

# Efficient mask region-based convolutional neural network-based architecture for COVID-19 detection from computed tomography data

Nader Mahmoud<sup>1,2</sup>, Ashraf B. El-Sisi<sup>1</sup>

<sup>1</sup>Computer Science Department, Faculty of Computers and Information, Menoufia University, Shebin El-Kom, Egypt

<sup>2</sup>Cybersecurity Department, Engineering and Information Technology College, Buraydah Private Colleges, Buraydah, Kingdom of Saudi Arabia

## Article Info

### Article history:

Received Jan 14, 2025

Revised Jun 5, 2025

Accepted Jun 30, 2025

### Keywords:

Chest computed tomography  
COVID-19

Deep learning

Fuzzy color technique

Mask region-based  
convolutional neural network

## ABSTRACT

The worldwide effect of the coronavirus disease (COVID-19) pandemic has been catastrophic, leading to a significant number of fatalities worldwide. In response to the outbreak, health care institutions have proposed the use of chest computed tomography (CT) as an important diagnosis tool for rapid diagnosis, leveraging deep learning approaches for disease detection. This paper aims to progress a robust methodology towards accurate diagnosis of COVID-19 based on deep learning approaches with chest CT images. We propose a mask region-based convolutional neural network (Mask R-CNN) model architecture that is well-trained and used to discriminate between COVID-19-infected and uninfected cases. In order to improve feature extraction, the proposed model incorporates a fuzzy color enhancement preprocessing technique that reduces image fuzziness and increases contrast. A publicly available chest CT dataset is considered for quantitative evaluation of the proposed architecture model, which includes various frontal image views of COVID-19 and non-COVID-19 cases. The proposed approach yielded an accuracy of 98.8% with 98.4% precision and 98.5% recall. Additionally, the proposed model architecture has been quantitatively evaluated in comparison with benchmark approaches, yielding superior performance in terms of conventional evaluation metrics.

This is an open access article under the [CC BY-SA](#) license.



## Corresponding Author:

Nader Mahmoud

Computer Science Department, Faculty of Computers and Information, Menoufia University

Shebin El-Kom, Egypt

Email: nader.mahmoud@ci.menofia.edu.eg

## 1. INTRODUCTION

The coronavirus disease (COVID-19) rising in Wuhan, China, 2019, has quickly evolved into a global health crisis. The virus spreads primarily through respiratory droplets, thus accurate and efficient diagnosis is vital in preventing transmission and managing patient care. The reverse transcription polymerase chain reaction (RT-PCR) was the principle diagnostic technique that has been used globally in order to detect viral ribonucleic acid (RNA) from nasopharyngeal swabs [1]. While RT-PCR is specific and widely adopted, it suffers from a number of limitations, such as long processing times, high dependency on specialized reagents and equipment, and moderate sensitivity rates, with studies reporting sensitivities as low as 71% [2]. Computed tomography (CT) imaging is a well-known medical imaging modality that allows for non-invasive visualization of internal body structures and is widely utilized in a variety of applications [3]. CT scans have demonstrated significant sensitivity when considered for detecting COVID-19-related pulmonary

abnormalities—reported to be as high as 98% [4]. CT scans are capable of exploiting characteristic features such as ground-glass opacities and other lung irregularities, even with patients who are asymptomatic or have negative RT-PCR. As a result, several healthcare institutions have adopted CT imaging as a complementary or alternative diagnostic tool, particularly during early-stage infections or in scenarios requiring rapid patient triage [5].

Despite its advantages, accurate interpretation of chest CT images is still very difficult, where diagnostic indicators can vary widely depending on the stage of infection and whether comorbidities are present. Moreover, manual interpretation is time-consuming and also subject to inter-observer variability that typically increases and becomes an issue in case of high patient volumes during the pandemic. Due to these limitations, there is increasing interest in using the application of artificial intelligence and deep learning to medical imaging, in order to provide fast, consistent, and scalable diagnostic assistance [6]. Several deep learning models have been researched recently for COVID-19 detection and segmentation tasks from chest X-ray and CT medical images. Preliminary research works have achieved promising classification results leveraging convolutional neural networks (CNNs) medical imaging datasets [7], [8]. However, the clinical relevance of these models was limited due to their lack of spatial awareness and inability to locate diseased tissues. Segmentation based frameworks, including U-Net, were proposed to tackle this problem, which allows for pixel-level detection and identification of infection regions. Despite the effectiveness of these models, they still struggle with generalizing to images with variable quality or subtle signs of disease.

Different advanced segmentation techniques have been researched; however, mask region-based convolutional neural network (Mask R-CNN) [9] has emerged as an efficient and powerful approach, which uniquely combines object detection and semantic segmentation in a unified framework [10], [11]. It has been studied and shown effective in a range of medical imaging tasks, including lung nodule detection [12], liver and multi-organ segmentation [13], breast tumor classification [14], and early cancer diagnosis [15]. Mask R-CNN is well-suited for identifying COVID-19-infected regions due to its ability to deliver accurate, instance-level, pixel-wise predictions. However, many of its current applications overlook the importance of preprocessing, especially in cases where low image contrast makes infected regions harder to identify.

To tackle these drawbacks, we propose a novel framework that brings together a Mask R-CNN-based segmentation model with fuzzy logic-based contrast enhancement. Fuzzy logic is well-suited to handle two prominent challenges in medical image analysis, which are uncertainty and ambiguity. The proposed approach transforms CT images into a fuzzy domain using adaptive fuzzifiers to identify crossover points, which is then followed by a contrast enhancement operator that amplifies high-intensity regions (e.g., lesions) while suppressing background noise. This preprocessing step significantly improves the quality of input images and enhances the segmentation accuracy of the proposed model. This paper makes the following key contributions:

- a. Introduce a fuzzy logic-based image contrast enhancement technique designed to highlight low-contrast COVID-19 infection regions in CT images of chest.
- b. Developing an enhanced Mask R-CNN segmentation framework that effectively leverages the contrast-enhanced input images, and allows for precise pixel-level infection detection.
- c. Extensive experimental validation on a public COVID-19 CT dataset, showing improved performance over baseline and state-of-the-art models in terms of standard performance metrics.

The remaining of the paper is organized as follows: section 2 provides an overview of related work in COVID-19 detection and segmentation from CT scans, along with a critical analysis of recent approaches. Section 3 details the proposed methodology, including the fuzzy enhancement process and the Mask R-CNN architecture. The experimental setup, performance metrics, and quantitative model evaluation on a public dataset are provided in section 4. The conclusion and potential future directions are discussed in section 5.

## 2. RELATED WORK

We present a survey of research work focused on deep learning-based detection. Our emphasis is on approaches that have significantly contributed to diagnostic tasks from CT and X-ray imaging modalities, and we highlight their methodologies, findings, and the remaining challenges our work aims to address. Jin *et al.* [16] developed a comprehensive deep learning system that performed lung segmentation and localized infectious slices for COVID-19 diagnosis. Despite, their work has achieved encouraging results, it relied on manually constructed pipelines for segmentation, which limited end-to-end automation. Inspired by the VGG architecture, Hu *et al.* [17] introduced a weakly supervised multiscale deep learning framework that effectively assimilates multi-scale lesion features. However, their approach did not leverage pixel-level segmentation capabilities and exhibited limitations in handling high inter-class similarities.

Polsinelli *et al.* [18] presented a SqueezeNet CNN architecture, in order to provide rapid inference for COVID-19 diagnosis from CT scans. However, their approach was efficient in terms of processing time, but because of the simplified architecture of the model, the level of accuracy suffered. On the other hand,

Biswas *et al.* [19] investigated an ensemble strategy combining successful network architectures including VGG-16, ResNet50 [20], and Xception [21]. Their model achieved remarkable generalization capabilities by utilizing transfer learning, however, the ensemble's computational demands were significantly higher, thus limiting its practicality in real-world clinical procedures. In contrast, Zhao *et al.* [22] have researched a different route through modifying the ResNet-v2 architecture. In order to enhance training reliability and overall model performance, they implemented weight standardization and replaced batch normalization with group normalization. Although this enhanced the robustness of their proposed system, it lacked explicit region-level segmentation, which is essential for medical diagnostic interpretability.

Significant advancements have been achieved in object detection frameworks in addition to classification focused research, beginning with R-CNN [23] and progressing through Fast R-CNN [24] and Faster R-CNN [25]. These models introduced region proposal mechanisms and bounding box prediction, which significantly improved detection accuracy. Region proposal networks (RPNs) were incorporated into Faster R-CNN, which resulted in significant speed and accuracy improvements. Due to robustness of Faster R-CNN, it served as the foundation for numerous subsequent enhancements [26] and remains top performer in several benchmarks. Recently, Mask R-CNN [9] has been presented, which extending the work of Faster R-CNN for both object detection and instant segmentation tasks in computer vision. The main contribution of Mask R-CNN is in its capability to perform pixel-wise segmentation in addition to object detection. The addition of an extra "mask head" branch improves precise segmentation masks for individual objects. The model also introduces two significant architectural which are alignments of regions of interest (RoI) corrected spatial misalignment and feature pyramid networks. RoI alignment tackles limitations of traditional RoI pooling through bilinear interpolation during the pooling phase. This leads to improved segmentation accuracy, especially when dealing with small object instances.

Despite these advancements, many existing research works either do not support fine-grained segmentation or struggle with effective CT image preprocessing for handling low contrast and image fuzziness. Additionally, a number of research works rely on relatively small datasets, which limits the generalizability of their results. Consequently, this paper introduces an efficient Mask R-CNN-based architecture tailored for COVID-19 detection in chest CT scans that effectively leverages the contrast-enhanced input images. A key novelty of the proposed architecture is the introduction of a fuzzy color enhancement preprocessing technique that significantly improves image clarity and contrast. We validate our method on a large, publicly available dataset and benchmark its performance against several state-of-the-art models, demonstrating clear advantages across standard evaluation metrics.

### 3. PROPOSED MODEL ARCHITECTURE

Figure 1 depicts the main steps of the proposed model architecture.

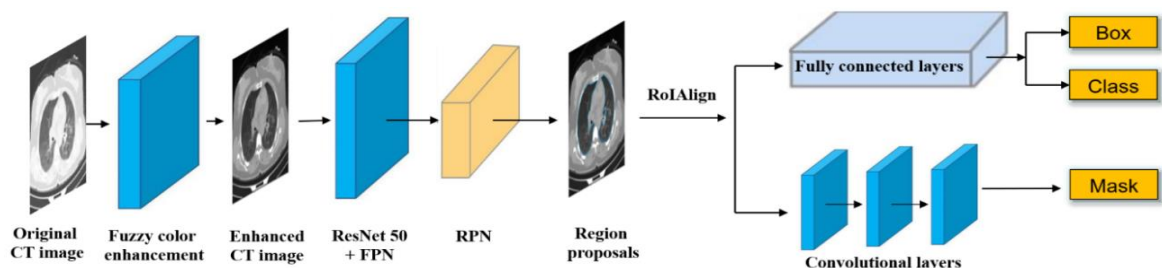


Figure 1. Proposed model architecture workflow

#### 3.1. Image preprocessing using fuzzy color enhancement

To address the challenges of CT images, such as low contrast, particularly in early-stage COVID-19 infections, we incorporated a fuzzy logic-based image enhancement technique [27] as a preprocessing step, detailed in Algorithm 1. This technique allows for significant improvement of brightness and contrast across CT images, therefore enabling more accurate segmentation in the subsequent Mask R-CNN stage. The fuzzy enhancement process begins by decomposing the image into overlapping fuzzy regions, each representing a distinct object that captures varying levels of uncertainty in brightness and contrast. Instead of assigning hard boundaries, each pixel is given a membership value for each region, based on how far away it is from the center of the region. These membership values influence the pixel's contribution during variance calculations, allowing the algorithm to better model uncertainty and reveal subtle structural details

that are critical for early diagnosis. Algorithm 1 details the enhancement process, which proceeds as follows:

- For each pixel, the mean and variance within its fuzzy neighborhood are computed, weighted by the membership degrees.
- These localized statistics are used to adjust pixel intensity, emphasizing regions of interest and suppressing background noise.
- The enhanced image is reconstructed by aggregating the results from all fuzzy windows, producing a globally consistent but locally adaptive contrast enhancement.

#### Algorithm 1. Fuzzy color-based image enhancement

**Input:** RGB image  $I$ , Window size  $w$ , Fuzzifier parameters  $\alpha, \beta$

**Output:** Enhanced image  $I_{\text{enhanced}}$

```

1. for each color channel  $c \in \{R, G, B\}$ , do:
2.    $I_c \leftarrow$  extract the channel  $c$  from  $I$ .
3.    $I_c^{\text{enhanced}} \leftarrow$  zero matrix of same size as  $I_c$ .
4.   for each pixel  $p$  in  $I_c$  do:
5.      $W_p \leftarrow$  extract  $w \times w$  window centered at  $p$ .
6.     Initialize  $\text{total\_weight} \leftarrow 0$ ,  $\text{weighted\_sum} \leftarrow 0$ 
7.     for each pixel  $q \in W_p$ , do:
8.        $d \leftarrow \text{EuclideanDistance}(p, q)$ .
9.        $\mu \leftarrow \exp(-\alpha \cdot d^\beta)$ 
10.      {Membership degree}
11.       $\text{total\_weight} \leftarrow \text{total\_weight} + \mu$ .
12.       $\text{weighted\_sum} \leftarrow \text{weighted\_sum} + \mu \cdot I_c(q)$ .
13.    End for
14.     $\text{mean} \leftarrow \frac{\text{weighted\_sum}}{\text{total\_weight}}$ 
15.     $\text{weighted\_variance} \leftarrow 0$ 
16.    for each pixel  $q \in W_p$ , do:
17.       $d \leftarrow \text{EuclideanDistance}(p, q)$ .
18.       $\mu \leftarrow \exp(-\alpha \cdot d^\beta)$ .
19.       $\text{weighted\_variance} \leftarrow \text{weighted\_variance} + \mu \cdot (I_c(q) - \text{mean})^2$ .
20.    End for
21.     $\text{variance} = \frac{\text{weighted\_variance}}{\text{total\_weight}}$ 
22.     $I_c^{\text{enhanced}}(p) = \text{Enhance}(I_c(p), \text{mean}, \text{variance})$ 
23.  End for
24.  Store  $I_c^{\text{enhanced}}$  in output channels
25. End for
26.  $I_{\text{enhanced}} = \text{Merge}(I_R^{\text{enhanced}}, I_G^{\text{enhanced}}, I_B^{\text{enhanced}})$ 
Return  $I_{\text{enhanced}}$ 

```

In this paper, the algorithm was applied separately to each of the three color channels (RGB) of the CT image. Though CT images are often greyscale, some datasets store them in three-channel format; thus, our method handles each channel in parallel and merges the outputs to reconstruct the final image. In contrast to conventional techniques for enhancing image contrast, including histogram equalization and its adaptive variant, CLAHE, fuzzy enhancement offers two main advantages. Firstly, it efficiently models uncertainty, which is essential for medical images, where pathological characteristics are frequently subtle. Secondly, it adjusts locally, enhancing visibility without introducing any artifacts or noise. Figure 2(a) showing improved contrast in lung regions potentially affected by infection. Figure 2(b) illustrates the effect of fuzzy color enhancement on a sample CT scan,

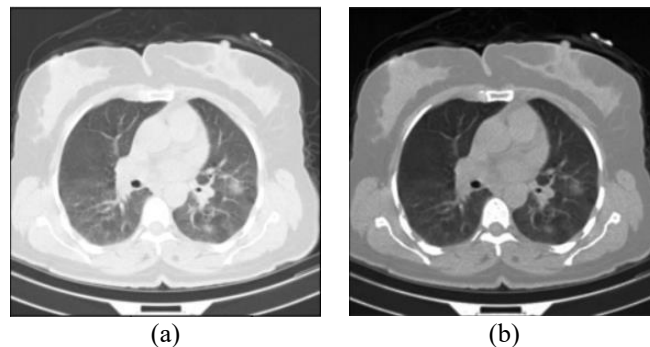


Figure 2. Fuzzy CT image enhancement (a) original CT image and (b) fuzzy colored enhanced image

### 3.2. Mask R-CNN-based segmentation and classification

In the second step of our pipeline, we apply a Mask R-CNN framework outlined in Algorithm 2 to detect and segment infected regions from preprocessed CT images, as shown in Figure 1. Mask R-CNN adds on top of Faster R-CNN a parallel branch for predicting pixel-wise segmentation masks, allowing simultaneous object detection and instance segmentation. The architecture consists of three core components, as described in Algorithm 2:

- a. **Backbone feature extractor (ResNet50 + FPN):** We utilize a ResNet50 model as the backbone for feature extraction due to its ability to balance between depth and computational efficiency. ResNet50 is composed of 48 convolutional layers, along with a max-pooling layer and a global average pooling layer. Aiming to enhance feature representation across multiple scales, we incorporate a feature pyramid network (FPN) [25] on top of ResNet50. FPN constructs a feature pyramid that captures both low-level and high-level features, which enhances the ability to detect lesions in CT images of different sizes and scales. COVID-19 lesions, e.g. ground-glass opacities, can appear at different sizes and intensities, sometimes scattered in the lungs. By combining high and low resolution features, FPN aids in the detection of both larger abnormalities and tiny early-stage lesions. This is essential for reliable detection across disease progression.
- b. **Region proposal network (RPN):** The RPN generate region proposals through applying a small network over the feature maps generated by the FPN. The coordinates of each anchor box are fine-tuned and given an objectness score. By reducing the number of possible areas where infection might exist, the RPN helps to improve the efficiency of the subsequent classification and segmentation.
- c. **Region of interest alignment (RoIAlign):** To accurately map the region proposals onto the feature map, we employ RoIAlign [28]. This method outperforms conventional RoI pooling by employing bilinear interpolation instead of coarse quantization. This ensures spatial alignment is preserved, which is especially critical for medical imaging where pixel-level accuracy is crucial. The RoIAlign process involves the steps detailed Algorithm 2. Inputs are feature maps and region proposals generated by FPN and RPN, respectively. Region proposals are subdivided into equal-sized grids to extract features from the matching regions in the input feature map. The aligned features from these grids represent the characteristics of each proposal.

After acquiring the spatially aligned features, a fully convolutional network [9] is used to generate binary masks for each proposed region. In parallel, a classification branch classifies each region into one of the predefined categories (e.g., infected vs. non-infected). The classification pipeline begins with convolutional layers followed by fully connected layers which encode spatial and contextual information from the RoI-aligned maps. Combining these elements allows the proposed model to accurately segment COVID-19 lesions from CT images in an end-to-end manner, classify them, and enable localization. This capability addresses the limitation of earlier CNN-based models, which lacked spatial localization and relied on whole-image classification, therefore reducing clinical interpretability.

#### Algorithm 2. COVID-19 lesion detection using mask R-CNN

**Input:** Preprocessed CT image  $I$

**Output:** Region-level classifications and segmentation masks

```

1. Extract feature maps from  $I$  using ResNet50 backbone
2. Construct multi-scale feature maps using Feature Pyramid Network (FPN)
3. Generate region proposals  $\mathcal{R}$  using RPN on FPN features
   for each feature map level  $P_i$  do:
       Slide a small  $3 \times 3$  window over  $P_i$ 
       for each window position:
           a. Generate anchor boxes with multiple scales and aspirations
           b. Predict the objectness score for each anchor (foreground/background)
           c. Regress bounding box offsets for anchor refinement
       End for
   End for
4. For each region proposal  $r \in \mathcal{R}$  do
   a. Apply RoIAlign on  $r$  to extract a fixed-size feature map
   b. Classify the region  $r$  using a CNN classifier
   c. Generate a binary segmentation mask using a parallel FCN mask branch
   End for
5. Aggregate classification scores and masks for final prediction
Return Predicted region classes and corresponding segmentation masks

```

## 4. RESULTS AND DISCUSSION

In this section, we provide a thorough description of the experimental setup, including details of the dataset used, the employed evaluation metrics, and the methodology followed for performance evaluation.

Furthermore, we present a rigorous quantitative evaluation of the proposed model architecture, comparing its results against established benchmark models to demonstrate its effectiveness and superiority.

#### 4.1. Experimental setup and dataset

The proposed model architecture was carried out and evaluated on a PC with an Intel Core i5 7th generation processor, 8 GB of RAM, a GTX 1080 GPU with 4 GB of RAM, and running a Windows operating system. We utilized the publicly available COVIDx-CT dataset [29], [30] for the evaluation, which consists of a 104,009 chest CT slice collected from 1,489 patient cases. This dataset aggregates chest CT examinations from multiple hospital cohorts across China as part of the China Consortium of Chest CT Image Investigation (CC-CI). The imaging data include chest CT volumes representing three categories: COVID-19 pneumonia, common pneumonia resulting from non-COVID-19 infections, and normal (uninfected) cases.

Figure 3 shows image samples for each of the three dataset categories: Figure 3(a) shows a COVID-19 case, Figure 3(b) shows a non-COVID-19 pneumonia case, and Figure 3(c) shows a normal case from a healthy individual. The COVIDx-CT dataset is chosen due to its high quality, well-organized structure, and readiness for direct use in benchmarking evaluation. Its standardized format and comprehensive coverage of infection types make it particularly suitable for consistent evaluation and comparison of deep learning model performance.

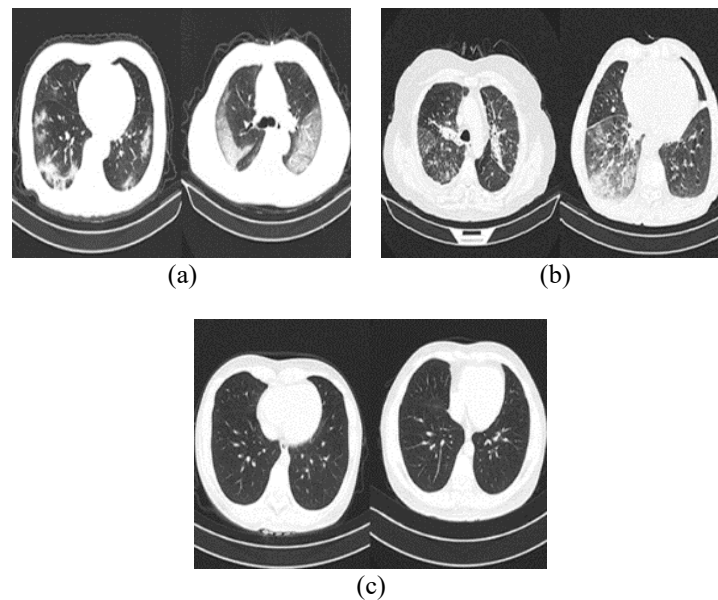


Figure 3. Samples of COVIDx-CT dataset (a) COVID-19 and pneumonia case, (b) non-COVID-19 pneumonia case, and (c) shows a normal case from a healthy individual

We followed the distribution proposed in [29], thus, we adopted a data split strategy that divides the COVIDx-CT dataset into 60% training–20% test–20% validation, respectively. To avoid data leakage and maintain the integrity of the evaluation, care was taken to ensure that CT scans from a single patient were allocated just to one of the sets. The resulting distribution is summarized in Table 1 and visually represented in Figure 4. This splitting approach consistently produces the most consistent and accurate results performance across all evaluated benchmarking models. To further ensure fairness and minimize selection bias, the data split was randomized for each experimental run while maintaining the overall distribution proportions. This strategy contributed to a robust and unbiased evaluation of the proposed architecture performance.

Table 1. COVIDx-CT dataset distribution

Type	Normal	Pneumonia	COVID-19	Total
Train	28202	21568	12635	62405
Validation	8658	7251	4893	20802
Test	8658	7251	4893	20802

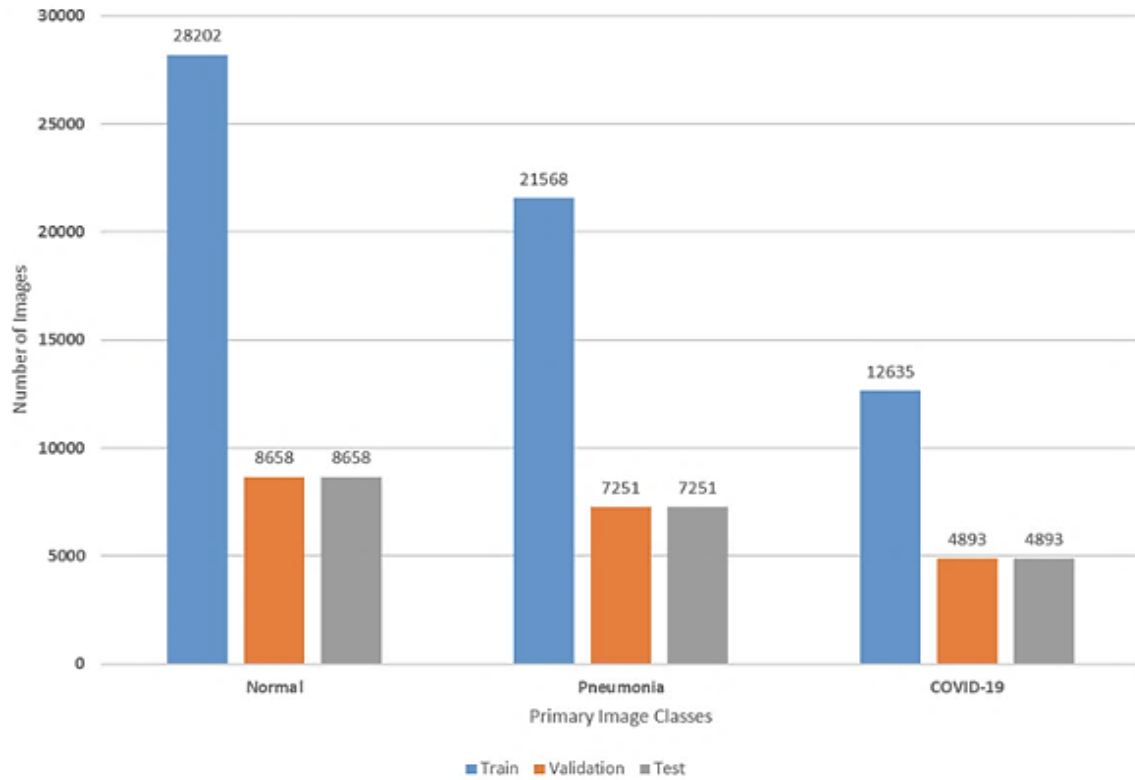


Figure 4. Dataset distribution

#### 4.2. Evaluation metrics

To evaluate the performance of the proposed model architecture, we employ standard evaluation metrics, including Precision, Recall, F1-score, and Accuracy:

- Precision: The proportion of true positive COVID-19 cases predicted by the model to the total number of model predictions (*i.e.*, true positives (TP) and false positives (FP)). This metric provides insights about the rate of false positives. The lower the number of false positives, the higher the yield.

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

- Recall: This is the sensitivity of the model. It is the ratio of predicted TP to the total number of actual positive instances, which includes true positives and false negatives.

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

- F1-score: Takes into account both false positives and false negatives by averaging the precision and recall metrics. It is important in cases where the distribution of classes is unequal.

$$F1 - Score = \frac{2 * Precision * Recall}{Precision + Recall} \quad (3)$$

- Accuracy: This is the most often used and straightforward categorization metric. It is calculated as the number of correct predictions divided by the number of samples. Although high accuracy is typically desirable, in some cases where the class distribution is not symmetric, it may not be an informative evaluation. In these scenarios, precision, recall, and F1-score offer a more thorough assessment of model performance.

$$Accuracy = \frac{TP+TN}{Total\ samples} \quad (4)$$

### 4.3. Performance evaluations and benchmarking

The proposed model architecture has undergone training for 100 epochs. The performance metrics of the proposed architecture are quantitatively summarized in Table 2 with benchmarking models. We have considered various benchmark models that use different deep learning models as their backbone. We are mainly interested in benchmark models that apply to CT images. Most of the benchmark approaches have evaluated their architectures on small datasets; in contrast, we have performed this comparison evaluation on larger datasets [29], [30]. We used the same distribution of datasets into train, validation, and test as discussed in section 4.2 for all benchmark approaches.

The results in Table 2 clearly show that the proposed Mask R-CNN model significantly outperforms all benchmark architectures. The yielded gains in accuracy, recall, and F1-score point to the proposed model's adaptability in COVID-19 detection with minimum false positives and false negatives. That is particularly vital in clinical procedures where early and accurate diagnosis can significantly affect treatment planning and isolation protocols. The ability to distinguish between COVID-19 and other pneumonias further enhances its utility in real-world applications. Unlike previous models which are typically trained on limited data, our proposed model shows consistent performance over a large dataset, reflecting its generalizability. The superior model performance in detecting pathogenic variations in CT images comes from the integration of FPN and fuzzy preprocessing, which ensure better spatial representation and well-contrasted image inputs.

Additionally, to assess the contribution of key architectural components, we conducted an ablation study, as presented in Table 3. Specifically, we evaluated the effects of fuzzy color enhancement, FPN, and the RoIAlign process. For comparison, we used a baseline model based on ResNet [20], which excludes these enhancements. The baseline achieved a precision of 93.1% and an F1-score of 91.4%, whereas our complete model attained 98.4% precision and a 97.4% F1-score. The gained performance improvement emphasizes the importance of every incorporated component. Hence, fuzzy preprocessing enhances contrast and structural visibility in CT scans. FPN enables effective multi-scale feature extraction, and RoIAlign ensures accurate spatial alignment of features during segmentation. These developments taken together help to create a more robust and clinically relevant COVID-19 detection system.

Table 4 evaluates the proposed model architecture in the case of using different ResNet variants, such as ResNet41, ResNet50, and ResNet101 as our backbone networks. According to the comparative evaluation, ResNet50 offers the best balance between performance and computational efficiency. It achieves high scores across all evaluation metrics, with a precision of 98.43%, recall of 98.51%, F1-score of 97.46%, and accuracy of 98.82%. On one hand, ResNet41 enables faster processing times at the expense of accuracy and recall, where the number of convolutional layers is reduced. On the other hand, ResNet101 does not bring too many improvements in terms of accuracy compared to ResNet50, despite its network depth defined by the quantity of convolutional layers and longer processing time. The marginal gain in accuracy (only 0.12% over ResNet50) does not justify the additional cost in most practical scenarios. Therefore, ResNet50 emerges as the most suitable backbone for our architecture, offering an optimal trade-off between speed, model size, and detection performance, making it well-suited for deployment in real-world clinical settings.

Table 2. Comparison of the proposed model architecture and benchmark architecture

Model	Precision	Recall	F1	Accuracy
VGG 16-based model [7]	93.20%	94.10%	93.60%	94.10%
Bayesian-based model [31]	92.40%	92.80%	92.10%	92.50%
DensNet-based model [32]	93.30%	94.10%	92.50%	93.70%
ResNet-based model [8]	93.80%	93.40%	93.60%	93.90%
<b>Proposed model architecture</b>	<b>98.40%</b>	<b>98.50%</b>	<b>97.40%</b>	<b>98.80%</b>

Table 3. Ablation study: performance comparison with and without key enhancements

Architecture	Precision	Recall	F1	Accuracy
Baseline ResNet [20]	93.1%	93.6%	91.4%	93.6%
Proposed full model	98.4%	98.5%	97.4%	98.8%

Table 4. Proposed model architecture evaluation with respect to different backbone layers

Backbone	Precision	Recall	F1	Accuracy
ResNet41	93.73%	90.13%	94.64%	93.72%
ResNet50	98.43%	98.51%	97.46%	98.82%
ResNet101	98.89%	98.81%	97.95%	98.94%



Figure 5 depicts the confusion matrix for the proposed model architecture. The analysis of misclassifications reveals that 33 non-COVID-19 pneumonia images were misclassified as normal, and 13 COVID-19 images were misclassified as normal. Further, the proposed model misclassified 42 COVID-19 cases as non-COVID-19 pneumonia, while 54 non-COVID-19 cases were misclassified as COVID-19. This indicates difficulty in distinguishing between these two clinically similar conditions, which is important in clinical procedures where correct treatment depends on proper classification. The misclassifications of COVID-19 as non-COVID-19 pneumonia (false negatives) are a critical concern, as they may lead to delayed or improper treatment decisions. Conversely, the misclassifications of non-COVID-19 pneumonia as COVID-19 (false positives) could result in unnecessary isolation, though they pose a lower risk in terms of public health management. Considering the model sensitivity and specificity, the yielded sensitivity detection for COVID-19 is 98.87%, while the specificity is 99.44%. Some misclassifications result from the radiological similarities between COVID-19 and other forms of pneumonia, which make distinction difficult.

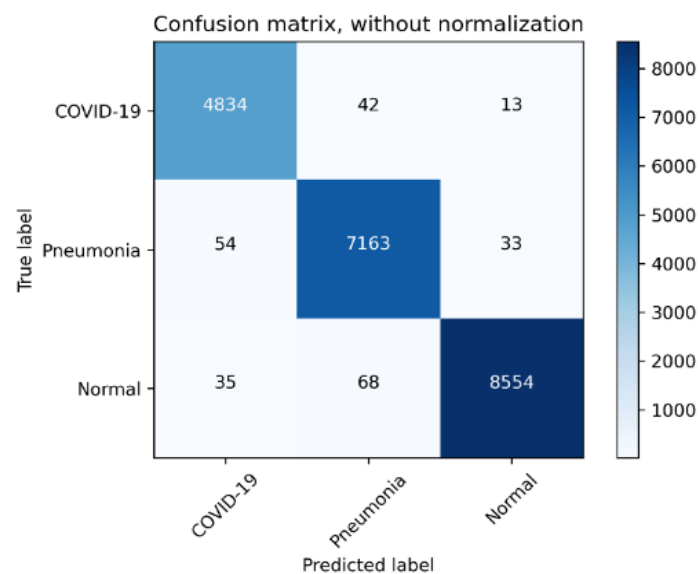


Figure 5. Confusion matrix representing the classification results of the proposed model on the dataset [29]

## 5. CONCLUSION

This paper proposed an effective Mask R-CNN-based architecture for the diagnosis of COVID-19 from chest CT images. The proposed architecture incorporates a novel image enhancement technique based on fuzzy logic, yielding notable improvements in contrast and reduces CT image ambiguity. Moreover, additional components are integrated, such as FPN and RoIAlign, which enhance spatial feature representation and classification accuracy. The presented architecture is quantitatively assessed on a general dataset, hence yields a superior performance over several benchmark models with an accuracy of 98.8%. These findings show the model's reliability and robustness, making it a promising candidate for clinical support in the early detection of COVID-19. Apart from its immediate application to COVID-19 diagnosis, the model's modular design and adaptability imply a great possibility for extension to other chest-related diseases, such as non-COVID-19 pneumonia or lung fibrosis. Furthermore, its performance on a large dataset enhances its generalizability and practical preparation for application in medical imaging systems. In summary, the effectiveness of the proposed architecture demonstrates how tailored preprocessing and deep learning architectures can significantly improve disease detection accuracy. Future work could explore incorporating additional clinical metadata—such as patient symptoms, medical history, or laboratory results, which could further improve model's capability to differentiate between visually similar CT scans and improve diagnostic precision.

## FUNDING INFORMATION

The authors declare that no funding was received for this study.

## AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Nader Mahmoud	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓		✓	
Ashraf B. El-Sisi		✓				✓	✓	✓	✓	✓	✓	✓	✓	✓

C : Conceptualization

M : Methodology

So : Software

Va : Validation

Fo : Formal analysis

I : Investigation

R : Resources

D : Data Curation

O : Writing - Original Draft

E : Writing - Review & Editing

Vi : Visualization

Su : Supervision

P : Project administration

Fu : Funding acquisition

## CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

## DATA AVAILABILITY

The dataset used in this study is available at <https://github.com/haydengunraj/COVIDNet-CT>, reference number [30].




## REFERENCES

- [1] T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim, and U. Rajendra Acharya, "Automated detection of COVID-19 cases using deep neural networks with X-ray images," *Computers in Biology and Medicine*, vol. 121, Jun. 2020, doi: 10.1016/j.compbiomed.2020.103792.
- [2] F. Khatami *et al.*, "A meta-analysis of accuracy and sensitivity of chest CT and RT-PCR in COVID-19 diagnosis," *Scientific Reports*, vol. 10, no. 1, Dec. 2020, doi: 10.1038/s41598-020-80061-2.
- [3] M. Ahmad *et al.*, "Deep belief network modeling for automatic liver segmentation," *IEEE Access*, vol. 7, pp. 20585–20595, 2019, doi: 10.1109/access.2019.2896961.
- [4] T. Ai *et al.*, "Correlation of chest CT and RT-PCR testing for coronavirus disease 2019 (COVID-19) in China: A report of 1014 cases," *Radiology*, vol. 296, no. 2, pp. E32–E40, Aug. 2020, doi: 10.1148/radiol.2020200642.
- [5] J. Shatri *et al.*, "The role of chest computed tomography in asymptomatic patients of positive coronavirus disease 2019: A case and literature review," *Journal of Clinical Imaging Science*, vol. 10, p. 35, Jun. 2020, doi: 10.25259/jcis\_58\_2020.
- [6] S. Siddique and J. C. L. Chow, "Machine learning in healthcare communication," *Encyclopedia*, vol. 1, no. 1, pp. 220–239, Feb. 2021, doi: 10.3390/encyclopedia1010021.
- [7] K. M. M. Uddin *et al.*, "Feature fusion based VGGFusionNet model to detect COVID-19 patients utilizing computed tomography scan images," *Scientific Reports*, vol. 12, no. 1, Dec. 2022, doi: 10.1038/s41598-022-25539-x.
- [8] Y. Song *et al.*, "Deep learning enables accurate diagnosis of novel coronavirus (COVID-19) with CT images," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 18, no. 6, pp. 2775–2780, Nov. 2021, doi: 10.1109/TCBB.2021.3065361.
- [9] J. W. Johnson, "Automatic nucleus segmentation with Mask-RCNN," in *Advances in Computer Vision*, Springer International Publishing, 2019, pp. 399–407, doi: 10.1007/978-3-030-17798-0\_32.
- [10] K. Lin *et al.*, "Face detection and segmentation based on improved mask R-CNN," *Discrete Dynamics in Nature and Society*, vol. 2020, pp. 1–11, May 2020, doi: 10.1155/2020/9242917.
- [11] H. Almubarak, Y. Bazi, and N. Alajlan, "Two-stage Mask-RCNN approach for detecting and segmenting the optic nerve head, optic disc, and optic cup in fundus images," *Applied Sciences*, vol. 10, no. 11, p. 3833, May 2020, doi: 10.3390/app10113833.
- [12] S. Mulay, G. Deepika, S. Jeevakala, K. Ram, and M. Sivaprakasam, "Liver segmentation from multimodal images using HED-Mask R-CNN," in *Multiscale Multimodal Medical Imaging*, Springer International Publishing, 2019, pp. 68–75, doi: 10.1007/978-3-030-37969-8\_9.
- [13] J.-H. Shu, F.-D. Nian, M.-H. Yu, and X. Li, "An improved Mask R-CNN model for multiorgan segmentation," *Mathematical Problems in Engineering*, vol. 2020, pp. 1–11, Jul. 2020, doi: 10.1155/2020/8351725.
- [14] J.-Y. Chiao, K.-Y. Chen, K. Y.-K. Liao, P.-H. Hsieh, G. Zhang, and T.-C. Huang, "Detection and classification the breast tumors using mask R-CNN on sonograms," *Medicine*, vol. 98, no. 19, p. e15200, May 2019, doi: 10.1097/md.00000000000015200.
- [15] T. Shibata, A. Teramoto, H. Yamada, N. Ohmiya, K. Saito, and H. Fujita, "Automated detection and segmentation of early gastric cancer from endoscopic images using Mask R-CNN," *Applied Sciences*, vol. 10, no. 11, p. 3842, May 2020, doi: 10.3390/app10113842.
- [16] C. Jin *et al.*, "Development and evaluation of an artificial intelligence system for COVID-19 diagnosis," *Nature Communications*, vol. 11, no. 1, Oct. 2020, doi: 10.1038/s41467-020-18685-1.
- [17] S. Hu *et al.*, "Weakly supervised deep learning for COVID-19 infection detection and classification from CT images," *IEEE Access*, vol. 8, pp. 118869–118883, 2020, doi: 10.1109/access.2020.3005510.
- [18] M. Polsinelli, L. Cinque, and G. Placidi, "A light CNN for detecting COVID-19 from CT scans of the chest," *Pattern Recognition Letters*, vol. 140, pp. 95–100, Dec. 2020, doi: 10.1016/j.patrec.2020.10.001.
- [19] S. Biswas, S. Chatterjee, A. Majee, S. Sen, F. Schwenker, and R. Sarkar, "Prediction of COVID-19 from chest CT images using an ensemble of deep learning models," *Applied Sciences*, vol. 11, no. 15, p. 7004, Jul. 2021, doi: 10.3390/app11157004.




- [20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [21] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul. 2017, pp. 1800–1807, doi: 10.1109/cvpr.2017.195.
- [22] W. Zhao, W. Jiang, and X. Qiu, "Deep learning for COVID-19 detection based on CT images," *Scientific Reports*, vol. 11, no. 1, Jul. 2021, doi: 10.1038/s41598-021-93832-2.
- [23] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Jun. 2014, pp. 580–587, doi: 10.1109/CVPR.2014.81.
- [24] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8691 LNCS, no. PART 3, 2014, pp. 346–361, doi: 10.1007/978-3-319-10578-9\_23.
- [25] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017, doi: 10.1109/TPAMI.2016.2577031.
- [26] A. Shrivastava, A. Gupta, and R. Girshick, "Training regionbased object detectors with online hard example mining," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 761–769, doi: 10.1109/CVPR.2016.89.
- [27] V. Patrascu and V. Buzuloiu, "Color image enhancement using the support fuzzification in the framework of the logarithmic model," in *Signal and Image Processing at LAPI Research Seminar*, 2005, doi: 10.13140/2.1.3014.6562.
- [28] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," in *2017 IEEE International Conference on Computer Vision (ICCV)*, Oct. 2017, doi: 10.1109/iccv.2017.322.
- [29] H. Gunraj, L. Wang, and A. Wong, "COVIDNet-CT: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest CT images," *Frontiers in Medicine*, vol. 7, Dec. 2020, doi: 10.3389/fmed.2020.608525.
- [30] H. Gunraj, "COVID-Net open source initiative - COVID-Net CT," *GitHub*, 2020. <https://github.com/haydengunraj/COVIDNet-CT> (accessed Sep. 27, 2024).
- [31] P. Tabarisaadi, A. Khosravi, and S. Nahavandi, "A deep Bayesian ensembling framework for COVID-19 detection using chest CT images," in *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Oct. 2020, pp. 1584–1589, doi: 10.1109/smc42975.2020.9283003.
- [32] N. Hasan, Y. Bao, A. Shawon, and Y. Huang, "DenseNet convolutional neural networks application for predicting COVID-19 using CT image," *SN Computer Science*, vol. 2, no. 5, Jul. 2021, doi: 10.1007/s42979-021-00782-7.

## BIOGRAPHIES OF AUTHORS



**Nader Mahmoud**    has received the B.Sc. from Computer Science Department, Faculty of Computers and Information (FCI), Menofia University, Egypt, in 2009. He has obtained the M.Sc. degree in 2015 under joint supervision between IRCAD Institute, France, and Computer Science Department at FCI, Menofia University, Egypt. He has awarded the Ph.D. degree from Strasbourg University, France. His research interests include biomedical imaging, computer vision, deep learning, augmented/mixed reality, camera tracking and mapping, and biometrics. He has published many articles in international conferences and journals. He has served as reviewer of more than 50 articles for many prestigious journals and conferences. He can be contacted at email: [nader.mahmoud@ci.menofia.edu.eg](mailto:nader.mahmoud@ci.menofia.edu.eg).



**Ashraf B. El-Sisi**    received the B.Sc. and M.Sc. degrees in electronic engineering and computer science engineering from Menoufia University, Faculty of Electronic Engineering in 1989 and 1995, respectively, and received his Ph.D. degree in computer engineering and control from Zagazig University, Faculty of Engineering in 2001. His current research interests include cloud computing, privacy-preserving data mining, and intelligent systems. He can be contacted at email: [ashraf.elsisi@ci.menofia.edu.eg](mailto:ashraf.elsisi@ci.menofia.edu.eg).