Journal of Theoretical and Applied Information Technology
15th March 2022. Vol.100. No 5
2022 Little Lion Scientific

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

MULTI-CLASSIFICATION MODEL FOR COVID-19 PREDICTION USING IMBALANCED X-RAY DATASET BASED ON TRANSFER LEARNING AND CLASS WEIGHTING-SMOTE METHOD

OLA M. EL ZEIN[1*], MONA M. SOLIMAN [2], A. K. ELKHOLY [3], NEVEEN I. GHALI [4]

[1,3] Faculty of Science, Al-Azhar University (Girls branch), Cairo, Egypt.
[2] Faculty of Computers and Artificial Intelligence, Cairo University, Cairo, Egypt.
[4] Faculty of Computers and Information Technology, Future University in Egypt, Cairo, Egypt.
[1]olaelzin@azhar.edu.eg, [2]mona.solyman@fci-cu.edu.eg, [4]neveen.ghali@fue.edu.eg

ABSTRACT

This paper proposes a deep neural networks model to predict COVID-19 patients automatically based on chest X-ray images. The model is trained using imbalance dataset with a new hybrid balancing technique proposed to solve this problem. The Deep Convolutional Neural VGG-16 is trained and utilized to extract features from a given chest X-ray image after some preprocessing steps. To overcome the data imbalance issue, a new hybrid Class Weights-SMOTE is applied to the extracted feature vector and compared with traditional balancing techniques. The feature vector is then classified utilizing a Fine-tuning VGG-16. The model provides a multi-classification for the input x-ray images into COVID-19, Normal, and Pneumonia. Comparison with existing methods shows that the proposed model achieves a superior classification accuracy and outperforms all other models, providing 98% accurate prediction and improving the model's performance on minority-class samples to achieve high accuracy 100%. The findings of this study could be useful for diagnosing COVID-19 from chest X-ray images.

Keywords: *Coronavirus, Gaussian filter, Imbalanced data, Class Weight, SMOTE*

1. INTRODUCTION

Coronavirus disease, generally referred to as COVID-19, may be a quite respiratory organ infection illness. It is caused by a novel virus known as severe acute respiratory syndrome coronavirus two (SARS- CoV-2). Coronaviruses are a family of viruses that are well-known to cause diseases such as the common cold, severe acute respiratory syndrome (SARS), and Middle East respiratory syndrome (MERS) [1, 2].

The coronavirus disease was first discovered in the city, China, in December 2019 [3], to control the large spread of theCOVID-19 virus, it requires correct detection and treatment. Reverse transcriptase-polymerase chain reaction (RT-PCR) is the commonplace diagnostic examination for COVID-19. The high generality of PCR for diagnosis of COVID-19 is because of its high properties and sensitivity (i.e., over 90%) [4, 5], though there are limitations of the COVID-19 PCR technique as taking a long time, expensive, and limits the number of tools available due to the long production time [6]. Taking into consideration the serious rates of growth of COVID-19, a quicker and cheaper testing mechanism is needed to deal with this disease outbreak. Researchers have found that radiological screening such as computed tomography (CT) and X-rays have high performance in COVID-19 diagnosis and can be an efficient tool for large-scale examination [7].

Chest X-ray is the most efficient method for diagnosing COVID-19, which plays an animated role in clinical nursing and medical specialty analysis [8]. It is fast, cheap, and gives the patient a lower radiation dose compared to magnetic resonance imaging (MRI) and computed tomography (CT) [9]. However, making an accurate diagnosis from X-ray images requires expert knowledge and experience [10]. However, the Coronavirus disease pandemic is speedily raising the need for knowledge in this field. It has improved perception and focused attention on the need for automatic detection techniques based on intelligent methods to give

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

assistance to diagnosing patients at the appropriate time with acceptable accuracy [11].

Artificial intelligence (AI) and machine learning techniques would help experts to detect COVID-19 carefully and quickly. Over recent years, machine learning techniques have been rapidly improved and combined into computer-aided design systems (CAD) to supply careful and quick diagnosis. The salient success of AI brings more signs of progress in medical image analysis. The capability of an efficient AI model is strongly dependent on learning from a sufficient number of training samples [12].

Deep learning (DL) is one of the popular research fields in AI which allows the creation of an end-to-end model to obtain confident outcomes. This is done utilizing input images without any manual feature extraction and classification from images. DL makes a breakthrough in disease diagnosis by performing a difficult diagnosis for radiology experts using massive amounts of images [13].

Methods such as Convolutional Neural Networks (CNN) usually require huge amounts of data for training. Labeling this data by specialists is both expensive and time-consuming. Sometimes, the amount of data is not enough for the training process. The shortage of data is considered an obstacle to CNN in practical applications. Transfer learning (TL) seemsto solve this problem. The TL technique is utilized to supplya pre-trained structure in a knowledge base that can from the same or another space, taking advantage of the information obtained to resolve new issues more quickly and viably. The biggest advantage of utilizing the transfer learning method is that it allows the training of data with fewer datasets andrequires fewer calculation costs [14].

Classification as being a main task in DL is a predictive modeling issue that includes assigning a class label to each sample [15]. One of the main challenges in classification tasks is the class imbalance, where the contribution of the minority-class samples to the final network model is certainly much smaller than that of the majority-class samples, Imbalanced classifications lead to a challenge for predictive modeling as most of the machine learning algorithms utilized for classification were designed for the assumption of an equal number of samples for each class. This results in models that have poor predictive performance, especially for the minority class. This is an issue because usually, the minority class is more important, and therefore the issue is more sensitive to classification errors for the minority class than the majority class [16]. Different balancing techniques like resampling methods (such as RUS (Random Under Sampling), ROS (Random Over Sampling), and SMOTE (Synthetic Minority Oversampling Technique)) and class weighting are introduced to handle imbalanced datasets.

This paper introduces a new multi-classification model for chest X-ray images using an imbalanced dataset. The model proposes a new hybrid technique between class weighting and the SMOTE method to solve the imbalanced dataset problem. The model is able to classify chest X-ray images into normal, COVID-19, and pneumonia infection images. The model consists of four main phases: data preprocessing, feature extraction using fine-tuned VGG-16, solving imbalanced dataset problems using a hybrid technique, and finally the classification phase using a fully connected neural network.

A comparative comparison between the proposed hybrid technique using class-weighting and SMOTE method against different balancing techniques on the testing dataset is introduced to prove the efficiency of the proposed technique in solving the imbalance dataset problem. Results show that the model's performance on minority-class samples has a large development after applying a hybrid balancing technique. It improves the COVID-19's predicted accuracy from 95.7% to 100%.

The main contributions of the proposed model can be outlined as follows:

- It can provide an acceptable accuracy and a fast classifier with chest X-ray images into Covid-19, normal, and pneumonia.
- The GF technique is applied to remove the noise that exists in the image and provide an acceptable diagnosis of COVID-19 diseases.
- A new hybrid technique of imbalanced data is applied to balance the samples of COVID-19, normal, and pneumonia classes and improve the accuracy of the minority class to reach 99.1%.
- A comparative study between the proposed balancing hybrid technique and four traditional balancing techniques is introduced to evaluate the effect of the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

proposed technique on the classification performance.

- The proposed method has proved its superiority to state-of-the-art models for COVID-19 diagnosis in terms of overall accuracy, precision, recall, and F-score measurements.

The remainder of this paper is organized as follows: Related work approaches typically utilized for the diagnosis of COVID-19 chest X-ray are introduced in Section 2. In Section 3, the proposed method steps based on fine-tuned VGG-16 and a hybrid balanced technique are proposed. Section 4 shows the experimental results with a comparative analysis of them with state-of-the-art models. Finally, conclusions and future work are introduced in Section 5.

2. RELATED WORK

The classification and diagnosis of COVID-19 from chest X-ray images have attracted more attention from researchersto diagnose and predict this disease. In this section, state-of- the-art models with diagnosis methods based on deep learning architecture are discussed.

Hemdan et al., [17] proposed a framework, called COVIDX-Net that can aid radiologists in diagnosing COVID-19 patients utilizing X-rays. The framework is evaluated utilizing a dataset of 50 X-ray images split into two classes: 25 COVID-19-positive images and 25 COVID-19-negative images. The images utilized were resized to 224×224 pixels. The COVIDX-Net framework applies seven deep learning models: MobileNet, ResNet-v2, Inception-ResNet- v2, Xception, Inception-v3, DenseNet, and modified VGG-19.Their evaluation results show that the VGG-19 and DenseNet models achieved high performances with an F-score of 91% for COVID-19 cases.

In addition, Hassanien et al., [18] presented a classification method that utilizes multi-level thresholding and an SVM to detect and diagnose COVID-19 in lung X-ray images. Their method was tested on 40 contrast-enhanced lung X-ray images (15 healthy and 25 COVID-19-infected regions) with a resolution of 512× 512 pixels. Their classification method has given a sensitivity of 95.76%, a specificity of 99.7% and an accuracy of 97.48%.

Chaudhary et al., [19] proposed the EfficientNet CNN model for the detection of COVID-19 from chest X-ray images. This work utilized the open-source COVIDx dataset.It has approximately 14000 X-Ray images with three classes: normal, pneumonia, and COVID-19. This work achieved a test accuracy of equal to 95% and a sensitivity value of 100% for COVID-19.

Hasan et al., [20] proposed a deep learning-based model to predict COVID-19 patients utilizing pneumonia chest X-ray images. Some deep learning features such as AveragePooling2D, flatten, dense, Image Data Generator, and dropout are utilized to preprocess the data, and a CNN-based VGG16 model is utilized to classify chest X-ray images of COVID-19 patients and predict COVID-19 patients with pneumonia. The model predicted pneumonia with an average accuracy of 91.69%, a sensitivity of 95.92%, and a specificityof 100%.

All of the previous research use TL to provide either a binary or multi-classification for COVID-19 diseases utilizing chest X-ray images. The proposed studies (Hemdan et al., [17], Hassanien et al., [18]) considered a fairly large number of chest X-ray samples to validate the CNN models. Chaudhary et al., [19] and Hasan et al., [20] utilized larger datasets to validate their models, but these datasets of Hasan et al. [20] led to the class imbalance problem. The main limitation of the model [20] is that it was validated on a very low count of COVID-19 samples (i.e. 460). This model achieves a poor predictive performance for the COVID-19 class, while the Pneumonia class achieves a high predictive performance due to the imbalanced data. The proposed model provides a solution for a multi-classification problem of COVID-19 with the imbalanced dataset.

3. THE PROPOSED METHOD

This work proposes a new hybrid balancing method using class-weighting and SMOTE techniques for introducing a multi-classifier model of chest x-ray images. The model utilizes a fine-tune of the VGG-16 network to extract the main features of X-ray images. A set of preprocessing steps is proposed to enhance the input images fed into the model. The proposed method includes four main phases: image preprocessing using Gaussian smoothing filtering, resizing of images, normalizing of images. The second phase is feature extraction phase using the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

fine-tuning VGG-16 model, followed by a new hybrid balancing technique for handling the imbalanced training data problem, and finally a multi-classification phase using the CNN model. Figure 1 shows the architecture of the proposed method. More details about the proposed methods are shown in the following subsections.

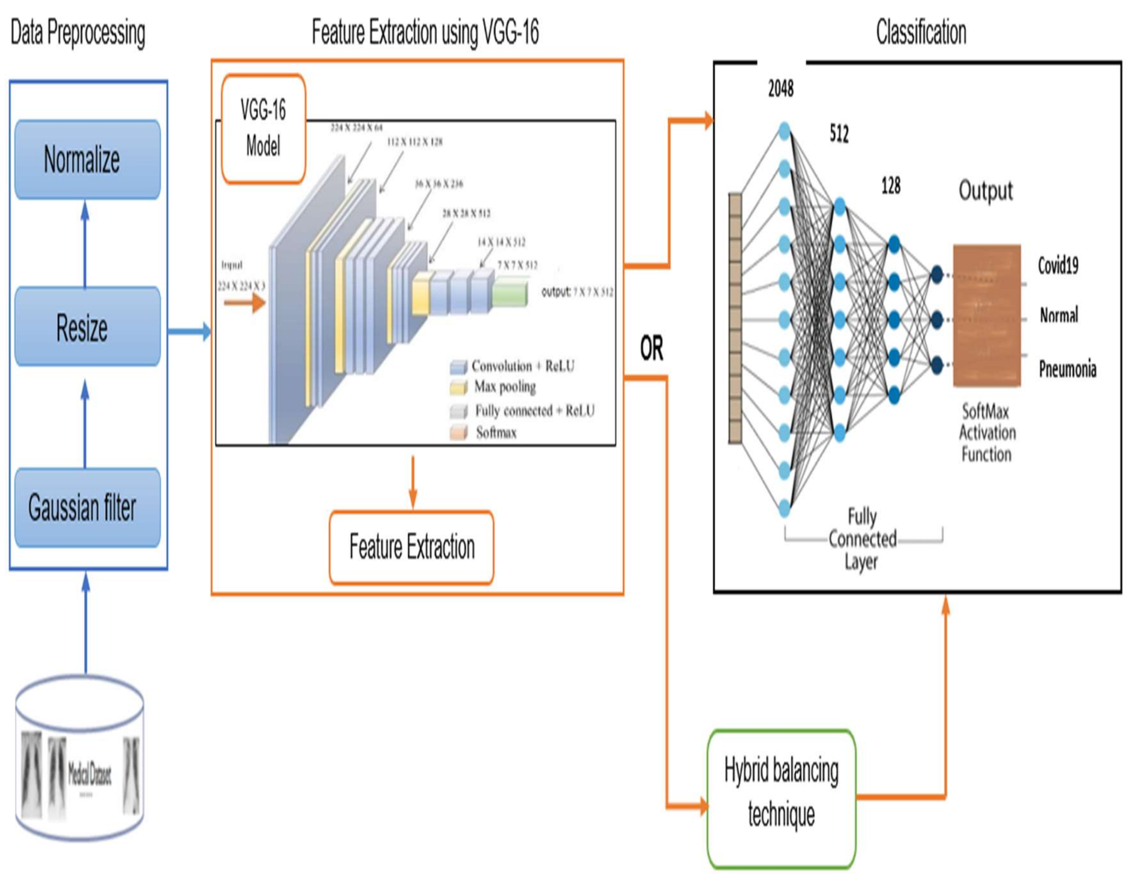


Fig. 1: Architecture Of The Proposed Model

3.1 Description of Data

The benchmark data utilized in this paper has 6432 chest X-ray images with three categories (COVID-19, Normal and Pneumonia). This dataset [21] is available for free from Kaggle. The images are also labelled as 0, 1 and 2 for COVID- 19, normal and pneumonia images, respectively. Images are collected from various publicly available resources [22, 23, and 24]. Figure 2 shows a few sample chest X-ray images from the dataset. Table 1 indicates the description of the utilized dataset. As shown in Table I, the dataset is split into training and testing sets. The proportion of data assigned to training for COVID-19, Normal, and Pneumonia is imbalanced. This issue will be solved by using a new hybrid balancing technique as shown in the following.

TABLE 1: Description of utilized dataset.

Category	Training set	Testing set
COVID-19	460	116
Normal	1266	317
Pneumonia	3418	855
Total	5144	1288
Percentage	80%	20%

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

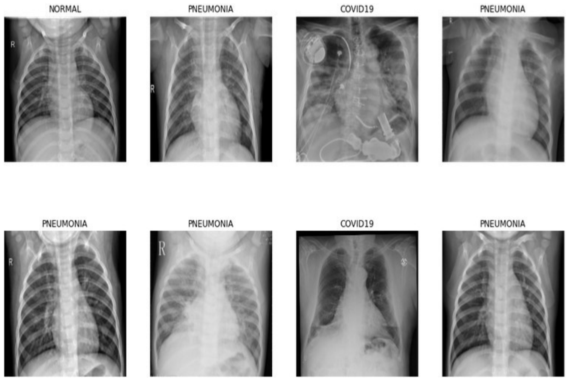


Fig. 2: Samples of Chest X-ray Images

3.2 Data Preprocessing

This section gives a detailed description of the methods utilized in the preprocessing step.

Gaussian filter-preprocessing: The Gaussian filter is a kind of image preprocessing that applies a filter on an imageto remove noise and smooth regions of images [25]. This filter takes the surrounding pixels (the number of which is specified by the size of the filter) and returns a single number computed with a weighted average based on the normal distribution. Gaussian filtering is highly effective in removing noise from the image. In this paper, the size of the filter that would be utilized to reduce image noise is the value inside 3×3. The two-dimensional Gaussian filter is computed as follows:

$$G(x, y) = \frac{1}{2\pi\, \sigma^2} e^{\frac{-x^2+y^2}{2\sigma^2}} \tag{1}$$

Where (x, y) indicates the Cartesian coordinates of the image that displays the dimensions of the window, and σ is the

3.3 Feature Extraction Using VGG-16

CNN is a DL architecture that can determine and classify features in images. It has a multi-layer neural network designed to analyze specific inputs and do specific tasks such as image classification, segmentation, and object detection. Also, DL healthcare applications such as medical images can utilize CNNs. The basic idea of CNN is to get features from the input (usually an image) at higher layers and to integrate them into more complex features at lower layers. Because of its multilayered architectural design, although it is computationally excessive, training such networks on a huge database takes several days.

standard deviation of a Gaussian function [26]. Figure 3 shows a few sample chest X-ray images after applyinga Gaussian filter.

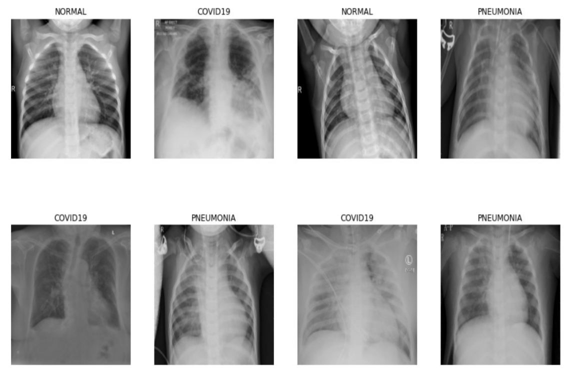


Fig. 3: The Impact Of Using GF On Chest X-Ray Dataset.

Resizing: All the input X-ray images are of different sizes and shapes, which increases the difficulty of efficient classification. In order to efficiently perform classification tasks, image preprocessing is performed. In the proposed method, all input images are resized to 224×224×3 pixelsto be suitable for the VGG-16 model.

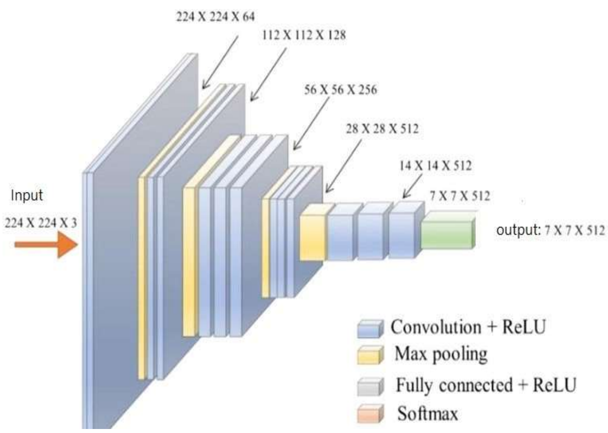
Normalization: Normalization of data is an important step and is generally utilized to keep numerical stability in the CNN architectures. With normalization, the CNN model is learned faster and more stable for the gradient descent. Therefore, in this research, the pixel values of the input images after applying a Gaussian smoothing filter have been normalized between the ranges of 0–1. The images utilized in the datasets are gray-scale images and the rescaling was achieved by multiplying each pixel value by 1/255.
So, such a CNN has two Major parts: A convolution layer that divides the various features of the image for analysis, and a fully connected layer that takes the output of the convolution layer and predicts the best classification for the image [27].

The most famous model that can be used for transfer learning is the VGG-16. The VGG network is proposed in 2014 by Karen Simonyan and Andrew Zisserman. VGG-16 is a CNN architecture that consists of 13 convolutions, rectification, pooling, and 3 fully connected layers [26]. The first part of the convolutional layer utilize 64 convolution kernels. The second part of the convolutional layer utilize 128 convolution kernels,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

the third part of the convolutional layer utilize 256 convolution kernels, and the last two parts of the convolutional layer utilize 512 convolutional kernels. The convolution network utilizes a 3×3 window size filter and a 2×2 pooling network.

In this work, the trained weights of the VGG-16 network are loaded from the Keras library and deleted the classification output layers. After applying GF and normalization, images of size 224×224×3 are fed into the pre-trained VGG-16 model, which then extracts features automatically. These features, 7×7×512 may represent the colour and the shape descriptors like circularity, roundness, etc. The output features of the VGG-16 model are reshaped into a vector of 25088. The architecture of VGG-16 without the fully connected layers is shown in Figure 4.



3.4 Hybrid Balancing Technique Class Weight-SMOTE

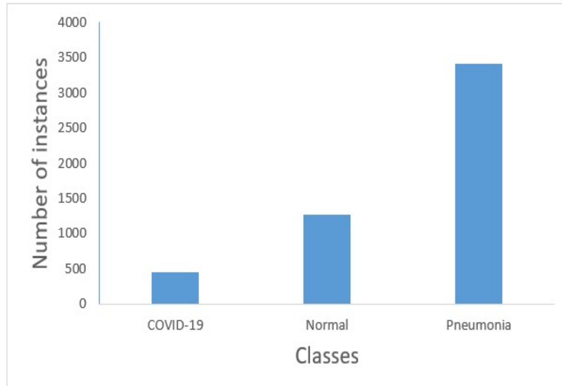
Fig. 4: VGG-16 Architecture

Different techniques such as resampling methods and class weighting methods are utilized here to handle imbalanced dataset issues.

The first technique is resampling methods, which consist of two categories: under sampling methods and oversampling methods. Under sampling methods such as the RUS method decrease the number of majority classes by removing random images from the majority class in the training dataset until the classes become balanced. Therefore, the training process can be easily implemented, and it enhances the issues associated with run time and storage [29]. Oversampling methods such as ROS and SMOTE methods increase the number of the minority class in the training dataset. ROS increases the number of minority classes by repeating the randomly selected set of instances from the minority class so that the majority class does not have over existence during the training process [30]. ROS can increase the likelihood of occurring overfitting, since it makes exact copies of the minority class examples. SMOTE creates new artificial samples based on the feature space similarities between existing minority classes by introducing non-repeated minority classes. The introduction of the new examples effectively serves to alter the bias of the learner, forcing a lot of general bias, mainly for the minority class. The new minority samples are created out of existing minority class imbalances utilizing the k-NN algorithm. The neighbors from the k-NN are randomly chosen based on the number of over-sampling that is needed [31, 32].

The second technique is class weights, which gives a weight for each class, which focuses more

An imbalanced dataset usually refers to an issue with classification issues where the number of instances in each class is not represented equally. Due to the differences betweeneach class, the algorithm tends to be biased towards most of the existing values. It is not difficult to get good



accuracy on these issues, but it does not mean that the model is good. The dataset [21] utilized here is imbalanced since the number of samples for COVID-19, Normal, and Pneumonia classes are equal to 460, 1266, and 3418 respectively. Figure 5 shows the distribution of the training dataset for each class.

Fig. 5: The Classes' Distribution Of Training Dataset.

As shown in Figure 5, the training dataset [21] utilized here includes more Pneumonia images than Normal and COVID-19 images, which may cause the model to be biased towards Pneumonia image detection than COVID-19 and Normal.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

attention on the minority classes such that the final result is a classifier that can learn equally from all classes. This is done to limit the size of the weight update in favor of the majority class. Small class weights and large class weights are assigned to the majority class and minority class, respectively. The balanced class weight for each class is calculated as follows [33]:

$$w_j = \frac{s}{n * x_j} \quad (2)$$

Where w_j is the weight for each class j, s is the total number of training dataset, n the number of classes, and x_j is the total number of each class j.

In this paper, a new hybrid balancing technique, Class Weight-SMOTE, is proposed. It suggests first using class weights to update the weights of the classes and increase the weights of the minority class using the previous equation (2) of balanced class weights. High weights are applied to the COVID-19 class, whereas low weights are applied to the other two classes. There are very few examples of the COVID-19 class, so the weight update is needed. Then, the SMOTE technique is utilized to generate a new synthetic examples along the line between the minority example and it's selected nearest neighbors to oversample the minority class to balance the class distribution. Thus, the overfitting problem is avoided and causes the decision boundaries for the minority class to spread further into the majority class space.
The difference in weights will influence the classification of the classes during the training phase. The main purpose is to punish the misclassification made by the minority class by setting a higher class weight and at the same time reducing weight for the majority class by utilizing a class weights technique. So that the network pays more attention to them when training. At the same time, avoid the overfitting by applying SMOTE to generate new artificial samples for the minority class as required. The addition of these synthetically generated minority class samples makes the class distributions a lot more balanced.
The combination of SMOTE and Class Weights have a positive effect on the minority class accuracy and is able to improve the performance of the model than plain SMOTE and Class Weights.
Algorithm 1 shows the hybrid Class Weight-SMOTE algorithm.

Algorithm 1 Hybrid Class Weight-SMOTE algorithm.

Input: Imbalanced train chest X-ray Dataset.
Method:
Compute the weight for each class.
Select Minority class sample, apply k-Nearest Neighbor.Compute the difference and gap.
Compute the artificial data in feature space.
Output: obtain the balanced dataset.

Different balancing techniques such as RUS, ROS, SMOTE, class weight are applied in this paper. After applying RUS, there will be 460 COVID-19 images, 460 Normal images, and 460 Pneumonia images, whereas applying ROS and SMOTE techniques, there will be 3418 COVID-19 images, 3418 Normal images, and 3418 Pneumonia images. Figure 6 shows the class distribution of the training dataset after applying RUS, ROS, and SMOTE techniques. When applying the class weight, the values of the updated weights of the COVID-19, Normal, and Pneumonia classes are 3.73, 1.35, and 0.5 respectively. One disadvantage of RUS approach is that some useful information might be lost from the majority class due to the under sampling. The disadvantage of the ROS technique is that it causes overfitting due to the repeating of instances which does not add any actual data to the training set. Experimentally, SMOTE has been shown to perform well against random over sampling due to synthetic minority class instances being added to the training set by creating a new dataset.

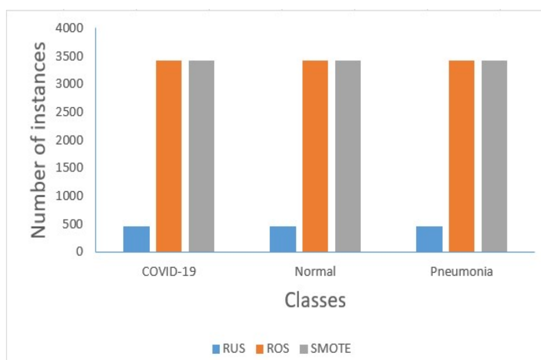


Fig. 6: The classes' distribution of training dataset after applying RUS, ROS, and SMOTE techniques.

After applying a new hybrid balancing technique, the values of the updated weights of the COVID-19, Normal, and Pneumonia classes are 3.73, 1.35,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

and 0.5 respectively, while the number of instances in the COVID-19, Normal, and Pneumonia classes is 3418, 3418, and 3418 respectively.
Figure 7 shows the class distribution of the training dataset after applying a hybrid balancing technique.

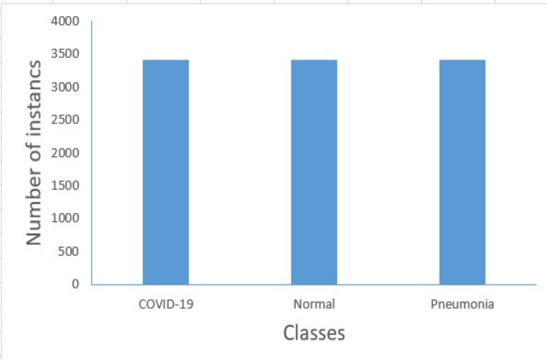


Fig. 7: The classes' distribution of training dataset after applying Hybrid balancing technique.

3.5 Classification

After the features were extracted automatically using pre-trained VGG-16 and a balanced training dataset, the classifier was used for the prediction and classification of chest X-ray images.
The fine-tuned VGG-16 model is used for the prediction and classification of COVID-19 images. During fine-tuning, the following modifications are performed on the model for retraining the model using the dataset [19]. The classification output layers of the VGG-16 network are removed and fine-tuned for VGG-16 by retraining only the last three layers of the VGG-16, while freezing the rest of the layers of the model to update its weights with the used dataset [19]. Then, the redesigned new classifier part of the model uses 2048, 512, and 128 neurons in its hidden dense layers. The output layer of the model has 3 neurons to classify images into three classes (COVID-19, Normal, and Pneumonia). Each dense layer (except the output layer) of the classifier part is followed by a Dropout with a rate of 0.4. The Dropout layer works by randomly dropping neurons during training. In the forward pass, the activations of the "dropped neurons" are neglected and have no contributions to weight update in the backward pass where the Dropout layer gave better results. Dropout reduces the capability of the model while training and guides the model during training against overfitting. The dense layer utilizes Rectified Linear Units (ReLU) activation function while the output layer utilizes softmax activation function for multi-classification. The mathematical computation of the softmax activation function is as follows [34]:

$$\text{Softmax}\,(x_i) = \frac{e^{x_i}}{\sum_{j=0}^{n} e^{x_i}} \quad (3)$$

Where x_i denotes the input vector and n represents the number of classes.

For training the model, a categorical cross-entropy loss function is utilized. The categorical cross-entropy function measures the performance of a classification model whose output (class score) is a probability value between 0 and 1. Categorical cross-entropy is calculated as [34]:

$$L(y, p) = -\sum_{j=0}^{n} y_j \log(p_j) \quad (4)$$

Where y_j is the actual value and p_j is the predicted value.

In the training step, because of the VGG-16 was pre-trained, all the above layers of CNN architecture except the last three layers freeze to update their weights with the utilized dataset. The Optimizer, learning rate, and batch size are the most crucial hyper-parameters for tuning the DL models. Hence, the training model was performed using the Adam optimizer with categorical cross-entropy as the loss function. The learning rate of the model is equal to 0.0001. The model was trained for 50 epochs, and the trained and validated batch size was set as 32.

4. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, the results obtained from several experiments are introduced. Overall experimental analysis for COVID-19 prediction from chest X-ray images utilizing a pre-trained VGG-16 model and a hybrid balancing technique is presented. A comparative analysis of fine-tuning VGG-16 plus different balancing techniques and fine tuning VGG-16 without different balancing methods is carried out. The results obtained from the proposed method are compared with recent state-of-the-art approaches. Finally, the most effective performing model is obtained.

The proposed model was trained on a colaboratory (Colab) where a Google search project is created to supply everybody with free GPU resources for their deep learning projects

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

and research. Each user has presently specified 12GB of RAM, and it will be up to 25GB. Google Colab gives a single 12GB NVIDIA Tesla K80 GPU, and it can be utilized constantly for up to 12 hours.

4.1 Evaluation Metrics of Model Performance

Deep transfer model performance is assessed using various evaluation parameters such as test accuracy, precision measure,recall measure, and F measure [35, 36].

The precision determines the count of models predicted correctly from all positive classes. It is calculated as shown as follows:

$$Precision = \frac{TP}{(TP+FP)} \tag{5}$$

A recall is the percentage of actual positive samples that are correctly predicted. It can be calculated as follows:

$$Recall = \frac{TP}{(TP+FN)} \tag{6}$$

The F score is a measurement of how close precision andrecall are, and it is calculated as shown as follows:

$$F\text{-}score = 2*\frac{(Precision * Recall)}{(Precision + Recall)} \tag{7}$$

Finally, the accuracy in which is a measure of the model's performance, and is computed by dividing the number of correct images by the total number of test images. It can be calculated as shown as follows:

$$Accuracy = \frac{(TN+TP)}{(TP+TN+FP+FN)} \tag{8}$$

Where TP is the number of True Positive samples (i.e the model correctly predicts the positive category), TN is the number of True Negative samples (i.e the model correctly predicts the negative category), FP is the number of False Positive samples (i.e the model incorrectly predicts the positive category), and FN is the number of False Negative samples from a confusion matrix (i.e the model incorrectly predicts the negative category the negative result is false).

4.2 Testing Accuracy and Confusion Matrix

Testing accuracy is an evaluation that determines the accuracy and performance of the proposed model.

The confusion matrix is a tabular method of visualizing the performance of the prediction model. Each entry in a confusion matrix refers to the count of predictions created by the model where it can be classified into the classes correctly or incorrectly. The confusion matrix and analysis of the matrix are provided in Figure 8 and Table 2 to test the performance of the proposed model.

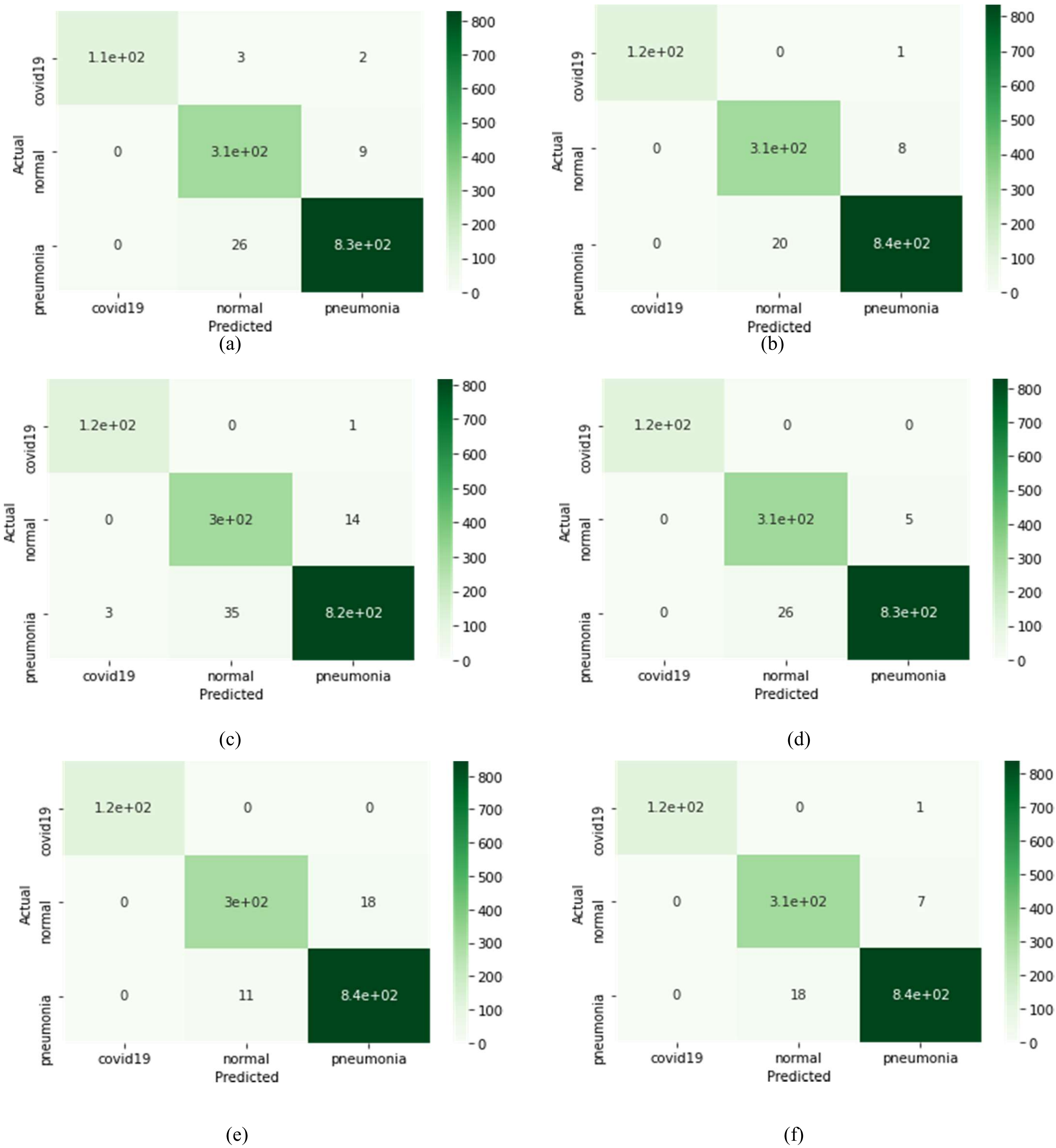


Fig. 8: Confusion matrix for (a) CNN without balancing technique, (b) CNN-Class Weights, (c) CNN-RUS, (d) CNN-ROS, (e) CNN-SMOTE, (f) Proposed model

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

TABLE 2: Performance matrices for the proposed models.

Models	Classes	Precision	Recall	F-score	Accuracy
CNN	COVID-19	1.00	0.96	0.98	0.957
	Normal	0.91	0.97	0.94	0.972
	Pneumonia	0.99	0.97	0.98	0.97
CNN-Class Weights	COVID-19	1.00	0.99	1.00	0.991
	Normal	0.94	0.97	0.96	0.975
	Pneumonia	0.99	0.98	0.98	0.977
CNN-RUS	COVID-19	0.97	0.99	0.98	0.991
	Normal	0.90	0.96	0.93	0.956
	Pneumonia	0.98	0.96	0.97	0.959
CNN-ROS	COVID-19	1.00	1.00	1.00	1.00
	Normal	0.92	0.98	0.95	0.984
	Pneumonia	0.99	0.97	0.98	0.969
CNN-SMOTE	COVID-19	1.00	1.00	1.00	1.00
	Normal	0.95	0.94	0.95	0.943
	Pneumonia	0.98	0.99	0.98	0.987
Proposed model	COVID-19	1.00	0.991	1.00	0.991
	Normal	0.95	0.98	0.96	0.978
	Pneumonia	0.99	0.98	0.98	0.980

Six confusion matrices for different balancing techniques, the hybrid proposed model CNN-Class Weight-SMOT and the VGG-16 based model without balancing technique for COVID-19 disease multi-classifications are shown in figure 8. Table 2 displays a detailed analysis of the proposed model compared with the traditional balancing techniques and CNN VGG16-based model without the balancing technique. The comparative study is based on precision, recall, F-score, and accuracy. Table 2 shows that for the COVID-19 instances, the CNN-based hybrid Class Weight-SMOTE obtained 100% precision, 99% recall, and a 100 percent F-score. The recall value (99%) indicates that the total number of false negatives is very low, whilst the precision value (100%) indicates that the total number of real negatives is very high. For the COVID-19 cases, the CNN-ROS and CNN-SMOTE models achieved 100% precision, 100% recall, and a 100% F-score. The precision value (100 %) means that the total of the true negatives is very high and there is no false positive value, but the recall value (100 %) suggests that there are no false-negative samples. When compared to other proposed strategies, the CNN-ROS and CNN-SMOTE models achieved the highest performance for COVID-19, reaching 100%, but CNN-ROS causes overfitting because the repeating of samples to balance the minority class samples.

The normal classification recorded 95% precision, 98% recall, and a 96% F-score for the hybrid model. While the pneumonia cases obtained 99% precision, 98% recall, and a 98% F-score, this shows that the hybrid model achieved high performance for normal and pneumonia classes compared with the other models in table 2. CNN-ROS and CNN-SMOTE models achieve the highest precision, recall, and F-score for COVID-19 cases, while they achieve low values for the precision and F-score of normal cases. As shown in table 2, CNN-ROS and CNN-SMOTE models achieve high accuracy of 100% for COIVD-19 compared with other proposed models, but the hybrid model achieves high accuracy for normal and pneumonia classes, with values of 97.8% and 98%, respectively, compared with the rest of the models in table 2. The main objective of this paper is to achieve a good results in detecting COVID-19 cases while not detecting false COVID-19 cases and making the classes more balanced. The empirical results in Table 2 reveal that the CNN-based hybrid Class Weight-SMOTE technique makes the distribution of classes more balanced than other proposed techniques and avoids the overfitting problem in classification. Also, the precision, recall, and F-score for the CNN-based hybrid Class-Weights-SMOTE technique have high or equal values compared with other proposed techniques for all classes. The proposed model is able to improve the predicted accuracy of COVID-19 from 95.7% to 100%. From the obtained results, it can be seen that the proposed model can be utilized as an alternate to RT-PCR, which takes around 4–10 hours to predict COVID-19 patients. We can utilize RT-PCR testing in those few cases where the proposed hybrid model is unsure about reducing the chances of errors.

4.3 Comparative Analysis

The above discussion proved that the proposed method CNN-Class Weights-SMOTE

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

represents so well in diagnosing COVID-19 from Chest X-rays. This section shows the comparison of the proposed model's performance of CNN- Class Weight-SMOTE with an existing method [20] that has the same dataset.

Table 3 and Figure 9 show the comparison results of testing accuracy between the proposed models based on a new hybrid Class Weight-SMOTE technique and traditional balancing techniques with an existing method [20]. As shown in Table 3 and Figure 9, the proposed CNN-Class Weights-SMOTE, and CNN-ROS with 97.8%, 97.6%, respectively. Contrariwise, CNN without balancing technique, CNN-Gaussian filter (GF), CNN-RUS, and the existing method [20] are given the lowest scores compared with a hybrid model, since these techniques help to obtain 96.9%, 97%, 95.9%, and 91.69% of accuracy, respectively. CNN-GF outperforms CNN, which is trained on imbalanced data because GF contributes in improving the model's performance by reducing noise from images. On the other hand, the CNN-RUS model has the lowest accuracy since some essential information is lost when some samples from the majority class are removed, which could be significant for the induction process. Table 3 shows that compared with the performance of the model trained on the imbalanced dataset before, the model accuracy's has increased by approximately 6% and performs better than the other existing model for detection and classification of COVID-19 from chest X-ray images. Because after balancing the training dataset, the model learns data features without bias towards majority-class samples, it can better learn the data characteristics of minority-class samples. Since both types of class weights and the SMOTE balancing techniques are effective when used in isolation, it becomes more effective when both types of methods are used together.

TABLE 3: Comparison Between Proposed CNN Model's Accuracy And The Existing Models

Models	Accuracy
CNN	96.8%
CNN-GF	97%
CNN-Class Weights	97.8%
CNN-RUS	95.9%
CNN-ROS	97.6%
CNN-SMOTE	97.8%
Hasan et al [18]	91.69
Proposed model	98%

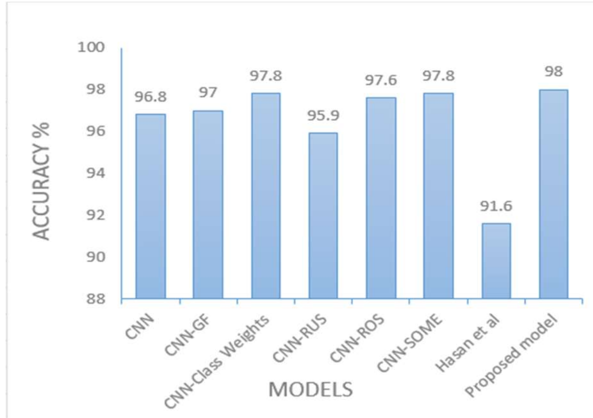


Fig. 9: Comparison Between Proposed CNN Model's AccuracyAnd The Existing Models.

Table 4 shows the comparison of the proposed model's CNN-Class Weights-SMOTE performance matrices Precision, Recall, F-score with the existing method that has the same data. The performance matrices of the proposed model give the highest value compared to the [20] model. As can be seen in Table 4, the precision of the COVID-19 prediction for the Hasan el al [20] model is 99%, recall is 81%, and F- score is 89%, while the precision for normal case detection is 83%, recall is 91%, F1-score is 87%, and the precision for pneumonia prediction is 95%, recall is 93%, and F-score is 94%. The precision of COVID-19 prediction for the proposed model is 100%, recall is 99%, and F-score is 100%. Whilethe precision of normal case detection is 95%, recall is 98%, F-score is 96%, and the precision of pneumonia prediction is 99%, recall is 98%, and F-score is 98%. Therefore, it is concluded that utilizing the Hasan el al [20], there are higher chances of being wrong in diagnosing COVID-19 predictions compared with normal and pneumonia cases, whereas utilizing the proposed model creates a balance between classes and there are fewer chances of being wrong in diagnosing COVID- 19 predictions compared with normal and pneumonia cases. Figure 10 and Figure 11 show a comparison of recall and F-score measures between Hasan el al [20] and the proposed model. Figure 10 shows that recall at 99% percent for the COVID-19 samples is the best because only one sample of COVID-19 was missed. This means that our model is really good correctly predicted all the COVID-19 patients in the test set and is able to improve the recall further by 18%. F-score is the go to metric when it comes to class imbalance problems and depends heavily on how imbalanced our training dataset. As shown in Figure 11, F-score achieves 100% percent

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

for the COVID-19 samples which means our training dataset is more balanced and our model achieved a 11% F-score improvement (from 0.89 to 1.00). Figure 12 shows an example of correctly predicted model images. As shown in Figure 12, the proposed model predicts three different classes of chest X-ray images, where the predicted class and target class are the same. As a result, it can be told that the proposed model could successfully identify each class due to the obtained results of three classes achieving high results as shown in figure 10 and figure 11.

TABLE 4: Comparison Between Proposed CNN Model and the existing models

Models	Classes	Precision	Recall	F-score
Hasan et al [18]	COVID-19 Normal Pneumonia	0.99 0.83 0.95	0.81 0.91 0.93	0.89 0.87 0.94
Proposed model	COVID-19 Normal Pneumonia	1.00 0.95 0.99	0.991 0.98 0.98	1.00 0.96 0.98

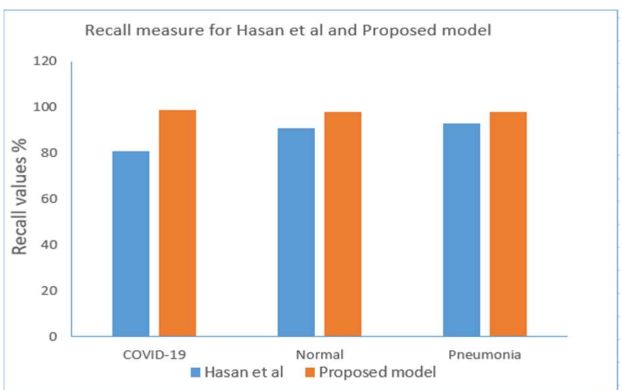


Fig. 10: Comparison Of Recall Measure Between Hasan Et Al [20] And The Proposed Model

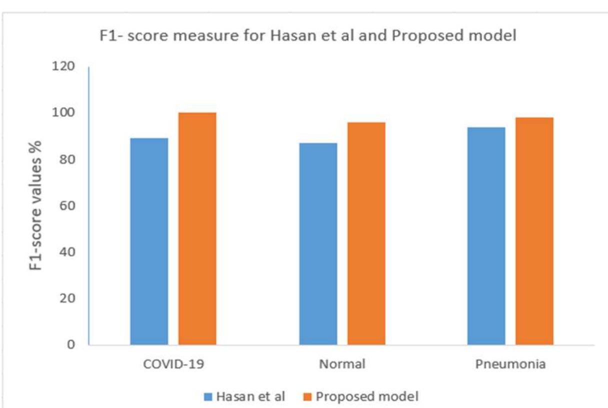


Fig. 11: Comparison Of F-Score Measure Between Hasan Et Al [20] And The Proposed Model

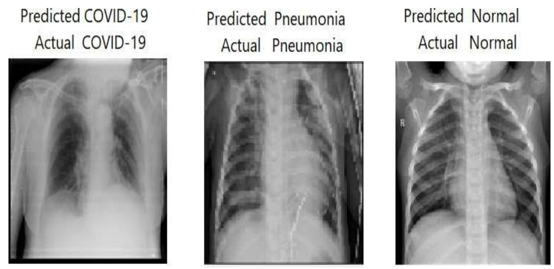


Fig. 12: The correctly predicted model's

5. CONCLUSIONS

Early prediction of COVID-19 patients is essential to prevent the spread of the disease to other people. In this work, an automatic CNN model based on chest X-ray images affected by a new hybrid Class Weights-SMOTE balancing techniques is proposed to predict COVID-19 patients. VGG-16 was trained on the imbalanced dataset and the balanced dataset after the hybrid balancing technique. The experiments have shown that the proposed model CNN-Class Weights-SMOTE is superior to the most state-of-the-art models in solving the COVID-19 prediction task, concerning different accuracy measures (e.g. accuracy, precision, recall, F-score), and has achieved the highest average accuracy of 98%. In conclusion, results proved that the proposed model can effectively enhance the neural network's ability to identify minority-class samples in image classification tasks of deep learning and achieved high accuracy for COVID-19 100%.

In future work, we plan to expand our research with other pre-trained CNNs to solve multi-classification problems. Expand this research to measure the grade of COVID-19 that helps the physician know the patient's condition. Also, Hyperparameter optimization is an active research point that required for automatic parameter selection rather than manual. We also expand this research to utilize more than pre-trained CNN architectures.

REFERENCES:

[1] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu, et al, "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China", The Lancet, Vol. 395, No. 10223, 2020, pp. 497-506.

[2] K. Elasnaoui, Y. Chawki, "Using X-ray images and deep learning for automated

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

detection of coronavirus disease", Journal of Biomolecular Structure and Dynamics, Vol. 39, No. 10, 2020, pp. 3615-3626.

[3] World Health Organization, "Laboratory testing for coronavirus disease 2019 (COVID-19) in suspected human cases: interim guidance, 2 March 2020", World Health Organization, World Health Organization2020.

[4] F. Shi, J. Wang, J. Shi, Z. Wu, Q. Wang, Z. Tang, K. He, Y. Shi, D. Shen, "Review of artificial intelligence techniques in imaging data acquisition", segmentation and diagnosis for COVID-19, IEEE Rev Biomed Eng, Vol. 14, , 2021 pp. 4-15. http://dx.doi.org/ .2020.2987975.

10.1109/RBME

[5] Y. Fang, H. Zhang, J. Xie, M. Lin, L. Ying, P. Pang and W. Ji, "Sensitivity of chest CT for COVID-19: comparison to RT-PCR", Radiology, Vol. 296, No. 2, 2020, pp. E115-E117.

[6] T. Ai, Z. Yang, H. Hou, C. Zhan, C. Chen, W. Lv, Q. Tao, Z. Sun and L. Xia, "Correlation of chest ct and rt-pcr testing for coronavirus disease 2019 (covid-19) in china: A report of 1014 cases", Radiology, Vol. 296, No. 2, 2020, pp. E32-E40.

[7] A. S. Albahri, J. R. Al-Obaidi, A. A. Zaidan, O. S. Albahri, Rula A. Hamid, B. B. Zaidan, A. H. Alamoodi and M. Hashim,"Multi- biological laboratory examination framework for the prioritization of patients with covid-19 based on integrated AHP and group VIKOR methods", International Journal of Information Technology and Decision Making, Vol. 19, No. 05, 2020, pp. 1247-1269.

[8] G. D. Rubin, C. J. Ryerson, L. B. Haramati, N. Sverzellati, J. P. Kanne, S. Raoof, N.W. Schluger, A. Volpi, J.-J. Yim, I. B. Martin et al.,"The role of chest imaging in patient management during the covid-19 pandemic: a multinational consensus statement from the fleischner society", Radiology, Vol. 296, No. 1, 2020, pp. 172–180.

[9] Z.Y. Zu, M.D. Jiang, P.P. Xu, W. Chen, Q.Q. Ni, G.M. Lu, L.J. Zhang, "Coronavirus disease 2019 (COVID-19): a perspective from China", Radiology, Vol. 296, No. 2, 2020, pp. E15-E25.

[10] O. M. El Zein, M. M. Soliman, A. K. Elkholy and N. I. Ghali, "Transfer Learning Based Model for Pneumonia Detection in Chest X-ray Images", International Journal of Intelligent Engineering and Systems, Vol. 14, No.5, 2021, pp. 56-66.

[11] Shankar K, Lakshmanaprabu S. K, A. Khanna, S. Tanwar, J.J. Ro-drigues, N.R. Roy, "Alzheimer detection using Group Grey Wolf Optimization based features with convolutional classifier", Computer and Electrical Engineering, Vol. 77, 2019, pp. 230-243.

[12] A. Hosny, C. Parmar, J. Quackenbush, L. H. Schwartz, and H. J. W. L. Aerts, "Artificial intelligence in radiology", Nature Reviews Cancer, Vol. 18, No. 8, 2018, pp. 500-510.

[13] K. E Asnaoui, Y. Chawki and A. Idri, "Automated Methods for Detection and Classification Pneumonia Based on X-Ray Images Using Deep Learning", Artificial Intelligence and Block chain for Future Cybersecurity Applications, Springer, Vol. 90, 2021, pp. 257-284.

[14] K. Weiss, T.M. Khoshgoftaar and D. Wang, "A survey of transfer learning", Journal of Big Data, Vol. 3, No. 9, 2016, pp. 1-40.

[15] G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, C. I. Sanchez, "A survey on deep learning in medical image analysis", Medical Image Analysis, Vol. 42, 2017, pp. 60-88.

[16] K. Teh, P. Armitage, S. Tesfaye, D. Selvarajah and I. D. Wilkinson, "Imbalanced learning: Improving classification of diabetic neuropathy from magnetic resonance imaging", PLoS ONE, Vol. 15, No. 12, 2020, pp. 1-15.

[17] E. E.-D. Hemdan, M.A. Shouman, and M. E. Karar, "COVIDX-Net: A Framework of Deep Learning Classifiers to Diagnose COVID-19 in X-Ray Images", ArXiv, Vol. 2003.11055, 2020, pp. 1-14.

[18] L.N. Mahdy, K.A. Ezzat, H.H. Elmousalami, H.A. Ella and A.E. Hassanien, "Automatic X-ray COVID-19 Lung Image Classification Sys- tem based on Multi-Level Thresholding and Support Vector Machine", medRxiv, 2020, pp. 1-8.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[19] Y. Chaudhary, M. Mehta, R. Sharma, D. Gupta, A. Khanna, and J. J. P. C. Rodrigues. “Efficient-CovidNet: Deep Learning Based COVID-19 Detection from Chest X-Ray Images”, 2020 IEEE International Conference on E-health Networking, Application and Services (HEALTH- COM), 2021, pp. 1-6.

[20] M. D. K. Hasan, S. Ahmed , Z. M. E. Abdullah , M. M. Khan , D. Anand, A. Singh , M. AlZain , and M. Masud, “Deep Learning Approaches for Detecting Pneumonia in COVID-19 Patients by Analyzing Chest X-Ray Images”, Hindawi, Mathematical Problems in Engineering Vol. 2021, 2021, pp. 1-8.

[21] https://www.kaggle.com/prashant268/chest-xray-covid19-pneumonia.

[22] https://github.com/ieee8023/covid-chestxray-dataset.

[23] https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia.

[24] https://github.com/agchung.

[25] A. Chikhalekar, “Analysis of image processing for digital x-ray”, International Research Journal of Engineering and Technology (IRJET), Vol. 3, No. 3, 2016, pp. 1364-1368.

[26] G. Deng and L. Cahill, “An Adaptive Gaussian filter for noise reduction and edge detection”, In 1993 IEEE Conference Record Nuclear Science Symposium and Medical Imaging Conference, Vol. 3, 1993, pp. 1615-1619.

[27] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, “A Survey of the Recent Architectures of Deep Convolutional Neural Networks”, Journal of Artificial Intelligence Review, Vol. 53, 2020, pp. 1-87.

[28] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition”, 3rd International Conference on Learning Representations, ICLR 2015, pp. 1-14.

[29] M. A. Tahir, S. Asghar, A. Manzoor and M. A. Noor, “A classification model for class imbalance dataset using genetic programming”, IEEE Access, Vol. 7, 2019, pp. 71013-71037.

[30] A. Liu, J. Ghosh and C. E. Martin, “Generative oversampling for mining imbalanced datasets”, In: DMIN, 2007, p. 66-72.

[31] V. Chawla, K.W. Bowyer, L.O. Hall and W.P. Kegelmeyer. “SMOTE: synthetic minority over-sampling technique”, Journal of artificial intelligence research, Vol. 16, 2002, pp. 321-357.

[32] C. Kumari, M. Abulaish and N. Subbarao, “Using SMOTE to deal with class-imbalance problem in bioactivity data to predict mTOR inhibitors”, SN Computer Science, Vol. 1, No. 3, 2020, p. 1–7.

[33] https://www.analyticsvidhya.com/blog/2020/10/improve-class-imbalance-class-weights/

[34] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, ” Deep Learning”, Cambridge, MIT press, Vol. 1, No. 2, 2016, pp. 1-800.

[35] J. Han, M. Kamber and J. Pei, “Data mining: concepts and techniques”, 2012.

[36] Y. Kim. “Convolutional Neural Networks for Sentence Classification”, Proceedings of the 2014 Conference on Empirical Methods in Natural Language, 2014, pp. 1746-1751.