone	ResNet-50, ResNet-101
Num_Classes	2 + 1
Image_Mın_Dım	512
Image_Max_Dım	512
Learnıng_Rate	0.0001
Learnıng_Momentum	0.9

Table 4 Mask R-CNN test results

Training	Backbone	Epoch	mAP (%)
Pre-trained	ResNet-50	30	87.84
	ResNet-101	30	88.61
Final training	ResNet-50	50	97.72
	ResNet-101	50	95.65

There are many hyperparameters that need to be adjusted when training a faster R-CNN model. Exploring all configurations is nearly impossible in terms of time and computational resources. Therefore, default and obvious configurations are adhered to. Specifically, num_workers and test_num_workers (number of workers to load data) 3, rpn_sigma (Sigma parameter for RPN's localization loss) 3, roi_sigma (Sigma parameter for ROI localization loss) 3, learning rate 0.0001, learning decay (learning decay

parameter) 0.01, num_classes (number of classes + BG) is set to 2 + 1, optimizer man. All other configurations are kept by default in faster R-CNN. The configurations of the study are given in Table 2. The predictions and colour masks made as a result of the training are shown in Fig. 5.

Faster R-CNN determines if there is an object in the box and then tries to classify the object in the box. As a result of the training, a mAP value of 93.86% is obtained. The faster R-CNN test mAP plot is shown in Fig. 6.

The ROI classification loss is 0.0612 per ROI. How successfully the model labels a predicted box with the appropriate class is measured by "class loss" (loss for box classification). Faster R-CNN ROI classification loss plot is shown in Fig. 7. The classification loss in the RPN (regional offer network) is 0.0324. The loss of "objectivity" measures how good the RPN is at labelling junction boxes as foreground (classes) or background. Classification loss graph for faster R-CNN in RPN is given in Fig. 8. The total training loss is 0.4885. This is the weighted sum of the individual losses calculated during each iteration, and its graph is shown in Fig. 9. The localization loss for ROI is 0.3304, and the localization loss for RPN is 0.0491. Localization losses for ROI and RPN are shown in Figs. 10 and 11, respectively.

3.2 Mask R CNN

Mask R-CNN used in this study was produced by Matterport Inc. It is the version with the MIT License published by [35]. This version is based on open ski-libraries such as Keras and Tensorflow. As the backbone, both the ResNet-50 feature pyramid network model and the ResNet-101 feature pyramid network model are used. Three thousand two hundred images of 512 × 512 size are used in model
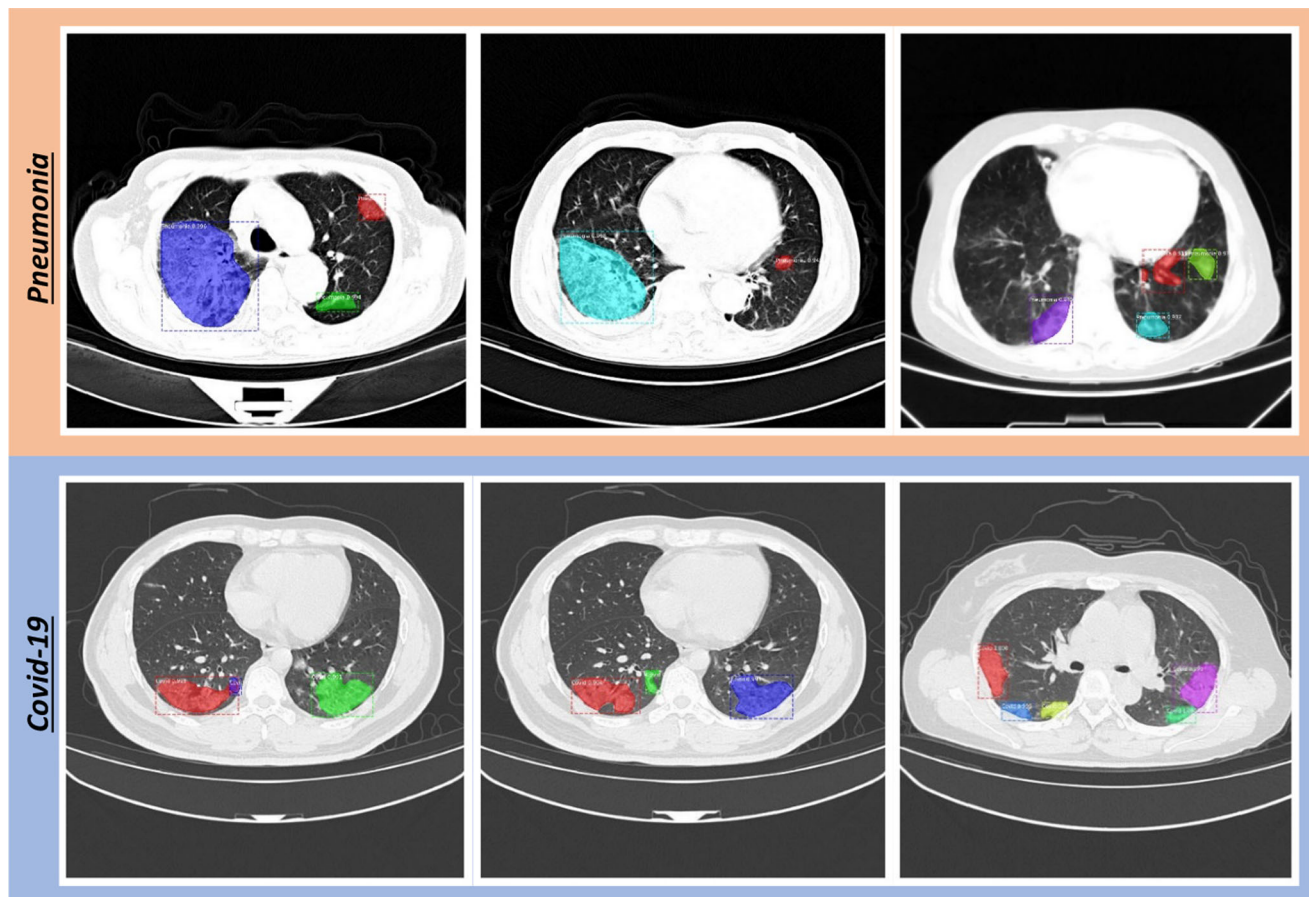
Fig. 12 Mask R-CNN test results

training, and 800 images of 512×512 size are utilized in training validation.

The number of training steps per period has been determined as 100. This is not required to match the size of the training set. TensorBoard updates are saved at the end of each period, so setting this value to a lower value will result in more frequent TensorBoard updates. Validation statistics are also calculated at the end of each term and can take some time, so choose it too small to avoid spending too much time on validation statistics. The number of images to train on each GPU is set to 1. At this stage, a GPU of about 12 GB can process two images, typically 1024 x 1024 pixels. The GPU should be adjusted for memory and image sizes. For best performance, select the highest number that the GPU can handle.

The number of test steps to be run at the end of each training period is determined to be 250. A larger number improves the accuracy of the validation statistics, but slows down the training. The minimum probability value to accept a detected sample is 0.85. ROIs below this threshold are skipped. The learning rate and momentum are determined as 0.0001 and 0.9, respectively. The configurations of the study are given in Table 3.

The training is repeated twice using different backbone architectures, ResNet-50 and ResNet-101. Instead of training the network from start to finish, pre-training is performed with 150 epochs using MSCOCO weights. Using the weights obtained from this preliminary training, the final training is completed with 250 epochs. In this way, more successful results are obtained in the final training by using the weights obtained from the pre-training. The results of the training using the fivefold cross-validation strategy are given in Table 4. Cross-validation is a statistical resampling method used to evaluate the performance of a network model on data in the most objective and accurate way possible [36]. The predictions and colour masks made as a result of the training are shown in Fig. 12. Loss graphs for ResNet-50 (5-fold) and ResNet-101 (5-fold) results are given in Figs. 13 and 14, respectively.

The segmentation quality was assessed using the intersection over union (IoU) metric, which was defined in Eq. (11). The IoU, also known as the Jaccard index, is one of the most widely used methods for measuring similarity between finite sample sets (or Jaccard similarity coefficient). IoU is typically expressed as the intersection (AB) divided by the union (AB) for two finite sample sets A and
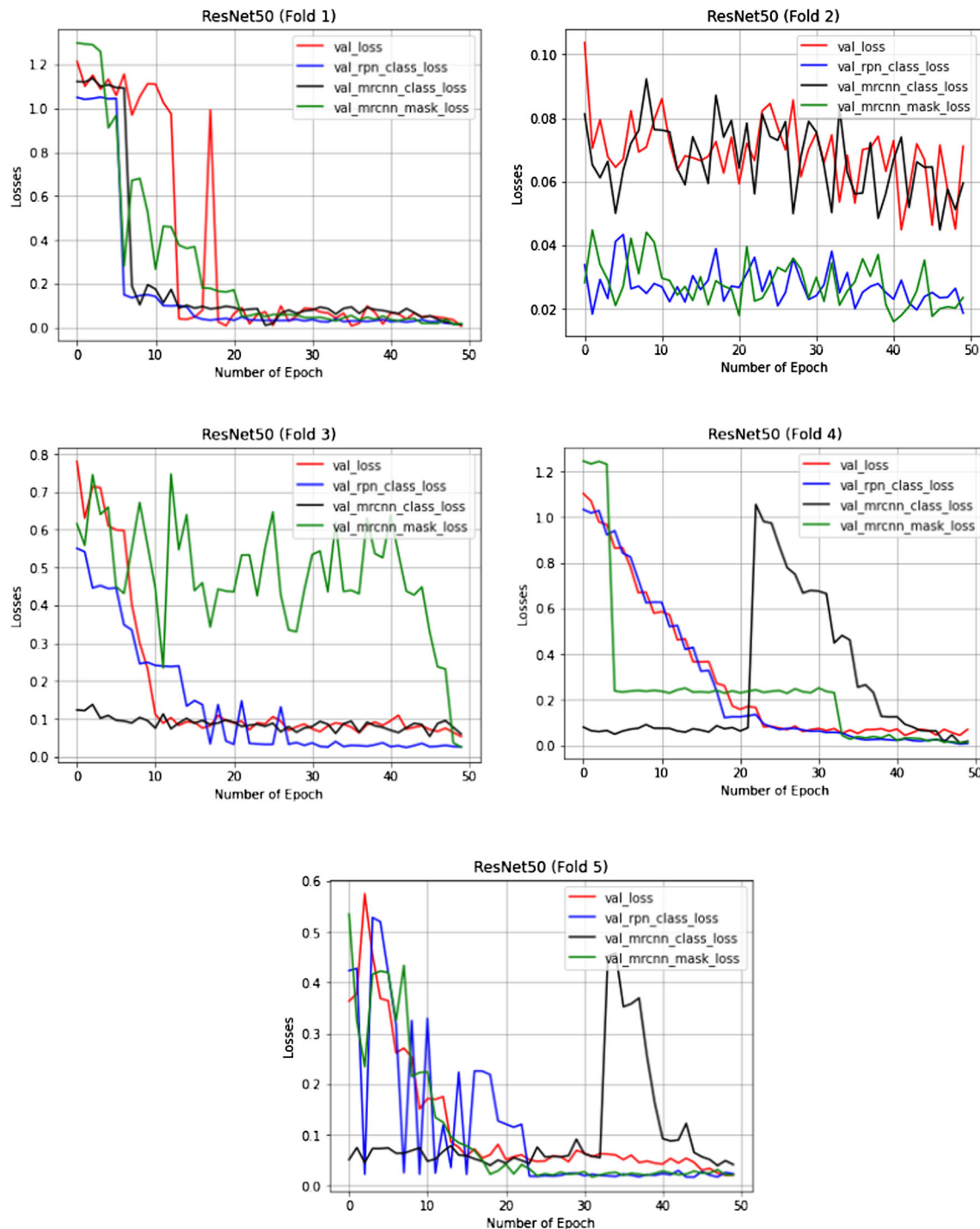
Fig. 13 Loss graphs for ResNet-50 (five fold)

B [37]. In this study, IoU is shown in Fig. 15, and rate of IoU is 0.8694.

$$IoU(A,B) = \frac{A \cap B}{A \cup B} \tag{11}$$

Algorithm losses should be minimized in the mask R-CNN and faster R-CNN training process, and the errors in segmentation should be minimized. As a result of the tests performed by trying different iterations during the training process for faster R-CNN, it is observed that the errors in 4100 iterations could not be reduced any less. For mask R-CNN, the loss rates are minimized after 250 epochs.
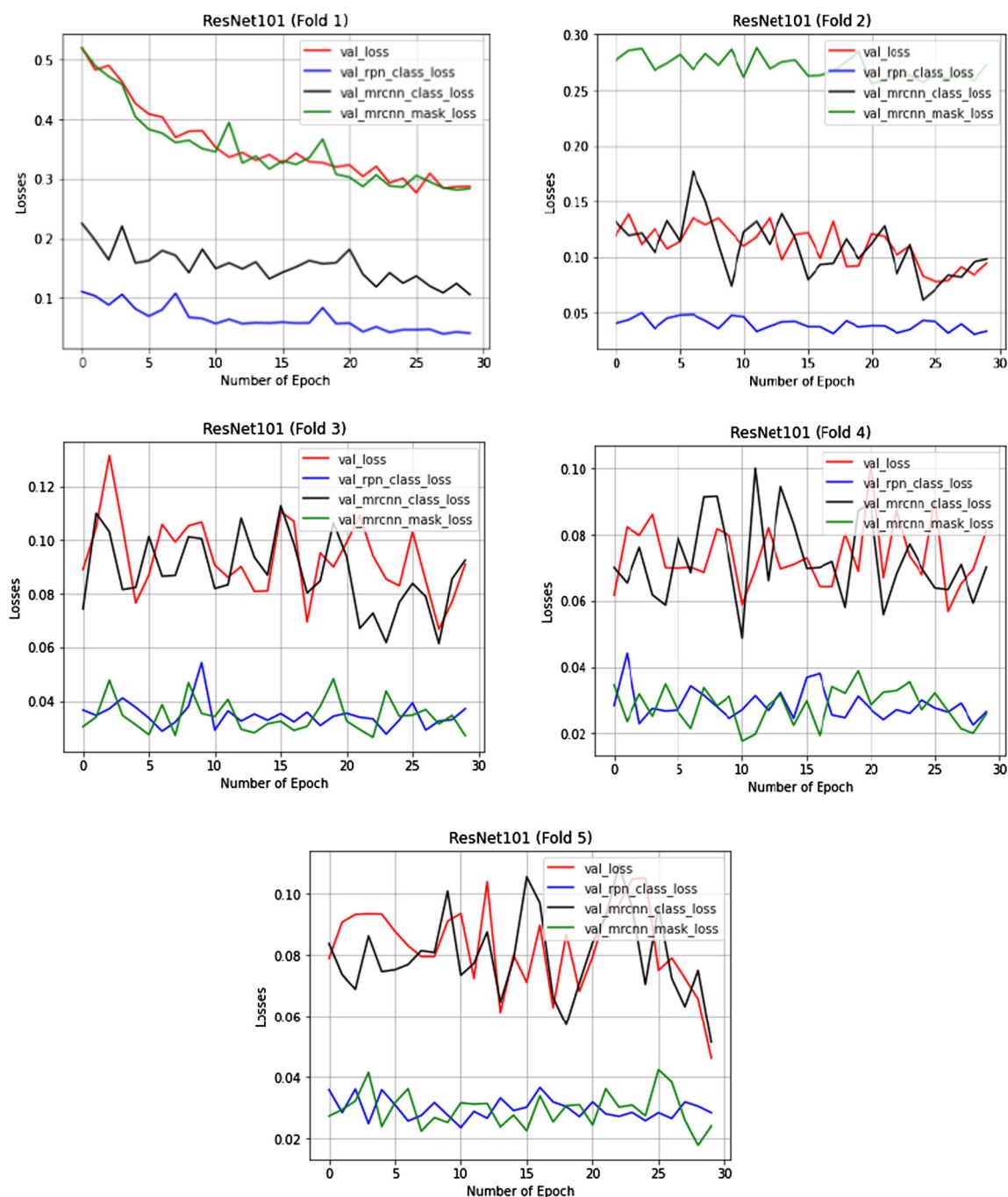
Fig. 14 Loss graphs for ResNet-101 (five fold)

In this study, region-based segmentation of COVID-19 and pneumonia infection in images obtained by computed tomography is performed using the mask R-CNN algorithm based on the CNN architecture, which is a deep learning model. The COVID-19 and pneumonia segmentations obtained from the images are detected with different colours thanks to the sample segmentation feature in the mask R-CNN algorithm, and the error rates are minimized as a result of the tests.

As a result of the training for the mask R-CNN algorithm, 97.72% and 95.65% mAP values are obtained from ResNet-50 and ResNet-101 network backbones, respectively. It has been observed that the model with the ResNet-50 backbone has a slightly lower computational load than the ResNet-101 backbone, and at the same time, the results obtained with the ResNet-50 as a result of the training are more successful. When the faster R-CNN
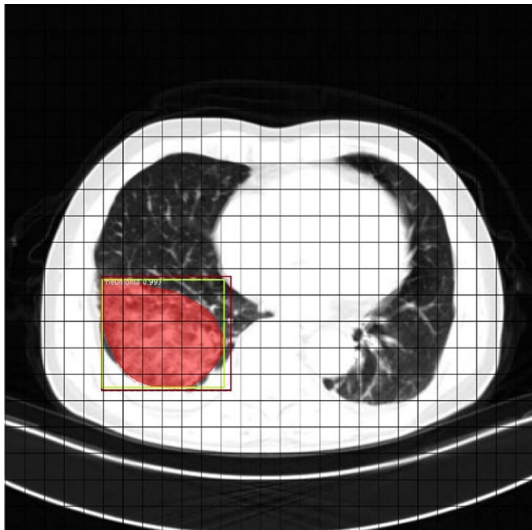
Fig. 15 An example of IoU results from used data

Fig. 16 ROC graph of VGG-16

Table 5 Test results of models

Models	Backbone	Map (%)
Mask R-CNN	ResNet-50	97.72
	ResNet-101	95.65
Faster R-CNN	VGG-16	93.86

Table 7 Comparison with other studies

No	Models	Data	Performance(%) or AP	Study
1	Mask R-CNN	8009	83 mAP	[38]
2	Faster R-CNN	8009	87 mAP	[38]
3	Mask R-CNN	3000	42 AP	[39]
4	Mask R-CNN	3000	35 AP	[17]
5	Faster R-CNN	6012	39.23 mAP	[40]

Table 6 Classification report of VGG-16

	Precision	Recall	F1-score	Support	
COVID-19	92	97	95	400	VGG-16
Pneumonia	97	92	95	400	
accuracy			95	800	
Macro avg	95	95	95	800	
Weighted avg	95	95	95	800	

algorithm is trained, 93.86% class accuracy is reached. Test results of models are given in Table 5.

The dataset used in this study has also been implemented to the classification process with the VGG-16 architecture, and the classification report are given in Table 6. With this model, the accuracy value of 94.87–95% F1-score is reached in this study, and the ROC graph of the model is shown in Fig. 16.

Patients with COVID-19 can be diagnosed using CT scan images. It can identify internal flaws, wounds, part dimensions, tumour, etc. It provides a thorough overview of a specific area that can be detected. A CT scan seems to be a reliable alternative to the current RT-PCR method. It
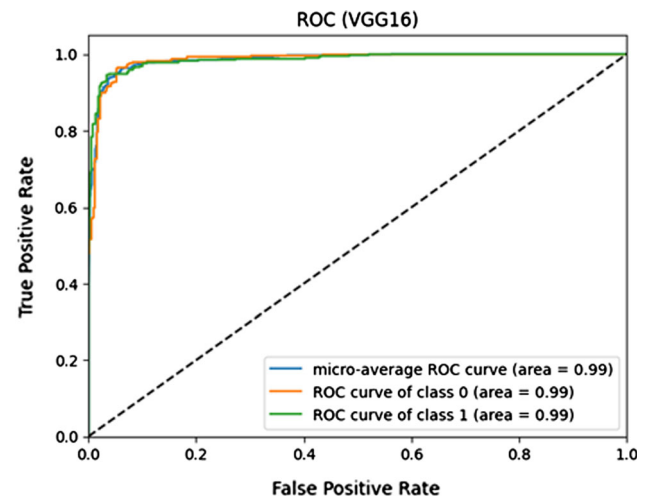
works well for categorizing COVID-19 patient images as well. It guarantees prompt and accurate delivery of results. Early disease or virus treatment is frequently the result of early detection. When the studies on the diagnosis of COVID-19 in the literature are examined, the faster R-CNN and mask R-CNN models are not popular. Therefore, the work carried out is important in one respect. If a comparison is made with the various studies given in Table 7, the results of my proposed models are satisfactory. In actuality, models are not directly comparable due to the variation in sample size and COVID-19 prevalence.

4 Discussion

This study's primary objective is to develop a cost-effective diagnostic method capable of rapidly identifying COVID-19 patients from CT images. Due to the lack of a larger dataset to date, this study was conducted by preparing a dataset containing COVID-19 chest X-ray images and labelling these images. This can be considered a limitation of our study. Plans for the future include validating the proposed framework against a large dataset containing additional COVID-19 chest X-ray images. In

addition, we plan to train our framework on a dataset containing CT images of COVID-19 patients and evaluate the performance of the method when trained on X-ray images. The lungs can be precisely identified using our method without any false alarms. To improve accuracy, numerous experiments are run on high-quality images from the dataset.

5 Conclusion

In this study, a method for automatic detection, segmentation, and classification of COVID-19 lesions with CT images is proposed. This article aims to develop a method for accurately diagnosing COVID-19 using CT images and deep learning techniques. Using CT images obtained from Yozgat Bozok University, the presented method commences with the creation of an initial dataset containing 4000 CT images. In this article, COVID-19 detection is performed for the faster R-CNN model. In this method, 93.86% accuracy rate is obtained by using VGG-16 backbone. In addition to the accuracy score, the classification loss per ROI (region of interest), the classification loss in RPN (Region Proposal Network) and the total training loss are calculated and given in the results part. Mask R-CNN method is trained using the ResNet-50 and ResNet-101 architectures in the training section. For these architectures, mAP values of 97.72% and 95.65% are obtained, respectively. Results for fivefolds of the classification methods used in this study are obtained using the cross-validation method. The article also presents the results by making comparisons with other studies. In this study, a deep learning algorithm based on VGG-16 was used to accurately identify and classify COVID-19 patients. When the performance of this algorithm is analysed, it can be said that the VGG-16 algorithm can extract confidential information from CT scan images to identify COVID-19 patients. The VGG-16 model used has an excellent accuracy of 94.87%, and the classification report and ROC graph are given to the study. One of the current limitations in this field of study is the scarcity of publicly available datasets of COVID-19 CT images and labelled data. In future studies, new architectural solutions, pre-processing algorithms and high-accuracy results can be suggested with a more comprehensive dataset by reducing losses. This study will enable researchers to develop more reliable methods and also to compare methods. The interpretability of artificial intelligence methods should be demonstrated and verified with the help of expert radiologists. It is highly recommended that clinicians, radiologists, and artificial intelligence engineers work together to develop interpretable and reliable deep learning solutions that can be easily implemented in hospitals. Otherwise, despite numerous efforts on a global scale, it will take time for these technologies to be used in hospitals to help humanity.

Declarations

Conflict of interest The authors declare no conflict of interest.

References

1. Sohrabi C, Alsafi Z, O'Neill N, Khan M, Kerwan A, Al-Jabir A, Iosifidis C, Agha R (2020) World Health Organization declares global emergency: a review of the 2019 novel coronavirus (COVID-19). Int J Surg 76:71–76
2. Podder S, Bhattacharjee S, Roy A (2021) An efficient method of detection of COVID-19 using Mask R-CNN on chest X-Ray images. AIMS Biophys 8(3):281–290
3. Shen D, Wu G, Suk HI (2017) Deep learning in medical image analysis. Annu Rev Biomed Eng 19:221
4. Ker J, Wang L, Rao J, Lim T (2017) Deep learning applications in medical image analysis. Ieee Access 6:9375–9389
5. Shah V, Keniya R, Shridharani A, Punjabi M, Shah J, Mehendale N (2021) Diagnosis of COVID-19 using CT scan images and deep learning techniques. Emerg Radiol 28(3):497–505
6. Ouchicha C, Ammor O, Meknassi M (2020) CVDNet: a novel deep learning architecture for detection of coronavirus (COVID-19) from chest x-ray images. Chaos, Solitons Fractals 140:110245
7. Singh D, Kumar V, Kaur M (2020) Classification of COVID-19 patients from chest CT images using multi-objective differential evolution–based convolutional neural networks. Eur J Clin Microbiol Infect Dis 39(7):1379–1389
8. Minaee S, Kafieh R, Sonka M, Yazdani S, Soufi GJ (2020) Deep-COVID: predicting COVID-19 from chest X-ray images using deep transfer learning. Med Image Anal 65:101794
9. Ardakani AA, Kanafi AR, Acharya UR, Khadem N, Mohammadi A (2020) Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: results of 10 convolutional neural networks. Comput Biol Med 121:103795
10. Şahin ME, Ulutas H, E., Yuce. (2021) A deep learning approach for detecting pneumonia in chest X-rays. Avrupa Bilim ve Teknoloji Dergisi 28:562–567
11. Fan X, Feng X, Dong Y, Hou H (2022) COVID-19 CT image recognition algorithm based on transformer and CNN. Displays 72:102150
12. Pathak Y, Shukla PK, Tiwari A, Stalin S, Singh S (2020) Deep transfer learning based classification model for COVID-19 disease. Irbm 43(2):87–92
13. Sahin ME (2022) Deep learning-based approach for detecting COVID-19 in chest X-rays. Biomed Signal Process Control 78:103977
14. Jaiswal A, Gianchandani N, Singh D, Kumar V, Kaur M (2021) Classification of the COVID-19 infected patients using DenseNet201 based deep transfer learning. J Biomol Struct Dyn 39(15):5682–5689

15. Kc K, Yin Z, Wu M, Wu Z (2021) Evaluation of deep learning-based approaches for COVID-19 classification based on chest X-ray images. SIViP 15(5):959–966

16. Karhan Z, and Fuat AKAL (2020). COVID-19 classification using deep learning in chest X-ray images. In 2020 medical technologies congress (TIPTEKNO) pp 1–4. IEEE

17. Ter-Sarkisov A (2022) COVID-ct-mask-net: Prediction of COVID-19 from ct scans using regional features. Appl Intell 52(9):9664–9675

18. Zheng C, Deng X, Qiang F, Zhou Q, Feng J, Ma H, Liu W, Wang X (2020) Deep learning-based detection for COVID-19 from chest CT using weak label. MedRxiv 395(10223):497

19. Ucar F, Korkmaz D (2020) COVIDiagnosis-Net: deep Bayes-SqueezeNet based diagnosis of the coronavirus disease 2019 (COVID-19) from X-ray images. Med Hypotheses 140:109761

20. Ozyurt F, Tuncer T, Subasi A (2021) An automated COVID-19 detection based on fused dynamic exemplar pyramid feature extraction and hybrid feature selection using deep learning. Comput Biol Med 132:104356

21. He K, Gkioxari G, Dollár P, and Girshick R. (2017). Mask r-cnn. In proceedings of the IEEE international conference on computer vision, pp. 2961–2969

22. Lin K, Zhao H, Lv J, Li C, Liu X, Chen R, Zhao R (2020) 2020. Discrete dynamics in nature and society, Face detection and segmentation based on improved mask R-CNN

23. Liu M, Dong J, Dong X, Yu H, and Qi L. (2018). Segmentation of lung nodule in CT images based on mask R-CNN. In 2018 9th international conference on awareness science and technology (iCAST) pp. 1–6. IEEE.

24. Mulay S, Deepika G, Jeevakala S, Ram K, Sivaprakasam M (2019) Liver segmentation from multimodal images using HED-mask R-CNN. International workshop on multiscale multimodal medical imaging. Springer, Cham, pp 68–75

25. Shu JH, Nian FD, Yu MH, Li X (2020) An improved mask R-CNN model for multiorgan segmentation. Math Probl Engi 2020:1–11

26. Nguyen DH, Le TH, Tran TH, Vu H, Le TL, and Doan HG. (2018). Hand segmentation under different viewpoints by combination of Mask R-CNN with tracking. In 2018 5th Asian conference on defense technology (ACDT), pp 14–20. IEEE.

27. Shibata T, Teramoto A, Yamada H, Ohmiya N, Saito K, Fujita H (2020) Automated detection and segmentation of early gastric cancer from endoscopic images using mask R-CNN. Appl Sci 10(11):3842

28. Cao G, Song W, and Zhao Z. (2019). Gastric cancer diagnosis with mask R-CNN. In 2019 11th international conference on intelligent human-machine systems and cybernetics (IHMSC) Vol 1, pp 60–63. IEEE

29. Chiao JY, Chen KY, Liao KYK, Hsieh PH, Zhang G, Huang TC (2019) Detection and classification the breast tumors using mask R-CNN on sonograms. Medicine 98(19):e15200

30. Zaidi SSA, Ansari MS, Aslam A, Kanwal N, Asghar M, Lee B (2022) A survey of modern deep learning based object detection models. Digital Signal Process 126:103514

31. Girshick R, Donahue J, Darrell T, and Malik J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp 580–587

32. He K, Zhang X, Ren S, Sun J (2015) Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Trans Pattern Anal Mach Intell 37(9):1904–1916

33. Girshick R. (2015). Fast r-cnn. In Proceedings of the IEEE international conference on computer vision, pp 1440–1448

34. Ren S, He K, Girshick R, Sun J (2015) Faster r-cnn: Towards real-time object detection with region proposal networks. Adv Neural Inf Process Syst. https://doi.org/10.1109/TPAMI.2016.2577031

35. Waleed A. 2017. Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow; 2017. Avaiblable at https://github.com/matterport/Mask_RCNN.

36. Berrar D (2019) Cross-Validation Call for Papers for Machine Learning journal: Machine Learning for Soccer View project Cross-validation. Elsevier, Amsterdam, pp 542–545

37. Zhou D, Fang J, Song X, Guan C, Yin J, Dai Y, and Yang, R. (2019). Iou loss for 2d/3d object detection. In 2019 international conference on 3D vision (3DV), pp 85–94. IEEE.

38. Tahir H, Iftikhar A, and Mumraiz M. (2021). Forecasting COVID-19 via registration slips of patients using resnet-101 and performance analysis and comparison of prediction for COVID-19 using faster r-cnn, mask r-cnn, and resnet-50. In 2021 international conference on advances in electrical, computing, communication and sustainable technologies (ICAECT), pp 1–6. IEEE

39. Ter-Sarkisov A (2020) Single-shot lightweight model for the detection of lesions and the prediction of COVID-19 from chest CT scans. Medrxiv 39(8):2638

40. Yao S, Chen Y, Tian X, Jiang R (2021) Pneumonia detection using an improved algorithm based on faster r-cnn. Comput Math Methods Med 2021:1–13