Article

Detecting Coronavirus from Chest X-rays Using Transfer Learning

Abeer Badawi *[ID] and **Khalid Elgazzar** [ID]

Faculty of Engineering and Applied Science, Ontario Tech University, Oshawa, ON L1G 0C5, Canada; khalid.elgazzar@ontariotechu.ca
* Correspondence: abeer.badawi@ontariotechu.ca

Abstract: Coronavirus disease (COVID-19) is an illness caused by a novel coronavirus family. One of the practical examinations for COVID-19 is chest radiography. COVID-19 infected patients show abnormalities in chest X-ray images. However, examining the chest X-rays requires a specialist with high experience. Hence, using deep learning techniques in detecting abnormalities in the X-ray images is presented commonly as a potential solution to help diagnose the disease. Numerous research has been reported on COVID-19 chest X-ray classification, but most of the previous studies have been conducted on a small set of COVID-19 X-ray images, which created an imbalanced dataset and affected the performance of the deep learning models. In this paper, we propose several image processing techniques to augment COVID-19 X-ray images to generate a large and diverse dataset to boost the performance of deep learning algorithms in detecting the virus from chest X-rays. We also propose innovative and robust deep learning models, based on DenseNet201, VGG16, and VGG19, to detect COVID-19 from a large set of chest X-ray images. A performance evaluation shows that the proposed models outperform all existing techniques to date. Our models achieved 99.62% on the binary classification and 95.48% on the multi-class classification. Based on these findings, we provide a pathway for researchers to develop enhanced models with a balanced dataset that includes the highest available COVID-19 chest X-ray images. This work is of high interest to healthcare providers, as it helps to better diagnose COVID-19 from chest X-rays in less time with higher accuracy.

Keywords: COVID-19; AI models; coronavirus detection; transfer learning; deep neural networks; chest X-rays

1. Introduction

Coronavirus disease (COVID-19) is a serious and contagious disease that has spread around the world since December 2019 [1]. The Worldometer [2] is a website developed by a team of developers and researchers to provide information about world numerical statistics. The website reported a worldwide total of 183,194,939 cases and 3,966,106 deaths from COVID-19 pandemic by the end of June 2021. The symptoms of COVID-19 include fever, cough, dyspnea, and fatigue [3]. Nevertheless, one of the acute symptoms is feeling chest pain and difficulties in breathing [3]. Nearly half of the cases with COVID-19 have an abnormal chest X-ray [4]. Chest imaging represents a very important role in the early diagnosis and treatment of patients suspected for COVID-19. Chest X-ray imaging is used to efficiently screen the patient's chest [5]. Furthermore, the improvements in deep learning applications in recent years have helped in accurately detecting COVID-19 from the chest X-ray [6]. Deep learning is a type of machine learning that simulates how humans learn certain types of information. We use it to analyze and identify patterns from data such as radiological tasks. Deep learning algorithms show promising results in extracting information from medical images and X-rays [7]. Therefore, we highlight the use of deep learning models in detecting COVID-19 cases from chest X-ray images. The automated prediction of COVID-19 from a chest X-ray will help doctors instantly detect the disease and

take actions. This research proposes a new transfer learning method using deep learning by adding a head or final layer to fit three pre-trained models, namely, DenseNet201, VGG16, and VGG19. We evaluate the performance by calculating the accuracy, precision, recall, specificity, and F1-score for both binary and multi-class classification. The proposed framework points out that applying deep learning models on chest X-ray images obtains reliable results to predict COVID-19.

Furthermore, we evaluate our models on a balanced dataset that we collected from normal, pneumonia, and COVID-19 chest X-ray images. This dataset overcomes two major drawbacks in the previous works, which are an imbalanced dataset and a small number of COVID-19 images. The COVID-ChestXray-15k dataset is collected from eleven different sources with a total of 5000 images of normal chest X-ray, 5000 pneumonia, and 4420 COVID-19 images before data augmentation and 5000 images after data augmentation for a total of 15,000 images. In summary, the contributions of this paper are as follows:

- We propose new modified three pre-trained deep learning models with transfer learning based on Dense-Net201, VGG16, and VGG19 to detect COVID-19 from X-ray images.
- We introduce a balanced dataset named COVID-ChestXray-15k, collected from eleven available datasets. We also use different data augmentation techniques to create this balanced dataset by increasing the COVID-19 images from 4420 to 5000 images. This provides a dataset with a total of 15,000 images (5000 normal, 5000 pneumonia and 5000 COVID-19).

The remainder of this paper is structured as follows. The related work is highlighted in Section 2. Section 3 outlines the methodology of the proposed method. In Section 4, we show and compare the results of this work from different aspects. Section 5 discusses the results and highlights the major findings. Lastly, we conclude the research in Section 6 and propose future directions.

2. Related Work

Since the COVID-19 pandemic appeared, several kinds of research were published on medical analysis, artificial intelligence, and data mining related to COVID-19 and how to diagnose it. In this literature review, we will focus on two different aspects: the available chest X-ray images datasets to detect COVID-19 cases using binary classification (COVID-19 vs. non-COVID-19) and multi-class classification (COVID-19, pneumonia, and normal), and the state-of-the-art deep learning models used to classify the chest X-ray images. Several researchers have recently provided a survey discussing the published datasets and deep learning models for chest X-ray images [8,9]. Alafif et al. [9] summarizes the top performing machine and deep learning techniques for diagnosing COVID-19 using chest X-ray.

Abbas et al. [10] trained a DeTraCResNet18-based binary model with 196 images (COVID-19 = 105, normal = 80, SARS = 11) to detect COVID and the model achieved 95.12% accuracy, 97.91% sensitivity, 91.87% specificity, and 93.36% precision. Maguolo and Nanni [11] collected 339,271 images (144 COVID-19, 339,127 pneumonia) and applied the AlexNet algorithm to detect COVID and have achieved 99.97% AUC. Hall et al. [12] pre-trained ResNet-50 on a total of 455 chest X-ray (135 of COVID-19 and 320 of viral and bacterial pneumonia). They obtained 89.2% accuracy and 95% AUC. Authors in [13] collected 3905 X-ray images (450 COVID-19, 3455 non-covid) and classified them using pre-trained MobileNet-v2 to achieve 99.18% accuracy. Alam et al. [14] trained 5090 images (1979 COVID-19, 3111 normal) using (CNN+HOG) + VGG19 pre-trained model and have reached accuracy of 99.49%. In [15], the authors collected 6926 images (2589 COVID-19, 4337 normal) and used a convolutional neural network to classify the images and achieved 94.43% accuracy. In [16], the authors used 610 images (305 COVID-19 and 305 normal) to train a CNN model using transfer learning and achieved an accuracy of 97.4%. Authors in [17] collected 900 images (500 COVID-19, 400 normal) and proposed a CoreDet deep learning model that achieved 99.1% accuracy. In [18], the authors collected 3252 images

(371 COVID-19, 2882 normal) and used the AlexNet model to classify the images and achieved 99.16% accuracy.

The authors in [19] compared VGG16, VGG19, InceptionResNet V2, InceptionV3, and Xception. They collected 327 images (125 COVID-19, 152 normal, 50 pneumonia) and achieved an accuracy of 84.1%, with 87.7% sensitivity and 97.4% AUC. The authors in [20] introduced 16,756 chest X-ray scans (358 from COVID-19 patient cases, 8066 patient cases with no pneumonia, and 5538 patients cases who have non-COVID19 pneumonia). They trained COVID-Net to achieve 92.4% accuracy and 91.0% sensitivity. In [21], the authors collected 2971 images (285 COVID-19, 1341 normal, 1345 pneumonia) and used CNN to achieve 94.03% accuracy. Chowdhury et al. [22] trained 2905 images (219 COVID-19, 1341 normal, 1345 pneumonia) using Parallel-dilated CNN and achieved an accuracy of 96.58%. Murugan et al. [23] proposed a balanced dataset of 2700 images (900 COVID-19, 900 normal, 900 pneumonia) and also proposed a E-DiCoNet model that achieved 94.07% accuracy. The authors in [24] collected 6100 images (225 COVID-19, 1583 normal, 4292 pneumonia) and classified the images using CNN and achieved 98.50% accuracy. Hussain et al. [17] collected 1300 images (500 COVID-19, 400 normal, 400 pneumonia) and used CoreDet to achieve 94.2% accuracy for multi-class classification. Ibrahim et al. [18] also proposed a multi-class classification using 7331 images (371 COVID-19, 2882 normal, 4078 pneumonia) and trained AlexNet to achieve 94.00% accuracy.

Most of the works in the literature review have used chest X-ray images and deep learning to diagnose COVID-19. This highlights the importance of chest X-ray images in diagnosing COVID-19 and helping doctors to detect COVID from chest X-ray faster. However, we noticed many limitations in previous works, such as imbalanced datasets and the small number of COVID-19 images to classify, which significantly impacts the performance of these models and provides a false impression on their success. This work proposes a new pre-trained deep learning model with transfer learning that we apply on Dense-Net201, VGG16, and VGG19 with high performance results. Furthermore, we use data augmentation techniques to create a balanced dataset that overcomes the imbalanced datasets limitation. Lastly, to overcome the small datasets problem, we collect a large dataset with 15,000 images (5000 normal, 5000 pneumonia, and 5000 COVID-19) to classify the images using deep learning.

3. Materials and Methods

This section describes the dataset and discusses the data preprocessing steps and data augmentation techniques we used. We also explain the pre-trained deep learning models and the transfer learning architecture. Lastly, we explain the performance evaluation metrics we used to evaluate our models.

3.1. Dataset Description

In this study, we use chest X-ray images from normal, pneumonia and COVID-19 cases. We collect eleven publicly available sub-databases to create one database called the COVID-ChestXray-15k dataset with a total of 4420 COVID-19 images before data augmentation, 5000 Pneumonia images, and 5000 normal images. We combine and modify eleven different public data from ChestX-ray8 dataset [25], Chest X-ray Images (pneumonia) dataset [26], BIMCV-COVID19 dataset [27], COVID-19 Image Data Collection [28], Figure 1 COVID -19 Chest X-ray Dataset Initiative [29], ActualMed COVID-19 Chest X-ray Dataset Initiative [30], SIRM COVID-19 database [31], Twitter COVID-19 CXR Dataset [32], Covid19 Image Repository [33], COVID-CXNet [34], and MOMA- Dataset [35]. We choose these eleven datasets because they are open source and fully available to the researchers as shown in Table 1:

- Normal images:
  1—ChestX-ray8 dataset [25], with a total of 5000 images.
- Pneumonia images:

2—Chest X-ray Images (Pneumonia) dataset [26], with a total of 4237 images, and 763 images from ChestX-ray8 dataset. [25].

- COVID-19 images:
  3—BIMCV-COVID19 dataset [27], with a total of 2473 images.
  4—COVID-19 Image Data Collection [28], with a total of 208 images.
  5—COVID-19 data from Figure 1 COVID-19 Chest X-ray Dataset [29], with a total of 55 images.
  6—COVID-19 data from the ActualMed COVID-19 Chest X-ray Dataset [30], with a total of 238 images.
  7—SIRM database [31], with a total of 68 images.
  8—Twitter data [32], with a total of 37 images.
  9—COVID-19 Repository [33], with a total of 243 images.
  10—COVID-CXNet [34], with a total of 877 images.
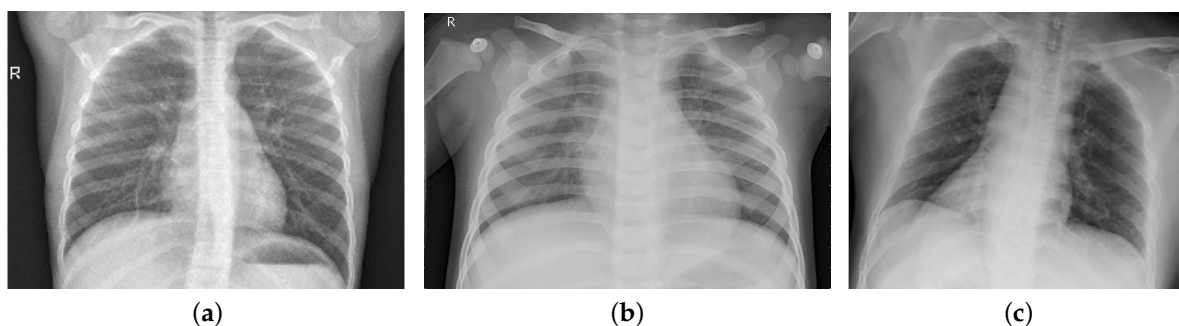 11—MOMA-Dataset [35], with a total of 221 images.

(**a**)   (**b**)   (**c**)

Figure 1. (**a**) Normal chest X-ray image from the dataset. (**b**) Pneumonia chest X-ray image from the dataset. (**c**) COVID-19 chest X-ray image from the dataset.

Table 1. COVID-ChestXray-15k dataset description.

Classes	Number of Images	Datasets
Normal	5000 images	[25]
Pneumonia	5000 images	[25,26]
COVID-19	4420 images (5000 after data augmentation)	[27–35]
Total	15,000 images	[25–35]

3.2. Data Preprocessing and Augmentation

We utilize data augmentation to improve the performance of the deep learning models for small datasets and create a balanced dataset. After data augmentation, the COVID-19 images increased to 5000 from 4420 COVID-19 samples. We use three image augmentation techniques (rotation, distortion, and flipping) to generate the images. The rotation operation for image augmentation is done by rotating the images clockwise and counterclockwise with a maximum of 10 degrees, and then we randomly distort the image. Lastly, we flip the images with a probability of 0.5 horizontally and vertically. An example of the dataset is shown in Figure 1 for normal, pneumonia and COVID-19 images. We apply some preprocessing steps before training the images to prepare it for classification. We convert the images to greyscale, resize them to 224×224 and convert them to the array dataset. To train the model, we divide the images into three classes, normal, Pneumonia, and COVID-19 images, with labels 0, 1, and 2. Lastly, we perform one-hot encoding on all the labels.

3.3. Pre-Trained Deep Learning Models

We chose three well known deep learning models as classifiers for our experiments: VGG16, VGG19, and DenseNet201. All the models are available in TensorFlow and Keras

libraries. We use these models as the base models and apply a new untrained head to each one of them. VGG16 [36] is a convolutional neural network (CNN) architecture with two convolution filter layers (3×3) and one pooling layer repeated three times. Then, three convolution filter layers (3×3) and one pooling layer were repeated two times. Lastly, the head of the architecture consists of three fully connected layers and SoftMax output. VGG19 [36] is a convolutional neural network (CNN) architecture with two convolution filter layers (3×3) and one pooling layer repeated three times. Then, four convolution filter layers (3×3) and one pooling layer were repeated two times. Finally, the head of the architecture consists of three fully connected layers and softmax output. DenseNet201 [37] is a densely connected convolutional Network. The layers in DenseNet have access to the original input image, which results in less computation. The architecture of DenseNet201 consists of four parts. The first one contains a 7×7 convolution layer followed by a 2×2 max-pooling layer, followed by a dense block of 1×1 convolution and 3×3 convolution repeated six times. The second part consists of a 1×1 convolution layer followed by a 2×2 max-pooling layer, followed by a dense block of 1×1 convolution and 3×3 convolution repeated 12 times. The third part contains the same layers as part two, but repeated 48 times. The fourth part consists of the same layers of parts two and three, but repeated 32 times. Classification layers or the head layer consist of 7×7 global average pooling and 1000 fully connected layers with SoftMax.

3.4. Transfer Learning

Transfer learning is one of the popularly used techniques nowadays in deep learning. It allows us to train small datasets with less time, and this is achieved by gaining information from pre-trained models on large datasets and transferring it to our model. This case occurs a lot in medical data such as images due to the small datasets available. By using transfer learning, we can train deep learning models on small datasets without overfitting. We remove the pre-trained network's final layers, which is important to fit with the new classification problem. Then, we replace it with new layers that fit with the new classes of our problem. We also adjust the average pooling to 4×4, fully connected network dimension to 64, and 0.5 dropout layer. The final layer consists of two class heads; the normal chest X-ray images and COVID-19 images with a binary cross-entropy loss function. Furthermore, we create another final layer consisting of three class heads for the normal chest X-ray images, pneumonia, and COVID-19 images with a categorical cross-entropy loss function. Figure 2 shows the head we added to the pre-trained models in detail.
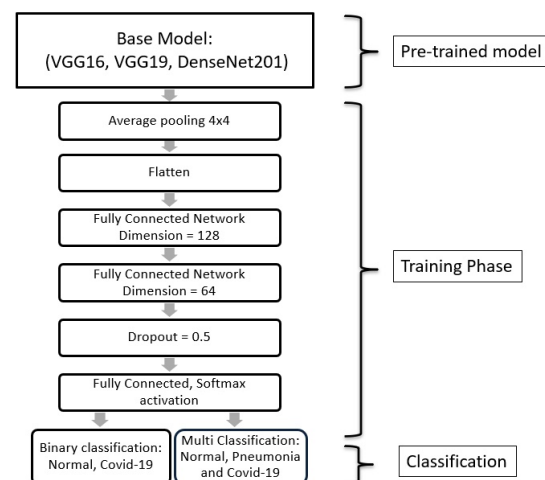

Figure 2. The proposed head architecture for the pre-trained models.

3.5. Performance Evaluation Metrics

We used a number of performance evaluation metrics to evaluate the performance of proposed models including accuracy, recall, precision, F1-score, and specificity, as shown

in Table 2. The true-positive and true-negative refer to the numbers of normal and COVID-19 images that are correctly classified. The false-negative and false-negative present the numbers of normal and COVID-19 images that have been wrongly classified.

Table 2. Performance metrics for the proposed model.

Performance Metric	Formula
Accuracy	$(TP + TN)/(TP + TN + FP + FN)$
Precision	$TP/(TP + FP)$
Recall	$TP/(TP + FN)$
F1-score	$2 * (Precision * Recall)/(precision + recall)$
Specificity	$TN/FP+TN$

4. Results

4.1. Experimental Setup

We use Python, TensorFlow, Keras, Sklearn, Open CV, matplotlib, Pandas, and NumPy libraries. We train all the models with 15 epochs. We optimize with Adam optimizer and learning rate of 0.0001, Batch size of 8. The machine we use to run all the codes is Intel with core i7 and an 8th generation CPU processor. We run the experiments for both binary and multi-class classifications to evaluate the performance on two different settings. The split rate of the data is 64% for training, 20% for testing, and 16% for validation dataset. We also make sure that we process each image in the pipeline exactly once, and we divide them between train, test, and validate sets.

4.2. Performance of Binary Classification

We train three pre-trained models with the new untrained head on the training and validation data using binary classification with class 0 indicating normal and class 1 for COVID-19. Figure 3 shows the plot of the accuracy and loss function on the training and validation data for the DenseNet201 model versus the number of epochs. We can observe some instability in the DenseNet201 model with a noticeable difference between the train and validation data outputs. The accuracy of the training data is 98.02%, and the loss is 0.029 at epoch 15. The best validation data accuracy is 94.66%, and the loss is 0.1324 at epoch 4. This result indicates that the model did not learn enough information while learning to predict the validation data. Figures 4 and 5 show the accuracy and loss results of the VGG16 and VGG19 models, respectively. The models show promising results and some stability between the train and validation data after epoch number 7. For the VGG16 model, the train accuracy is 99.30%, and the loss is 0.021 at epoch 15. The validation accuracy is 98.75%, and the loss is 0.036 at epoch 15. On the other hand, the VGG19 model achieves an accuracy of 99.02%, and the loss is 0.026. For the validation, accuracy is 98.59%, and the loss is 0.033. This result indicates that the model can successfully classify the validation data from learned information from training data.

Figure 3. Accuracy and loss function of the pre-trained model DenseNet201 on the training and validation datasets for binary classification.
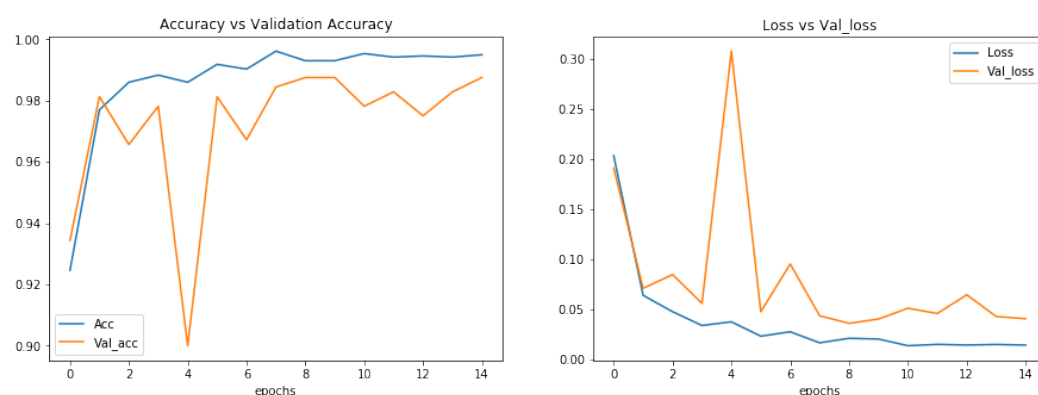

Figure 4. Accuracy and loss function of the pre-trained model VGG16 on the training and validation datasets for binary classification.

Figure 5. Accuracy and loss function of the pre-trained model VGG19 on the training and validation datasets for binary classification.

4.3. Testing Binary Classification

We evaluate all models on the test set and present the results in Table 3. The results of the DenseNet201 model are: 94.24% for precision, recall with 89.34%, F1-score with 91.72%, accuracy with 91.75%, and specificity 78.00%. The overall performance decreases when comparing validation and train sets with the test set data for the DenseNet201 model. This decrease is due to the specificity result, which indicates that the true negative prediction is low, reflecting on the total performance. The VGG16 shows promising results with a precision of 99.57%, recall with 99.64%, F1-score with 99.60%, accuracy with 99.62%, and specificity with 99.67%. Finally, the VGG19 model results 98.94% for precision, recall with 98.94%, F1-score with 98.94%, accuracy with 99.00%, and specificity with 98.66%. Both

models show stable results close to the train and validation set results. Considering the three models, VGG16 obtained the highest overall performance, with a slight difference from the VGG19 model.

Table 3. Performance evaluation of the three pre-trained models for binary classification.

Network	Precision	Recall	F1-Score	Accuracy	Specificity
DenseNet-201	94.24%	89.34%	91.72%	91.75%	78.00%
VGG16	**99.57%**	**99.64%**	**99.60%**	**99.62%**	**99.67%**
VGG19	98.94%	98.94%	98.94%	99.00%	98.66%

4.4. Performance of Multi Class Classification

We train DenseNet201, VGG16, and VGG19 models on the train and validation data using multi-class classification with class 0 to normal, class 1 to COVID-19, and class 2 to Pneumonia. Figure 6 shows the plot of the accuracy and loss function on the train and validation data for the DenseNet201 model versus the number of epochs. We also observe instability in the DenseNet201 model with a remarkable difference between the train and validation data outputs. The accuracy of the training data is 95.04%, and the loss is 0.015. The validation data accuracy is 85.15%, and the loss is 0.352. This result indicates that the model did not obtain sufficient information while learning to predict the new validation data. Figures 7 and 8 show the accuracy and loss results of the VGG16 and VGG19 models. For the VGG16 model, the Train accuracy is 96.40%, and the loss is 0.12. The validation accuracy is 94.25%, and the loss is 0.16. On the other hand, the VGG19 model achieves an accuracy of 94.72%, and the loss is 0.152. For the validation, accuracy is 94.03%, and the loss is 0.156. This result shows how the VGG16 and VGG19 models can obtain high results for the validation dataset from the training dataset. We assume that DenseNet201 performance is considered unstable compared to VGG16 and VGG19 because DenseNet201 contains less parameters, so the model needs more epochs to learn.
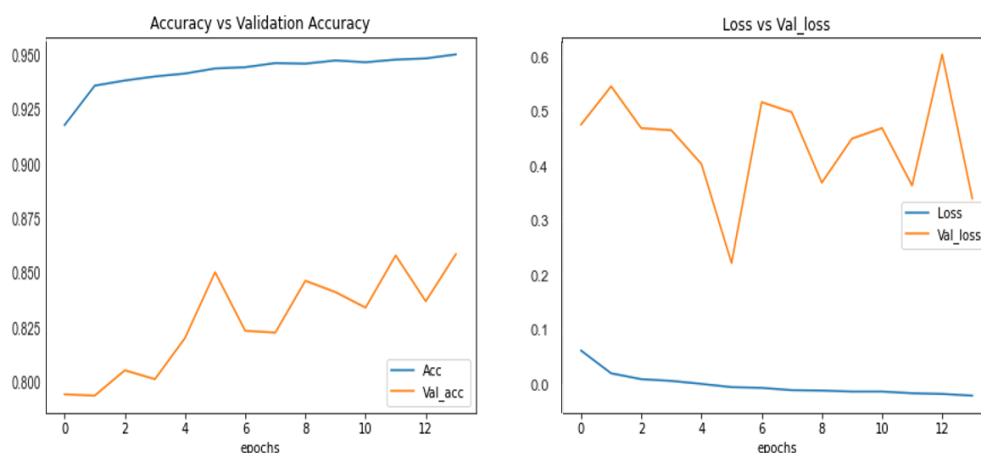

Figure 6. Accuracy and loss function of the pre-trained model DenseNet201 on the training and validation datasets for multi-class classification.
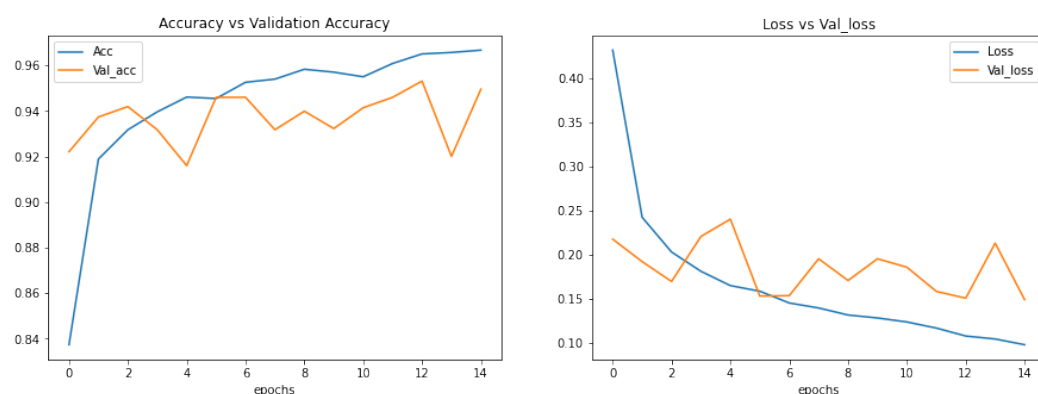
Figure 7. Accuracy and loss function of the pre-trained model VGG16 on the training and validation datasets for multi-class classification.
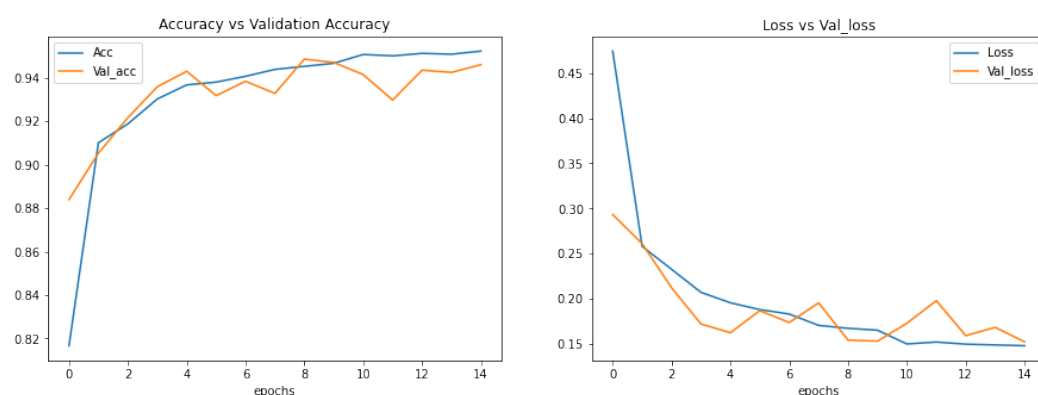

Figure 8. Accuracy and loss function of the pre-trained model VGG19 on the training and validation datasets for multi-class classification.

4.5. Testing Multi Class Classification

We evaluate the three models on the test set and present the results for the multi-class classification in Table 4. The results of the DenseNet201 model are: 94.07% for precision, recall with 88.30%, F1-score with 89.44%, accuracy with 91.97%, and specificity 86.30%. The overall performance increases when comparing validation set with the test set data for the DenseNet201 model. The VGG16 shows promising results with a precision of 95.48%, recall with 95.41%, F1-score with 95.41%, accuracy with 95.48%, and specificity with 95.37%. Finally, the VGG19 model results 95.01% for precision, recall with 95.41%, F1-score with 95.41%, accuracy with 95.48%, and specificity with 95.37%. Both models show stable results close to the train and validation set results. Considering the three models, VGG16 obtained the highest overall performance, with a slight difference from the VGG19 model. Furthermore, to ensure the efficiency of the model. We predict a random sample from the original dataset using the VGG16 pre-trained model. The results are shown in Figure 9, and the model accurately predicted all the random sample images as normal, pneumonia, or COVID-19 images, as shown in the true and predicted labels.

Table 4. Performance evaluation of the three pre-trained models for multi-class classification.

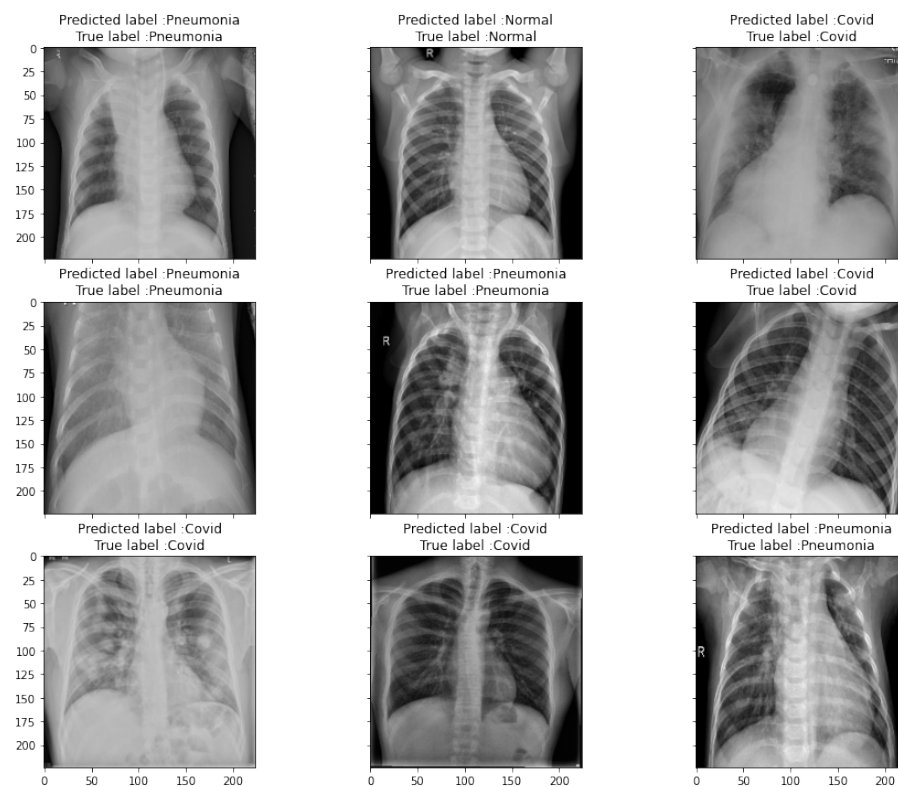| Network	Precision	Recall	F1-Score	Accuracy	Specificity
DenseNet-201	94.07%	88.30%	89.44%	91.97%	86.30%
VGG16	**95.48%**	**95.41%**	**95.41%**	**95.48%**	**95.37%**
VGG19	95.01%	94.95%	94.96%	95.03%	94.90%

Figure 9. Predicting a random sample from the original dataset using the VGG16 pre-trained model for multi class classification.

5. Discussion

As shown in Table 5, we compare our work to the state-of-the-art techniques found in the recent literature. We can claim with confidence that the proposed dataset presents the first balanced dataset with the largest number of COVID-19 cases. Imbalance in a dataset, especially if the number of COVID-19 images are small, does not provide a valid classification even if they obtained a high accuracy result. Only two authors [16,23] presented a balanced dataset, but with a small number of images compared to our dataset. It is also noticeable that our proposed transfer learning techniques produced the highest binary classification accuracy in detecting COVID-19 and normal images compared to the other techniques in the literature. However, for the multi-class classification, authors in [22,24] achieved higher accuracy, but the used datasets are imbalanced, and the COVID-19 images are 219 and 225, respectively. This indicates that the COVID-19 images are not enough to be correctly classified by the algorithm.

Our proposed research highlights two aspects that overcome the other recent works. First, we notice that the mentioned research contains a small number of COVID-19 images and imbalanced datasets from the collected datasets. This problem affects the results, especially to classify COVID-19 from other classes. Our paper directs this problem by creating a new balanced dataset called COVID-ChestXray-15k dataset with the highest COVID-19 images. This dataset contains COVID-19 X-ray images from eleven different sources, with a total of 5000 normal images, 5000 Pneumonia images, and 5000 COVID-19 images after data augmentation. Second, we introduce a transfer learning technique from different deep learning algorithm approaches with promising results. We train, validate, and test VGG16, VGG19, and DenseNet201 pre-trained deep learning models. We propose a final or head layer for the pre-trained models that fit our data and achieve high performance. We achieve the highest accuracy compared to the performance shown in the literature review with an accuracy of 99.62% for binary classification. For the multi-class classification, we obtained an accuracy of 95.48%. Even though we cannot compare the published research because each research uses a different dataset and algorithms, this

problem occurs due to the rapid change in the chest X-ray datasets available online every day. We provide this comparison to highlight the previous research in this area, clarify our paper contribution, and the enhancement we made compared to the previous work, and explain the limitations of that research to resolve them in our research. Lastly, this study is comparable to the state-of-the-art results and can be trustworthy for future work, as it obtained the results on a large and balanced dataset.

Table 5. Comparison between our work with the state-of-the-art work for the COVID-19 detection using chest X-ray images.

Classes	Reference	Dataset	Techniques	Accuracy
2	[10]	196 images (COVID-19 = 105, normal = 80, SARS = 11)	DeTraCResNet18	95.12%
Binary	[11]	339,271 images (COVID-19 = 144, pneumonia = 339,127)	AlexNet	-
Classification	[12]	455 images (135 of COVID-19 and 320 of pneumonia)	pre-trained ResNet-50	89.2%
	[13]	3905 X-rays (450 COVID-19, 3455 non-covid)	pre-trained MobileNet-v2	99.18%
	[14]	5090 images (1979 COVID-19, 3111 normal)	(CNN+HOG) + VGG19 pre-trained model	99.49%
	[15]	6926 images (2589 COVID-19, 4337 normal)	Convolutional neural network	94.43%
	[16]	610 images (305 COVID-19 and 305 normal)	Transfer learning with CNN	97.4%
	[17]	900 (500 COVID-19, 400 normal)	CoreDet	99.1%
	[18]	3252 images (371 COVID-19, 2882 normal)	AlexNet	99.16%
	Proposed	**10,000 (5000 COVID-19, 5000 normal)**	**Transfer Learning (VGG16, VGG19, DenseNet201)**	**99.62%**
3	[19]	327 images (COVID-19 = 125, normal = 152, pneumonia = 50)	VGG16, VGG19, InceptionResNet, InceptionV3, Xception.	84.1%,
Multi-class	[20]	16,756 images (358 COVID-19, 8066 no pneumonia, 5538 non-COVID19)	COVID-Net	92.4%
Classification	[21]	2971 images (285 COVID-19, 1341 normal, 1345 pneumonia)	CNN	94.03%
	[22]	2905 images (219 COVID-19, 1341 normal, 1345 pneumonia)	Parallel-dilated CNN	96.58%
	[23]	2700 images (900 COVID-19, 900 normal, 900 pneumonia)	E-DiCoNet	94.07%
	[24]	6100 images (225 COVID-19, 1583 normal, 4292 pneumonia)	CNN	98.50%
	[17]	1300 images (500 COVID-19, 400 normal, 400 pneumonia)	CoreDet	94.2%
	[18]	7331 images (371 COVID-19, 2882 normal, 4078 pneumonia)	AlexNet	94.00%
	Proposed	**15,000 (5000 COVID-19, 5000 normal, 5000 pneumonia)**	**Transfer Learning (VGG16, VGG19, DenseNet201)**	**95.48%**

6. Conclusions

The prediction of COVID-19 using chest X-ray prevents the spread of the disease on the chest and detects the virus faster. In this study, we train, validate, and test three popular deep learning algorithms with transfer learning. We test DenseNet201, VGG16, and VGG19 as pre-trained models to classify chest X-ray images of COVID-19. The results show that the VGG16 pre-trained model achieves the highest accuracy among the three models with an accuracy of 99.62% on the test set. Furthermore, we repeat the same steps with multi-class classification for normal, pneumonia, and COVID-19 images. As a result, we attain an accuracy of 95.48%. Furthermore, this study introduces the COVID-ChestXray-15k balanced dataset collected from eleven different sources with a total of 5000 normal, 5000 pneumonia images, and 5000 COVID-19 chest X-ray images after using data augmentation. This dataset includes a large number of COVID-19 images compared to previous research to overcome the imbalanced dataset problem. In light of our findings, the proposed dataset can help researchers train machine learning and deep learning models with a balanced dataset that includes a high quantity of COVID-19 images. Furthermore, the obtained results can assist specialists in detecting COVID-19 from the chest X-ray in an earlier stage to make decisions faster. Future directions involve increasing the number of dataset images if any open-source data is available as well as extending the proposed data to include other chest X-ray images from other types of diseases.

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

References

1. Paules, C.I.; Marston, H.D.; Fauci, A.S. Coronavirus Infections—More Than Just the Common Cold. *JAMA* **2020**, *323*, 707. [CrossRef] [PubMed]
2. Coronavirus Cases. Available online: https://www.worldometers.info/coronavirus/ (accessed on 1 July 2021).
3. Bell, D.J. COVID-19: Radiology Reference Article. Available online: https://radiopaedia.org/articles/COVID-19-4 (accessed on 1 July 2021).
4. Rousan, L.A.; Elobeid, E.; Karrar, M.; Khader, Y. Chest X-ray Findings and Temporal Lung Changes in Patients with COVID-19 Pneumonia. *BMC Pulm. Med.* **2020**, *20*, 1–9. [CrossRef]
5. Wong, H.; Lam, H.; Fong, A.H.; Leung, S.T.; Chin, T.W.; Lo, C.; Lui, M.M.; Lee, J.; Chiu, K.W.; Chung, T.W.; et al. Frequency and Distribution of Chest Radiographic Findings in Patients Positive for COVID-19. *Radiology* **2020**, *296*, E72–E78. [CrossRef]
6. McBee, M.P.; Awan, O.A.; Colucci, A.T.; Ghobadi, C.W.; Kadom, N.; Kansagra, A.P.; Trid, A.S.; Auffermann, W.F. Deep Learning in Radiology. *Acad. Radiol.* **2018**, *25*, 1472–1480. [CrossRef] [PubMed]
7. Kim, M.; Yan, C.; Yang, D.; Wang, Q.; Ma, J.; Wu, G. Deep Learning in Biomedical Image Analysis. In *Biomedical Information Technology*; Academic Press: Cambridge, MA, USA, 2020; pp. 239–263.
8. Islam, M.M. Karray, F.; Alhajj, R.; Zeng, J. A Review on Deep Learning Techniques for the Diagnosis of Novel Coronavirus (COVID-19). *IEEE Access* **2021**, *9*, 30551–30572. [CrossRef]
9. Alafif, T.; Tehame, A.M.; Bajaba, S.; Barnawi, A.; Zia, S. Machine and Deep Learning towards COVID-19 Diagnosis and Treatment: Survey, Challenges, and Future Directions. *Int. J. Environ. Res. Public Health* **2021**, *18*, 1117. [CrossRef]
10. Abbas, A.; Abdelsamea, M.M.; Medhat Gaber, M. Classification of COVID-19 in Chest X-ray Images Using DeTraC Deep Convolutional Neural Network. *Appl. Intell.* **2021**, *51*, 854–864. [CrossRef]
11. Maguolo, G.; Nanni, L. A Critic Evaluation of Methods for COVID-19 Automatic Detection from X-ray Images. *Inform. Fusion* **2021**, *76*, 1–7. [CrossRef]
12. Hall, L.; Goldgof, D.; Paul, R.; Goldgof, G.M. Finding COVID-19 from Chest X-rays Using Deep Learning on a Small Dataset. *arXiv* **2020**, arXiv:2004.02060.
13. Apostolopoulos, I.D.; Aznaouridis, S.I.; Tzani, M.A. Extracting Possibly Representative COVID-19 Biomarkers from X-ray Images with Deep Learning Approach and Image Data Related to Pulmonary Diseases. *J. Med. Biol. Eng.* **2020**, *40*, 462–469. [CrossRef]
14. Alam, N.A.; Ahsan, M.; Based, M.A.; Haider, J.; Kowalski, M. COVID-19 Detection from Chest X-ray Images Using Feature Fusion and Deep Learning. *Sensors* **2021**, *21*, 1480. [CrossRef]
15. Duran-Lopez, L.; Dominguez-Morales, J.P.; Corral-Jaime, J.; Vicente- Diaz, S.; Linares-Barranco, A. COVID-XNet: A custom deep learning system to diagnose and locate COVID-19 in chest X-ray images. *Appl. Sci.* **2020**, *10*, 5683. [CrossRef]
16. Mahmud, T.; Rahman, M.A.; Fattah, S.A. CovXNet: A multidilation convolutional neural network for automatic COVID-19 and other pneumonia detection from chest X-ray images with transferable multi-receptive feature optimization. *Comput. Biol. Med.* **2020**, *122*, 103869. [CrossRef] [PubMed]
17. Hussain, E.; Hasan, M.; Rahman, M.A.; Lee, I.; Tamanna, T.; Parvez, M.Z. CoroDet: A deep learning based classification for COVID-19 detection using chest X-ray images. *Chaos Solitons Fractals* **2021**, *142*, 110495. [CrossRef] [PubMed]
18. Ibrahim, A.U.; Ozsoz, M.; Serte, S.; Al-Turjman, F.; Yakoi, P.S. Pneumonia Classification Using Deep Learning from Chest X-ray Images During COVID-19. *Cogn. Comput.* **2021**, 1–13. [CrossRef]
19. Moutounet-Cartan, P.G.B. Deep convolutional neural networks to diagnose COVID-19 and other pneumonia diseases from posteroanterior chest x-rays. *arXiv* **2020**, arXiv:2005.00845.
20. Wang, L.; Lin, Z.Q.; Wong, A. COVID-Net: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest X-ray Images. *Sci. Rep.* **2020**, *10*, 1–12. [CrossRef]
21. Ahammed, K.; Satu, M.S.; Abedin, M.Z.; Rahaman, M.A.; Islam, S.M.S. Early Detection of Coronavirus Cases Using Chest X-ray Images Employing Machine Learning and Deep Learning Approaches. *medRxiv* **2020**. [CrossRef]
22. Chowdhury, N.K.; Rahman, M.M.; Kabir, M.A. PDCOVIDNet: A parallel-dilated convolutional neural network architecture for detecting COVID-19 from chest X-ray images. *Health Inf. Sci. Syst.* **2020**, *8*, 1–14. [CrossRef]
23. Murugan, R.; Goel, T. E-DiCoNet: Extreme Learning Machine Based Classifier for Diagnosis of COVID-19 Using Deep Convolutional Network. *J. Ambient. Intell. Humaniz. Comput.* **2021**, *12*, 8887–8898. [CrossRef] [PubMed]
24. Sekeroglu, B.; Ozsahin, I. Detection of COVID-19 from Chest X-ray Images Using Convolutional Neural Networks. *SLAS Technol. Transl. Life Sci. Innov.* **2020**, *25*, 553–565. [CrossRef]
25. Wang, X.; Peng, Y.; Lu, L.; Lu, Z.; Bagheri, M.; Summers, R.M. ChestX-ray8: Hospital-Scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.
26. Kermany, D.S.; Goldbaum, M.; Cai, W.; Valentim, C.C.; Liang, H.; Baxter, S.L.; Zhang, K. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell* **2018**, *172*, 1122–1131. [CrossRef] [PubMed]
27. BIMCV. Available online: https://bimcv.cipf.es/bimcv-projects/bimcv-covid19/#1590858128006-9e640421-6711 (accessed on 1 July 2021).

28. Cohen, J.P.; Morrison, P.; Dao, L.; Roth, K.; Duong, T.Q.; Ghassemi, M. COVID-19 image data collection: Prospective predictions are the future. *arXiv* **2020**, arXiv:2006.11988.
29. Agchung. Available online: https://github.com/agchung/Figure1-COVID-chestxray-dataset (accessed on 1 July 2021).
30. Agchung. Available online: https://github.com/agchung/Actualmed-COVID-chestxray-dataset (accessed on 1 July 2021).
31. Redazione. COVID-19 DATABASE. Available online: https://www.sirm.org/category/senza-categoria/COVID-19/ (accessed on 1 July 2021).
32. Twitter COVID-19 CXR Dataset. Available online: http://twitter.com/ChestImaging/ (accessed on 1 July 2021).
33. Winther, H.B.; Laser, H.; Gerbel, S.; Maschke, S.K.; Hinrichs, J.B.; Vogel-Claussen, J.; Meyer, B.C. COVID-19 Image Repository. *Figshare Dataset* **2020**. [CrossRef]
34. Armiro. Available online: https://github.com/armiro/COVID-CXNet (accessed on 1 July 2021).
35. Shams, M.; Elzeki, O.; Abd Elfattah, M.; Hassanien, A. Chest X-ray images with three classes: COVID-19, Normal, and Pneumonia. *Mendeley Data* **2020**, V3. [CrossRef]
36. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
37. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.