applied sciences

MDPI

Article

Performance Analysis for COVID-19 Diagnosis Using Custom and State-of-the-Art Deep Learning Models

Ali Tariq Nagi [1], Mazhar Javed Awan [1], Mazin Abed Mohammed [2], Amena Mahmoud [3], Arnab Majumdar [4] and Orawit Thinnukool [5,*]

[1] Department of Software Engineering, University of Management and Technology, Lahore 54770, Pakistan; ali.tariq.nagi@gmail.com (A.T.N.); mazhar.awan@umt.edu.pk (M.J.A.)
[2] College of Computer Science and Information Technology, University of Anbar, Anbar 31001, Iraq; mazinalshujeary@uoanbar.edu.iq
[3] Computer Science Department, Faculty of Computers and Information, Kafrelsheikh University, Kafr Al-Sheikh 33516, Egypt; amena_mahmoud@fci.kfs.edu.eg
[4] Faculty of Engineering, Imperial College London, London SW7 2AZ, UK; a.majumdar@imperial.ac.uk
[5] College of Arts, Media and Technology, Chiang Mai University, Chiang Mai 50200, Thailand
* Correspondence: orawit.t@cmu.ac.th

Abstract: The modern scientific world continuously endeavors to battle and devise solutions for newly arising pandemics. One such pandemic which has turned the world's accustomed routine upside down is COVID-19: it has devastated the world economy and destroyed around 45 million lives, globally. Governments and scientists have been on the front line, striving towards the diagnosis and engineering of a vaccination for the said virus. COVID-19 can be diagnosed using artificial intelligence more accurately than traditional methods using chest X-rays. This research involves an evaluation of the performance of deep learning models for COVID-19 diagnosis using chest X-ray images from a dataset containing the largest number of COVID-19 images ever used in the literature, according to the best of the authors' knowledge. The size of the utilized dataset is about 4.25 times the maximum COVID-19 chest X-ray image dataset used in the explored literature. Further, a CNN model was developed, named the Custom-Model in this study, for evaluation against, and comparison to, the state-of-the-art deep learning models. The intention was not to develop a new high-performing deep learning model, but rather to evaluate the performance of deep learning models on a larger COVID-19 chest X-ray image dataset. Moreover, Xception- and MobilNetV2- based models were also used for evaluation purposes. The criteria for evaluation were based on accuracy, precision, recall, F1 score, ROC curves, AUC, confusion matrix, and macro and weighted averages. Among the deployed models, Xception was the top performer in terms of precision and accuracy, while the MobileNetV2-based model could detect slightly more COVID-19 cases than Xception, and showed slightly fewer false negatives, while giving far more false positives than the other models. Also, the custom CNN model exceeds the MobileNetV2 model in terms of precision. The best accuracy, precision, recall, and F1 score out of these three models were 94.2%, 99%, 95%, and 97%, respectively, as shown by the Xception model. Finally, it was found that the overall accuracy in the current evaluation was curtailed by approximately 2% compared with the average accuracy of previous work on multi-class classification, while a very high precision value was observed, which is of high scientific value.

Keywords: artificial Intelligence; chest X-ray; COVID-19; deep learning; convolutional neural networks; lung opacity; mobilenetv2; transfer learning; Xception; pneumonia

check for updates

Citation: Nagi, A.T.; Awan, M.J.; Mohammed, M.A.; Mahmoud, A.; Majumdar, A.; Thinnukool, O. Performance Analysis for COVID-19 Diagnosis Using Custom and State-of-the-Art Deep Learning Models. *Appl. Sci.* **2022**, *12*, 6364. https://doi.org/10.3390/app12136364

Academic Editor: Keun Ho Ryu

Received: 23 March 2022
Accepted: 19 June 2022
Published: 22 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The world has been afflicted with a disastrous pandemic that has affected the population of our planet on a very large scale and which was named COVID-19 by the World Health Organization (WHO) in February 2020 [1]. The virus originated in Wuhan City

in China and then spread throughout the entire globe with devastating effects, causing around 45 million deaths in a period of just 1.7 years, as per the World Health Organization report [2]. The number of victims of COVID-19 was about 0.22 billion across the world. Recovery from the illness can take as long as several weeks to months, with long-term effects causing organ damage [3,4]. Moreover, the patient must be placed in isolation during that period to prevent the spread of COVID-19. This not only consumed lives, but also hindered everyday office work, social and family routines, and paralyzed the world economy [5,6]. Researchers battled this disease since its emergence, looking for efficient and accurate detection methods, and uncovering solutions for its treatment, and vaccinations for human immunization [7–9].

Artificial intelligence geared with high-performance computers induced hope in medical science to diagnose diseases rapidly, which would otherwise take an enormous amount of time, if relying only on medical science and well-qualified and highly skilled doctors. In a study on breast cancer prediction, it was found that artificial intelligence could predict cancer at a rate higher than a well-organized evaluation of a panel of eleven pathologists. Hence, artificial intelligence would also help curtail medical costs incurred before treatment in hospitals and clinics [10–18]. Radiographic images, CT scans and chest X-rays show visual attributes exposing symptoms of COVID-19. The limited availability of expert radiologists, however, creates a bottleneck in the manual diagnosis of COVID-19. Therefore, the use of artificial intelligence should be, and already is, a viable source for the detection of COVID-19 using radiographic images [19–21].

This study aims to deploy a deep learning network and existing known models in convolutional neural networks, which are known to perform well on images, on a database of chest X-rays for COVID-19 detection to determine a viable architecture for developing a medical diagnosis system for accurate diagnosis of this virus. The study measured the performance gains of deep learning models on the largest COVID-19 chest X-ray image dataset, according to the best of the authors' knowledge [22].

Our research experimentation and analysis are different from previous research work in that they were performed on a larger database of chest X-ray images than used previously in the literature. The datasets utilized in the literature include far fewer COVID-19 chest X-ray images, ranging from 127 to 850 as per the authors reviewed in Section 2. This number is considerably lower compared with the number of samples available for the other cases; for example, in one study, the number of pneumonia cases exceeded 5000, while the number of COVID-19 cases was only 415 [23]. Here, a considerably low number of COVID-19 chest X-ray images and imbalance occurred. The study achieved the highest accuracy to date—as per the covered literature—in which 237 COVID-19, 1338 normal, and 1336 viral pneumonia cases were used; the authors specified that the number of COVID-19 cases was significantly lower and the problem was that of a highly imbalanced classification [24]. Moreover, in some cases in the literature, the authors balanced the classes by matching the number of samples of other classes to those of the COVID-19 samples, thereby achieving the maximum accuracy up to 99.90% using a very low number of chest X-ray test images; the total number of samples for COVID-19 chest X-rays for training and testing of only 404 [25]. Due to the limitations of past research work, the authors felt the need for a re-evaluation of the deep learning models on a larger dataset of COVID-19 chest X-ray images and the current study was implemented due to the availability of a database containing 3616 COVID-19 chest X-ray images, the highest ever worked upon, to the best of the authors' knowledge.

The main aim of the subject research work is the performance evaluation of deep learning models on a dataset with the largest number of COVID-19 images ever used, to the best of the authors' knowledge, to see the improvements in the performance metrics. The subject study addresses a major gap in research on COVID-19 detection and, in doing so, makes the following contributions:

- To the best of our knowledge, this study is the first that benchmarks deep learning models on the largest COVID-19 dataset.
- In this study, we developed a custom CNN model for benchmarking on the subject dataset.
- The custom CNN model outperformed MobileNetV2 in terms of precision, which is of high scientific value.
- The current study includes a detailed and comprehensive evaluation of the performance involving several performance metrics to observe the impact of increasing the COVID-19 dataset on the research work.

The subsequent sections include related work in Section 2, while Section 3, Methodology, describes the dataset and deep learning models utilized for the purpose of evaluation, as well as the experimental setup and evaluation metrics for benchmarking. Next, in Section 4, results are described with graphical elaborations. Finally, Sections 5 and 6 include discussion, conclusion, and limitations and future work.

2. Related Work

The current section presents a considerable amount of work, found in the literature, that explores the role of artificial intelligence in the automated diagnosis of COVID-19 using chest X-ray images, and distinguishing it from pneumonia and healthy cases.

A deep learning framework was developed, termed COVID-CheXNet that could detect COVID-19 using chest X-rays [26]. First, image pre-processing, using contrast-limited adaptive histogram equalization and a Butterworth Band-pass filter. Following this, two parallel deep learning networks were applied, one of which was ResNet34 and the other comprised the high-resolution network, HRNet. The image enhancement module consisted of applying the adaptive contrast enhancement method using CLAHE. The CLAHE method enhances the contrast and edges of the image. Moreover, a Butterworth filter was applied for noise reduction. The train–test distribution of the dataset was 70% to 30%. The ResNet model [27] was pre-trained on ImageNet, while the HRNet model was trained on the subject dataset. The two models, HRNet and Resnet34, combined score-level fusion and decision-level fusion. Using WSR in the score-level fusion, the best results were obtained with an accuracy of 99.99%, specificity of 100%, precision of 100%, F1-score of 99.99%, MSE of 0.011%, and RMSE of 0.012%, where WSR represents the Weighted Sum Rule in which weights are assigned to the individual scores and can be adjusted related to the performance of the individual level performance to obtain overall optimal performance.

A computer-aided diagnosis (CAD) scheme was developed that involved several image pre-processing algorithms: removing diaphragms from the X-ray images, normalizing the contrast-to-noise ratio of the image, generating three different images from the input dataset, and feeding them to VGG16 CNN. The test set and train set had a ratio of 90:10. The images were converted to binary form, followed by morphological operations and the higher intensity component; the diaphragm region was removed from the images. Before feeding the images to the CNN, three different images were generated after diaphragm removal. The first image was generated by applying bilateral filtering to remove the noise. To generate the second image, the histogram equation method was applied, normalizing the image, and thereby enhancing lung tissue patterns and characteristics concerning COVID-19. The last image was fed to the network as it was. The three images were resized to 224 $\times$ 224 pixels and given as inputs to the VGG16 model. The top layer of VGG16 was removed and two FC layers were added with 256 and 128 nodes using ReLU. Finally, a classification layer, Softmax, was added. The accuracy obtained was 94% in classifying three classes and the accuracy of diagnosing COVID-19 was 98.6%. The model was compiled with a batch size of 4, maximum epochs of 200, and an initial learning rate of 10^{-5}; with loss value [28].

An automated deep learning method was developed for an early diagnosis of COVID-19 from chest X-ray images using transfer learning. The models tested for evaluation were Inception–ResNetV2, InceptionNetV3 and NASNetLarge. The training and test dataset

ratio was 80:20. InceptionNetV3 gave the best performance with accuracies of 98.63% and 99.02% with and without using data augmentation on the training data. Inception–ResNetV2 and InceptionNetV3 were equal and produced the highest accuracy of 97.87% with data augmentation, while NASNetLarge exceeded in performance without a data augmentation strategy, obtaining an accuracy of 94.87% [29].

An approach for COVID-19 detection using radiography images and comprising three phases was proposed. The first phase detected the presence of pneumonia, while the second phase differentiated between COVID-19 and pneumonia. The third phase localized the areas generating a map where COVID-19 was present. The authors employed the VGG16 model using pre-trained weights on ImageNet; they froze the FC layer head and added new layers, namely, AveragePooling2D, flatten, dense, dropout, and a dense layer using the Softmax function. Data augmentation was applied by rotating the image by 15 degrees clockwise or counterclockwise. The authors used the gradient-weighted class activation mapping (Grad-CAM) algorithm to visualize the areas infected with COVID-19. Grad-CAM works by looking into the final convolutional layer and then examining the gradient information flowing into that layer, thereby generating a heatmap. A cross-validation technique was employed by the authors. The experiments were performed on a dataset of 6523 X-ray images, achieving an average accuracy of 97% [30].

A study [31] involving feature extraction using deep learning and transfer learning, while feeding the features into an SVM classifier to classify the COVID-19 and normal cases was performed. The dataset contained 180 COVID-19 and 200 normal cases. The deep learning models used were ResNet18, ResNet50, ResNet101, VGG16, and VGG19 [32]; different SVM kernel functions namely, linear, quadratic, cubic, and Gaussian were employed. The ResNet50 with the linear kernel function on SVM produced the best results, with an accuracy of 94.7%. Also, a customized CNN model was applied with end-to-end training on the dataset, producing an accuracy of 91.6%, and a fine-tuned ResNet50 model achieved an average accuracy of 92.6%. Following this, the authors deployed SVM on local texture descriptors, which achieved lower accuracy than the feature extraction using deep learning and then applying SVM. The proposed convolution model with 21 layers was trained with a learning rate of 0.001 and over 300 iterations, producing an accuracy of 91.58%.

A study was undertaken where an ensemble of three deep learning models was used. The CNN models used were DenseNet201, ResNet50V2, and InceptionV3. An ensemble of the prediction of these three models to generate the final classification. The proposed approach achieved higher accuracy than the other three individual techniques, giving an accuracy of 95.7%, and a sensitivity of 98% on the test set. Two architectures (one for binary classification and the other for multi-class (3 classes) classification) comprising an ensemble classification embedded with transfer learning and CNN layers were proposed. The variation between these two architectures was the usage of the Sigmoid function instead of Softmax in the classification layer. The accuracy achieved for multi-classification was 99.21%; for binary-classification, it was 96.15% [33].

A modified VGG-19 architecture with transfer learning was deployed for the prediction of COVID-19 based on chest X-rays, patient ID, sex, age, diagnosis (COVID-19 or not), survival (yes or no), scan view (e.g., posteroanterior, supine anteroposterior), timestamp, and location The dataset comprised 181 COVID-19 and 364 normal chest X-ray images. Transfer learning was deployed for the pre-modified part and achieved an accuracy of 96.3% [34].

A medical decision support system was developed using convolutional neural network with EfficientNetB4 as the base architecture for the diagnosis of COVID-19. The entire model was trained on chest X-ray images. The efficiency achieved for binary classification was 99.62%, and for multi-class classification it was 97.11%. A pre-trained Xception model with transfer learning was used with a custom deep CNN model. The obtained accuracy in the test phase was 97.4%. The dataset comprised 127 COVID-19 cases, 500 pneumonia, and 500 normal chest X-ray images [35].

In another study [36], a simple CNN, deep learning model was devised and a modified pre-trained AlexNet network was deployed for distinguishing between normal and COVID-19 cases. The CNN architecture compromised one convolutional layer with only 16 filters, one batch normalization layer, a RELU layer, two fully connected layers, and finally, a SoftMax layer for classification. The modified AlexNet architecture was established by replacing the last layers with the intended layers. The authors used 85 X-ray images and 203 CT scan images of COVID-19 and 85 X-ray and 153 CT scan images of normal cases. The accuracies obtained for the proposed model on X-ray images and CT scan images were 94% and 94.1%, respectively. Similarly, the accuracies obtained for the modified AlexNet Architecture were 98% and 82%.

A study [37] was made where a pre-trained ResNet50 architecture and an ensemble of pre-trained VGG16, ResNet50, and a custom CNN were deployed for differentiating between COVID-19 and pneumonia using chest X-ray images. The former achieved an accuracy of 89.2%, while the latter succeeded in achieving an accuracy of 91.24%. ResNet-18, ResNet50, SqueezeNet, and DenseNet-121 were used as pre-trained models in research using transfer learning where the last layer of the architectures was trained to distinguish between COVID-19 and non-COVID-19 chest X-ray images. The authors employed 184 COVID-19 and 5000 non-COVID-19 chest X-ray images for training and testing purposes. SqueezeNet and ResNet18 achieved a slightly better performance than the other models.

CoroNet, a deep learning architecture, was proposed to detect the subject virus. The architecture was based on Xception and deployed a dropout layer and two fully connected layers at the end of the model, with a total number trainable parameters of 33,969,964. The model achieved accuracy of 99%, 95% and 89.6% on 2-class, 3-class, and 4-class, respectively [38].

The pre-trained VGG16 architecture was deployed, with new layers added at the top, to detect and differentiate between COVID-19, bacterial pneumonia, and normal cases [39]. The dataset contained 224 chest X-ray images of COVID-19, 700 images of bacterial pneumonia, and 504 healthy cases. The authors achieved accuracies of 96% and 92.5% in two-class and three-class, respectively. The highest accuracy achieved so far on COVID-19 detection using lung X-ray images is 99.658% for 3-class classification, as per the covered literature [40].

The previous studies had limitations in that a xhigh-class imbalance was observed and, also, they included very few COVID-19 chest X-ray images. As already mentioned, the highest number of images for COVID-19 used was 850, whereas the number of COVID-19 samples in the subject research work is about 4.25 times this number (3616 images). Therefore, the previous work had limitations regarding the low number of COVID-19 datasets and class imbalance. The current work has more than four times the highest dataset of COVID-19 images ever utilized, and the authors were interested to see the performance results for such a contrast between the current dataset and the previous datasets.

3. Methodology

This section presents the architectures of deep learning models and the materials deployed for experimentation in this study. Section A presents the details regarding the dataset used, while Section B elaborates the deep learning models. Section C and Section D describe the experimental setup and evaluation metrics and tools.

3.1. Dataset

The dataset of chest X-rays utilized in this research work is available in Kaggle and was utilized in other research studies [41,42]. The database emerged from a small number of COVID-19 cases [42] and was constantly updated. The current, updated version of the database contains the largest number of COVID-19 chest X-ray images yet to be experimented upon in any research work to the best of the authors' knowledge. The current updated version of the database contains 3616 COVID-19 cases, 6012 lung opacity cases,

and 10,192 normal (healthy) images. Lung opacity represents the set where the opacity is formed by pneumonia and differentiation of COVID-19 from this class and the healthy (normal) class is provided in the subject study.

The database was developed by researchers from Qatar University (Doha, Qatar), University of Dhaka (Bangladesh), Pakistan and Malaysia, and with the cooperation from medical doctors. The database development was initiated from sources including the Italian Society of Medical and Interventional Radiology (SIRM) COVID-19 database, COVID-19 positive chest x-ray images from different articles [41], The Radiopedia, COVID-19 chest imaging is mentioned in the initial database research article. The current distribution also contains an additional class, lung opacity, which was recently added from the Radiological Society of North America (RSNA) CXR dataset with number of images, 6012 [23].

Samples of chest X-ray images for normal, COVID-19 and lung opacity cases used in the training of the deep learning models in our research work are shown in Figure 1.

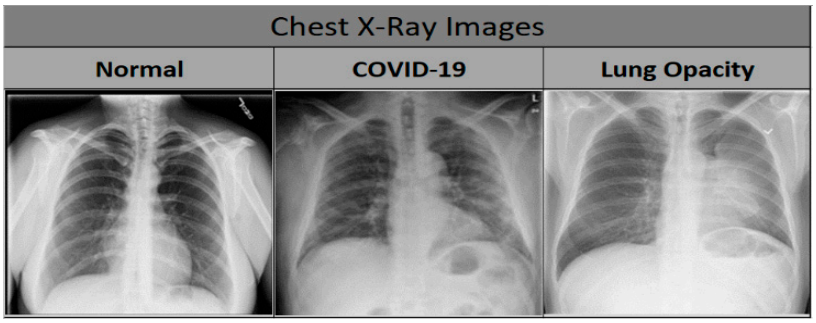


Figure 1. Sample chest X-ray images of normal, COVID-19 and lung opacity cases.

3.2. Approach

For the evaluation of the deep learning models on the subject dataset, the authors developed a custom deep learning model and used several reference state-of-the-art models employed in earlier COVID-19 chest X-ray studies. In addition to testing the known models, the authors first developed and tested a custom deep learning model. Among the state-of-the-art models, the chosen models were the MobileNetV2 model with additional layers on top, which gave the highest performance among other known models in previous research work, and the Xception model with some layers added on the top. The designated label for the custom deep learning model is Custom-Model. The overall system architecture is depicted in Figure 2.

The following models used for the performance evaluation of deep learning on the subject dataset.

3.2.1. Custom-Model

The custom model involves convolution, non-linearity, downsampling (pooling layer), and generic, multi-layer neural network. Before the architectural details at an abstract level are explored, an introduction to the underlying mathematical formations is helpful and is given below.

The convolution layer deployed in the network allows features to be learnt for images, which activate for edges along different angles. Diving deep into the network, the layers can learn to activate for complex patterns, e.g., shapes represented by a group of edges, while further depth explores features for meaningful objects. Convolution involves computations (summation of dot products) by spanning kernels or filters over the input image [43].

Let H^l, W^l, D^l, denote the input dimensions to the l-th convolution layer of order 3.

Normally, multiple kernels having a spatial span, $H \times W$. Let D be the number of kernels, while, for denoting the kernel, D^l represents the number of channels of input. Kernel f will be an order 4 tensor in $\mathbb{R}^{H \times W \times D^l \times D}$.

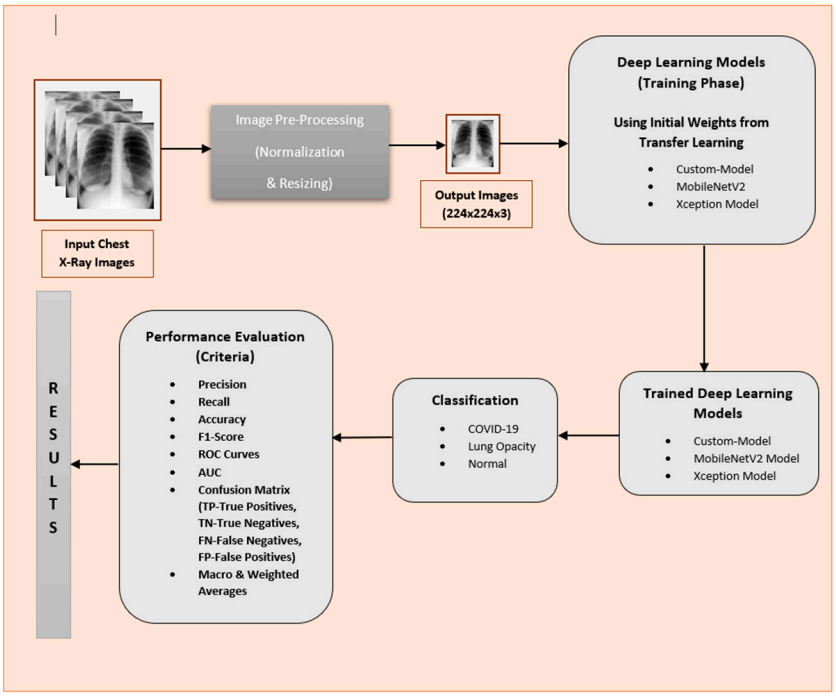


Figure 2. Overall system architecture.

The kernel spans across the entire image and the sum of products, HWD^l, of the corresponding elements ranging in the D^l channels yield the result at a specified spatial location of the output feature map. A stride corresponds to the number of pixels to be skipped while spanning at each convolution operation.

In terms of mathematical definition, while keeping the subject context, the convolution involves convolution operation of a kernel or filter over the layer of input data, which is defined as follows:

$$y_{i^{l+1}j^{l+1},d} = \sum_{i=0}^{H} \sum_{j=0}^{W} \sum_{d^l=0}^{D^l} f_{i,j,d^l,d} \times x^l{}_{i^{l+1}+i,\, j^{l+1}+j,\, d^l} \tag{1}$$

where the third dimension d^l ranges across the number of channels, and the first and third dimensions range across the height and width of the kernel, i.e.,

$$0 \leq i < H,\ 0 \leq j < W,\ 0 \leq d^l < D^l \text{ and } 0 \leq d < D.$$

and $y_{i^{l+1}j^{l+1},d}$ represents the output of the l-th which is the input to the $l+1$ layer.

It should be noted that specific values of these index variables refer to a specific point in a kernel and a bias term bd is usually added to $y_{i^{l+1}j^{l+1},d}$.

Introducing non-linearity introduces non-linear transformation in the model, helping recognize complex patterns. Nonlinearity was introduced with ReLU, which is used in several fields, such as image processing. It cuts off the inputs before feeding into the next layer. However, it does not involve any downsampling.

Mathematically, it can be represented as in Equation (2).

$$y_{i,j,d} = \max\left\{0,\ x^l{}_{i,j,\,d^l}\right\} \tag{2}$$

$$0 \leq i < H^l = H^{l+1},\ 0 \leq j < W^l = W^{l+1} \text{ and } 0 \leq d^l < D^l = D^{l+1}$$

A graphical representation is shown in Figure 3.

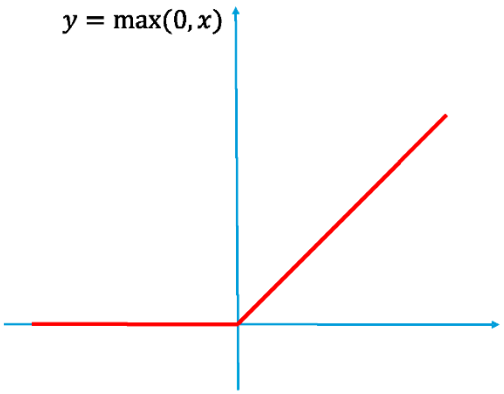


Figure 3. The ReLU function.

Downsampling was introduced in the model with a pooling layer, which reduces the dimension of the inputs by mapping a sub-region to a single value. The authors deployed the max pooling layer, which is widely used.

Let x^l be the input to the pooling layer in $\mathbb{R}^{H^l \times W^l \times D^l}$. Let $H \times W$ be the spatial extent of the pooling. Let $H^{l+1} \times W^{l+1} \times D^{l+1}$ the dimensions of the pooling layer output; in terms of mathematics, the operation is then described in Equation (3).

$$y_{i^{l+1} j^{l+1}, d} = \max_{0 \le i < H, 0 \le j < W} x^l_{i^{l+1} \times H + i,\ j^{l+1} \times W + j,\ d} \tag{3}$$

$$0 \le i^{l+1} < H^{l+1},\ 0 \le j^{l+1} < W^{l+1},\ 0 \le d < D^{l+1} = D^l.$$

After multiple layers of convolution, ReLU and pooling, the authors deployed fully connected layers transforming distributed features and using neural units where the output is obtained through the following function:

$$y_i = f(w_1 x_1 + w_2 x_2 + .. + w_n x_n + b) \tag{4}$$

where f is the activation function (ReLU in the subject case), x represents the inputs (after the flattening layer), w represents the weights associated with the neuron, and b is the bias.

The final layer of the neural network, the Softmax layer, deployed the Softmax activation function, which gives the probabilities of the classes for the final classification. Mathematically, it can be expressed as in Equation (5).

$$softmax(z_i) = \frac{e^{z_i}}{\sum_{i=1}^{M} e^{z_i}} \tag{5}$$

where z represents the outputs of the neurons of the final layer. The values from the neurons are normalized using the exponential function. M represents the number of neurons.

A custom CNN model, named Custom-Framed-CNN-Model, was developed where layers of convolution were stacked up following a pooling layer downsampling after each convolution layer, and increasing the number of filters of convolution layers, as shown in Unit-1 in Figure 4. The number of filters was increased to 512 in order to obtain more abstractions of high-level complex features in the COVID-19 chest X-ray images. Moreover, we maintained a low 3×3 filter size the stacked layers, as increasing the filter was shown to negatively affect the overall accuracy. After this, the feature maps were fed to a neural network containing 3 layers of neurons with dropout to avoid overfitting; finally, a classification layer was added as per Unit-2 Figure 4. The number of trainable parameters was 8,270,659.

The activation function used in the architecture was ReLU. The resulting architecture is shown in Figure 4.

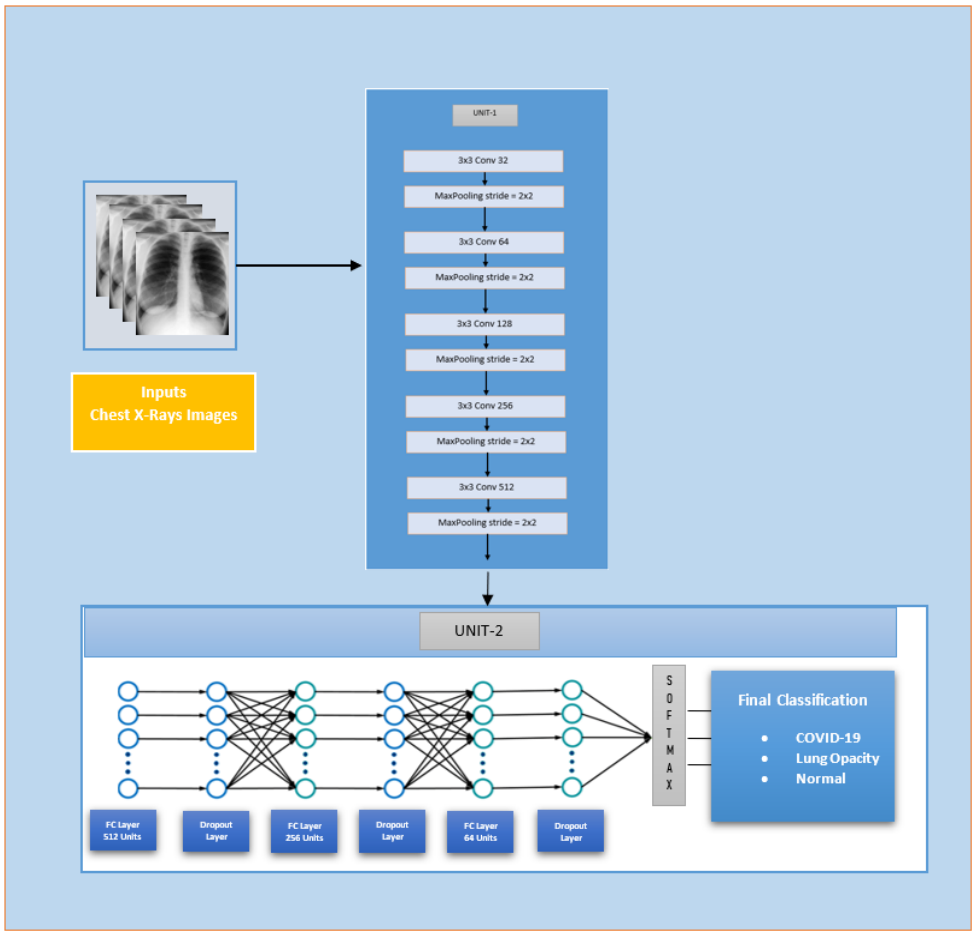


Figure 4. Custom-Model architecture.

3.2.2. Deep Learning Model-1 (Extended Mobilenetv2 Model)

The subject model was utilized from the literature [43] and involved performance comparisons for different state-of-the-art models, with some added layers for COVID-19 diagnosis, as in Figure 5. The only difference is the addition of a linear transformation with dense layer before each ReLU block in the subject experimentation to further increase the complexity of the model.

MobileNetV2 [44] was fit as a pre-trained model as it was found to achieve the highest performance compared with other models. The entire model was trained on the dataset using the initial pre-trained weights in the experiments. The same strategy as in the research work [45] was followed and included the same parameters, such as optimizer, learning rate, and the number of epochs. The number of trainable parameters was 8963.

3.2.3. Deep Learning Model-2 (Extended Xception Model)

The Xception model [46] added few layers used for as evaluation of the dataset. The global average pooling layer (for minimizing overfitting by downsampling) followed by a Softmax layer were used for modification. The resulting architecture is shown in Figure 6. The number of trainable parameters was 20,813,099.

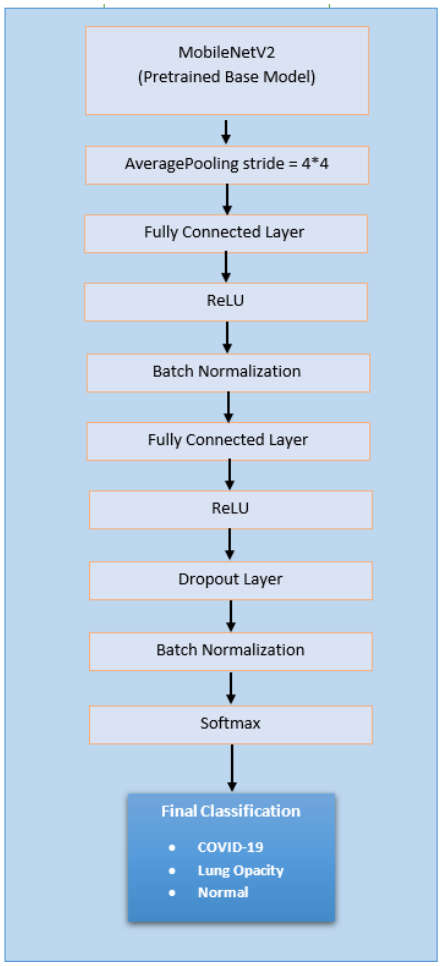


Figure 5. Deep learning Model-1 (extended Mobilenetv2 model).

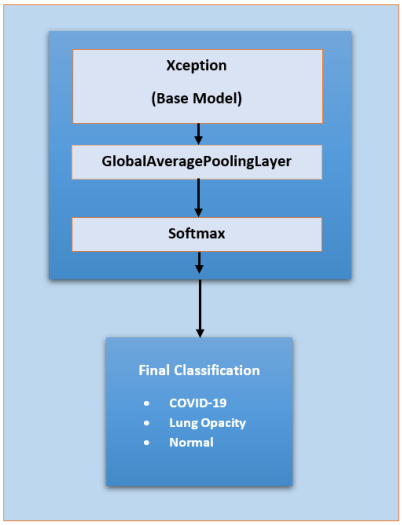


Figure 6. Xception-based architecture.

3.3. Experimental Setup

The setup, chosen for performance evaluation, is categorized showing hardware, operating system, data pre-processing, hyperparameters, etc., and is illustrated in the categories below.

3.3.1. System and Software Setup

The machine used for the experimental setup in our research is based on Processor, Intel(R) Xeon(R) Silver 4110 CPU with 16 GB RAM. The execution is based completely on CPUs rather than on GPU. We deployed our implementation in Python (Version 3.8.8). The research models were implemented using Keras implementation in Tensorflow 2.5.0. Moreover, Jupyter Notebook was used in Anaconda on the Windows 10 operating system [47].

3.3.2. Input Setup

The input images were resized to $224 \times 224 \times 3$ pixels. The images were re-scaled in the range [0, 1]. Image augmentation was applied to the images and included operations, such as shear, zoom, and horizontal flipping.

3.3.3. Hyperparameter Choice

It is essential to mention here that the authors used the initial weights from pre-trained state-of-the-art models on ImageNet using transfer learning and performed full training on the dataset. The initial and final learning rate, convergence, and patience represent initial training on the dataset and are only for the additional layers, base model was kept frozen to adjust the weights of those layers according to the new dataset and not starting with random weights. The base model represents complete training on our model with all layers being unfrozen.

The hyperparameters used in the research work are shown in Table 1. For MobileNetV2, the hyperparameters were kept the same as in a previous reference study [23] that achieved the highest performance among several different state-of-the-art models; this was to maintain a reference when comparing to other models. Some of the hyperparameters for the other two models were altered [48]. The final learning rate for all layers of Xception was kept to 1.00×10^{-5}. However, the number of epochs was reduced to 200 and 250 for the Custom-Model and Xception, respectively, with the learning rate 1.00×10. The latter models were found to converge within the specified epochs. A batch size of 32 was selected for the Custom-Model and Xception, as recommended in a previous study [49].

Table 1. Hyperparameter Comparison.

Architecture	Optimizer	Learning Rate	Batch Size	Image Shuffling in Batches	Number of Epochs for Convergence	Patience	Number of Epochs
Custom-Model	Adam	1.00×10^{-3}	32	Yes	N/A	N/A	200
MobileNetV2	Adam	1.00×10^{-5}	200	Yes	573	100	1000
Xception	Adam	Initial LR = 1.00×10^{-3}, Final LR = 1.00×10^{-5}	32	Yes	126 (Initial Convergence), 146 (Final Convergence)	50 (Initial Patience), 50 (Final Patience)	250

3.3.4. Train–Test Ratio

The training set comprised 70% of the images, while the tested images formed 30% of the dataset. The training and test set was evaluated on the images for generalization behavior.

3.3.5. Evaluation Metrics and Tools

The authors aimed to provide and contribute a thorough and comprehensive assessment of deep learning models on the dataset. The selected different evaluation criteria were precision, recall, F1 score, accuracy, confusion matrix, and region of convergence curves (ROC). Each evaluation criterion is detailed below.

1. Precision

Precision measures the accuracy of the positive diagnosed cases. In the subject context, it tells us how accurate the results are, if a person tests positive with COVID-19. In simple words, it is classified as:

$$\text{Precision} = \frac{\text{True Postive}}{\text{Total Classified Positive}} \tag{6}$$

The actual formula for the calculation in this research work is in (7).

$$\text{Precision} = \frac{\text{True Postive}}{\text{True Positive} + \text{False Positive}} \tag{7}$$

Let $C_{\downarrow}$ represent the subclasses of the class C, $C_{\downarrow}^{C}$ represent the positive classes predicted by the classifier, and $C_{\downarrow}^{d}$ represent the class labels for positive cases in the dataset; precision for a subclass can then be defined as follows:

$$\text{Precision}_{\downarrow} = \frac{\left| C_{\downarrow}^{C} \cap C_{\downarrow}^{d} \right|}{C_{\downarrow}^{C}} \tag{8}$$

2. Recall

This tells the percentage of true positive predicted by the actual number of true positives. It can be calculated using the formulae in (9) and (10).

$$\text{Recall} = \frac{\text{True Positive}}{\text{Total Actual Positive}} \tag{9}$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \tag{10}$$

As defined for precision, recall can similarly be defined using the same representations C, $C_{\downarrow}^{C}$. and $C_{\downarrow}^{d}$, as shown below.

$$\text{Recall}_{\downarrow} = \frac{\left| C_{\downarrow}^{C} \cap C_{\downarrow}^{d} \right|}{C_{\downarrow}^{d}} \tag{11}$$

3. F1 Score

This shows us the balance between precision and recall. The F1 score for each class is calculated employing the mathematical statement in (12) below.

$$\text{F1Score}_{\downarrow} = 2 \times \frac{\text{Precision}_{\downarrow} \times \text{Recall}_{\downarrow}}{\text{Precision}_{\downarrow} + \text{Recall}_{\downarrow}} \tag{12}$$

For COVID-19, the F1 score will measure the potency of the model to detect all the COVID-19 cases, to avoid other classes.

4. Accuracy

This is the percentage of correctly identified classes divided by the total number of cases. For a multiclass problem with classes i, mathematically, it is given in (5).

$$\text{Accuracy} = \frac{\text{True Positive}_{\text{i}} + \text{True Negative}_{\text{i}}}{\text{True Positive}_{\text{i}} + \text{True Negative}_{\text{i}} + \text{False Positive}_{\text{i}} + \text{ False Negative}_{\text{i}}} \tag{13}$$

5. ROC (Receiver Operating Characteristic) Curve

The ROC curve shows the performance of the classifier at different threshold settings. By varying threshold settings, TPR (true positive rate or recall or sensitivity) is plotted against the FPR (false positive rate or specificity).

AUC (area under the curve) represents the area under the ROC curve.

6. Confusion Matrix

This is an evaluation tool that shows the number corresponding to correct and incorrect predictions down to individual class levels. It shows metrics, such as true positive (TP), true negative (TN), false positive (FP), and false negative (FN) which were used in the performance assessment.

4. Results

This section elaborates the evaluation of the deep learning models on the dataset [30] relying on the above-mentioned metrics, namely, accuracy, precision, recall, and F1 score. Furthermore, we also show the ROC (receiver operating characteristic) curves, AUC (area under curve) and macro and weighted averages. The contemplated metrics plots and curves are presented in the subsequent sections.

4.1. Accuracy

The accuracy metric is plotted in Figure 7 for Xception, MobileNetV2, and the Custom-Model, the latter being used as a baseline for evaluating the dataset as it had the highest number of COVID-19 chest X-ray images.

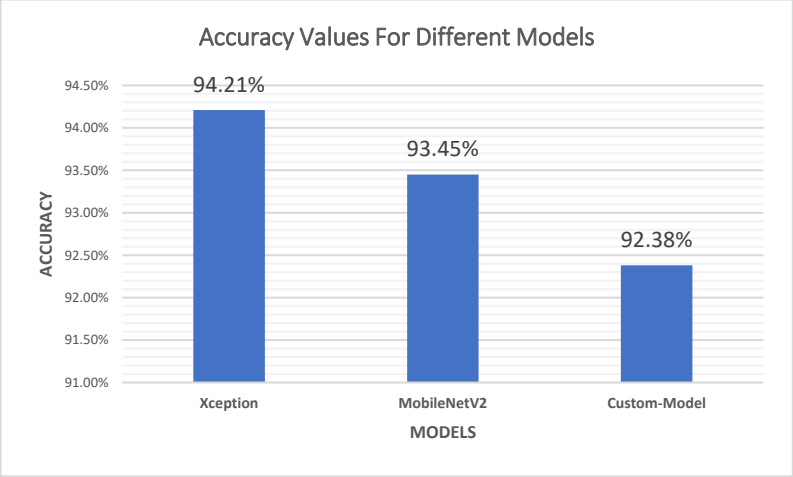


Figure 7. Accuracy plot for different models in terms of percentages.

From Figure 7, it is evident that the Xception model exceeds the other models in accuracy. The second model, which precedes, is the MobileNetV2 model. It is to be noted that the highest accuracy achieved from these models is 94.2%. Xception outclassed the other deep learning models, as indicated by Figure 7.

4.2. Precision

The precision plots are shown in Figure 8 which highlights the top performance of Xception model for correctly predicting the positive cases. It can be seen that the custom Custom-Model places second as the top-performing model, followed by MobileNetV2 model.

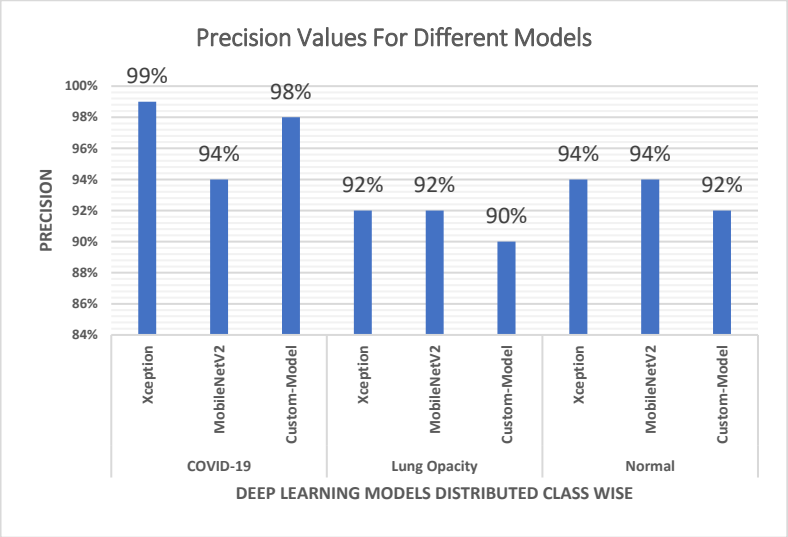


Figure 8. Precision plot for different models in terms of percentages.

A close look at the precision plot supports the view that the MobileNetV2 model performed significantly lower than the other models by a difference of 4–5% for the prediction of COVID-19 cases in precision values. Furthermore, it can be observed that both the Custom-Model and Xception model are more inclined towards making more COVID-19 true predictions when compared with the performance on predicting the normal and lung opacity cases, hence, providing better results for COVID-19 scenarios. The metrics values for different classes vary from 90% to 98% for the Custom-Model and from 92% to 99% in the case of the Xception model.

4.3. Recall

Looking at the recall values exhibited in Figure 9, the highest recall value is reached by both the Xception model and the MobileNetV2 model, both outperforming the Custom-Model. It can be seen that the highest obtained value is 95%, while the lowest is 92% for COVID-19.

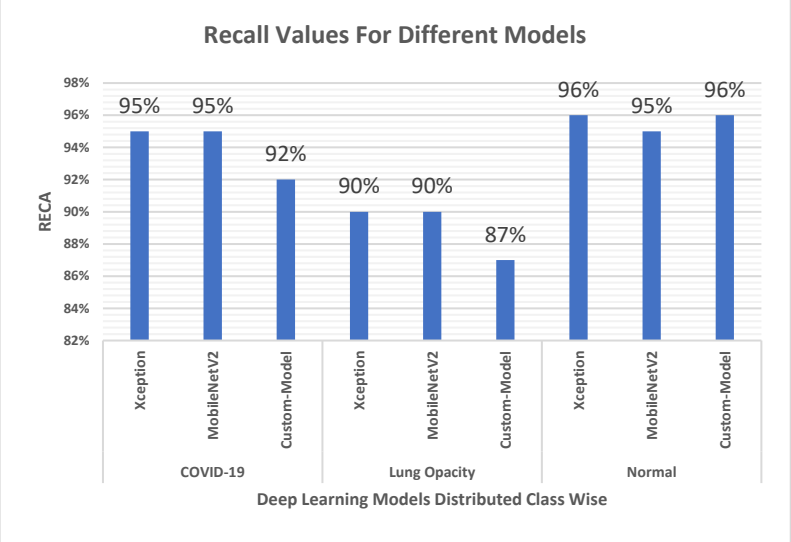


Figure 9. Recall plot for different models in terms of percentages.

Figure 9 also elaborates low recall values for identifying lung opacity when compared with normal and COVID-19 cases, resulting in more false negative cases, while the latter classes only have a slight difference in performance in terms of recall.

4.4. F1 Score

The F1 score is presented in Figure 10 and shows that the Xception model outperforms the other models, while they reached equivalent values. The Xception model exceeds the others by 2%. The results show that Xception is preferred over the other two models when a balance is to be kept between precision and recall values.

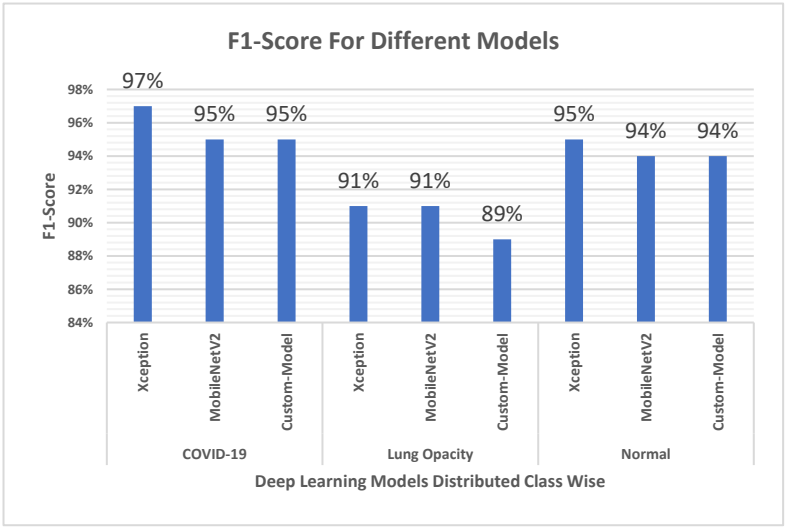


Figure 10. F1 score plot for different models in terms of percentages.

4.5. Confusion Matrix, ROC Curves and Area under Curve

The confusion matrices for the three classes are presented in Figures 11–13.

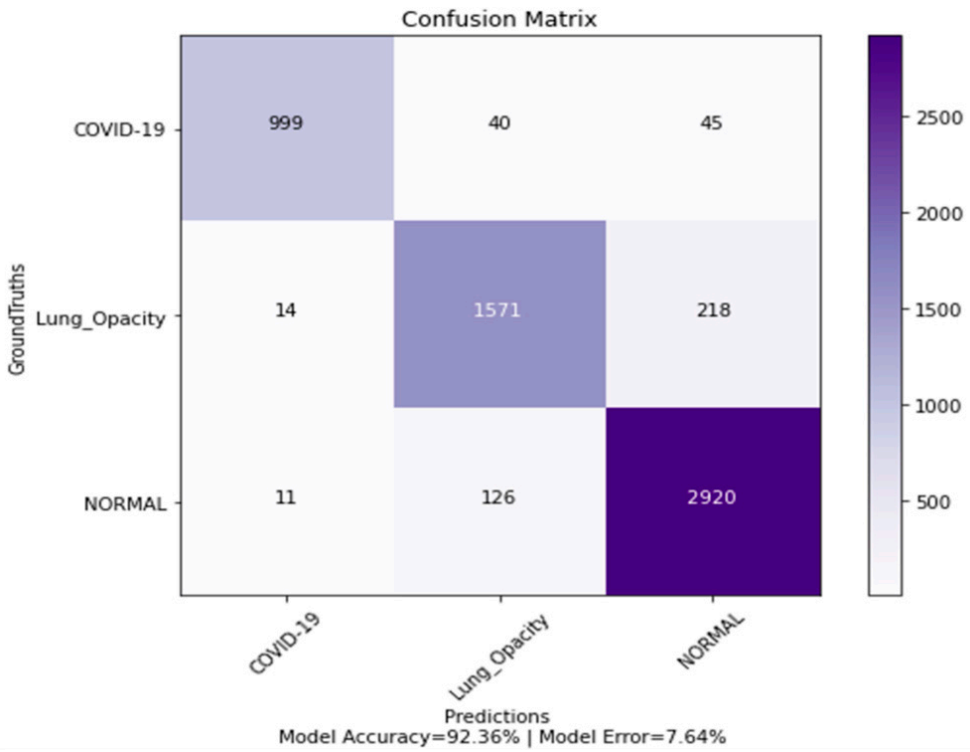


Figure 11. Confusion matrix for Custom-Model.

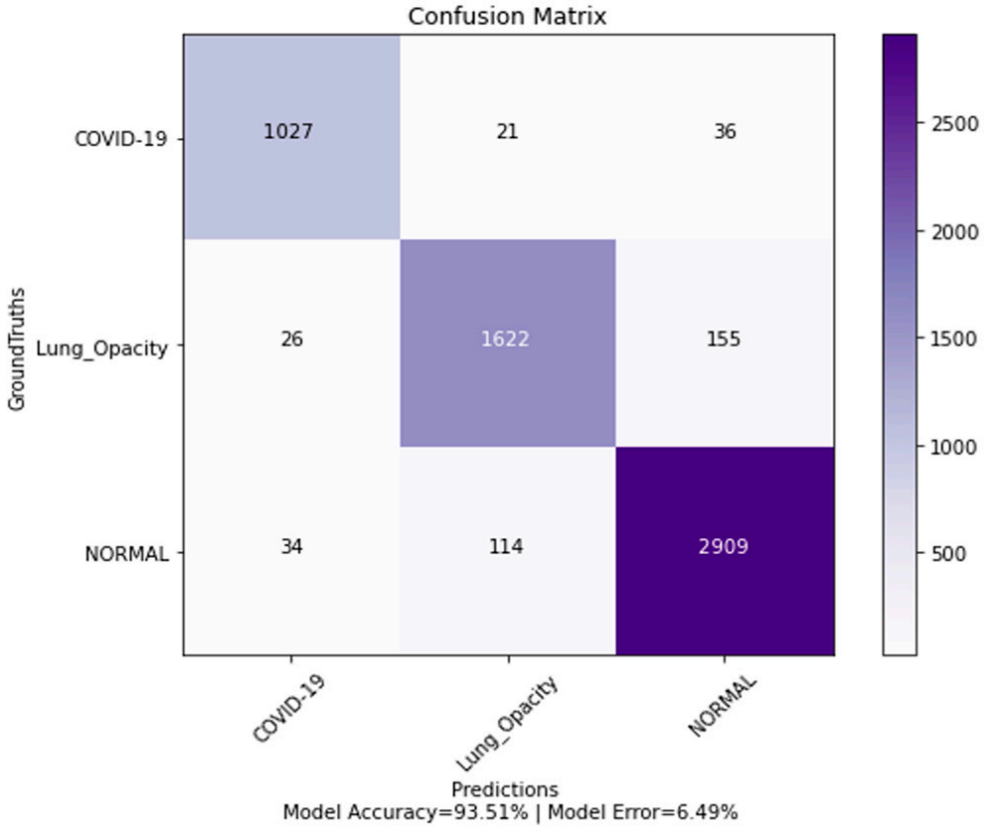


Figure 12. Confusion matrix for MobileNetV2 model.

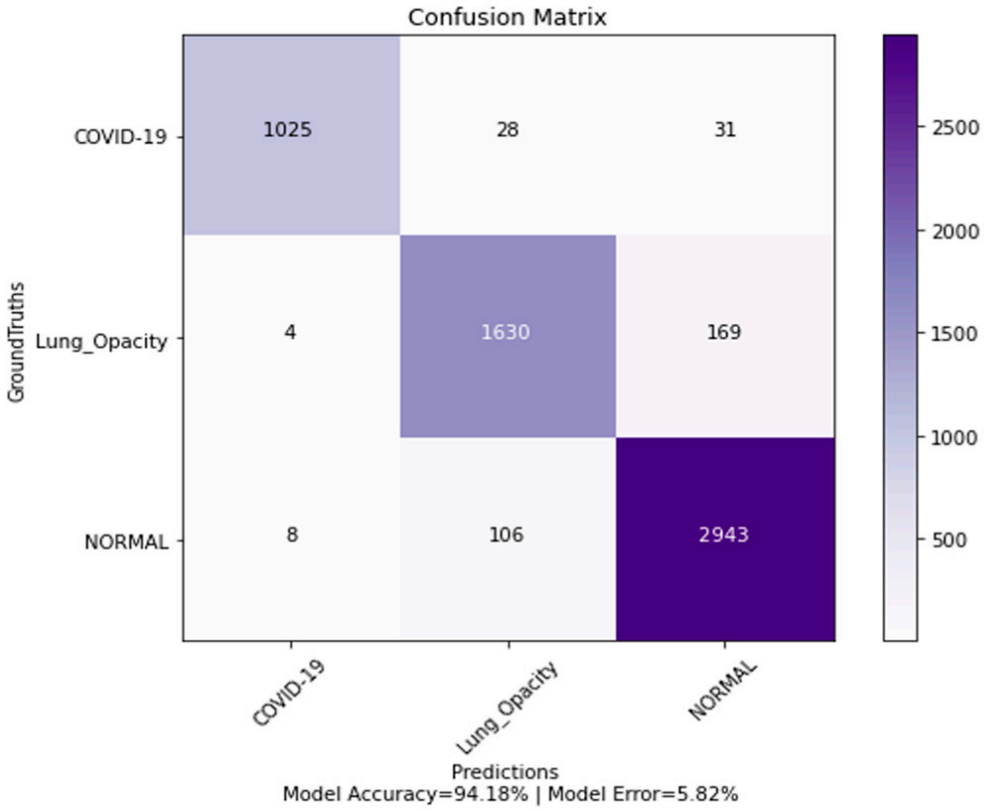


Figure 13. Confusion matrix for Xception model.

From the confusion matrix, it can be seen that the MobileNetV2 model was able to predict two more COVID-19 cases than Xception and Custom-Model. It is better than the Xception model in this regard. The total number of correctly predicted cases by the MobileNetV2 model is 1027, while in the case of Xception, it is 1025.

Moreover, the Xception model predicted the smallest number of false positive cases compared with Custom-Model and MobileNetV2 which performed least accurately in this regard. The total number of false positive cases for the Xception model is 12, for Custom-Model it is 25, and for MobileNetV2 it is 60. The number of false negative cases for the MobileNetV2 model is slightly lower than that of the Xception model, their values being 57 and 59, respectively.

The ROC curves shown below are plotted for Class 0 (COVID-19), Class 1 (lung opacity) and Class 2 (normal). The ROC curve for custom-model is shown in the Figure 14.

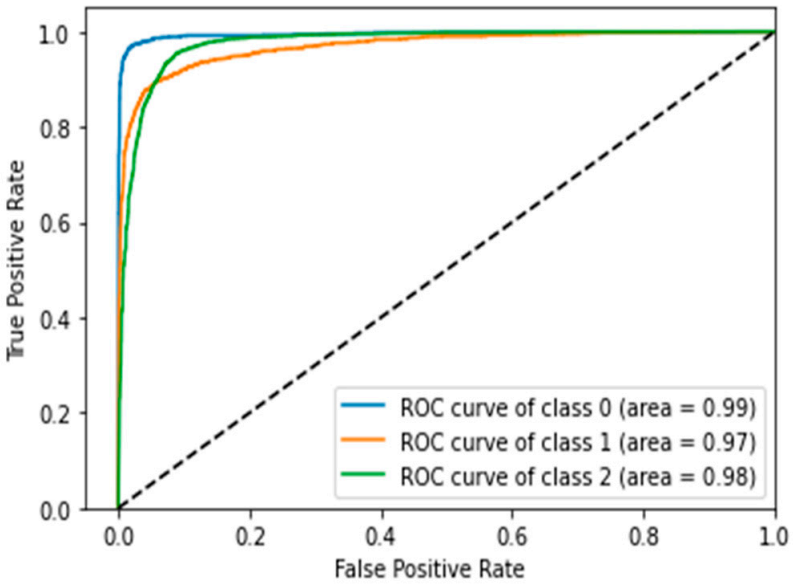


Figure 14. ROC curve for Custom-Model.

The ROC curve for MobileNetV2 is shown in the Figure 15.

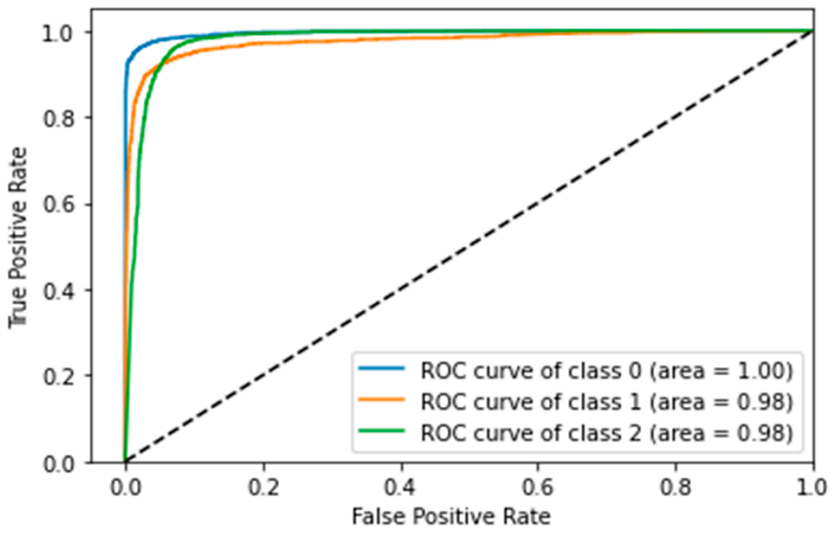


Figure 15. ROC curve for MobileNetV2 model.

The ROC curve for Xception model is shown in the Figure 16.

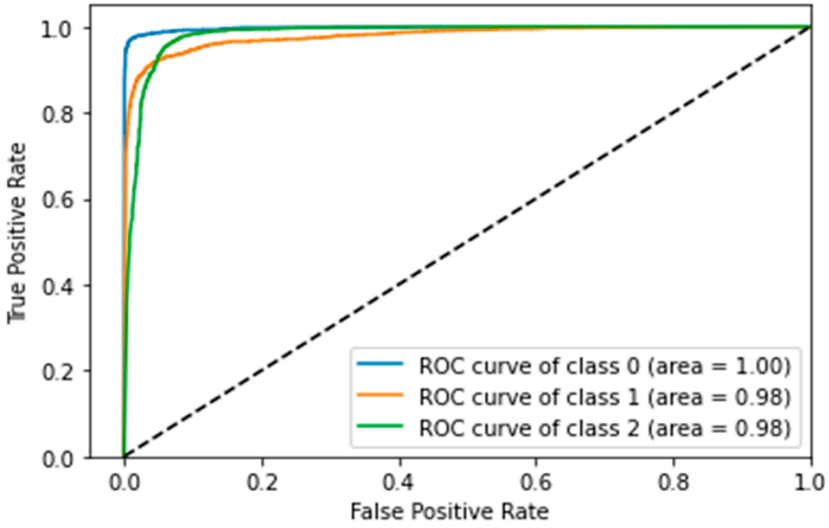


Figure 16. ROC curve for Xception model.

The curves are similar for all three models; however, the AUC for the MobileNetV2 model and the Xception model are equivalent, having a value of 1.00, while Custom-Model scored 0.99 for the COVID-19 class.

Further, performance differentiation can be seen in Table 2, which shows that the Xception model performed better than the other models in terms of macro and weighted averages for the COVID-19 class.

Table 2. Macro and weighted averages of different scores.

Architecture	Average Scheme	Precision	Recall	F1 Score
Custom-Model	Macro Average	0.93	0.92	0.92
MobileNetV2 Model		0.94	0.93	0.93
Xception Model		0.95	0.94	0.94
Custom-Model	Weighted Average	0.92	0.92	0.92
MobileNetV2 Model		0.93	0.94	0.93
Xception Model		0.94	0.94	0.94

5. Discussion

This study, firstly compared three deep learning models to evaluate their performance on a dataset with the highest number of COVID-19 chest X-ray images, to the best of the authors' knowledge. The first model, Custom-Model, was developed by the authors. Secondly, the MobileNetV2-based model [23], which comprised the MobileNetV2 model with added layers, and finally, the Xception model with added layers were evaluated using different metrics, specifically, accuracy, precision, recall, and F1 score. Moreover, ROC, AUC, and macro and weighted averages were also compared.

The Xception model was found to exceed the other models in terms of performance percentages. However, the confusion matrix demonstrated that the MobileNetV2 model correctly identified two more COVID-19 cases than did the Xception model. Also, it should be added that MobileNetV2 misclassified a far higher number of the other classes as COVID-19.

In terms of precision, the custom-developed Custom-Model performed better than the MobileNetV2 model by a difference of 4%, making it the best performer compared with various state-of-the-art models in a previous study [50]. The difference in precision values of the MobileNetV2 model and the Xception model is 5%. From the confusion matrix, it can be seen that MobileNetV2 wrongly predicted 60 non-COVID-19 cases to be positive, while this number was 12 for Xception, and 25 for Custom-Model. Precision gives more weighting

to Custom-Model and Xception when compared with MobileNetV2 because if more people are wrongly identified as positive for COVID-19 and sent to treatment facilities (hospitals). Further, as reported by The New York Times in the USA, the COVID-19 pandemic led to the flooding of hospitals with patients causing a lack of available beds, and of doctors and nurses. Therefore, in such an alarming situation, it is critical for the diagnostic tests to be more accurate in diagnosing positive results for COVID-19.

The MobileNetV2 model, however, exceeds slightly–by 2 cases–in recall (i.e., sensitivity) when compared with the Xception model. Custom-Model misclassified 85 COVID-19 cases as normal and lung opacity, Xception misclassified 59, and MobileNetV2 misclassified 57 COVID-19 cases. This reflects the case where false negatives are also important, as accrediting COVID-19 to other classes will deprive the patients of treatment and will also result in non-isolation of the patient, therefore, possibly causing an increase in the spread of the virus.

Overall, the Xception model with some added layers outperformed the two other models well and exceeds both of them in terms of F1 score. The ROC curves for the three models are comparable; however, MobileNetV2 and Xception exceed the third model in terms of the AUC.

The number of COVID-19 images from the literature, as covered by the authors, contained a minimum of 127 images to a maximum of 850 and few are high imbalance in terms of class. The highest accuracy found in the literature surveyed was a maximum of 99.9% for binary classification where the number of images of the other classes (400) were kept equal to the number of images of the subject virus. The maximum accuracy achieved on binary classification from the literature was 99.90%, while on multi-class classification, it was 99.658% on the test set. The image distribution for the former was 400 each for COVID-19 and pneumonia cases and, for the latter, it was 2911 images divided into 237 COVID-19 positive images, 1338 normal images, and 1336 viral pneumonia images. The current work utilized the Kaggle database, which contains 3616 COVID-19 chest X-ray images, and which is about 426% the size of the maximum number of COVID-19 images utilized in the covered literature.

For the comparison with previous research work, the authors used the given hyperparameters and the same extended architecture employed in a previous study using MobileVNet2 as a base model. The previous metrics evaluations were 99.658% for accuracy, and 1 for precision, recall, and F1 score from the literature work. From the analysis of the current dataset, the values for the same architecture were 93.45% for accuracy, 94% for precision, and 95% for both recall and F1 score. Drops of 5% and 6% were observed in the MobileNetV2 model during the subject evaluation, where accuracy and precision values dropped by approximately 6%, while a curtail of 5% was observed for recall and the F1 score. The Table 3 is shown the comparison of previous state-of the-art studies.

Table 3. State-of-the-art comparison COVID-19 studies.

Studies	Cases Number X-ray Datasets			Method Utilized	Accuracy (%) Binary Classification	Accuracy (%) Multi-Class Classification
	COVID-19	Pneumonia	Normal			
[23]	237	1336	1338	Extended MobileNetV2 architecture	N/A	99.66%
[26]	400	400		COVID-CheXNet architecture	99.90%	
[25]	415	5179	2880	VGG16 CNN		93.9
[29]	850	500	915	Inception ResNetV2, InceptionNetV3 and NASNetLarge	N/A	97.87%, 97.87% and 96.24%

Table 3. *Cont.*

Studies	Cases Number	X-ray Datasets		Method Utilized	Accuracy (%) Binary Classification	Accuracy (%) Multi-Class Classification
[31]	250	2753	3520	VGG16 with AveragePooling2D, Flatten, Dense, Dropout and a Dense layer using Softmax function	N/A	97%.
[32]	180		200	ResNet18, ResNet50, ResNet101, VGG16, and VGG19 with different SVM kernel functions (Linear, Quadratic, Cubic, and Gaussian) and a customized CNN model	92.6% (Pre-Trained Resnet 50 Model) 91.6% (Proposed CNN Model, Average accuracy using different SVM kernels)	
[33]	538		468	Ensemble of three models	95.70%	
[34]	401	401	401	Ensemble of two state-of-the-art classifiers with additional CNN layers	96.15%	99.21%
[27]	181		364	Modified VGG-19 architecture	96.30%	
[35]	404	404	404	Extended EfficientNetB4 architecture	99.62%	97.11%
[39]	127	500	500	Extended Xception architecture	N/A	94.59%
[39]	157	500	500	Extended Xception architecture	N/A	90.21%

Another comparison from a literature study where CoroNet, a deep learning architecture comprising Xception as proposed. The model achieved 94.59% accuracy, 95% precision, 96.9% recall, and 95.6% F1 score on a 3-class classification. The evaluation on another dataset in the same study demonstrated an accuracy of 90.21%, precision of 97%, recall of 89%, and an F1 score of 93%, with the dataset composition comprising 157, 500, and 500 chest X-ray images for COVID-19, normal, and pneumonia, respectively. The current study used an Xception-based model with a global average pooling layer and a classification layer (Softmax), and showed 94.21% accuracy, 99% precision, 95% recall, and a 97% F1 score. The comparison showed an increase of 4% in precision, while other values were competitive. An overall increase of 4% to 6% was observed for the Xception model when compared with the lower values of metrics from the above-mentioned literature work.

The summary of the literature research work and the chest X-ray image distribution for each class in the dataset is shown in Table 1. The highest accuracy achieved for COVID-19 detection using lung X-ray images is 99.658% for 3-class classification. However, the previous studies had limitations in that a high imbalance in class was observed and further, contained very few COVID-19 chest X-ray images. As already mentioned, the highest number of images for COVID-19 used was 850, whereas the number of COVID-19 samples in this research was about 426% of this number (3616 images). Therefore, the previous work limited in terms of the low number of COVID-19 cases in the dataset and the class imbalance. With this dataset, the authors achieved a maximum accuracy of 94.21%, which is greater than the accuracy achieved for the Xception-based model by 4%. However, it is less than the accuracy achieved in a literature research work by a maximum of approximately 4.4% using another model. Overall, it can be concluded that the accuracy achieved in the current work is reduced by approximately 2.2%, when compared for a 3-class classification.

Also, the comparison is limited by the number of deep learning models employed, i.e., only three in this study. Recent studies showed that a dataset, when collected from

different sources, can lead to dataset bias [51,52] where the model learns the source of the images rather than the disease itself. Images from different sources can have different characteristics, such as overall pixel intensities and corner labels, and may present electrodes with cables in the X-ray, bras, etc. The model will tend to learn these characteristics and associate these for the classification of the disease. For future work, the dataset should also be analyzed for bias. Moreover, other state-of-the-art models can be evaluated on the dataset utilized in the current study. Further, from the obtained performance of the custom CNN model, the authors would like to enhance it further to an advanced level where it would lead to a better diagnosing and demarcating of the subject virus cases from other cases.

6. Conclusions

Artificial intelligence was used to deploy experimentation on a dataset containing the largest number of COVID-19 images ever used, to the best of the authors' knowledge and based on a survey of the literature. For the subject purpose, a CNN model was created and deployed, and state-of-the-art deep learning models were selected, including one which achieved top performance in a previous study. The main purpose was to test the performance of deep learning algorithms on the updated COVID-19 database, and also to develop and test a custom deep learning model on the subject dataset. However, there was no intention to develop a highly remarkable deep learning model, but rather to test the performance on a relatively larger database of chest X-ray images. Moreover, the intention was to determine the performance gains on the utilized, larger COVID-19 chest X-ray image dataset.

The maximum performance from among the deployed models on the subject dataset was 94.206% in terms of accuracy, 99% in terms of precision, 95% in terms of recall, and 97% in terms of F1 score. Promising results were found for deep learning when deployed for the diagnosis of COVID-19 using chest X-rays. In terms of precision, it was found that 99% of the predicted positive COVID-19 cases were accurate, which is of high scientific value. Furthermore, the experimentation on the subject dataset showed that a higher accuracy in precision is obtained when compared with recall, demonstrating that the percentage of correctly identified positive cases than the percentage of total correctly identified COVID-19 cases out of overall positive COVID-19 cases in the dataset. The MobileNetV2 model succeeded in predicting slightly more cases of COVID-19 correctly, while producing far more wrong predictions of COVID-19 cases. Also, the MobileNetV2 model gave slightly lower misclassified COVID-19 cases when compared with Xception. In terms of accuracy, the current evaluation shows a decrease of 2% in accuracy when compared with the average accuracy of covered literature work for 3-class classification. Although this study offers a detailed and thorough insight into the performance results of a larger COVD-19 chest X-ray image dataset, we, nevertheless, believe that the performance analysis of deep learning models on COVID-19 will become increasingly mature with a competitive ratio of analyzed subject virus chest X-ray images to other, normal, and abnormal images in the dataset. It is essential to mention here that although we utilized a large database with the highest number of chest X-ray images of the subject virus, we still had a data imbalance where COVID-19 chest X-ray images formed 18% of the total images, and lung opacity and normal cases formed 30% and 52%, respectively.

Author Contributions: Conceptualization, investigation, methodology, resources, software, and writing–original draft: A.T.N., M.J.A. and O.T.; investigation, validation, review and editing: A.T.N., M.J.A. and O.T.; conceptualization, software, writing—review and editing: M.A.M., A.M. (Amena Mahmoud) and A.M. (Arnab Majumdar); writing—review and editing A.T.N., M.J.A., M.A.M., A.M. (Amena Mahmoud) and O.T. All authors have read and agreed to the published version of the manuscript.

Funding: The authors received no specific funding for this study.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: This dataset available online and anyone can be used.

Acknowledgments: This research work was partially supported by Chiang Mai University and University of Management and Technology.

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

References

1. Hasoon, J.N.; Fadel, A.H.; Hameed, R.S.; Mostafa, S.A.; Khalaf, B.A.; Mohammed, M.A.; Nedoma, J. COVID-19 anomaly detection and classification method based on supervised machine learning of chest X-ray images. *Results Phys.* **2021**, *31*, 105045. [CrossRef] [PubMed]
2. Awan, M.J.; Bilal, M.H.; Yasin, A.; Nobanee, H.; Khan, N.S.; Zain, A.M. Detection of COVID-19 in Chest X-ray Images: A Big Data Enabled Deep Learning Approach. *Int. J. Environ. Res. Public Health* **2021**, *18*, 10147. [CrossRef] [PubMed]
3. Haafza, L.A.; Awan, M.J.; Abid, A.; Yasin, A.; Nobanee, H.; Farooq, M.S. Big Data COVID-19 Systematic Literature Review: Pandemic Crisis. *Electronics* **2021**, *10*, 3125. [CrossRef]
4. Gupta, M.; Jain, R.; Arora, S.; Gupta, A.; Javed Awan, M.; Chaudhary, G.; Nobanee, H. AI-enabled COVID-19 Outbreak Analysis and Prediction: Indian States vs. Union Territories. *Comput. Mater. Contin.* **2021**, *67*, 933–950. [CrossRef]
5. Awan, M.J.; Imtiaz, M.W.; Usama, M.; Rehman, A.; Ayesha, N.; Shehzad, H.M.F. COVID-19 Detection by using Deep learning-based Custom Convolution Neural Network (CNN). In Proceedings of the 2021 International Conference on Innovative Computing (ICIC), Lahore, Pakistan, 9–10 November 2021; pp. 1–7.
6. Abdulkareem, K.H.; Mostafa, S.A.; Al-Qudsy, Z.N.; Mohammed, M.A.; Al-Waisy, A.S.; Kadry, S.; Lee, J.; Nam, Y. Automated System for Identifying COVID-19 Infections in Computed Tomography Images Using Deep Learning Models. *J. Healthc. Eng.* **2022**, *2022*, 5329014. [CrossRef] [PubMed]
7. Mahmoudi, R.; Benameur, N.; Mabrouk, R.; Mohammed, M.A.; Garcia-Zapirain, B.; Bedoui, M.H. A Deep Learning-Based Diagnosis System for COVID-19 Detection and Pneumonia Screening Using CT Imaging. *Appl. Sci.* **2022**, *12*, 4825. [CrossRef]
8. Ibrahim, D.A.; Zebari, D.A.; Mohammed, H.J.; Mohammed, M.A. Effective hybrid deep learning model for COVID-19 patterns identification using CT images. *Expert Syst.* **2022**, e13010. [CrossRef]
9. Alyasseri, Z.A.A.; Al-Betar, M.A.; Doush, I.A.; Awadallah, M.A.; Abasi, A.K.; Makhadmeh, S.N.; Alomari, O.A.; Abdulkareem, K.H.; Adam, A.; Damasevicius, R.; et al. Review on COVID-19 diagnosis models based on machine learning and deep learning approaches. *Expert Syst.* **2022**, *39*, e12759. [CrossRef]
10. Awan, M.J.; Rahim, M.S.M.; Salim, N.; Rehman, A.; Nobanee, H.; Shabir, H. Improved Deep Convolutional Neural Network to Classify Osteoarthritis from Anterior Cruciate Ligament Tear Using Magnetic Resonance Imaging. *J. Pers. Med.* **2021**, *11*, 1163. [CrossRef]
11. Ali, Y.; Farooq, A.; Alam, T.M.; Farooq, M.S.; Awan, M.J.; Baig, T.I. Detection of Schistosomiasis Factors Using Association Rule Mining. *IEEE Access* **2019**, *7*, 186108–186114. [CrossRef]
12. Vaishnav, P.K.; Sharma, S.; Sharma, P. Analytical review analysis for screening COVID-19 disease. *Int. J. Mod. Res.* **2021**, *1*, 22–29.
13. Tariq, A.; Awan, M.J.; Alshudukhi, J.; Alam, T.M.; Alhamazani, K.T.; Meraf, Z. Software Measurement by Using Artificial Intelligence. *J. Nanomater.* **2022**, *2022*, 7283171. [CrossRef]
14. Nagi, A.T.; Wali, A.; Shahzada, A.; Ahmad, M.M. A Parellel two Stage Classifier for Breast Cancer Prediction and Comparison with Various Ensemble Techniques. *VAWKUM Trans. Comput. Sci.* **2018**, *15*, 121. [CrossRef]
15. Naseem, U.; Rashid, J.; Ali, L.; Kim, J.; Emad Ul Haq, Q.; Awan, M.J.; Imran, M. An Automatic Detection of Breast Cancer Diagnosis and Prognosis based on Machine Learning Using Ensemble of Classifiers. *IEEE Access* **2022**. [CrossRef]
16. Awan, M.J.; Mohd Rahim, M.S.; Salim, N.; Rehman, A.; Nobanee, H. Machine Learning-Based Performance Comparison to Diagnose Anterior Cruciate Ligament Tears. *J. Healthc. Eng.* **2022**, *2022*, 255012. [CrossRef]
17. Sharma, T.; Nair, R.; Gomathi, S. Breast Cancer Image Classification using Transfer Learning and Convolutional Neural Network. *Int. J. Mod. Res.* **2022**, *2*, 8–16.
18. Allioui, H.; Mohammed, M.A.; Benameur, N.; Al-Khateeb, B.; Abdulkareem, K.H.; Garcia-Zapirain, B.; Damaševičius, R.; Maskeliūnas, R. A multi-agent deep reinforcement learning approach for enhancement of COVID-19 CT image segmentation. *J. Pers. Med.* **2022**, *12*, 309. [CrossRef]
19. Shamim, S.; Awan, M.J.; Mohd Zain, A.; Naseem, U.; Mohammed, M.A.; Garcia-Zapirain, B. Automatic COVID-19 Lung Infection Segmentation through Modified Unet Model. *J. Healthc. Eng.* **2022**, *2022*, 6566982. [CrossRef]
20. Albahli, A.S.; Algsham, A.; Aeraj, S.; Alsaeed, M.; Alrashed, M.; Rauf, H.T.; Arif, M.; Mohammed, M.A. COVID-19 public sentiment insights: A text mining approach to the Gulf countries. *CMC Comput. Mater. Contin.* **2021**, *67*, 1613–1627. [CrossRef]
21. Obaid, O.I.; Mohammed, M.A.; Mostafa, S.A. Long Short-Term Memory Approach for Coronavirus Disease Predicti. *J. Inf. Technol. Manag.* **2020**, *12*, 11–21.
22. 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*, 245. [CrossRef] [PubMed]

23. Ben Jabra, M.; Koubaa, A.; Benjdira, B.; Ammar, A.; Hamam, H. COVID-19 diagnosis in chest X-rays using deep learning and majority voting. *Appl. Sci.* **2021**, *11*, 2884. [CrossRef]
24. Chandra, T.B.; Verma, K.; Singh, B.K.; Jain, D.; Netam, S.S. Coronavirus disease (COVID-19) detection in chest X-ray images using majority voting based classifier ensemble. *Expert Syst. Appl.* **2021**, *165*, 113909. [CrossRef] [PubMed]
25. Heidari, M.; Mirniaharikandehei, S.; Khuzani, A.Z.; Danala, G.; Qiu, Y.; Zheng, B. Improving the performance of CNN to predict the likelihood of COVID-19 using chest X-ray images with preprocessing algorithms. *Int. J. Med. Inform.* **2020**, *144*, 104284. [CrossRef]
26. Al-Waisy, A.S.; Al-Fahdawi, S.; Mohammed, M.A.; Abdulkareem, K.H.; Mostafa, S.A.; Maashi, M.S.; Arif, M.; Garcia-Zapirain, B. COVID-CheXNet: Hybrid deep learning framework for identifying COVID-19 virus in chest X-rays images. *Soft Comput.* **2020**. [CrossRef]
27. Awan, M.J.; Rahim, M.S.M.; Salim, N.; Mohammed, M.A.; Garcia-Zapirain, B.; Abdulkareem, K.H. Efficient Detection of Knee Anterior Cruciate Ligament from Magnetic Resonance Imaging Using Deep Learning Approach. *Diagnostics* **2021**, *11*, 105. [CrossRef]
28. Vaid, S.; Kalantar, R.; Bhandari, M. Deep learning COVID-19 detection bias: Accuracy through artificial intelligence. *Int. Orthop.* **2020**, *44*, 1539–1542. [CrossRef]
29. Albahli, S.; Albattah, W. Detection of coronavirus disease from X-ray images using deep learning and transfer learning algorithms. *J. X-ray Sci. Technol.* **2020**, *28*, 841–850. [CrossRef]
30. Brunese, L.; Mercaldo, F.; Reginelli, A.; Santone, A. Explainable deep learning for pulmonary disease and coronavirus COVID-19 detection from X-rays. *Comput. Methods Programs Biomed.* **2020**, *196*, 105608. [CrossRef]
31. Ismael, A.M.; Şengür, A. Deep learning approaches for COVID-19 detection based on chest X-ray images. *Expert Syst. Appl.* **2021**, *164*, 114054. [CrossRef]
32. Awan, M.J.; Masood, O.A.; Mohammed, M.A.; Yasin, A.; Zain, A.M.; Damaševičius, R.; Abdulkareem, K.H. Image-Based Malware Classification Using VGG19 Network and Spatial Convolutional Attention. *Electronics* **2021**, *10*, 2444. [CrossRef]
33. Das, A.K.; Ghosh, S.; Thunder, S.; Dutta, R.; Agarwal, S.; Chakrabarti, A. Automatic COVID-19 detection from X-ray images using ensemble learning with convolutional neural network. *Pattern Anal. Appl.* **2021**, *24*, 1111–1124. [CrossRef]
34. Gianchandani, N.; Jaiswal, A.; Singh, D.; Kumar, V.; Kaur, M. Rapid COVID-19 diagnosis using ensemble deep transfer learning models from chest radiographic images. *J. Ambient. Intell. Humaniz. Comput.* **2020**. [CrossRef] [PubMed]
35. Marques, G.; Agarwal, D.; Díez, I.D.L.T. Automated medical diagnosis of COVID-19 through EfficientNet convolutional neural network. *Appl. Soft Comput.* **2020**, *96*, 106691. [CrossRef]
36. Maghdid, H.S.; Asaad, A.T.; Ghafoor, K.Z.; Sadiq, A.S.; Mirjalili, S.; Khan, M.K. Diagnosing COVID-19 pneumonia from X-ray and CT images using deep learning and transfer learning algorithms. *Multimodal Image Exploit. Learn.* **2021**, *11734*, 99–110.
37. Hall, L.O.; Paul, R.; Goldgof, D.B.; Goldgof, G.M. Finding COVID-19 from Chest X-rays using Deep Learning on a Small Dataset. *arXiv* **2020**, arXiv:2004.02060.
38. Minaee, S.; Kafieh, R.; Sonka, M.; Yazdani, S.; Soufi, G.J. Deep-COVID: Predicting COVID-19 from chest X-ray images using deep transfer learning. *Med. Image Anal.* **2020**, *65*, 101794. [CrossRef]
39. Khan, A.I.; Shah, J.L.; Bhat, M.M. CoroNet: A deep neural network for detection and diagnosis of COVID-19 from chest X-ray images. *Comput. Methods Programs Biomed.* **2020**, *196*, 105581. [CrossRef]
40. Pandit, M.K.; Banday, S.A. SARS n-CoV2-19 detection from chest X-ray images using deep neural networks. *Int. J. Pervasive Comput. Commun.* **2020**, *16*, 419–427. [CrossRef]
41. Rahman, T.; Chowdhury, D.M.; Khandakar, A. COVID-19 Radiography Database. 2021. Available online: https://www.kaggle.com/tawsifurrahman/covid19-radiography-database (accessed on 19 March 2022).
42. Chowdhury, M.E.; Rahman, T.; Khandakar, A.; Mazhar, R.; Kadir, M.A.; Mahbub, Z.B.; Islam, K.R.; Khan, M.S.; Iqbal, A.; Al Emadi, N. Can AI help in screening viral and COVID-19 pneumonia? *IEEE Access* **2020**, *8*, 132665–132676. [CrossRef]
43. Mujahid, A.; Awan, M.J.; Yasin, A.; Mohammed, M.A.; Damaševičius, R.; Maskeliūnas, R.; Abdulkareem, K.H. Real-Time Hand Gesture Recognition Based on Deep Learning YOLOv3 Model. *Appl. Sci.* **2021**, *11*, 4164. [CrossRef]
44. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.-C. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520.
45. Wu, J. Introduction to convolutional neural networks. *Natl. Key Lab. Nov. Softw. Technol. Nanjing Univ. China* **2017**, *5*, 495.
46. Albawi, S.; Bayat, O.; Al-Azawi, S.; Ucan, O.N. Social touch gesture recognition using convolutional neural network. *Comput. Intell. Neurosci.* **2018**, *2018*, 697310. [CrossRef] [PubMed]
47. Awan, M.; Rahim, M.; Salim, N.; Ismail, A.; Shabbir, H. Acceleration of knee MRI cancellous bone classification on google colaboratory using convolutional neural network. *Int. J. Adv. Trends Comput. Sci.* **2019**, *8*, 83–88. [CrossRef]
48. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.
49. Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M.; et al. Tensorflow: A system for large-scale machine learning. In Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI '16), Savannah, GA, USA, 2–4 November 2016; pp. 265–283.

50. Bengio, Y. Practical recommendations for gradient-based training of deep architectures. In *Neural Networks: Tricks of the Trade*; Springer: Berlin/Heidelberg, Germany, 2012.
51. Dhont, J.; Wolfs, C.; Verhaegen, F. Automatic COVID-19 diagnosis based on chest radiography and deep learning—Success story or dataset bias? *Med. Phys.* **2021**, *49*, 978–987. [CrossRef]
52. López-Cabrera, J.D.; Orozco-Morales, R.; Portal-Diaz, J.A.; Lovelle-Enríquez, O.; Pérez-Díaz, M. Current limitations to identify COVID-19 using artificial intelligence with chest X-ray imaging. *Health Technol.* **2021**, *11*, 411–424. [CrossRef]