Transfer learning to detect COVID-19 automatically from X-ray images, using convolutional neural networks.

Mundher Taresh, Ningbo Zhu and Talal Ahmed Ali Ali

College of Information Science and Engineering, Hunan University, Hunan, 400013, Chang Sha, China Correspondence should be addressed to Mundher Taresh; mundhert@hnu.edu.cn

Emails: quietwave@hnu.edu.cn (Ningbo Zhu), taaw2012@hnu.edu.cn (Talal Ahmed Ali Ali)

Abstract

Novel coronavirus pneumonia (COVID-19) is a contagious disease that has already caused thousands of deaths and infected millions of people worldwide. Thus, all technological gadgets that allow the fast detection of COVID- 19 infection with high accuracy can offer help to healthcare professionals. This study is purposed to explore the effectiveness of artificial intelligence (AI) in the rapid and reliable detection of COVID-19 based on chest X-ray imaging. In this study, reliable pre-trained deep learning algorithms were applied to achieve the automatic detection of COVID-19-induced pneumonia from digital chest X-ray images.

Moreover, the study aims to evaluate the performance of advanced neural architectures proposed for the classification of medical images over recent years. The data set used in the experiments involves 274 COVID-19 cases, 380 viral pneumonia, and 380 healthy cases, which was derived from several open sources of X-Rays, and the data available online. The confusion matrix provided a basis for testing the post-classification model. Furthermore, an open-source library PYCM was used to support the statistical parameters. The study revealed the superiority of Model vgg16 over other models applied to conduct this research where the model performed best in terms of overall scores and based-class scores. According to the research results, deep Learning with X-ray imaging is useful in the collection of critical biological markers associated with COVID-19 infection. The technique is conducive for the physicians to make a diagnosis of COVID-19 infection. Meanwhile, the high accuracy of this computer-aided diagnostic tool can significantly improve the speed and accuracy of COVID-19 diagnosis.

Keywords: Covid-19 · classification · X-ray image · Deep learning

Introduction

The ongoing COVID-19 pandemic is a continuing coronavirus disease pandemic of 2019 (COVID-19) caused by extreme acute respiratory coronavirus syndrome 2 (SARS-CoV-2). Back in December 2019, the outbreak was first reported in Wuhan, China. Afterwards, it was declared a global public health emergency by the World Health Organization on January 30th, 2020, and then a pandemic on March 11th of the same year [1].

Governments are scrambling to shut down borders, monitor communications closely, trace the infected, isolate the suspected cases. Nevertheless, the number of people getting affected by the virus is still soaring in most countries, and it is expected to continually increase before the medicine/vaccine is made available after numerous clinical trials. As for such a situation, the right situation must be understood to make the right decisions. Therefore, multiple testing is a priority to be addressed and has already started in most countries.

Nevertheless, it can be appreciated that these experiments are vitally important, but it takes time to be performed with absolute accuracy. It has the potential to pose a risk, which is because if the infected are not detected promptly, it leads to passing on the infection to others, which could lead to an explosive rise. It can result in devastation, especially in those densely-populated countries.

The standard real-time COVID-19 test is called RT-PCR (Polymerase chain reaction) test that is purposed to determine the presence of antibodies against the virus [2]. Furthermore, The molecular testing of respiratory samples is recommended for the identification and laboratory confirmation of COVID-19 infection. However, it takes much time and likely to produce false-negative outcomes, as well [3]. Meanwhile, large-scale COVID-19 tests cannot be conducted in many developing countries due to its high cost. Where the immediate diagnosis depends on the symptoms appear.

Nevertheless, the work currently carried out allows the disease to be diagnosed in a way that is cheaply affordable, fast, and effortless for clinical application. Besides, it shows various advantages, the most important of which is that it removes the possibility that medics come into contact with the source of infection. This is particularly important amid the fast spread of the disease, especially in those countries with weak healthcare conditions. Even if the X-ray image does not decide the correct treatment, initial screening of cases is of benefit in a timely application of quarantine. Thus, the development of rapid and accurate diagnostic methods to contain the disease has become an urgent need.

Artificial Intelligence (AI) has recently been widely employed to accelerate biomedical research. AI was used in many applications, such as image detection, data classification, image segmentation, using deep learning approaches [4, 5]. People infected with COVID-19 may suffer from pneumonia as the virus spreads to the lungs. Numerous profound learning studies have detected the disease using a chest X-ray imaging approach. [6]

The American College of Radiology (ACR) advised against the use of CTs and x-rays as a first-line diagnostic or screen tool for COVID-19 diagnosis. It was indicated that images could only show the signs of an infection. These symptoms may be triggered by other factors [7]. However, there have been plenty of studies where artificial intelligence was applied to test COVID19 based on chest X-ray images [8–12]. Despite the satisfactory results achieved, the dilemma is that researchers have mixed cases of healthy and deficient cases of pneumonia, since the model then tends to ignore the contrast between groups between these two groups, and the accuracy achieved will not be a reliable test. When the 'Healthy and 'Pneumonia' classes are combined as a single class, the separability will disappear, thus making the findings misleading. In this case, validation is required if the groups are combined.

Therefore, the combination of healthy and pneumonia cases is not considered as an appropriate decision. Currently, the major challenge is to develop an algorithm that is capable of identifying

a patient with COVID-19 infection by examining chest x-ray images with viral pneumonia. [10, 13]. Nowadays, many radiology images have been commonly used for the identification of COVD19. Hemdan et al. [11] used deep learning models in X-ray images to diagnose COVID-19 and suggested a COVIDX-Net model consisting of seven CNN models.

Wang and Wong [12] presented a deep residual architecture called COVID-Net .it is one of the early works done on COVID-19, which uses a deep neural network to classify chest X-ray images into three categories (COVID-19, Healthy, Non-COVID-19). COVID-Net achieved an accuracy of 92.4%. Ioannis et al. [10] evaluated various state-of-art deep architectures on chest X-ray images. With transfer learning their best model, VGG19 managed to achieve an accuracy of 93.48% and 98.75% for 3-class and 2-class classification tasks respectively on a dataset consisting. Narin et al. [9] experimented with three different CNN models (ResNet50, InceptionV3, and Inception- ResNetV2), and ResNet50 achieved the best accuracy of 98% for 2-class classification. Since they did not include pneumonia cases in their experiment, it is unknown how well their model would distinguish between COVID-19 and other pneumonia cases. Ozturk et al. [14] proposed a CNN model based on DarkNet architecture to detect and classify COVID-19 cases from X-ray images. Their model achieved binary and 3-class classification accuracy of 98.08% and 87.02%, respectively, on a dataset consisting of 125 COVID-19, 500 Pneumonia, and 500 healthy chest X-ray images. Li and Zhu [15] presented a novel mobile AI approach for CXR based COVID-19 screening called COVID-MobileXpert to be reliably deployed at mobile devices for point-of-care testing. Afshar et al. [16] proposed a framework based on Capsule Network, known as the COVID-CAPS, for COVID-19 identification using X-ray images. The proposed COVID-CAPS achieved 95.7% accuracy, 90% sensitivity, 95.8% specificity, and 0.97Area Under the Curve (AUC). Farooq and Hafeez [17] presented COVID-ResNet for the classification of COVID-19 and three other infection types. COVID-ResNet was trained on a publicly available dataset. Ferhat Ucar, [18] introduced a COVID-19 detection AI model, COVIDiagnosis-Net, based on deep SqueezeNet with Bayes optimization. The implemented deep learning model has obtained an accuracy performance of 98.3%. Asif Iqbal Khan [19] proposed an in-depth learning approach to detect COVID-19 cases from chest radiography images. The proposed method (CoroNet) is a convolutional neural network designed to identify COVID-19 cases using chest X-ray images. The experimental results indicated that the suggested model achieved an overall accuracy of 89.6%. Suat Toraman et al. [20]proposed a Convolutional CapsNet for the detection of COVID-19 disease by using chest X-ray images with capsule networks. In this study, we used COVID-19 chest images data set, viral pneumonia chest images, and healthy chest images, to evaluate the effectiveness of the state-of-the-art pre-trained Convolutional Neural Networks with regard to the automatic diagnosis of COVID-19 from chest X-rays. An automated prediction of COVID-19 was proposed that applied pre-trained transfer models on Chest X-ray images based on a deep convolution neural network.A series of 1034 chest X-rays images are stored and used for training and evaluation of the CNNs to achieve such a purpose. As the size of the COVID-19 related samples is small (274 images), transfer learning is considered to be a preferred strategy for training the deep CNNs. A combination of pre-trained models VGG16, DenseNet121, InceptionV3, InceptionResNetV2, MobileNet, DenseNet169, NASNetLarge, and Exception were employed to achieve a higher prediction accuracy for a small X-ray dataset.

Methodology

Our methodology is illustrated in Figure 1. It involves the following steps:

chest X-ray image, transfer learning neural networks (VGG16, DenseNet121, InceptionV3, InceptionRes- NetV2, MobileNet, DenseNet169, NASNetLarge, and Exception), and feature

extraction. Then, it will be explained in more detail in the following subsections.

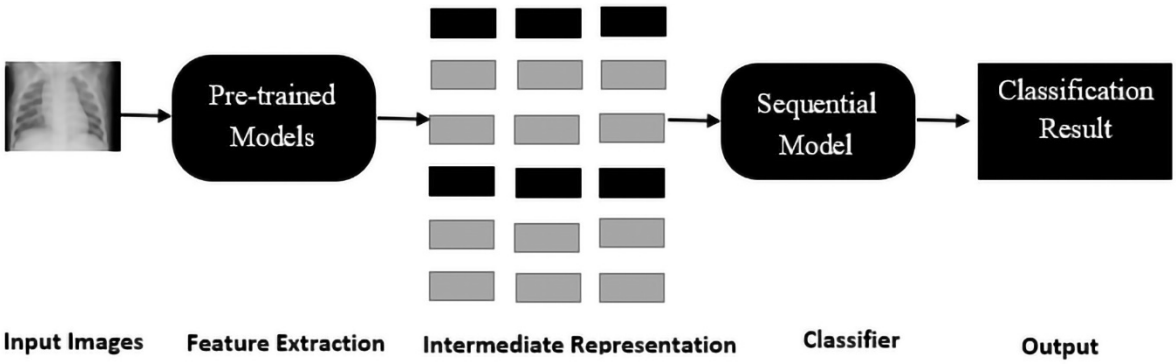

Figure 1: Outline of the methodology

Dataset of the study

Multiple sources of X-Rays were obtained from open sources [21], and the data made available online [22] for research purposes. The study was conducted by detailing the dataset for the confirmed COVID-19 cases, the cases of viral pneumonia infection (excluding COVID-19), and 'healthy' cases. The study goal was addressed by running Convolution Neural Networks on the problem of classification and construction of the appropriate algorithm. While the algorithm is required to be highly accurate as the health of people are endangered. The data set applied in the experiments involves 274 COVID-19 cases, 380 viral pneumonia cases, and 380 healthy. The data set was prepared and verified as reliable by reviewing it with chest specialists. Taking into account those cases of viral pneumonia should be free of any COVID-19 cases.

In summary, the collected COVID-19 cases are considered ideal for case studies. The X-ray images were rescaled to 224 × 224. CNN is capable of ignoring the insignificant variations in position. It searched for the patterns not only to a specific position of the image but also to moving patterns. The dataset was split into two sets of training (70%) and validation (30%), where the training set was applied to train the classifiers, and the validation set was applied to choose the most satisfactory performance.

Pre-trained model

In this study, the transfer learning technique was applied that was introduced by using ImageNet data to resolve inadequate data and the time for preparation. The weights trained on ImageNet were downloaded for each model. The feature maps were treated as input size in the applied layers training process. Besides, since the convolution base was run on the small data set and the extracted features were taken as input, it worked in a highly efficient way. Thus, for fine-tuning, a brief description was made of the CNNs employed for automatic detection. Table 1 shows the CNNs applied for the classification function and criteria for transfer learning. The parameters were determined after several experiments. The number of possible choices was limitless; their contribution to improving efficiency could be explored in future research. The parameter called Layer frozen refers to the number of untrainable layers starting from the bottom of the CNN, which is good because their weights are not expected to change during the process of model training. The last activation feature map in the pre-trained model provides us with the bottleneck features, which can then be flattened and fed into a fully connected deep neural network classifier. The other layers closer to the output features were trained to allow the extraction of more information from the late coevolutionary layers. In particular, the rectified linear unit (ReLU) activated all of the convolutions layers [23]. A dropout layer [24] was added to prevent the occurrence of overlapping [25] for neural networks using tow hidden layers. The CNNs were

compiled using the RMSprop(lr=1e-5) optimization method [26]. The training lasted fifteen epochs, with the batch size set to 32. The consumed time to start the first epoch and the total training time for each model are shown in Table 1.

Table 1: The parameters of CNNs and computational time in seconds.

Pre-trained models	Frozen layers	Time for training	
		first epoch	Total
InceptionV3	230	5	23
Exception	116	5	61
InceptionResNetV2	682	14	84
MobileNet	67	2	30
VGG16	18	3	45
DenseNet169	575	12	68
NASNetLarge	819	20	146
DenseNet121	407	9	65

METRICS

In this paper, the performance of classification models for identification COVID-19+ based on popular pre-trained models was evaluated. The proposed deep transfer learning models were trained separately using the Python programming language. All experiments were conducted with Tesla K80 GPU graphics card on Google Collaboratory with Windows 10 operating system. In this study, the confusion matrix provides a common basis for the performance of classifiers to be evaluated. The literature on performance metrics based on confusion matrices is plentiful and diversified and includes both frequent proposals for new statistics and the development of statistical models for their estimation. Here, we introduce an open-source Python library known as PyCM [27]. It is not only a Python-based multi-class confusion matrix library purposed to support both the input and direct matrix data vectors but also a useful tool for evaluating the post-classification model that supports overall statistics parameters and class-based statistics parameters.

Confusion matrix

A confusion matrix was introduced to analyze whether the prediction is consistent with the actual results. The confusion matrix is an effective method in assessing the classifier for its performance in classifying multi-class objects. This study focused on the general properties of the learning algorithm to address the problems with multi-class classification and measure quality with the confusion matrix. The instances in a predicted class represent each row of the matrix, while each column represents the instances in an actual class. The confusion matrix is regarded as one of the accurate measurements that provide more insight into the achieved validation accuracy. The three classes are investigated with the eight types of deep transfer learning. Nevertheless, such an assessment remains unclear, and the need for quantitative evaluation cannot be avoided [28].

Overall performance statistics

The most used metric for reporting multi-class classification output is its sample accuracy (ACC), that is defined as the number of correct predictions in all classes k, as divided by the number of examples, n. Despite the conceptual simplicity, it is known that the assessment of performance using sample accuracy alone is likely to result in misinterpretation. Since the

accuracy does not take into account the degree of a class difference [29–32], it is limited to being interpreted concerning accuracy based on a data set. The average F1 scores per-class, known as macroF1, provide a commonly used way to solve the constraint mentioned above. Multi-class and multi-label classification problems are often assessed by the "Macro F1" metric [33] and are computed as simple arithmetic means. As one of the dominant metrics in the remote sensing region, the Kappa coefficient [34] provides another solution to overcome the limitation on sample accuracy. For a given confusion matrix, the degree of general agreement is quantified $C \in G^{M\times M}$.

As with the $F1_{macro}$ and K_c, the class imbalance of the data is taken into account. Nonetheless, the number of errors may be invariant and do not necessarily represent one intuitively considered predictive power [35]. An alternative is a macro-averaged accuracy (Acc_{macro}) [36], which is defined as the arithmetic average of the partial accuracies of each class. Besides, other metrics were also computed, such as Overall Matthews Correlation Coefficient ($MCC_{overall}$), which can be extended to multiple categories [37, 38], Hamming loss ($L_{Hamming}$), which is the fraction of wrong labels to the total number of labels, and true negative rate (TNR_{macro}).

Class-based performance statistics

Since this study is mainly purposed to support the detection of COVID-19 infection, the accuracy is related only to COVID-19. Based on those metrics, what was calculated includes the accuracy(ACC), MCC, AUC, Geometric mean of specificity and sensitivity (GM), Error Rate(ERR), and the specificity (TNR)of the model.

Precision-recall metrics

Precision-Recall's metric was also employed to estimate the quality of output for the classifier. Precision-Recall curves are deemed more informative when binary classifiers are evaluated on imbalanced data sets using such performance measures as precision and recall metrics. A high area under the curve of a precision-recall curve can be detected with either high precision or high recall, suggesting either a low false-positive rate or a low false-negative rate. The high scores for both indicate that the classifier is restoring not only accurate results (high precision)
but also a majority of all positive results (high recall)—moreover, the higher f1-score, the more consistent the classification model. Given the limitation on single metrics-precision, recall, and f1-score, an average precision score and precision-recall to each class were adopted to assess the overall capacity. Herein, average precision (AP) is involved in measuring the classifier for its accuracy using a weighted mean of precision achieved at each threshold. Furthermore, the output is binarized if the precision-recall curve and average precision were extended to multi-class classification. The precision-recall curve can be plotted along with F1-score ISO curves by considering each element of the label indicator matrix, which is regarded as a binary prediction (micro-averaging).

Results And Discussion

The confusion matrix is shown in Figure 2, based on the unseen images for (COVID-19) out of 274 instances of images, (healthy) out of 380 instances of images, and (viral pneumonia) out of 380 instances of images. Both false-negative and false-positive could affect medical decisions negatively. A false-positive result is produced when an individual is inaccurately assigned to a class, such as a healthy individual categorized as COVID-19 patient. False-negative results when an individual falling into a given class is excluded from such a group. As confirmed by the confusion matrix results, there was a consistency between the predicted and actual results,

implying a better performance of the model in the classification of multi-class objects, as shown in Fig. 2.

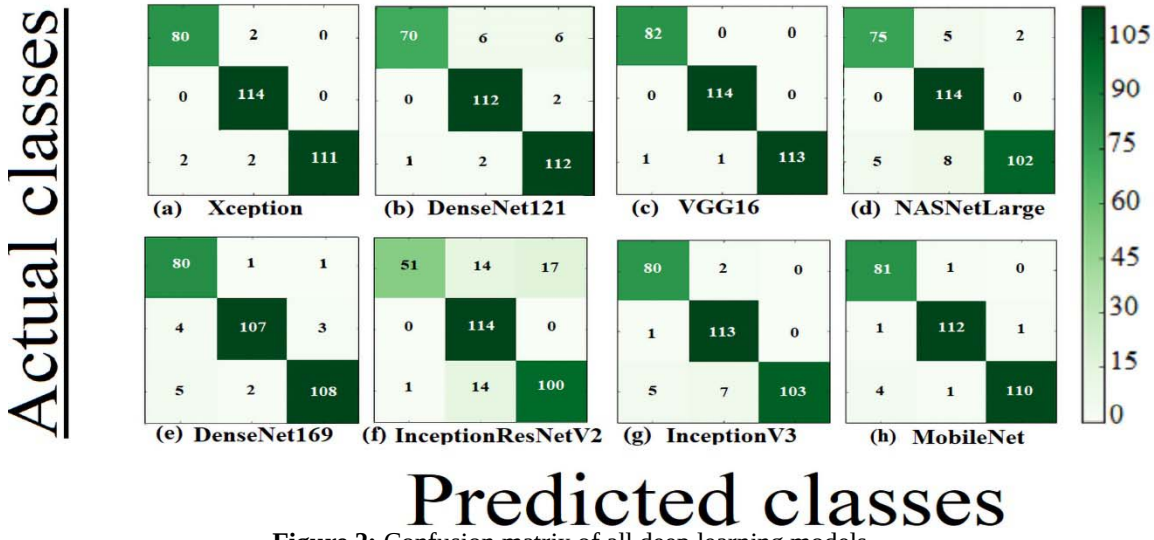

Figure 2: Confusion matrix of all deep learning models

Furthermore, the performance of the prediction model was assessed in the testing sets against the overall scores. The suggested parameters were selected depending on some characteristics of the input, or imbalance multi-classes classification in our case. Tables 2 and 3 show the overall and class-based parameters, respectively, as calculated from the confusion matrix.

Table 2 shows the most satisfactory performance at a macro accuracy of 99.57%, F 1macro of 99.01%, and Kappa of 99.03% for the VGG16 classifier. The lowest performance values have been yielded at a macro accuracy of 90.14%, F 1macro of 83.80%, and Kappa of 77.23% for InceptionResNetV2. Consequently, the VGG16 model demonstrates its superiority to the other models. All classifiers work well to produce high performance. Due to imbalanced data, the problem remains to determine which of these classifiers performs better in confirming COVID-19 infection.

Table 2: overall statistical parameters of different classification models (%)

classifiers	Acc_{macro}	$F1_{macro}$	$L_{Hamming}$	K_c	$MCC_{overall}$	TNR_{macro}
Xception	**98.71**	**98.02**	**1.93**	**97.07**	**97.10**	**99.03**
DenseNet121	96.36	94.18	5.47	91.66	91.80	97.14
VGG16	**99.57***	**99.01 ***	**0.64***	**99.03 ***	**99.03***	**99.69***
NASNetLarge	95.71	93.45	6.43	90.24	90.45	96.73
DenseNet169	96.57	94.75	5.15	92.23	92.28	97.50
InceptionResNetV2	90.14	83.80	14.79	77.23	78.31	92.22
InceptionV3	96.79	95.17	4.82	92.70	92.86	97.60
MobileNet	98.29	97.34	2.57	96.10	96.13	98.76

Any misdiagnosis may lead to severe consequences, especially concerning COVID-19 cases. As indicated by some of the results obtained from the confusion matrix, the classifier was capable of verifying all positive cases (COVID-19). However, it was wrong to consider some negative cases as positive cases. Besides, some of the classifiers were ineffective in verifying all positive cases, and they were indicated as negative cases. Therefore, it is necessary to calculate the parameters

of the COVID-19 class, which is the target set for this paper. Table 3 shows the performance of the classifiers concerning only COVID-19 cases as our target class using one versus all approach, as supported in PyCM as well.

Table 3: OverallStatistical Parameters of Different Classification Models (%)

COVID-19 Classifiers	ACC	AUC	ERR	GM	PPV	F1	TNR
Xception	98.71	98.34	1.29	98.34	97.56	97.56	99.13
DenseNet121	95.82	92.47	4.18	92.19	98.59	91.50	99.56
VGG16	**99.69 ***	**99.78***	**0.32***	**99.78***	**98.80***	**99.39***	**99.56***
NASNetLarge	96.14	94.64	3.86	94.59	93.75	92.59	97.82
DenseNet169	96.46	96.82	3.54	96.81	89.89	93.57	96.07
InceptionResNetV2	89.71	80.88	10.29	78.69	98.08	76.12	99.56
InceptionV3	97.43	97.47	2.57	97.47	93.02	95.24	97.38
MobileNet	98.07	98.30	1.93	98.30	94.19	96.43	97.82

As shown in Table 3, concerning ACC, ERR, F1score, MCC, and AUC, the VGG16 classification model is statistically superior to the other classification models. In addition to the parameters mentioned above, the prediction model was also evaluated from the perspectives of precision-recall metrics. Figure 3 shows the average precision score for the classifiers, and Figure 4 shows the extension of the precision-recall curve to multi-classes. As revealed by the precision-recall curves shown in Fig. 3, the prediction model proposed by us not only yielded a high average precision score in VGG16, Xception, and Mobilenet (AP = 0.99) (Fig. 3a,c,h) but also achieved a better performance in the detection of COVID-19 cases regarding the extension of the precision-recall curve to multi-classes (Fig. 4a,c,h).

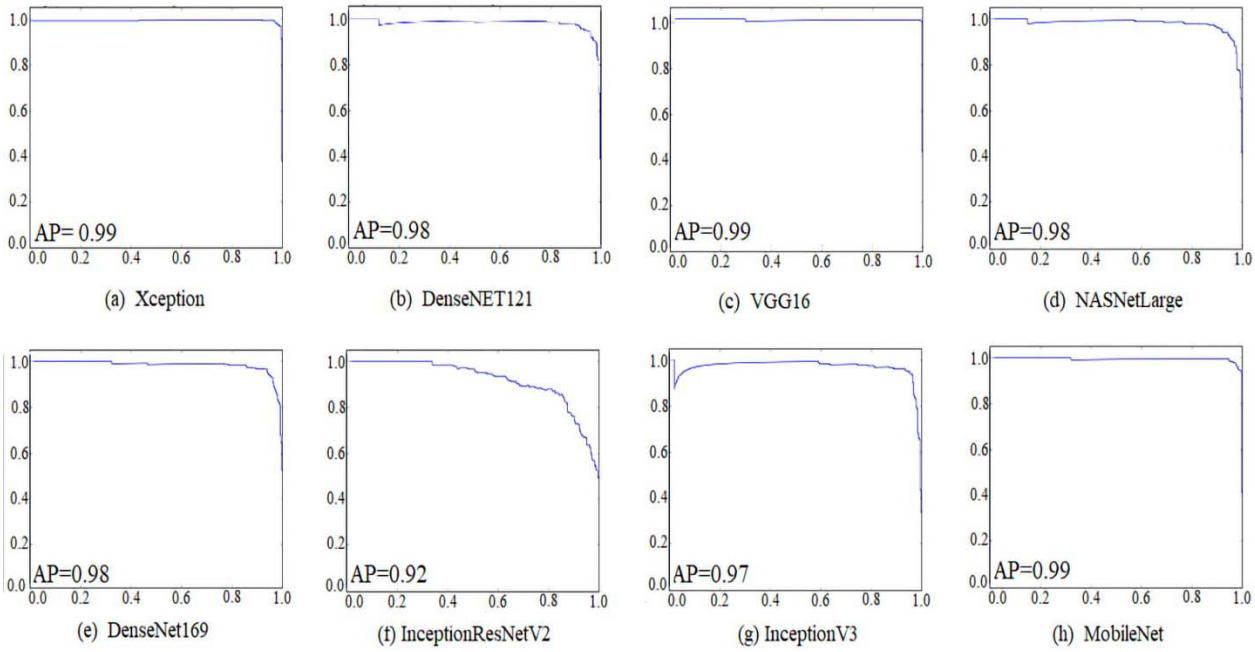

Figure 3: the average precision curves for the classifiers

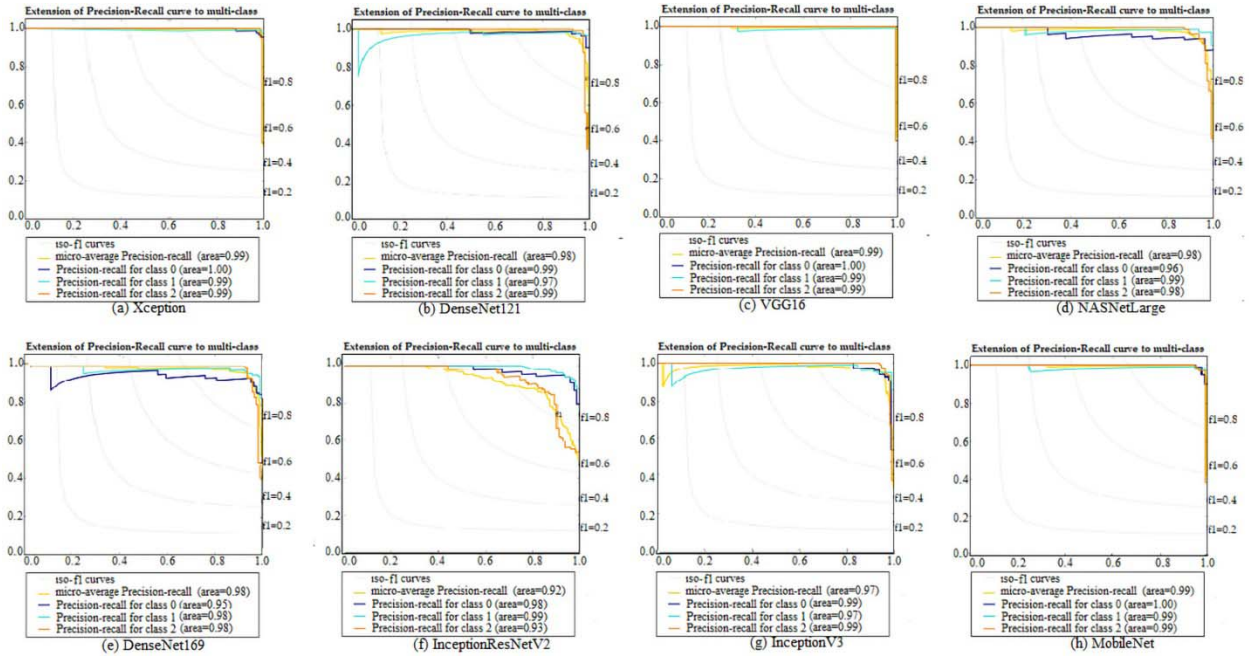

Figure 4: the precision-recall curve to multi-classes

Comparison between the State-of-the-Art methods

The AI techniques regarding the image classification approaches can help in early diagnose of the disease. Considering AI, CNN methods achieve better and faster results compared to the traditional diagnosis methods. In this paper, a rapid, robust, and efficient COVID-19 diagnosis method is proposed. The proposed method performs the X-ray images into multi-class as Healthy, Pneumonia, and COVID-19. The general performance comparison of our study with the state-of-art methods is given in this section to evaluate the proposed CNN model. In the model evaluations, the related studies depend on the multi-class classification of the chest X-ray images with various AI techniques. Table 4 shows the comparison results with the related studies uses similar data sets. While Table 5 shows, the performance values of the listed studies are given in terms of COVID-19 class accuracy.

Sethy and Behera [8] employed the ResNet50 CNN model along with SVM for the detection of COVID-19 cases from chest X-ray images. CNN model acts as a feature extractor, and SVM serves the purpose of the classifier. Their model achieved an accuracy of 95.38% on the 2-class problem. Narin et al. [9] used chest X-ray images coupled with the ResNet50 model to achieve a 98 % COVID-19 detection accuracy. Ioannis et al. [10] established the Deep Learning model using 224 confirmed COVID-19 images. Their model has achieved performance rates of 98.75 %, and 93.48 % respectively for the two and three classes. Wang and Wong [12] proposed a deep COVID-19 detection model (COVID-Net) that achieved 92.4% accuracy in the classification of four classes (healthy, non- COVID-19 pneumonia, and COVID-19). A CNN model proposed based on the DarkNet architecture by Ozturk et al. [14] has achieved performance rates of 98.08% and 87.02%, respectively, for two and three classes. The COVIDiagnosis-Net model, which was proposed by Ferhat Ucar et al. [18], has achieved a performance accuracy of 98.3%. Hemdan et al. [11] introduced COVIDX-Net to detect COVID-19 in X-ray images. They got 90% accuracy by using 25 COVID- 19 positive and 25 healthy images. Asif Iqbal Khan [19] proposed a model called CORONET on chest X-ray images. CoroNet model managed to achieve an accuracy of 99% and 95% for 3-class and 2-class classification tasks respectively on a data set consisting of 224 COVID-19, 700 pneumonia, and 504 healthy X-ray images. Toraman S et al. [20], proposed a Convolutional CapsNetfor the

detection of COVID-19 disease by using chest X-ray images with capsule networks. Their proposed method achieved an accuracy of 97.24% and 84.22%for binary class and multi-class, respectively. Amid the performance metrics that Tables 4 and 5 give, our model outperforms similar studies that use chest X-rays in the diagnosis of the COVID-19.

Table 4: The general comparison of the proposed method between state-of-the-art methods.

Study	Method Used	Classes	ACC	TNR	F1	MCC	Kappa
Wang and Wong [12]	COVID-Net	3	92.4		90		
Ucar et al. [18]	Bayes-SqueezeNet	3	98.3	99.1	98.3	97.4	
Ioannis et al. [10]	VGG-19	3	93.48	98.75			
Ioannis et al. [10]	MobileNet v2	3	94.72	96.46			
Sethy and Behra [8]	ResNet50+SVM	3			95.52	.9141	.9076
Ozturk [14]	DarkCOVIDNet	3	87.02	92.18	87.37		
Asif Iqbal Khan [19]	CORO NET	3	95	97.5	95.6		
S.Toraman et al. [20]		3	84.22	91.79	84.21		
This study	**VGG16**	**3**	**99.57**	**99.68**	**99.36**	**99.03**	**99.03**

Table 5: COVID-19 class comparison of the proposed method between the state-of-the-art methods

Study	Method Used	Classes	ACC	F1	TNR
Narin et al. [9]	RESNET 50	2	98	98	100
Ioannis et al. [10]	VGG16	2	98.75		98.75
Sethy and Behra [8]	ResNet50+SVM	2	95.38		93.47
Ozturk [14]	DarkCOVIDNet	2	98.08	96.51	95.3
Hemdan et al. [11]	VGG16	2	90	91	80
Asif Iqbal Khan [19]	CoroNet	2	99	98.5	98.6
S.Toraman et al [20]	CapsNet	2	97.24	97.24	97.04
This study	**VGG16**	**2**	**99.78**	**99.75**	**99.56**

To the best of our knowledge, the proposed model reveals perfect and outstanding classification performance for the diagnosis COVID-19 with chest X-rays. A speedy and smooth implementation characterizes the work that we carried out. The promising and encouraging results of deep learning models in the detection of COVID-19 from radiography, images indicate that deep learning has a more significant role to play in fighting this pandemic soon.

Conclusion

In this study, the overall and class-based parameters, respectively, were computed from the confusion matrix. Ac- cording to the research results, the high performance can be achieved in multiclass-classification for all classifiers. Due to imbalanced data, however, the problem remains to identify which of these classifiers performs better in confirming COVID-19 cases. Any misdiagnosis may lead to severe consequences, especially concerning COVID- 19 cases. Therefore, the parameters of the COVID-19 class were calculated in the study. The study revealed the superiority of Model vgg16 to other models applied in this research where the model achieved the highest values in terms of overall scores and based-class scores.

The study demonstrated that deep Learning with X-ray imaging might extract significant biological markers related to the COVID-19 disease. The technique is helpful to physicians in diagnosing COVID-19 patients. Meanwhile, the high accuracy of this computer-aided diagnostic

tool can contribute to a significant improvement in the speed and accuracy of COVID-19 diagnosis.

For future studies, it is necessary to address other shortcomings. In particular, a more detailed analysis requires a more massive amount of patient data, especially those associated with COVID-19. Furthermore, such effective deep learning models as VGG16, and GoogLeNet, have been trained on more than a million images, which are barely available in the medical domain. Besides, there is a possibility that is training deep neural networks with limited data available results in over-fitting and hinders good generalization.

References

[1] World Health Organization Director. “General’s opening remarks at the media briefing on COVID-19”, 03 2020. [Online]. Available: https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-03-2020.

[2] Victor M Corman, Christian Drosten et al. “Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR”. *Eurosurveillance*, vol. 25, no. 3, 2000045, 2020.

[3] Xingzhi Xie, Zheng Zhong and Wei Zhao. “Chest CT for typical 2019-nCoV pneumonia: relationship to negative RT-PCR testing”. *Radiology*, vol. 296, no. 2, 41–45, 2020.

[4] Mesut Toğaçar, Burhan Ergen and Zafer Cömert. “Application of breast cancer diagnosis based on a combination of convolutional neural networks, ridge regression and linear discriminant analysis using invasive breast cancer images processed with autoencoders”. *Medical Hypotheses*, vol. 135, 109503, 2020. [Online]. Available: https://dx.doi.org/10.1016/j.mehy.2019.109503.

[5] Xiaolong Liu, Zhidong Deng and Yuhan Yang. “Recent progress in semantic image segmentation”. *Artificial Intelligence Review*, vol. 52, no. 2, 1089–1106, 2019.

[6] Amit Kumar Jaiswal, Prayag Tiwari et al. “Identifying pneumonia in chest X-rays: A deep learning approach”. *Measurement*, vol. 145, 511–518, 2019.

[7] American College Radiology. “ACR Recommendations for the use of Chest Radiography and Computed Tomography (CT) for Suspected COVID-19 Infection”, 03 2020. [Online].Available: https://www.acr.org/Advocacy-and-Economics/ACR-Position-Statements/Recommendations-for-Chest-Radiography-and-CT-for-Suspected-COVID19-Infection.

[8] Prabira Kumar Sethy and Santi Kumari Behera. "Detection of coronavirus disease (covid-19) based on deep features". *Preprints 2020030300*,2020.

[9] Ali Narin, Ceren Kaya and Ziynet Pamuk. "Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks". *arXiv preprint arXiv:2003.10849*, 2020.

[10] Ioannis D, Apostolopoulos et al. "Covid-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks". *Physical and Engineering Sciences in Medicine*, vol. 43, no. 2, 635–640, 2020.

[11] Ezz El-Din Hemdan, Marwa A. Shouman and Mohamed Esmail Karar. "Covidx-net: A framework of deep learning classifiers to diagnose covid-19 in x-ray images". *arXiv preprint arXiv:2003.11055* , 2020.

[12] Linda Wang and Alexander Wong. "_A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest X-Ray Images". *arXiv preprint arXiv:2003.09871*, 2020.

[13] Saleh Albahli. "A Deep Neural Network to Distinguish COVID-19 from other Chest Diseases using X-ray Images". *Current Medical Imaging*, vol. 16, 1–11, 2020.

[14] Tulin Ozturk, Muhammed Talo et al. "Automated detection of COVID-19 cases using deep neural networks with X-ray images". *Computers in Biology and Medicine*, vol. 121, 103792, 2020. [Online]. Available: 10.1016/j.compbiomed.2020.103792.

[15] Xin Li and Dongxiao Zhu. "Covid-Xpert: An ai powered population screening of covid-19 cases using chest radiography images". *arXiv preprint arXiv:2004.03042*,2020.

[16] Parnian Afshar et al. "Covid-caps: A capsule network-based framework for identification of covid-19 cases from x-ray images.". *arXiv preprint arXiv:2004.02696*, 2020.

[17] Muhammad Farooq and Abdul Hafeez. "Covid-Resnet: A deep learning framework for screening of covid19 from radiographs". *arXiv preprint arXiv:2003.14395*,2020.

[18] Ferhat Ucar and Deniz Korkmaz. "COVIDiagnosis-Net: Deep Bayes-SqueezeNet based diagnosis of the coronavirus disease 2019 (COVID-19) from X-ray images". *Medical Hypotheses*, vol. 140, 109761, 2020.

[19] Asif Iqbal Khan, Junaid Latief Shah and Mohammad Mudasir Bhat. "CoroNet: A deep neural network for detection and diagnosis of COVID-19 from chest x-ray images". *Computer Methods and Programs in Biomedicine*, vol. 196, 105581, 2020.

[20] Suat Toraman, Talha Burak Alakus and Ibrahim Turkoglu. "Convolutional Capsnet: A novel artificial neural network approach to detect COVID-19 disease from X-ray images using capsule networks". *Chaos, Solitons & Fractals*, vol. 140, 110–122, 2020.

[21] Joseph Paul Cohen, Paul Morrison and Lan Dao. "COVID-19 image data collection". 2020. [Online]. Available: https://github.com/ieee8023/covid-chestxraydataset.

[22] Tawsifur Rahman et al. "COVID-19 Radiography Database", 2020. [Online]. Available: https://www.kaggle.com/tawsifurrahman/covid19-radiography-database.

[23] Vinod Nair and Geoffrey E. Hinton. "Rectified linear units improve restricted Boltzmann machines". In *Proceedings of the 27th International Conference on Machine Learning*, ICML'10, pages 807–814, Haifa, Israel, June 2010.

[24] Geoffrey E. Hinton et al. "Improving neural networks by preventing co-adaptation of feature detectors". *arXiv preprint arXiv:1207.0580*, 2012.

[25] Douglas M. Hawkins. "The Problem of Overfitting". *Journal of Chemical Information and Computer Sciences*, vol. 44, no. 1, 1–12, 2004.

[26] Tijmen Tieleman and Geoffrey Hinton. "Lecture 6.5-RmsProp: Divide the gradient by a running average of its recent magnitude". *COURSERA Neural Networks Mach. Learn*, vol. 4, no. 2, 26–31, 2012.

[27] Sepand Haghighi, Masoomeh Jasemi et al. "PyCM: Multi-class confusion matrix library in Python". *Journal of Open Source Software*, vol. 3, no. 25, 729, 2018.

[28] Henry Carrillo, Kay H. Brodersen and José A. Castellanos. "Probabilistic performance evaluation for multi-class classification using the posterior balanced accuracy". In *Armada M., Sanfeliu A., Ferre M. (eds) ROBOT2013: First Iberian Robotics Conference*, volume 252, pages 347–361. Springer, Cham, 2014.

[29] Rehan Akbani, Stephen Kwek and Nathalie Japkowicz. "Applying support vector machines to imbalanced datasets". In *Boulicaut JF., Esposito F., Giannotti F., Pedreschi D. (eds) Machine Learning: ECML 2004. ECML 2004. Lecture Notes in Computer Science*, volume 3201, pages 39–50. Springer, Berlin, Heidelberg, 2004.

[30] Kay H. Brodersen, Christoph Mathys et al. "Bayesian mixed-effects inference on classification performance in hierarchical data sets". *The Journal of Machine Learning Research*, vol. 13, no. 1, 3133–3176, 2012.

[31] Nitesh V. Chawla, Kevin W. Bowyer et al. "SMOTE: synthetic minority over-sampling technique". *Journal of artificial intelligence research*, vol. 16, 321–357, 2002.

[32] Nathalie Japkowicz and Shaju Stephen. "The class imbalance problem: A systematic study1". *Intelligent Data Analysis*, vol. 6, no. 5, 429–449, 2002.

[33] Juri Opitz and Sebastian Burst. "Macro F1 and Macro F1". *arXiv preprint arXiv:1911.03347*, 2019.

[34] Jacob Cohen. "A coefficient of agreement for nominal scales". *Educational and psychological measurement*, vol. 20, no. 1, 37–46, 1960.

[35] Ryuei Nishii and Shojiro Tanaka. "Accuracy and inaccuracy assessments in land-cover classification". *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 1, 491–498, 1999.

[36] Claude Sammut and Geoffrey I. Webb. *Encyclopedia of machine learning*. eds. Encyclopedia of machine learning, 2011. Springer Science & Business Media.

[37] Jan Gorodkin. "Comparing two K-category assignments by a K-category correlation coefficient". *Computational Biology and Chemistry*, vol. 28, no. 5-6, 367–374, 2004.

[38] Brian W. Matthews. "Comparison of the predicted and observed secondary structure of T4 phage lysozyme". *Biochimica et Biophysica Acta (BBA)-Protein Structure*, vol. 405, no. 2, 442–451, 1975.