European **Rev**iew for **Med**ical and **Pharmacol**ogical **Sci**ences 2024; 28: 1213-1226

Accurate diagnosis of COVID-19 from lung CT images using transfer learning

H.G. TAS[1], M.B.H. TAS[2], B. IRGUL[3], S. AYDIN[3], U. KUYRUKLUYILDIZ[4]

[1]Department of Anesthesiology and Reanimation, Faculty of Medicine, Erzincan Binali Yıldırım University, Erzincan, Turkey
[2]Computer Engineering, Faculty of Engineering, Erzincan Binali Yıldırım University, Erzincan, Turkey
[3]Department of Radiology, Faculty of Medicine, Erzincan Binali Yıldırım University, Erzincan, Turkey
[4]Department of Anesthesiology and Reanimation, Faculty of Medicine, Erzincan Binali Yıldırım University, Erzincan, Turkey

Abstract. – OBJECTIVE: In this study, it is aimed to classify data by feature extraction from tomographic images for the diagnosis of COVID-19 using image processing and transfer learning.

MATERIALS AND METHODS: In the proposed study, CT images are made better detectable by artificial intelligence through preliminary processes such as masking and segmentation. Then, the number of data was increased by applying data augmentation. The size of the dataset contains a large number of images in numerical terms. Therefore, the results of the models are more reliable. The dataset is split into 70% training and 30% testing. In this way, different features of the applied models were found, and positive effects were achieved on the result. Transfer Learning was used to reduce training times and further increase the success rate. To find the best method, many different pre-trained Transfer Learning models have been tried and compared with many different studies.

RESULTS: A total of 8,354 images were used in the research. Of these, 2,695 consist of COVID-19 patients and the remaining healthy chest tomography images. All of these images were given to the models through masking and segmentation processes. As a result of the experimental evaluation, the best model was determined to be ResNet-50 and the highest results were found (accuracy 95.7%, precision 94.7%, recall 99.2%, specificity 88.3%, F1 score 96.9%, ROC-AUC score 97%).

CONCLUSIONS: The presence of a COVID-19 lesion in the images was identified with high accuracy and recall rate using the transfer learning model we developed using thorax CT images. This outcome demonstrates that the strategy will speed up the diagnosis of COVID-19.

Key Words:
Artificial intelligence, Transfer Learning, Thorax CT, COVID-19.

Introduction

COVID-19 is still a public health issue in some parts of the world. Infection symptoms can range from a simple cold to life-threatening illnesses. Coronavirus infections are frequently accompanied by respiratory symptoms. Symptoms of coronavirus infection include fever, cough, and difficulties with breathing, respiratory issues, weariness, and sore throat[1-3].

The diagnosis of COVID-19 can be made by various methods, but the accuracy and speed of these methods are limited. In recent years, artificial intelligence techniques have come to the fore as a new and rapid approach to the diagnosis of COVID-19. Machine learning techniques such as feature extraction and classification were used to identify COVID-19 symptoms in tomographic images. However, there is not enough COVID-19 data for these approaches, which may affect accuracy and reliability. In addition, the distribution of the sections taken should be diversified in order to evaluate limited sources and tomography images from every angle. The research' results could be improved by clearly identifying the sections infected with COVID-19 in tomographic images and then using these sections as input. Transferring each slice to the education model will reveal the accuracy and precision of the work.

Many studies[4,5] have been conducted on the use of deep learning in the interpretation of radiological images. Many researchers use and continue to study deep learning techniques for detecting and classifying COVID-19 using radiological images. Convolutional Neural Network (CNN) architectures are one of the most effective approaches among deep learning algorithms[4,5].

Corresponding Author: Baris Irgul, MD; e-mail: barisirgul@gmail.com; baris.irgul@erzincan.edu.tr

This study aims to extract and classify features from tomographic images for the diagnosis of COVID-19 using Transfer Learning. Transfer Learning is a Deep Learning technique in which a model is trained and developed for one task and then reused in a second related task. It refers to the situation where what is learned in one environment is used to improve optimization in another environment[6]. We conducted many experiments with pre-trained neural networks and used multiple models to obtain the best results for COVID-19 detection. There are two major benefits when using transfer learning: the results are achieved faster, and when architectures known to function well are employed, the results are obtained with greater performance.

Materials and Methods

Deep learning models often use Transfer Learning techniques to achieve better performance results in classification tasks. This is due to the preprocessing studies and large data sizes involved in the datasets used. In this study, accuracy, recall, sensitivity, specificity, F1-score, and ROC-AUC scores obtained from artificial intelligence were used to measure performance indicators. These metrics are indispensable in artificial intelligence studies. It is a situation that cannot be determined exactly whether the amount of data considered is high or low. Neural networks created in datasets are in connection with neurons. In deep learning, the features that pass through these neural networks are extracted from a numerical value called weight. After navigating the entire neuron network, these numerical values evaluate the highest circulation network in a learning function and reveal metrics for the desired goal. These metrics can vary with activation functions and optimizers. While it can cause unwanted events such as overfitting in high datasets, low datasets are natural processes that are expected to be insufficient in feature extraction. It is really tough to locate an additional strategy that can be used other than data duplication of a small dataset. However, if large datasets are employed, the solutions to this problem become more numerous and simpler. For this reason, since it is impossible to make a statistical preliminary inference, it is not possible to make a power analysis and give a clear number or a minimum number.

The study's flow is shown in Figure 1. The data to be collected will be pre-processed first, followed by data derivation/duplication. The data will then be divided into three categories: training, testing, and validation. The categorization process will then be carried out utilizing transfer learning.

The working flow of the study is as follows:

- The tomography images required for the dataset are collected. It should be done carefully to ensure optimal results.
- To improve the data, pre-processing studies such as masking, segmentation, data augmentation, and image resizing are performed using the obtained data.
- The dataset was partitioned among 70 percent training and 30 percent testing. The test group was likewise split by 20% validation.
- Data was trained using a variety of transfer learning algorithms, including VGG16, ResNet50, and InceptionResNetV2.
- Evaluation metrics were employed to compare the results.

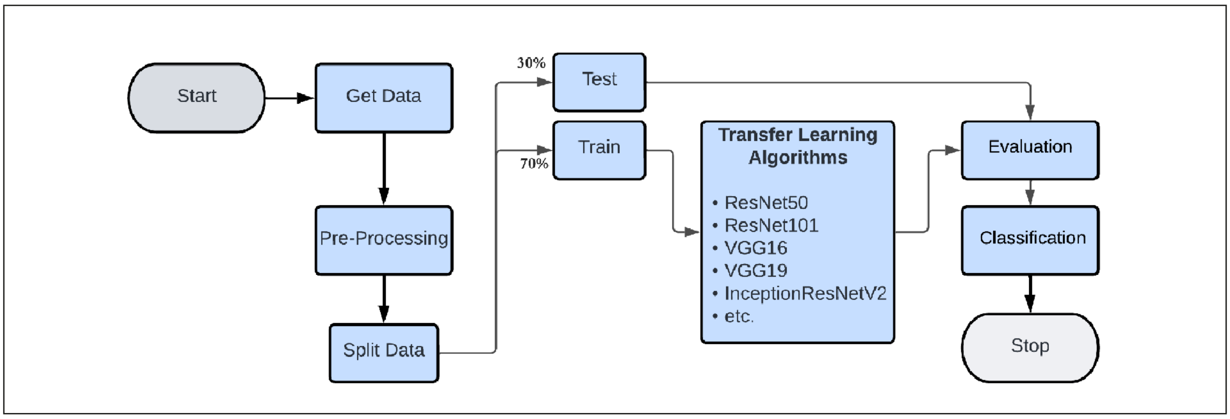

Figure 1. Flow chart of the study.

Python, an open-source language, and its libraries were used in the study. The tensor and Compute Unified Device Architecture (CUDA) operations required for the classification of the images have been done with TensorFlow version 2.4.0 (Google Inc., Mountain View, CA, USA) and the libraries it contains. Image processing studies are done with the help of the OpenCV library. The operating system Windows 10 is preferred. The system that runs the developed models:

- CPU: Intel i7 12700 (Santa Clara, CA, USA)
- GPU: Nvidia RTX 3080 12GB (Santa Clara, CA, USA)
- RAM: 32 GB 3600 MHz.

Dataset

The dataset was compiled from 29 patients and contains 8,354 CT images. The images in the dataset are in two classes: COVID-19 and non-COVID-19 images. The images used contain all CT sections of a subject. Without any elimination process, all sections of the patient are used in the dataset. The dataset was used as 70% training, 30% testing, and 20% of the training group as validation. Some sample images in the dataset are given in Figure 2. The images were obtained from absolutely random patients, and each area was diagnosed by a specialized doctor. To avoid remaining in low numbers and to produce more accurate and consistent results, the validation data was extracted from the training data. We attempted to ensure that the results produced by experimenting with more images would be more accurate in this manner.

Pre-Processing

Undesirable parts in the image were separated by image processing methods. In this method,

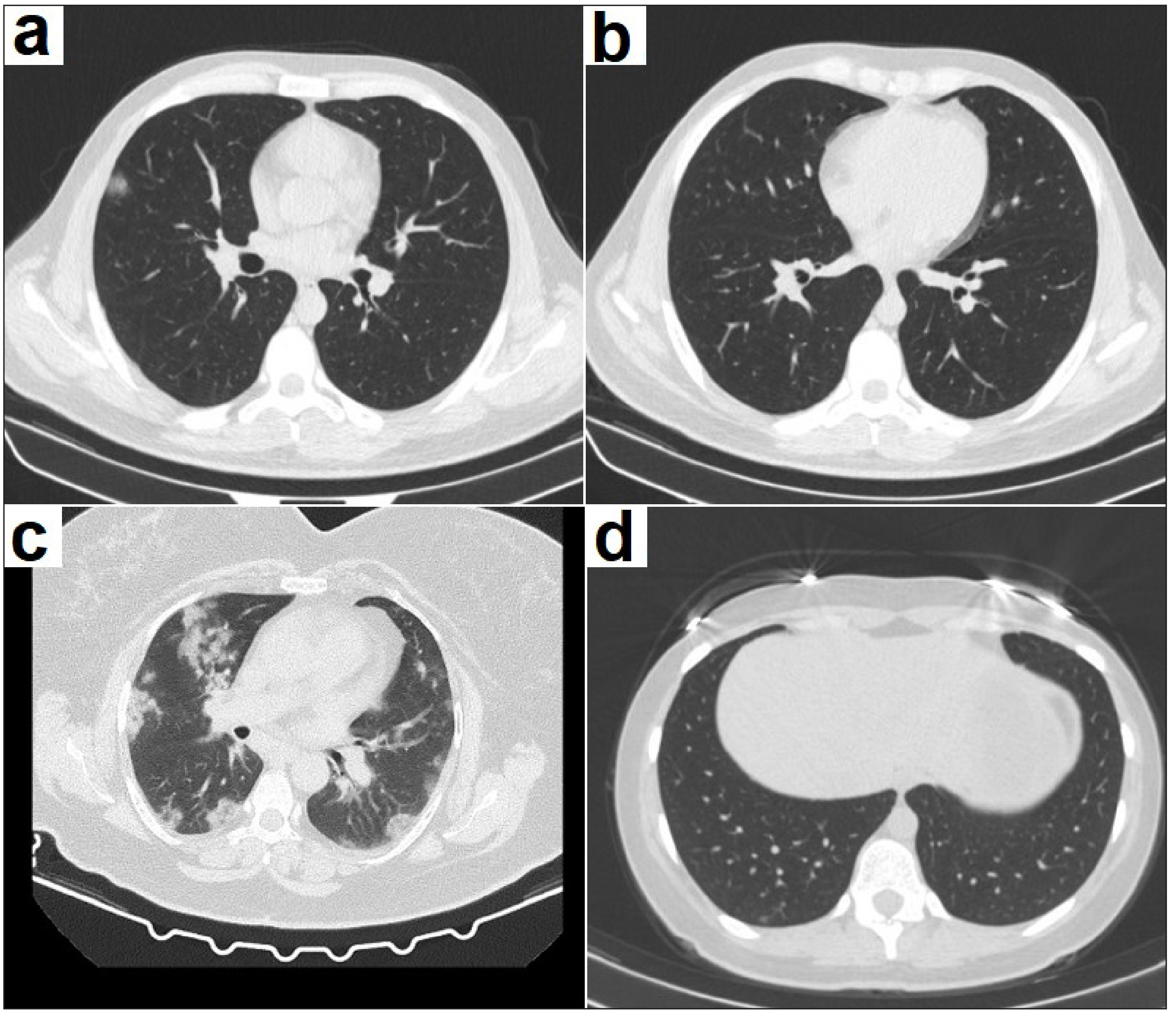

Figure 2. Sample images of the dataset. **a**, and **c**, COVID-19; **b**, and **d**, non-COVID-19 samples.

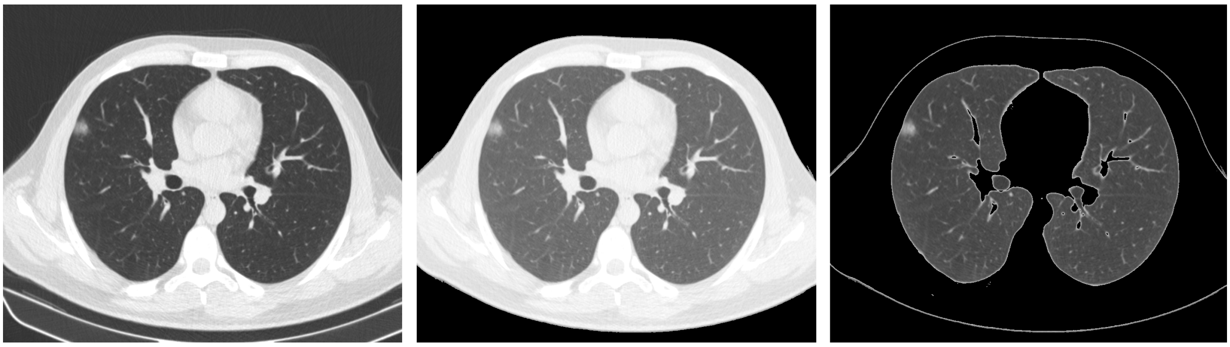

Figure 3. Preprocessing work on images.

the training data can generate superior outcomes because deep learning algorithms may be able to reduce unwanted or useless characteristics if there are extraneous details. These efforts must be made to prevent this from occurring and to improve the learning process. Figure 3 shows the pre-processing work that is expected to be completed. Unnecessary details on the outside of the body can be removed by masking the image first[7] and then segmentation[8] will ensure that only the areas with lung tissue remain in the image. Threshold differences were used to make changes. In this way, the opportunity to distinguish between color contrasts in the original image was provided. Therefore, a significant difference was created in the real image with the threshold image value without any distortion in the image. Then, the same color tones on the masked image with the colors close to black in the real image were detected, and the lung parts were separated from each other. Figure 3a is the original image, Figure 3b is the masked image, and Figure 3c is the segmented version.

Deep Learning

A classification approach based on artificial intelligence was used in the investigation. There are numerous ways to accomplish this. In recent years, numerous techniques, including Deep Learning[9-11] and its sub-branches, Convolutional Neural Networks (CNN)[9,12-14] and accordingly transfer learning have been used. These methods are frequently used because they give very good results, especially on images. Transfer learning was used within the scope of the study. Transfer learning is also a CNN-based training method. The distinction from standard CNN is that it is one of the primary reasons why it is preferred since it uses previously trained models and produces high-performance results in less time. A classical CNN architecture is given in Figure 4.

In CNN models, data is given to an input layer. Then, the features are extracted from the convolutional layers and transferred to the layers called pooling. The goal here is to lower the amount of data that can have more qualities and deliver good results in a shorter amount of

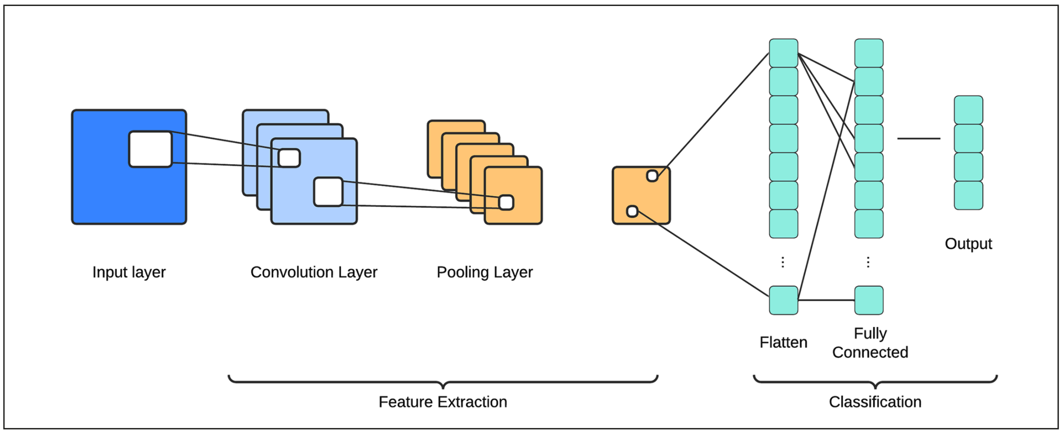

Figure 4. A general CNN architecture.

time. Following these operations, a classification procedure is performed in the smoothing layers, and the result determined by the classifier is revealed in the output layer. The convolution layer computes the result of multiplying the filters (kernel) by resizing a window in the input data. This procedure shrinks the amount of the supplied data and highlights specific aspects. After receiving the output of the convolution layer, the activation layer applies it to a specific activation function (e.g., ReLU) and therefore, conducts a non-linear operation on the network's outputs. The pooling layer is used to further reduce the size of the input data. This layer performs size reduction by maximizing or averaging features within a given window size. Fully connected layers convert the entire feature map into a single vector and pass the results to the next layer. The mathematical formulation of CNNs is formed by bringing these layers together. In particular, the output of a CNN is usually calculated as a classification result using the SoftMax activation function. The mathematical formulation of a CNN is fairly complex; however, the essential processes are typically represented as follows: a convolution operation calculates the result of multiplying input data by a window of a specific size. This procedure can be described numerically as follows[15]:

$$Y(i,j) = sum\ (sum\ (X\ (i{:}i+f-1,\ j{:}\ j+f-1)*\ W)) \quad (1)$$

Here, X is the matrix representation of the input data, W is the window or filter matrix, Y is the output of the convolution operation. and , represent the positions in the output matrix, while represents the window size. This formulation is valid only for the convolution operation. CNN is a complex artificial neural network that includes many different mathematical operations and therefore it is not possible to express its entire formulation with a single equation. However, the convolution operation is one of the most basic mathematical operations of CNN, and this formulation refers to several basic matrix multiplications used in the convolution layer.

Transfer Learning

Transfer Learning has become an important topic in machine learning research in recent years. This technique is the reuse of a pre-trained model for a similar task. This approach can be very useful in problems with limited datasets or insufficient computational resources. Transfer Learning has several different methods, such as creating a new model using part of the pre-trained model or using the outputs of the pre-trained model using a completely different model. Studies[16-18] show that Transfer Learning is particularly effective in image classification, object detection and natural language processing. In a study using multiple criteria, it was observed that Transfer Learning outperformed many tasks. As a result, Transfer Learning is regarded as a critical technique in machine learning research.

VGG16-VGG19

The VGG model was designed to improve performance in the 2014 ImageNet Large Scale Visual Recognition Challenge (ILSVRC) competition by Simonyan and Zisserman[14]. The model is a multilayer convolutional neural network and includes 16 convolutional layers and 3 fully connected layers. VGG19 and VGG16 are convolutional neural network models with similar structures, but VGG19 is an expansion of the VGG16 model and shows a few differences. VGG19 has three layers more structure than the VGG16 model due to its deeper network structure. VGG16 consists of 16 layers, while VGG19 consists of 19 layers. VGG19 has the capacity to learn more complex features due to its deeper network structure. VGG19 has three additional convolution layers from the VGG16 model. The structure of these architectures is shown in Figure 5.

ResNet50- ResNet101- ResNet152

It proposes a "residual learning" strategy that allows deep networks to be trained more readily and fast, addressing a prevalent challenge in deep learning approaches. This approach allows the network to transfer information learned from previous layers to later layers, allowing the network to be deeper and perform better. This approach, combined with the ResNet architecture, has formed the basis of successful deep learning models used in many fields[19]. The general architectures of the ResNet model are shown in Figure 6. ResNet50, ResNet101, and ResNet152 are variations of ResNet to achieve deeper networks using "big blocks" of varying numbers and structure. The key distinction between these variants is the number of blocks in the network and the type of layers contained within these blocks. ResNet50 contains 50 blocks, and each block is a "bottleneck" structure with 1×1, 3×3, and 1×1 convolution layers.

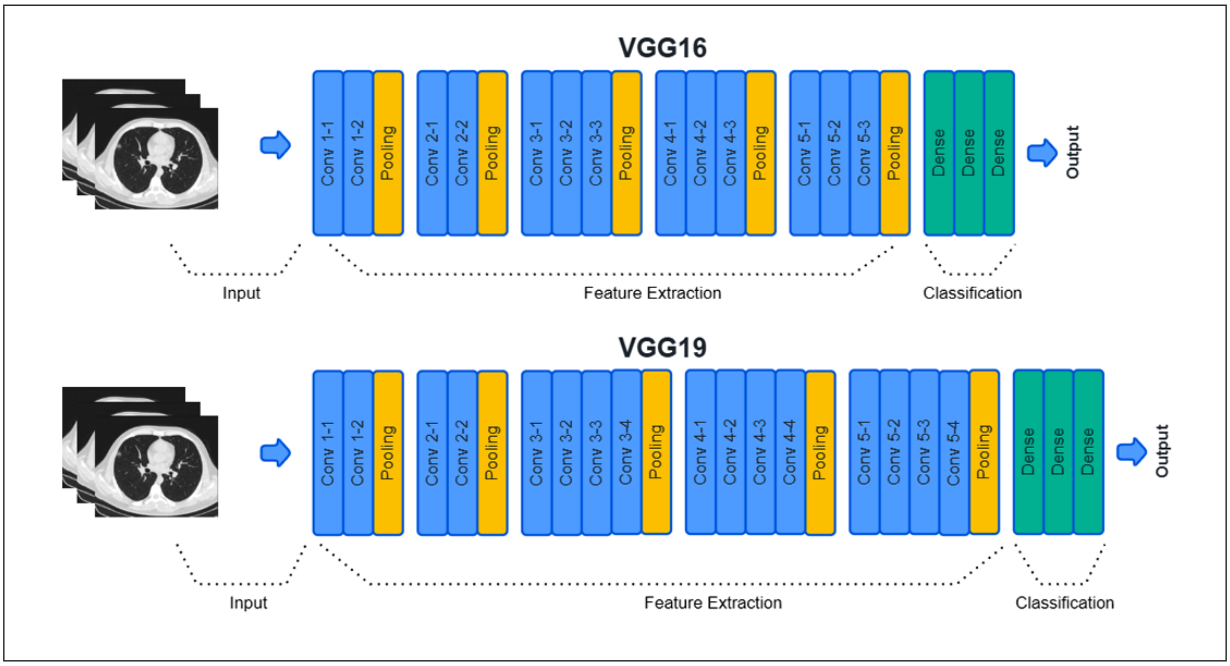

Figure 5. Overview of VGG16 and VGG19 architectures.

It covers the development of ResNet, one of the most popular architectures of deep learning. In the first version of ResNet (ResNet-18, ResNet-34), each block takes the output of the previous block as input and transfers the output directly to the next block with an additional "skip connection". This process complicates the training of deeper networks. He et al[20] proposes a new method called "identity mappings" to overcome this problem.

NASNetLarge

The NASNetLarge model, which is used based on NASNet, is designed to efficiently optimize hyperparameters (weights, layers, etc.) and architectural designs automatically. NASNetLarge is part of the NASNet model family and combines many different types of cells. These cells are made up of learnable blocks and are optimized for the function they are used for. As an illustration, normal cells carry out feature extraction for the following block, whereas reduction cells are used to decrease the depth of features and speed up feature extraction. Architecture, hyperparameters, and cell types are automatically optimized using a technique called NAS (Neural Architecture Search). This improves the model's efficiency and performance. It is also feasible to retrain the model for different tasks using pre-trained weights *via* transfer learning. It was said that it demonstrated great classification performance in experiments performed on the ImageNet (Stanford, CA, USA, https://www.image-net.org/index.php) dataset[21].

InceptionResNetV2

InceptionResNetV2 (https://keras.io/api/applications/inceptionresnetv2/) is a deep learning model developed by Google. This model includes the residual learning architecture (ResNet) features as well as the Inception network. As a result, a model with high success rates was developed, while it was also made faster and more efficient by using fewer parameters. The model is designed as a multilayer neural network and is often used for image classification tasks. Features of InceptionResNetV2 include filters at multiple scales, parallel structures, side-by-side connections, and short-path connections. The model has more layers and parameters than previous versions and gives better results on more complex datasets. However, these extra layers and parameters require more computational power, data, and training time. InceptionResNetV2 shows high success rates in a variety of tasks, especially complex tasks such as visual recognition, object detection, and segmentation. Therefore, it is used in many applications, especially in the fields of artificial intelligence and image processing[22].

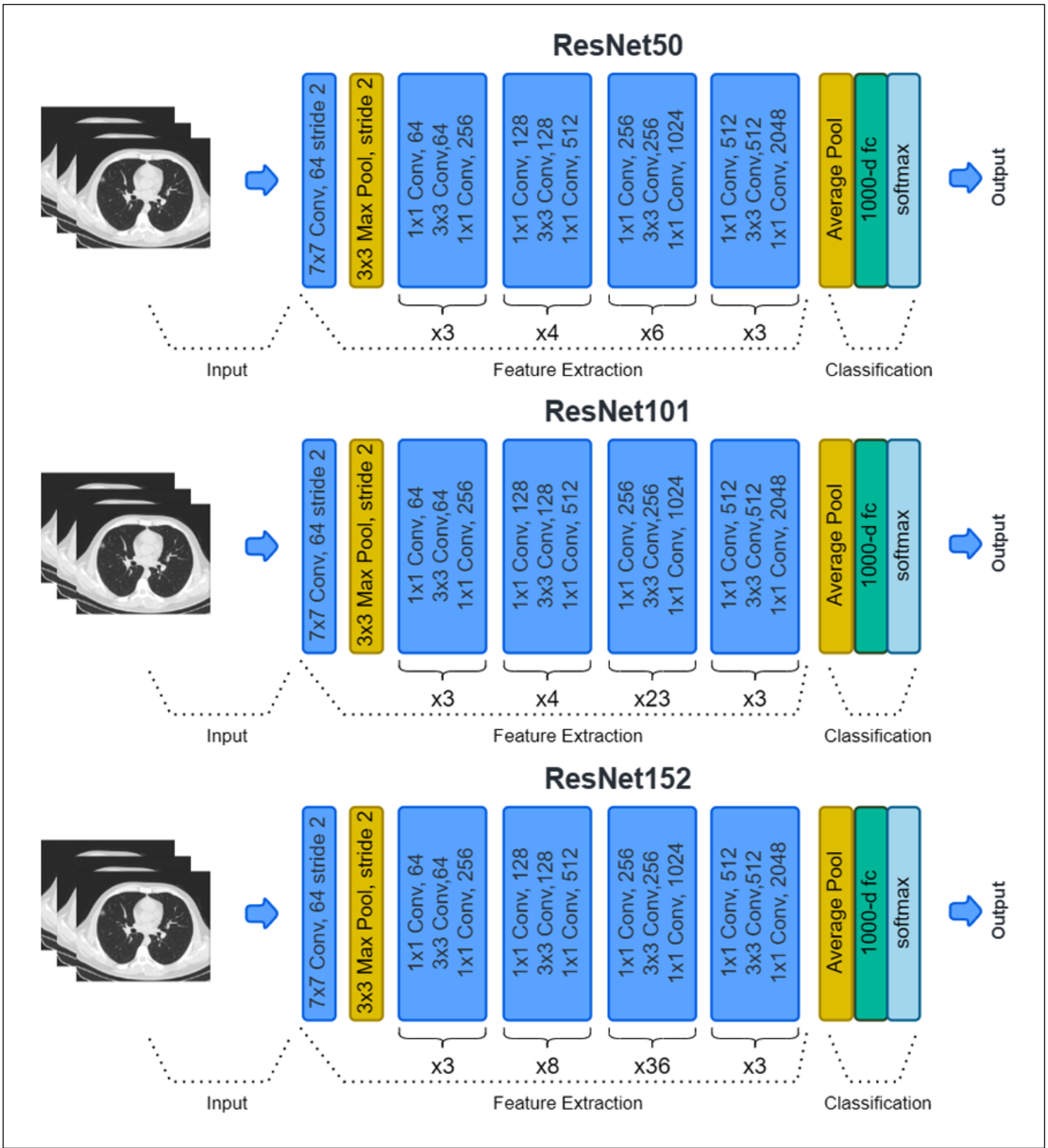

Figure 6. Overview of ResNet50, ResNet101 and ResNet152 architectures.

InceptionV3

Szegedy et al[23] proposed a better architecture by addressing the problems in previous versions of the Inception architecture. First, research has discovered that models with fewer parameters generally have higher accuracy rates. As a result, it has been discovered that it has a lighter architecture and hence has a favorable effect on the results. In addition, the slowdown of the learning rate, which is another problem encountered in the learning process of deeper networks, is also included in the study. To solve this problem, it is recommended to increase the learning rate by considering other layers besides the output, before making the network deeper. The article also examines various factors that are important for proper training of networks. The proposed architecture is named Inception V3, and high accuracy rates have been achieved in the ImageNet dataset. It is said that due to the lightness of the mesh, it is also possible to use it on less-equipped devices.

DenseNet121- DenseNet169- DenseNet201

Densely Connected Convolutional Networks (DenseNet) is a deep learning architecture published by Huang et al[24] in 2016. DenseNet is a more densely connected CNN architecture that employs a customized module. DenseNet comes in three different variations: DenseNet121, DenseNet169, and DenseNet201. These names indicate the number of layers for each model. DenseNet also makes training of the model more stable by using some techniques such as class-weighted cross-entropy and gradient clipping. DenseNet architectures are shown in Figure 7.

Evaluation Metrics

Precision is defined as the proportion of real positive samples to categorized positive samples. It

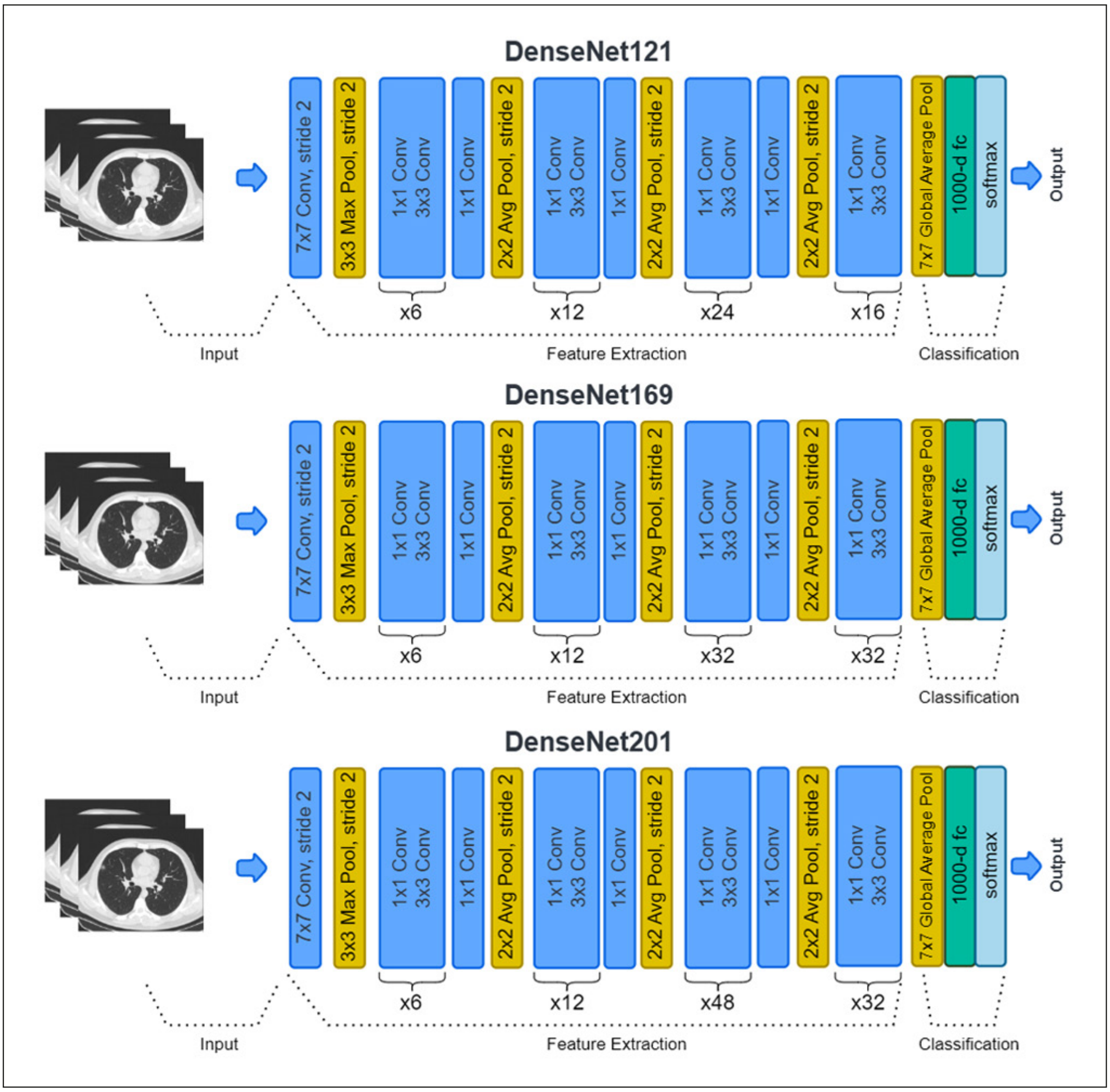

Figure 7. DenseNET architectures overview.

shows the percentage of true positive cases within the positive sample. Precision is concerned with lowering the quantity of false positive samples.

$$Precision = TP / (TP + FP) \tag{2}$$

Recall refers to the ratio of true positive samples to all positive samples. That is, it shows the success of detecting all true positive samples. Recall focuses on reducing the number of false negative samples.

$$Recall = TP / (TP + FN) \tag{3}$$

F1-score is the harmonic mean of precision and recall. If precision and recall are both high, the F1-score will also be high.

$$F1\text{-}score = 2 * (Precision * Recall) / (Precision + Recall) \tag{4}$$

Accuracy refers to the ratio of correctly classified samples to the total number of samples. While this metric is suitable for balanced datasets, it may be insufficient for unbalanced datasets.

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \tag{5}$$

Specificity refers to the proportion of true negative samples in all negative samples. That is, it shows the success of detecting all true negative samples.

$$Specificity = TN / (TN + FP) \tag{6}$$

The ROC curve is used to visually evaluate the performance of the classifiers. The ROC curve calculates the accuracy and precision of the classifiers using different threshold values. The area under the ROC curve (AUC) can be used as a metric to measure the performance of the classifier.

The metrics above are prevalent metrics used to measure the performance of classification models. These indicators are used to assess the study's success criterion.

Results

In the study, method approaches were examined using the same parameters. For this reason, the basic transfer learning steps have not been changed. The performances of the architectures were examined by changing only the optimizer and activation functions. Batch size 16, final dense layer 128, and image resolution 512x512 were used in all models. The best results were obtained with Stochastic Gradient Descent (SGD)[25]. In addition, Adam[26], Adagrad[27], and RMSprop[28] were tried to find the best optimizer. ReLU[29] and Leaky ReLU[30] gave the best results in terms of activation functions. Models were run at 50 epochs. The results are shown in Table I.

These results were obtained when the images were applied at the end of the pre-processing process and at a high resolution, such as 512x512. Since it is not possible to test all the parameters,

Table I. The results of the models at the end of the study.

	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	F1-Score (%)	ROC-AUC (%)
VGG16	88.9	88	96.7	72.4	92.1	91
VGG19	89.1	88.2	96.8	72.9	92.3	92
ResNet50	**95.7**	**94.7**	**99.2**	**88.3**	**96.9**	97
ResNet101	94.5	93.7	**99.2**	86	96.3	**98**
ResNet152	92.5	91.5	98	80.9	94.6	95
DenseNet121	88.5	86.2	98.6	67.1	92	91
DenseNet169	86.5	85.1	97	65.5	90.7	91
DenseNet201	86.5	85.6	97.9	62.6	90.7	90
InceptionResNetV2	78.5	78.6	93.8	46.4	85.5	85
Inceptionv3	84.7	81.5	98.3	53	89.1	87
NASNetLarge	80.4	84.2	87.5	65.5	85.8	84

ROC, Receiver Operating Characteristic; AUC, Area Under the Curve.

the augmented results were obtained by making corrections. The highest results were obtained with ResNet50.

We may compare the results of the operations conducted by looking at the confusion matrix and ROC Curve values. Based on our findings, the number of situations in which persons who were not genuinely sick were classified as patients (false positive) was fairly low. On the other hand, the rate of evaluation of people with COVID-19 disease as normal by the model is very low. Considering the ROC Curve values, if we evaluate the approach of the method to the problem, a very good ratio has been achieved (Figure 8). A comparison of all models is given in Figure 9.

Discussion

The main distinguishing feature of this study, as compared to others[6-13], is that it yields high-performance results. Finding a recall value of 99.2% clearly shows that the system is working according to a suitable principle. In addition, ROC-AUC scores also support this situation. Since datasets with low data are generally used in other studies, the results of success in solving the problem also include the luck factor. When datasets containing larger data are used, it is possible to obtain results by minimizing the chance factor. The images used in the study were improved by using image processing techniques. Masking and segmentation operations have made better feature extraction by removing unnecessary details and increasing the efficiency of the results found. The highest result was obtained with ResNet50. Confusion Matrix and ROC are given in Figure 8. All chest CT images of each patient were used in the complete dataset, and no images were extracted.

Singh et al[31] classified chest CT images with Artificial Neural Network (ANN), Adaptive Neural Fuzzy Inference System (ANFIS) CNN, and CNN models they created. They separated different percentages as test and train and applied these methods for each. According to the results obtained, approximately 93.5% success was achieved. The test data remained low due to the lack of images in the dataset. It is provided to calculate the performances from the given confusion matrices. Compared to our own study, a very low dataset was used, and no segmentation operations were performed. When performance metrics are compared, it is seen that they get lower results.

Wang et al[32] predict that artificial intelligence methods can be used to extract the radiological features of COVID-19 and provide a clinical diagnosis before pathogenic test results. As a dataset, 1,065 cases of COVID-19 and CT images of patients diagnosed with typical viral pneumonia were collected, and an algorithm was established by changing the transfer learning model. In the study, 89.5% accuracy, 88% specificity, and 87% sensitivity were obtained in internal validation, while external validation results were 79.3% accuracy, 83% specificity, and 67% sensitivity. A lower dataset than our study was used. Lower performance metric results were found. Masking is done and not segmented properly. Images were obtained using a threshold.

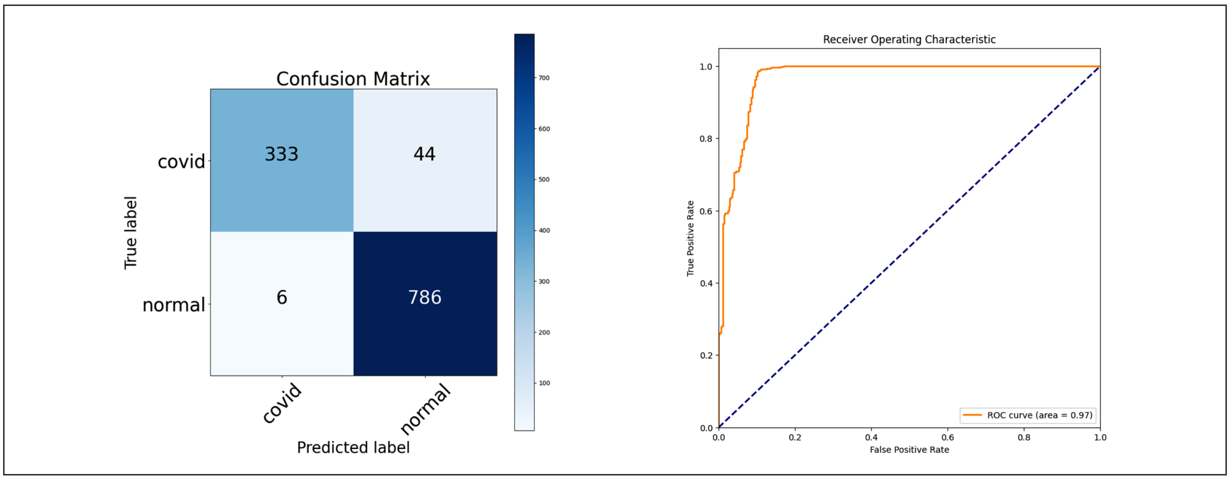

Figure 8. According to the results found, the confusion matrix and ROC-Curve.

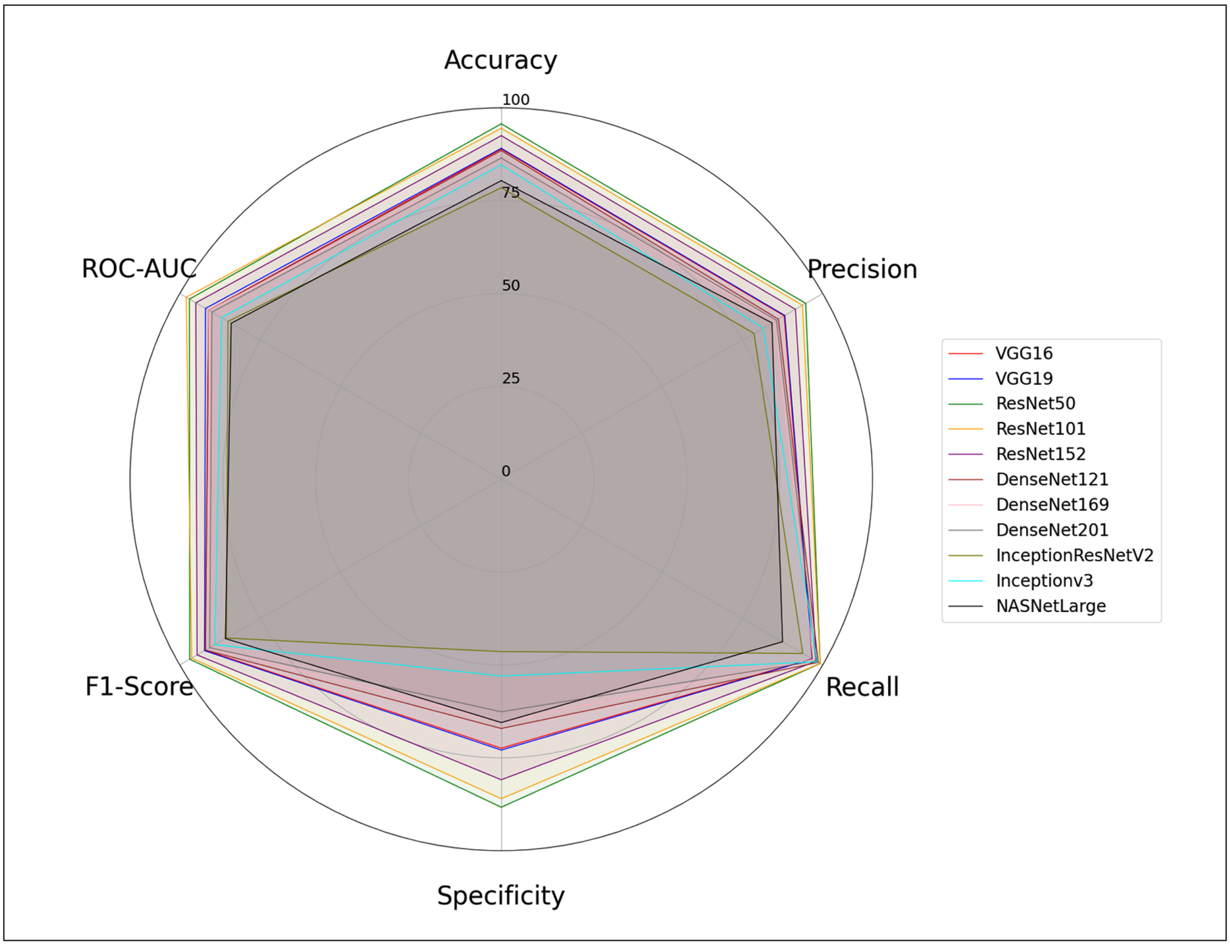

Figure 9. Comparison of all models.

Ardakani et al[33]'s paper highlights a fast and valid way for diagnosing COVID-19 utilizing artificial intelligence. A total of 1,020 CT scans of 108 patients from the COVID-19 group and 86 patients with atypical and viral pneumonia from the non-COVID-19 group were analyzed. 10 different artificial neural networks, namely ResNet-101 and Xception, were used. ResNet-101 was able to distinguish COVID-19 cases from non-COVID-19 cases with 99.02% specificity, while Xception was able to discriminate with 100% specificity. The performance of the radiologists was moderate, with a specificity of 83.33% and a sensitivity of 89.21%, with an AUC of 0.873. It got the highest results with ResNet. If we compare the studies, a lower dataset and the use of raw images are striking differences. Many types of architecture have been used and compared.

Vinod and Prabaharan[34] proposed a model based on information available from WHO, and chest X-ray images were collected from Kaggle and CT scan images. With the decision tree classifier, feature extractions of images are classified and separated for training and testing. CT scan images detected COVID-19 with a 93% recall and 82% accuracy. When the studies were compared, lower results were obtained due to differences such as low dataset usage, lack of masking, and segmentation.

In their work, Mei et al[35] used 512x512 pixel images and performed segmentation operations with CNN, Multilayer Perceptron (MLP), and Joint operations, achieving accuracy rates of 79.6%, 74.2%, and 83.5%, respectively. In this study, transfer learning was not used. When the differences, such as the differences in the dataset are taken into consideration, the results appear to be lower in our study.

In a study by Chen et al[36], transfer learning was used in conjunction with regional differences in CT images to obtain 86.8% success. Data augmentation was used in their study and com-

Table II. The Ct value of YAP1 was detected by qRT-PCR after transfection with si-YAP1.

Reference	Data	Method	Accuracy (%)
Singh et al[31]	Chest CT Scan	Custom CNN	~93.5
Wang et al[32]	Chest CT Scan	Modified *Inception* Transfer Learning	92.4
Ardakani et al[33]	Chest CT Scan	*ResNet50*	94.1
Vinod et al[34]	Chest CT Scan	Decision Tree	82
Mei et al[35]	Chest CT Scan	CNN, MLP, Joint	79.6, 74.2, 83.5
Chen et al[36]	Chest CT Scan	*ResNet50*	86.8
Carvalho et al[37]	Chest CT Scan	CNN, *XGBoost*	95
Gifani et al[38]	Chest CT Scan	Transfer Learning	85
Our Work	**Chest CT Scan**	**Transfer Learning**	**95.7**

CT, Computerized Tomography; CNN, Convolutional Neural Network; MLP, Multilayer Perceptron.

parisons were done utilizing a variety of transfer learning architectures. The ResNet-50 algorithm produced the best results, performing at a level of 87.3%. With the approach they advised, they were unable to pass the transfer learning test.

Carvalho et al[37] obtained 95% accuracy by using CNN and XGBoost in their study. A total of 708 images were studied, of which 312 were with COVID-19 and 396 were with non-COVID-19. Threshold is used, but segmentation operations are not performed. It includes using CNN to extract features from CT images with and without COVID-19 and using XGBoost for data classification. Results 95.07% accuracy, 95.09% recall, 94.99% sensitivity, 95% F-score, 95% AUC, and 90% Kappa values were found. In their study[37], a very low number of data was used compared to our study, and the results obtained are lower for these reasons.

According to Gifani et al[38] used 15 pre-trained convolutional neural network (CNN) architectures (EfficientNets(B0- B5), NasNetLarge, NasNetMobile, InceptionV3, ResNet-50, SeResnet 50, Xception, DenseNet121, ResNext50 and Inception_resnet_v2) were used and fine-tuned for the target task. Next, recognition performance is further improved by creating an ensemble method based on majority voting of the best deep transfer learning outcomes. The study shows that deep learning methods can be used successfully in the diagnosis of COVID-19, with the results obtained using many different architectures. Majority voting of 5 deep transfer learning models created with EfficientNetB0, EfficientNetB3, EfficientNetB5, Inception_resnet_v2 and Xception showed higher results in terms of sensitivity (0.854), accuracy (0.85), and precision (0.857) in the diagnosis of COVID-19 with data from CT scans compared to other models. Compared to our study, a dataset containing more data was not used. This can be explained by the lower results, as the images can remove unnecessary features because segmentation processes are not performed.

Cumulative Accuracy representations are shown in Table II.

Limitations

Our research has some limitations. Although the existence of a COVID-19 lesion can be seen in each tomography section, evaluating all sections does not allow for disease staging. In this regard, clinical assessment is required for confirmation. Using larger capacity computers, performance rate, and speed can be boosted. More accurate findings can be obtained by using computed tomography images with thinner slices and higher dpi values.

Conclusions

A total of 8,354 images were used in the study. Of these, 2,695 are composed of COVID-19, and the remaining are healthy chest CT images. All of these images were given to the models by masking and segmentation processes. Transfer Learning produces better outcomes while consuming less time. It can be implemented into big systems and produces results in a short period of time. A wide variety of pre-trained models were used. In this way, it has been determined which model works better for this problem. Compared with previous studies[6-13], the differences are revealed. As a result of the study, it was determined that the best model was ResNet50, and 95.7% accuracy, 94.7% precision, 99.2% recall, 88.3% specificity, 96.9% F1-Score, and finally, 97% ROC-AUC score were found. It is thought that the results obtained are satisfactory since it contains a large amount of data.

In future studies, a model will be designed that will label diseased regions with different disease and/or internal organ images and present the location and location of the diseased region to the user with performance information. For this, the differences will be revealed by using models with lightweight architecture and without using segmentation. Depending on the course of the study, different models will be tried and compared again.

Conflict of Interest

The authors declare that they have no conflict of interests.

Acknowledgements

The authors acknowledge all healthcare and secretariat staff.

Funding

No funding has been received for the study.

Informed Consent

Informed consent was obtained from all subjects involved in the study.

Authors' Contribution

Conceptualization, H.G.T and M.B.H.T.; methodology, H.G.T.; software, B.I.; validation, B.I., M.B.H.T and U.K.; formal analysis, U.K.; investigation, S.A.; resources, H.G.T..; data curation, M.B.H.T.; writing-original draft preparation, M.B.H.T.; writing-review and editing, B.I..; visualization, S.A.; supervision S.A..; project administration, H.G.T. All authors have read and agreed to the published version of the manuscript.

Data Availability

Data generated or analyzed during the study are available from the corresponding author by request.

Ethics Approval

Ethics committee approval number: EBYU-KAEK-2023/08/6, Date: 13.04.2023. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

ORCID ID

Hakan Gokalp Tas: 0000-0001-5680-9544
Mehmet Bilge Han Tas: 0000-0001-6135-1849
Baris Irgul: 0000-0002-7551-8686
Sonay Aydin: 0000-0002-3812-6333
Ufuk Kuyrukluyildiz: 0000-0001-6820-0699

References

1) Umakanthan S, Sahu P, Ranade AV, Bukelo MM, Rao JS, Abrahao-Machado LF, Dahal S, Kumar H, Kv D. Origin, transmission, diagnosis and management of coronavirus disease 2019 (COVID-19). Postgrad Med J 2020; 96: 753-758.
2) Stokes EK, Zambrano LD, Anderson KN, Marder EP, Raz KM, El Burai Felix S, Tie Y, Fullerton KE. Coronavirus Disease 2019 Case Surveillance - United States, January 22-May 30, 2020. MMWR Morb Mortal Wkly Rep 2020; 69: 759-765.
3) Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X, Cheng Z, Yu T, Xia J, Wei Y, Wu W, Xie X, Yin W, Li H, Liu M, Xiao Y, Gao H, Guo L, Xie J, Wang G, Jiang R, Gao Z, Jin Q, Wang J, Cao B. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020; 395: 497-506.
4) Kogilavani SV, Prabhu J, Sandhiya R, Kumar MS, Subramaniam U, Karthick A, Muhibbullah M, Imam SBS. COVID-19 Detection Based on Lung Ct Scan Using Deep Learning Techniques. Comput Math Methods Med 2022; 7672196: 1-13.
5) Akgül İ, Kaya APV, Ünver E, Karavaş E, Baran A, Tuncer S. Covid-19 detection on x-ray images using a deep learning architecture. J Eng Res 2022; 11: 15-26.
6) Gao Y, Mosalam KM. Deep transfer learning for image-based structural damage recognition. Comput -Aided Civ Infrastruct Eng 2018; 33: 748-768.
7) Aziz MA, Kanazawa H, Murakami Y, Kimura F, Yamaguchi M, Kiyuna T, Yamashita Y, Saito A, Ishikawa M, Kobayashi N, Abe T, Hashiguchi A, Sakamoto M. Enhancing automatic classification of hepatocellular carcinoma images through image masking, tissue changes and trabecular features. J Pathol Inform 2015; 6: 26.
8) Müller D, Rey IS, Kramer F. Automated chest ct image segmentation of covid-19 lung infection based on 3d u-net. arXiv preprint arXiv: 200704774 2020.
9) LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015; 521: 436-444.
10) Hinton GE, Osindero S, Teh YW. A fast learning algorithm for deep belief nets. Neural Comput 2006; 18: 1527-1554.
11) Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput 1997; 9: 1735-1780.
12) LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proc IEEE 1998; 86: 2278-2324.

13) Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Adv Neural Inf Process 2012; 25.

14) Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv: 14091556 2014.

15) Goodfellow I, Bengio Y, Courville A. Deep learning: MIT press; 2016.

16) Pan S, Yang Q. Engineering D. A survey on transfer learning. IEEE Trans Knowl Data Eng 2010; 22: 1345-1359.

17) Yosinski J, Clune J, Bengio Y, Lipson H. How transferable are features in deep neural networks? Adv Neural Inf Process 2014; 27.

18) Shin HC, Roth HR, Gao M, Lu L, Xu Z, Nogues I, Yao J, Mollura D, Summers RM. Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. IEEE Trans Med Imaging 2016; 35: 1285-1298.

19) He K, Zhang X, Ren S, Sun J, editors. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition; 2016.

20) He K, Zhang X, Ren S, Sun J, editors. Identity mappings in deep residual networks. Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part IV 14; 2016: Springer.

21) Zoph B, Vasudevan V, Shlens J, Le QV, editors. Learning transferable architectures for scalable image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition; 2018.

22) Szegedy C, Ioffe S, Vanhoucke V, Alemi A, editors. Inception-v4, inception-resnet and the impact of residual connections on learning. Proceedings of the AAAI conference on artificial intelligence; 2017.

23) Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z, editors. Rethinking the inception architecture for computer vision. Proceedings of the IEEE conference on computer vision and pattern recognition; 2016.

24) Huang G, Liu Z, Van Der Maaten L, Weinberger KQ, editors. Densely connected convolutional networks. Proceedings of the IEEE conference on computer vision and pattern recognition; 2017.

25) Bottou L, editor Large-scale machine learning with stochastic gradient descent. Proceedings of COMPSTAT'2010: 19th International Conference on Computational StatisticsParis France, August 22-27, 2010 Keynote, Invited and Contributed Papers; 2010: Springer.

26) Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv preprint arXiv:14126980 2014.

27) Duchi J, Hazan E, Singer Y. Adaptive subgradient methods for online learning and stochastic optimization. J Mach Learn Res 2011; 12: 2121-2159.

28) Tieleman T, Hinton G. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. COURSERA: Neural networks for machine learning 2012; 4: 26-31.

29) Nair V, Hinton GE, editors. Rectified linear units improve restricted boltzmann machines. Proceedings of the 27th international conference on machine learning (ICML-10); 2010.

30) Maas AL, Hannun AY, Ng AY, editors. Rectifier nonlinearities improve neural network acoustic models. Proc icml; 2013: Atlanta, GA.

31) Singh D, Kumar V, Vaishali, Kaur M. Classification of COVID-19 patients from chest CT images using multi-objective differential evolution-based convolutional neural networks. Eur J Clin Microbiol Infect Dis 2020; 39: 1379-1389.

32) Wang S, Kang B, Ma J, Zeng X, Xiao M, Guo J, Cai M, Yang J, Li Y, Meng X, Xu B. A deep learning algorithm using CT images to screen for Corona virus disease (COVID-19). Eur Radiol 2021; 31: 6096-6104.

33) Ardakani AA, Kanafi AR, Acharya UR, Khadem N, Mohammadi A. 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 2020; 121: 103795.

34) Vinod DN, Prabaharan SRS. Data science and the role of Artificial Intelligence in achieving the fast diagnosis of Covid-19. Chaos Solit Fractals 2020; 140: 110182.

35) Mei X, Lee HC, Diao KY, Huang M, Lin B, Liu C, Xie Z, Ma Y, Robson PM, Chung M, Bernheim A, Mani V, Calcagno C, Li K, Li S, Shan H, Lv J, Zhao T, Xia J, Long Q, Steinberger S, Jacobi A, Deyer T, Luksza M, Liu F, Little BP, Fayad ZA, Yang Y. Artificial intelligence-enabled rapid diagnosis of patients with COVID-19. Nat Med 2020; 26: 1224-1228.

36) Chen X, Yao L, Zhou T, Dong J, Zhang Y. Momentum contrastive learning for few-shot COVID-19 diagnosis from chest CT images. Pattern Recognit 2021; 113: 107826.

37) Carvalho ED, Carvalho ED, de Carvalho Filho AO, de Araújo FHD, Rabêlo RdAL, editors. Diagnosis of COVID-19 in CT image using CNN and XGBoost. 2020 IEEE Symposium on Computers and Communications (ISCC); 2020: IEEE.

38) Gifani P, Shalbaf A, Vafaeezadeh M. Automated detection of COVID-19 using ensemble of transfer learning with deep convolutional neural network based on CT scans. Int J Comput Assist Radiol Surg 2021; 16: 115-123.