Neural Computing and Applications manuscript No.
(will be inserted by the editor)

Improving effectiveness of different deep learning-based models for detecting COVID-19 from computed tomography (CT) images

Erdi Acar · **Engin Şahin** · **İhsan Yılmaz**

Received: date / Accepted: date

Abstract Computerized Tomography (CT)has a prognostic role in the early diagnosis of COVID-19 due to it gives both fast and accurate results. This is very important to help decision making of clinicians for quick isolation and appropriate patient treatment. In this study, we combine methods such as segmentation, data augmentation and the generative adversarial network (GAN) to improve the effectiveness of learning models. We obtain the best performance with 99% accuracy for lung segmentation. Using the above improvements we get the highest rates in terms of accuracy (99.8%), precision (99.8%), recall (99.8%), f1-score (99.8%) and roc acu (99.9979%) with deep learning methods in this paper. Also we compare popular deep learning-based frameworks such as VGG16, VGG19, Xception, ResNet50, ResNet50V2, DenseNet121, DenseNet169, InceptionV3 and InceptionResNetV2 for automatic COVID-19 classification. The DenseNet169 amongst deep convolutional neural networks achieves the best performance with 99.8% accuracy. The second-best learner is InceptionResNetV2 with accuracy of 99.65%. The third-best learner is Xception and InceptionV3 with accuracy of 99.60%.

Keywords COVID-19 · Computed tomography · Deep learning · Data augmentation · Lung segmentation · GAN

Erdi Acar
Department of Computer Engineering, Çanakkale Onsekiz Mart University, 17100 Çanakkale, Turkey

Engin Şahin (Corresponding author)
Department of Computer and Instructional Technologies Education, Çanakkale Onsekiz Mart University, 17100 Çanakkale, Turkey
ORCID: 0000-0002-8040-0519
E-mail: enginsahin@comu.edu.tr

İhsan Yılmaz
Department of Computer Engineering, Çanakkale Onsekiz Mart University, 17100 Çanakkale, Turkey

1 Introduction

Coronavirus disease (COVID-19) has led to huge public health problem in the international community since it is rapidly spreading all over the World. Although polymerase chain reaction (PCR) test is standard for confirming COVID-19 positive patients, medical imaging such as X-ray and non-contrast computed tomography (CT) plays an important role in COVID-19 detection. Since COVID-19 can be determined by the presence of lung ground-glass opacities on CT images, which is clearer and more precise according to X-ray images, in early stages, diagnosis of COVID-19 from CT will help decision making of clinicians for quick isolation and appropriate patient treatment [1]. However chest CT images take more time for the specialists to diagnose. Therefore, it is important to use CT images for automated diagnosis of COVID-19. For this aim it is purposed deep learning-based frameworks [2].

Chen et al. [3] determine COVID-19 or non-COVID-19 from CT images including 51 COVID-19 patients and 55 patients with other diseases using UNet++ segmentation model. The results of COVID-19 classification are 95.2% (accuracy), 100% (sensitivity), and 93.6% (specificity). A U-Net and 3D CNN models are used for lung segmentation and diagnosing of COVID-19, respectively, in Ref. [4]. The results of the model are 90.7% (sensitivity), 91.1% (specificity), and 0.959 (AUC). Jin et al. [5] determine COVID-19 or non-COVID-19 from CT images including 496 COVID-19 positive cases and 1385 negative cases using 2D CNN based model. The results of the model are 94.1% (sensitivity), 95.5% (specificity), and 0.979 (AUC). Also Jin et al. [6] propose ResNet50 for classification and UNet++ for segmentation using CT images of 1136 cases (i.e., 723 COVID-19 positives, and 413 COVID-19 negatives). The results of the model are 97.4% (sensitivity) and 92.2% (specificity). Most of the previous works on COVID-19 are given in Ref. [7].

The main contributions of this paper can be summarized as:

- It is difficult to identify COVID-19 in CT images due to infections caused by COVID-19 often occur in small regions of the lungs [1] and the large variation in both position and shape across different patients [8]. Therefore we will use segmentation model based on bidirectional ConvLSTM U-Net [9] and graph-cut image processing [10] to improve the effectiveness of learning models.
- To improve the effectiveness of learning models we use data augmentation operations such as Random distortion, Flip, Rotate and Zoom.
- Deep learning algorithms require a large data set for training. If the model is trained with a small data set, the model becomes overfitting [8]. For this aim we will use the generative adversarial network (GAN) to generate more images and overcome the overfitting problem.
- Due to the above improvements, we achieved a high success of 99.80% in detecting COVID-19.

The rest of the paper is organized as follows. In Section 2, the materials and methods used in the study are presented. Section 3 presents the results of different analyzes for different deep learning algorithms in the proposed frameworks. The comparisons with the other related studies and the discussion of the results are given in Sec. 4.

2 Material and methods

2.1 Dataset

We collect the CT images from different open sources in Refs. [11,12]. Some of the CT images collected from the data sources are ignored due to repetition or high correlation problems in the CT images on the source. The chest CT images in our dataset consist of 1232 COVID-19 and 1668 healthy images. From the CT images in our dataset, 986 COVID-19 images and 1334 healthy images are randomly selected for training set. 99 COVID-19 images and 133 healthy images are randomly selected for validation set. 246 COVID-19 images and 334 healthy images are randomly selected for testing set.

2.2 Improving the data

This section describes the steps (lung segmentation, pre-processing and generative adversarial network) which are carried out to make training better.

2.2.1 Lung segmentation

BConvLSTM U-Net [9] architecture is used for segmentation. 1606 training size and 1606 mask size CT images are used for the machine learning. The Adam algorithm [13] with learning rate: $1e-4$ is used for stochastic optimization. The learning rate is dynamically reduced by using the ReduceLROnPlateau in the Keras Api during the learning. The results accuracy: 0.9902 and error: 0.0325 are obtained. The architecture of obtaining the mask images from the CT images by using BConvLSTM U-Net model is given in Fig. 1. The samples and obtained mask images by using BConvLSTM U-Net model are given in Fig. 2. The sample images after applying graph-cut image processing [10] are given in Fig. 3.

2.2.2 Pre-processing

We use some pre-processing steps to optimize the training process in the training of deep learning models. These steps are resizing, image normalization and data augmentation. The images in the dataset vary in terms of resolution and size. The images of the relevant region are resized to 224×224 pixel. The intensity values of all images are normalized from $[0,255]$ to the standard normal distribution by min-max normalization to the intensity range of $[0,1]$ as follows.

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{1}$$

where x is the pixel intensity. x_{min} and x_{max} are minimum and maximum intensity values of the input image. This operation helps to speed up the convergence of the model by achieve a uniform distribution across the dataset. We use image data augmentation (DA) [14] for deep learning to improve the effectiveness of learning

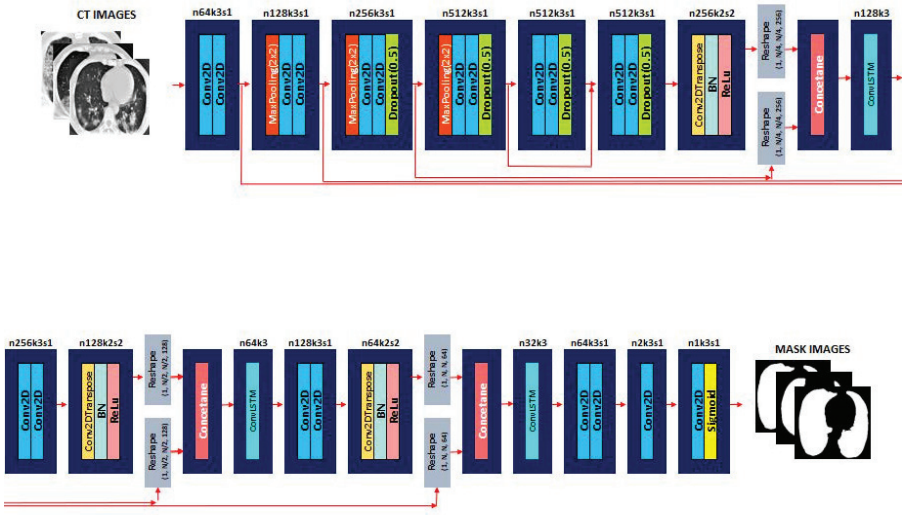

Fig. 1 The architecture of obtaining the mask images from the CT images by using BConvLSTM U-Net model

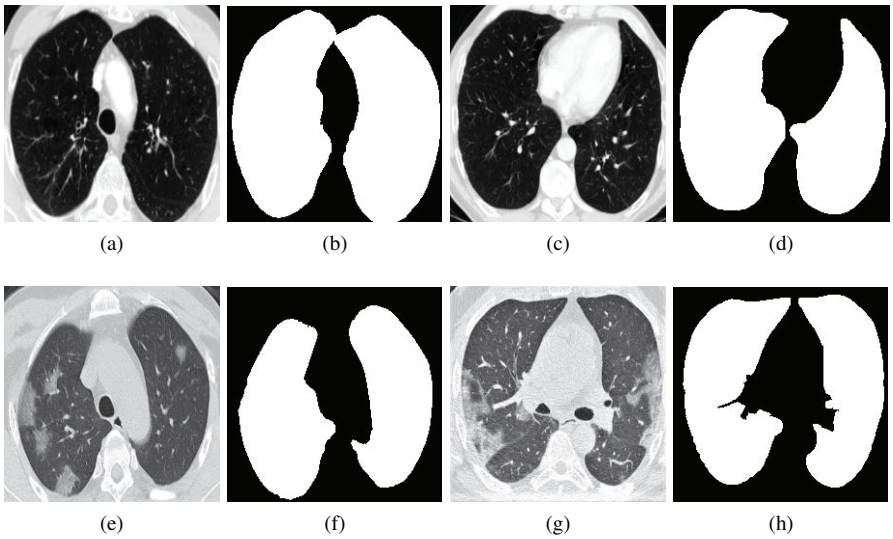

Fig. 2 The samples and obtained mask images by using BConvLSTM U-Net model **a, c, e, g** Sample images **b, d, f, h** Mask images

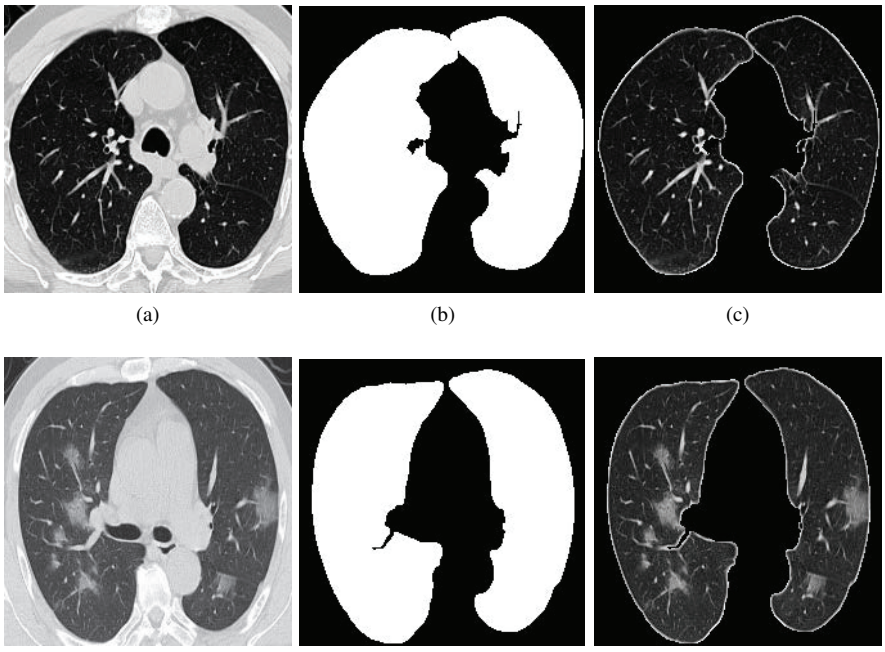

Fig. 3 The sample images after applying graph-cut image processing **a, d** Sample images **b, e** Mask images **c, f** Graph-cut images

models. Random distortion, Flip (left, right), Rotate and Zoom types of data augmentation are applied to the images.

2.2.3 Generative adversarial network (GAN)

Adam optimization algorithm with learning rate: 0.0001 is used for discriminator and generator training. The Least Squares function is used as the loss function in the study. The training epoch is 40000. 1232 real images with COVID-19 are used in the study. 3768 synthetic images with COVID-19 are produced as a result of the GAN training on real images with COVID-19 by using the generator. 1668 real healthy images are used in the study. 3332 synthetic healthy images are produced as a result of the GAN training on real healthy images by using the generator. The architecture and framework of GAN applied in the study are given in Figs. 4 and 5, respectively.The samples of the COVID-19 and healthy synthetic images produced after the GAN training are given in Figs. 6 and 7, respectively.

2.3 Experimental evaluation

In this paper, two types of experimental studies are carried out in frameworks with and without (only with original samples) GAN. The frameworks without and with

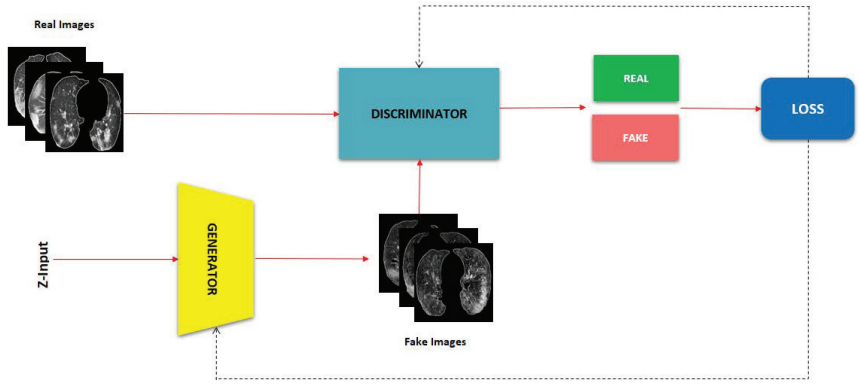

Fig. 4 The architecture of GAN

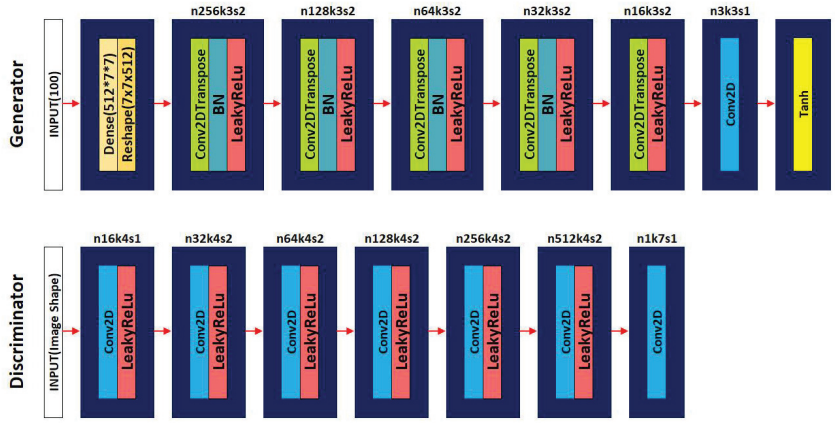

Fig. 5 The framework of GAN

GAN are given in Figs. 8 and 9, respectively. The numbers of the real images without GAN and the real and synthetic images produced with GAN in sets are shown in Table 1. In the framework with GAN, totally 7200 samples are used in the training set (including 3600 COVID-19 and 3600 healthy) are used to train the model by the validation set of totally 800 samples (400 COVID-19 and 400 healthy) and then test the model performance by the testing set of totally 2000 samples (1000 COVID-19 and 1000 healthy). To verify the validity of the operations, the data augmentation operations which are applied to both the original sets and the extended with GAN sets are shown in Table 2.

The methods followed in both frameworks are the same except for the datasets used. Keras Api is used for training, the Adam function is used for optimization in

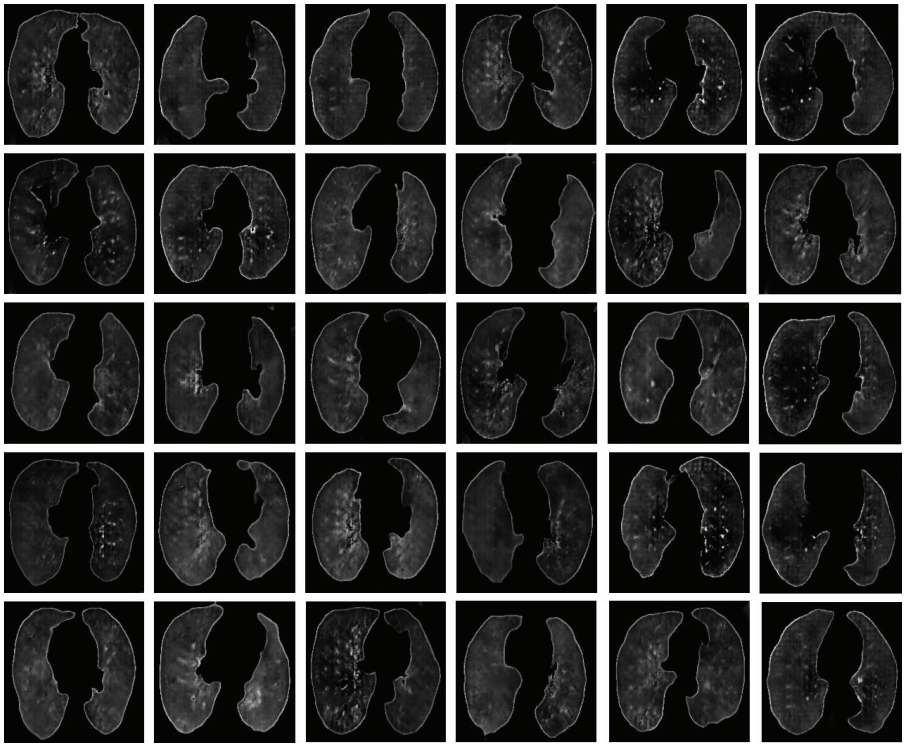

Fig. 6 The samples of the COVID-19 synthetic images produced after the training without GAN

Table 1 The numbers of the real images without GAN and the real+synthetic images produced with GAN in training, validation and testing sets

Framework	Types	Training set	Validation set	Testing set
Real images without GAN	COVID-19	986	99	246
	Healthy	1334	133	334
Real+Synthetic images produced with GAN	COVID-19	3600	400	1000
	Healthy	3600	400	1000

the both frameworks. The learning rate is started as 0.001. During the training, the learning rate is dynamically reduced by using the ReduceLROnPlateau in the Keras Api. The deep learning models, VGG16, VGG19, Xception, ResNet50, ResNet50V2, DenseNet121, DenseNet169, InceptionV3 and InceptionResNetV2 are used in the both frameworks. Dense(512, relu), BatchNormalization, Dense(256, relu), Dense (1, Sigmoid) layers are used in the full connected layer for classification, respectively.

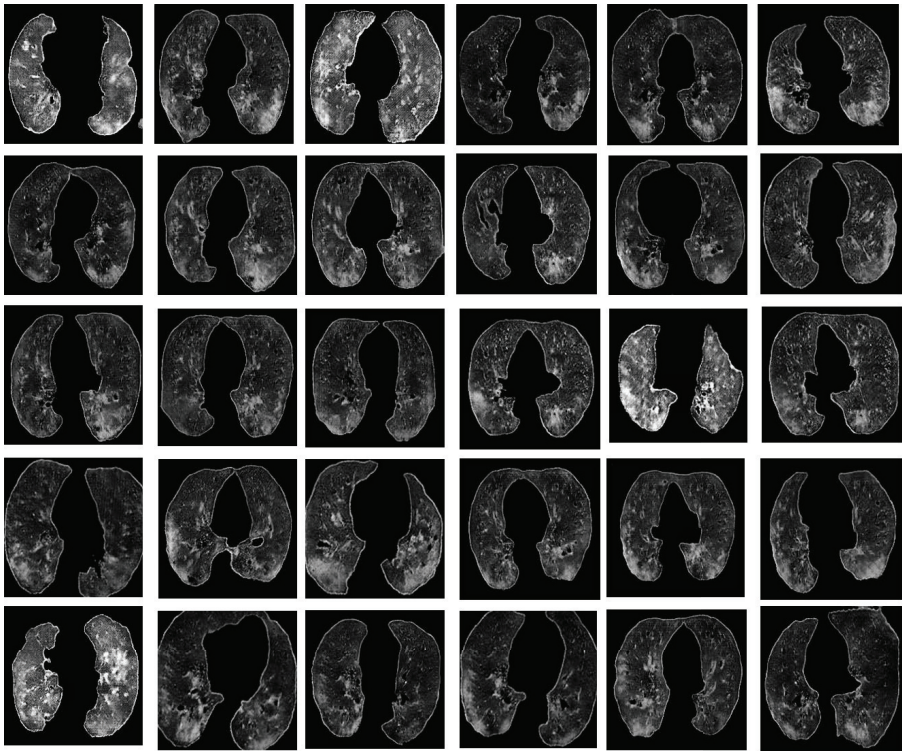

Fig. 7 The samples of the healthy synthetic images produced after the training with GAN

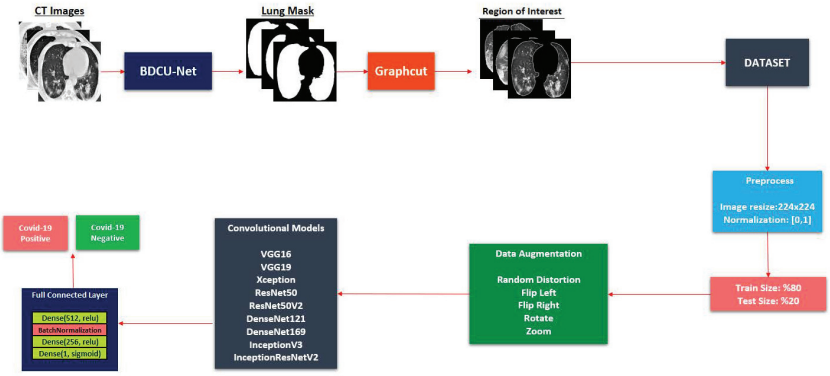

Fig. 8 The framework without GAN

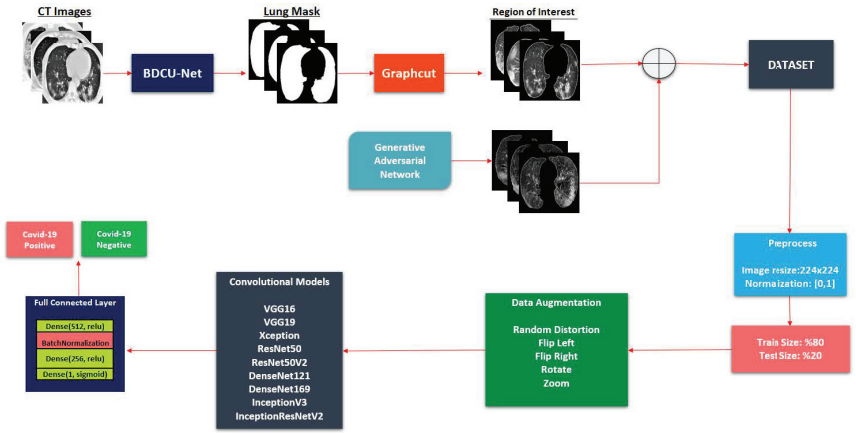

Fig. 9 The framework with GAN

Table 2 Types of Data Augmentation

Types	Parameters
Random distortion	probability=0.5 grid width=4 grid height=4
Flip (left, right)	probability=0.5
Rotate	probability=0.5 max left rotation=10 max right rotation=10
Zoom	probability=0.5 min factor=0.9 max factor=1.05

3 Results

Estimation performances of the methods in this study are measured with metrics such as accuracy, precision, recall and f1-score. All evaluation metrics were calculated as follows according to Kassani et al. [2].

Dividing the number of correctly classified cases into the total number of test images shows the accuracy and is calculated as follows.

$$\text{Accuracy} = \frac{t_p + t_n}{t_p + t_n + f_p + f_n} \tag{2}$$

where t_p is the number of instances that correctly predicted, f_p is the number of instances that incorrectly predicted, t_n is the number of negative instances that correctly predicted, f_n is the number of negative instances that incorrectly predicted.

The recall is used to measure correctly classified COVID-19 cases. Recall is calculated as follows.

Table 3 The comparative classification performance result of the deep learning models used in our framework without GAN

Model	Accuracy	Error	Precision	Recall	F1-score	Roc Auc
VGG16	0.974137	0.076263	0.967742	0.988024	0.977778	0.992941
VGG19	0.968955	0.075962	0.981407	0.964072	0.972810	0.997574
Xception	0.991379	0.022388	0.991045	0.994012	0.992526	0.999769
ResNet50	0.965517	0.1229	0.951149	0.991018	0.970674	0.985626
ResNet50V2	0.981034	0.045709	0.973607	0.994012	0.983704	0.999148
DenseNet121	0.982759	0.094841	0.982143	0.988024	0.985075	0.995211
DenseNet169	0.982759	0.098201	0.976471	0.994012	0.985163	0.989661
InceptionV3	0.993103	0.022390	0.994012	0.994012	0.994012	0.999647
InceptionResNetV2	0.989655	0.064712	0.985207	0.997006	0.991071	0.998406

Table 4 The comparative classification performance result of the deep learning models used in our framework with GAN

Model	Accuracy	Error	Precision	Recall	F1-score	Roc Auc
VGG16	0.9920	0.048799	0.991018	0.9930	0.992008	0.998045
VGG19	0.9885	0.127115	0.982231	0.9950	0.988574	0.997247
Xception	0.9960	0.013771	0.994024	0.9980	0.996008	0.999911
ResNet50	0.9920	0.026624	0.988095	0.9960	0.992032	0.999674
ResNet50V2	0.9935	0.021429	0.993994	0.9930	0.993497	0.999680
DenseNet121	0.9945	0.010011	0.993021	0.9960	0.994508	0.999948
DenseNet169	0.9980	0.005571	0.9980	0.9980	0.9980	0.999979
InceptionV3	0.9960	0.014410	0.996000	0.9960	0.9960	0.999903
InceptionResNetV2	0.9965	0.008892	0.996004	0.9970	0.996502	0.999966

$$\text{Recall} = \frac{t_p}{t_p + f_p} \tag{3}$$

The percentage of correctly classified labels in truly positive patients is defined as the precision and is calculated as follows.

$$\text{Precision} = \frac{t_p}{t_p + f_n} \tag{4}$$

The F1-score is defined as the weighted average of precision and recall combining both precision and recall, and is calculated as follows.

$$\text{F1-score} = 2 \times \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \tag{5}$$

In the studies, the operations are performed on 334 healthy and 246 with COVID-19 images for the framework without GAN, and 1000 healthy and 1000 with COVID-19 images for the framework with GAN. The confusion matrices of the our framework without GAN and with GAN are given in Figs. 10 and 11, respectively. The comparative classification performance results of the deep learning models used in the our framework without GAN are shown in Table 3. Likewise, the results in the our framework with GAN are shown in Table 4.

(a) VGG16 (b) VGG19 (c) Xception

(d) ResNet50 (e) ResNet50V2 (f) DenseNet121

(g) DenseNet169 (h) InceptionV3 (i) InceptionResNetV2

Fig. 10 The confusion matrices of the our framework without GAN

4 Discussion

The imaging methods, such as CT, play a more important role in making the initial diagnosis for COVID-19 than the Polymerase Chain Reaction (PCR) tests. The performance of deep learning models depends directly on the quality of the features determined from the image. The focus of this study, which is based on the success of deep learning models, is to achieve high accuracy results using features related to the classification of COVID-19 in CT imaging. The machine is trained by using BConvLSTM U-Net architecture for lung segmentation. In every experiment, the technique of reproducing original images with the GAN method is used to evaluate the performance of the classifiers.

Performance information of deep learning models such as VGG16, VGG19, Xception, ResNet50, ResNet50V2, DenseNet121, DenseNet169, InceptionV3 and InceptionResNetV2 without GAN and with GAN are shown in Tables 1and 2, respectively. In our studies, the highest accuracy rate is 99.80% with the DenseNet169 method, the highest precision rate is 99.80% with the DenseNet169 method, the highest recall rate is 99.80% with the DenseNet169 method, the highest f1-score is 99.80% with the DenseNet169 method and the highest roc acu is 99.9979% with the DenseNet169 method. All of the highest values are obtained with the DenseNet169 method. There-

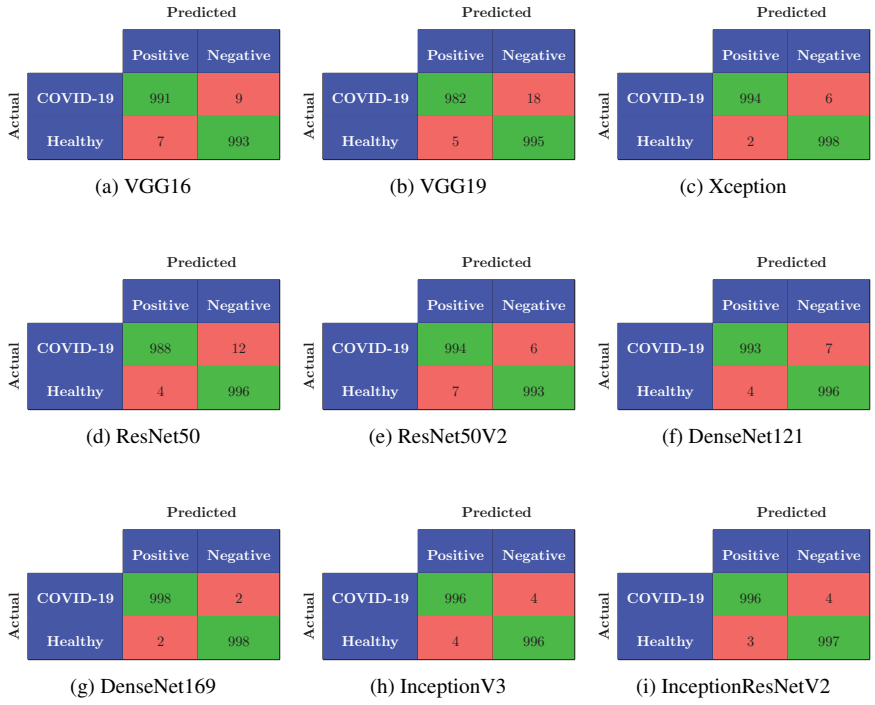

Fig. 11 The confusion matrices of the our framework with GAN

fore, we can say that the DenseNet169 method is more efficient than the other deep learning methods.

The image segmentation methods and the best accuracy results of the studies in COVID-19 applications with CT images and our method are shown in Table 5. The high number of samples used in deep learning and the segmentation architecture used are very important for higher accuracy. In our study, more samples are used than the existing studies.

It is clearly seen from the Table 5 that we obtained the highest accuracy value 99.80% in the literature along with the BConvLSTM U-Net method that we use for image segmentation as well as many examples.

In addition, Kassani et al. [2] perform accuracy, precision, recall and f1-score analyzes with different deep learning algorithms without image segmentation. The accuracy, precision, recall and f1-score analyzes with different deep learning algorithms in the our and Kassani et al.'s [2] frameworks are shown in Table 6.

It can be clearly seen from Table 6 that our framework is more successful than the Kassani et al.'s framework [2] in terms of accuracy, precision, recall and f1-score rates in all of the deep learning algorithms.

In conclusion, the highest available rates are achieved in terms of accuracy (99.80%), precision (99.80%), recall (99.80%), f1-score (99.80%) and roc acu (99.9979%)

Table 5 The image segmentation methods and the best accuracy results of the studies in COVID-19 applications with CT images

Literature	Images	Method	Accuracy(%)
Our proposed with GAN	1000 COVID-19 1000 Others	BConvLSTM U-Net	99.80
Chen et al. [3]	51 COVID-19 55 Others	UNet++	95.2
Zheng et al. [4]	313 COVID-19 229 Others	U-Net	90.8
Wang et al. [15]	44 COVID-19 55 Others	CNN	82.9

Table 6 The accuracy, precision, recall and f1-score analyzes with different deep learning algorithms in the our and Kassani et al.'s [2] frameworks

Model	Framework	Accuracy(%)	Precision(%)	Recall(%)	F1-score(%)
VGG16	Our proposed Ref. [2]	99.20 91.00	99.1018 94.00	99.30 94.00	99.2008 94.00
VGG19	Our proposed Ref. [2]	98.85 90.00	98.2231 94.00	99.50 94.00	98.8574 94.00
Xception	Our proposed Ref. [2]	99.60 96.00	99.4024 98.00	99.80 98.00	99.6008 98.00
ResNet50	Our proposed Ref. [2]	99.20 98.00	98.8095 96.00	99.60 96.00	99.2032 96.00
ResNet50V2	Our proposed Ref. [2]	99.35 96.00	99.3994 96.00	99.30 96.00	99.3497 96.00
DenseNet121	Our proposed Ref. [2]	99.45 99.00	99.3021 98.00	99.60 98.00	99.4508 98.00
InceptionV3	Our proposed Ref. [2]	99.60 95.00	99.60 99.00	99.60 99.00	99.60 99.00
InceptionResNetV2	Our proposed Ref. [2]	99.65 94.00	99.6004 96.00	99.70 95.00	99.6502 95.00

with deep learning methods in this paper. The our presented framework with GAN in this paper is more successful than all the existing studies that classify COVID-19 with the deep learning algorithms.

Funding

Not applicable

Availability of data and material

Not applicable

Code availability

Not applicable

Authors' contributions

All authors contributed to the study conception and design.
Conceptualization: Erdi Acar, Engin Şahin, İhsan Yılmaz; Data curation: Erdi Acar; Formal analysis: Erdi Acar, Engin Şahin; Investigation: Erdi Acar, Engin Şahin, İhsan Yılmaz; Methodology: Erdi Acar, İhsan Yılmaz; Resources: Erdi Acar; Software: Erdi Acar; Supervision: İhsan Yılmaz; Visualization: Engin Şahin; Writing - original draft preparation: Engin Şahin; Writing - review and editing: Engin Şahin;
All authors read and approved the final manuscript.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. He K, Zhao W, Xie X, Ji W, Liu M, Tang Z et al (2020) Synergistic learning of lung lobe segmentation and hierarchical multi-instance classification for automated severity assessment of COVID-19 in CT images. arXiv:2005.03832v1
2. Kassani SH, Kassasni PH, Wesolowski MJ, Schneider KA, Deters R (2020) Automatic detection of coronavirus disease (COVID-19) in x-ray and CT images: a machine learning based approach. arXiv:2004.10641v1
3. Chen J, Wu L, Zhang J, Zhang L, Gong D, Zhao Y et al (2020) Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography: a prospective study. medRxiv:2020.02.25.20021568
4. Zheng C, Deng X, Fu Q, Zhou Q, Feng J, Ma H et al (2020) Deep learning-based detection for COVID-19 from chest CT using weak label. medRxiv:2020.03.12.20027185
5. Jin C, Cheny W, Cao Y, Xu Z, Zhang X, Deng L et al (2020) Development and evaluation of an AI system for COVID-19 diagnosis. medRxiv:2020.03.20.20039834
6. Jin S, Wang B, Xu H, Luo C, Wei L, Zhao W et al (2020) AI-assisted CT imaging analysis for COVID-19 screening: Building and deploying a medical AI system in four weeks. medRxiv:2020.03.19.20039354
7. Shi F, Wang J, Shi J, Wu Z, Wang Q, Tang Z et al (2020) Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for COVID-19. arXiv:2004.02731
8. Khalifa NEM, Taha MHN, Hassanien AE, Elghamrawy S (2020) Detection of coronavirus (COVID-19) associated pneumonia based on generative adversarial networks and a fine-tuned deep transfer learning model using chest x-ray dataset, arXiv:2004.01184 (2020).
9. Azad, R, Asadi-Aghbolaghi M, Fathy M, Escalera S (2019) Bi-directional ConvLSTM U-Net with densley connected convolutions. arXiv:1909.00166

10. Massoptier L, Misra A, Sowmya A (2009) Automatic lung segmentation in HRCT Images with diffuse parenchymal lung disease using graph-cut. 24th International Conference Image and Vision Computing New Zealand.
11. Zhao J, Zhang Y, He X, Xie P (2020) COVID-CT-Dataset: A CT scan dataset about COVID-19. arXiv:2003.13865
12. Wang LL, Lo K, Chandrasekhar Y, Reas R, Yang J, Eide D et al (2020) CORD-19: The Covid-19 open research dataset. arXiv:2004.10706v2
13. Kingma DP, Ba J (2014) Adam: A method for stochastic optimization. arXiv:1412.6980
14. Shorten C, Khoshgoftaar TM (2019) A survey on image data augmentation for deep learning. Journal of Big Data 6(1):60.
15. Wang S, Kang B, Ma J, Zeng X, Xiao M, Guo J et al (2020) A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19). medRxiv:2020.02.14.20023028