Covid-19 CHEST X-RAY IMAGE CLASSIFICATION USING DEEP LEARNING

Team_1066_Covientist

gunther6, mb33, rk14, soumava2

ABSTRACT

Due to the rapid growth of Covid-19 cases around the world, it has become imminent to harness artificial intelligence in order to improve the existing medical diagnosis process related to this disease. Conflicting nature of Covid-19 and Pneumonia symptoms pose difficulty in identifying the actual presence of disease in Flu season. It was found in early studies that there are some abnormalities persisted in chest radiography images which presumably a crucial determining characteristic of the patients infected with Covid-19. Therefore, chest X-ray image classification has emerged as an alternative technique to aid medical diagnosis during pandemic time. However, it is cumbersome to manually detect Covid-19 from a set of images comprised of both Covid-19 and pneumonia cases thus deep learning technique has been proposed to build a state-of-the-art tool to enhance the current diagnosis process. We aim to implement a set of deep learning models on online available resources of 6432 images and further strengthen by utilizing data augmentation techniques to provide better generalization of the model during testing phase. Furthermore, we would select the model exhibiting best performance metrics during validation phase in order to make an optimistic outcome in detection of Covid-19 and distinguish them from normal/healthy and pneumonia affected chest X-ray images.

Keywords

Covid-19, Coronavirus detection, Deep learning, X-ray, Pneumonia, Classification.

1. INTRODUCTION

The Covid-19, a viral infection causes severe respiratory illness ranging from common cold to life threating disease like Severe Acute Respiratory Syndrome (SARS) and Middle East Respiratory Syndrome (MERS). According to the reports of WHO, common symptoms of Covid-19 are same as that of common flu, which include fever, tiredness, dry cough, and shortness of breath, aches, pains and sore throat. Due to the similar nature of flu symptoms, it is difficult to detect the virus at early stage and this cannot be treated by normal anti-biotics material.

The WHO approved method of testing corona virus is considered as the reverse transmission polymerase chain reaction (RT-PCR) method where the short sequences of DNA or RNA are analyzed and reproduced or amplified. However, there have been unforeseen challenges occurred while following this testing procedure:

1) It has been observed that negative results do not rule out the possibility of a person infected with Covid-19.

2) Limited availability of testing kits and screening workstation created roadblocks to carry out mass Covid-19 screening in every part of the world.

The above scenarios create massive implication to medical professionals and staffs to handle the exponential growth of cases with accurate testing measure and making subsequent medical decisions accordingly. Therefore, it has become imminent to come up with an alternative approach to support the existing medical diagnosis process. X-ray imaging is frequently used modality by medical practitioners to assert or to deny the possibility of any bacterial pneumonia infection. The same process can be applied to detect Covid-19 by finding out the anomaly between X-ray images of a Covid-19 infectant and a person suffering from bacteria\viral pneumonia. Easy availability of X-ray machines makes it purposeful for detecting Covid-19 cases in the absence of screening workbenches and kits. However, the biggest challenge is to manually examine each X-ray images and extract the findings lead to enormous time and presence of medical professionals. Thus, it is obvious that a computer-aided method driven by state-of-the-art deep learning models will improve the current Covid-19 diagnosis process with more accurate prediction and less time-consuming simulated tasks. There are numerous research efforts have already been conducted to understand the feasibility of implementing Covid-19 detection techniques by leveraging deep learning techniques pertain to image classification tasks.

Deep learning methods are proven to be useful in delivering high-quality results in conjunction with other advantages such as 1) maximum utilization of unstructured data, 2) elimination of additional cost, 3) reduction of feature engineering, and 4) elimination of explicit data labelling. Therefore, deep learning methods are often used to extract relevant features in order to classify objects using its autonomous nature.

Table 1 represents some of the previous research work done on Covid-19 using deep learning models. Similar to our project, some research works were oriented towards multi-class classification work as opposed to binary classification.

Table 1

Author(Year)	Data	Training model	Image classes	Accuracy

Apostolopoulos and Bessiana (2020)	Chest X-Ray	VGG19, Mobile Net	Covid-19, Pneumonia, Normal	97.8%
Ozturk (2020)	Chest X-Ray	DarkCovid Net	Covid+, Pneumonia, No-findings	87%
Ahuja (2020)	Chest X-Ray	ResNet 18	Covid-19 vs Covid-19-	99.40%
Govardhan Jain(2020)	Chest X-Ray	ResNet50 + ResNet 101	Covid-19 vs Covid-19-	98.93%

2. METHODOLOGY

The abstract pipeline of our image classification project depicts below:
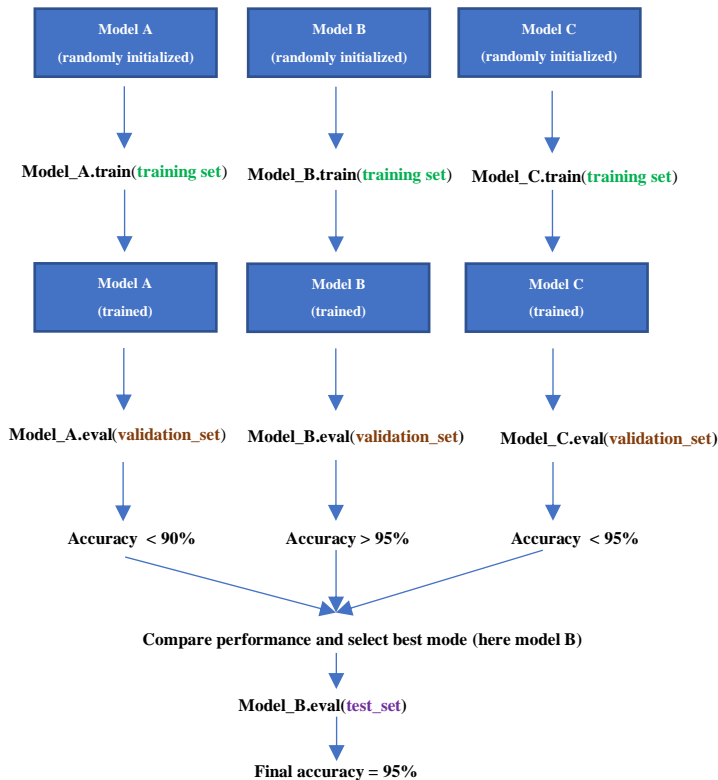
In a nutshell, following steps were performed throughout the exercise:

1) Radom model initialization (pre-trained and simple CNN model)

2) Trained each model with the training dataset separately.

3) Evaluate each trained model on validation set, adjusting hyperparameters to improve validation accuracy.

4) Finalize the model with best validation accuracy.

5) Evaluate the best model using test data and report final test accuracy.

2.1 Dataset and pre-processing

For this particular project, we used chest x-ray images extracted from Chest X-ray Kaggle dataset, a publicly available Covid-19 data repository that collected data from various public sources as well as through indirect collection from hospitals and physicians.

We applied further stratified resampling method to split the dataset into three different categories by introducing validation set which will be used to prevent model overfitting and accurately evaluate the model. Considering this approach, we decided to split the main dataset into 80% train data, 10% validation data and 10% test data. Here is a simple breakdown of the number of images segregated in the train, validation and test folders of the original dataset:

Classifier	Train	Validation	Test	Total
COVID-19	460	58	58	576
NORMAL	1266	158	159	1583
PNEUMONIA	3418	427	428	4273
TOTAL	5144	643	645	6432

Fig. 1 - 3 show images of all the three cases, that are considered for this project, such as normal\healthy, pneumonia and Covid-19.

All the original train, validation and test images were resized into standard size 224 x 224 in order to maintain uniformity in resolution. We firmly believed that down sampling input images to a smaller size would help as pre-trained model architectures expects input images of that size feed into the network. We also applied center crop to reduce background noises from the images.

Based on initial comprehensive analysis, we choose to apply the following image augmentation techniques in order to avoid data overfitting issue:

i) Rotation: Rotated images at various angles between -10º to +10º and added these augmented images in the train dataset.

ii) Gaussian Blur: Applied Gaussian filter of kernel size (3,3) on the training images in order to remove high frequency components.

iii) Flip: The trained images were randomly flipped horizontally and vertically to achieve data augmentation.

iv) RandomResizecrop: Images were cropped and resized randomly to overcome any unforeseen data overfitting issues.

However, none of the image augmentation techniques proven to be effective to achieve higher accuracy. Moreover, we experienced adverse impact of these techniques as the pre-trained models with augmented images incurred high loss during training phase. Presumably, image augmentation step failed to improve accuracy because unlike other images, X-ray image data does not require specific transformation

for better model training purpose. Rather it could have detrimental impact on training accuracy which we encounter during initial model training phase. Therefore, we decided to exclude image augmentation techniques outlined above from the data pre-processing step. We also tested image transformation into the gray format(one channel), but the best results were observed by keeping the RGB(three channel) format.

| Fig.1: Normal\Healthy | Fig.2: Pneumonia | Fig.3: COVID-19 |
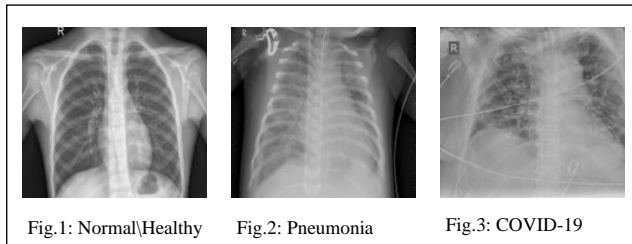
Note: There will be a plot added to the final version of this project paper in order to demonstrate how the training loss got improved after exclusion of certain image augmentation methods.

2.2 Model Selection

In recent years, the use of deep learning algorithms in general and convolutional neural networks (CNNs) led to many breakthroughs in a variety of computer vision applications like segmentation, recognition and object detection.

As per the proposed methodology, the CNN model selection process was twofold with more emphasis on selecting the final model with highest validation accuracy. Transfer learning coupled with development of our own CNN architecture was predominately the main approach that we followed during model selection phase.

Transfer Learning approach: Transfer learning method was selected as the first approach which demonstrates reusability of some pre-trained models specific to image classification task. We intended to pursue transfer learning for the following reasons: 1) The initial layers of the network are already trained, which are otherwise very hard to train due to the vanishing gradient problem. Eventually it saves train time and avoid huge computational power required for training the classification model. 2) Due to lack of Covid-19 X-ray images, transfer learning is the most imperative choice to train the model with comparatively low volume of training data. 3) Initial literature survey revealed that similar Covid-19 vs Pneumonia research work utilized pre-trained network to develop image recognition models that yielded higher accuracy. 4) Easy adaptability of pre-trained CNN architectures proven to be beneficial to fit training dataset comprised of standard size images.

As a result of an extensive research, following pre-trained models were selected for initial model building and accuracy check:1) **ResNet, 2) VGG, 3) Inception and 4) EfficientNet.**

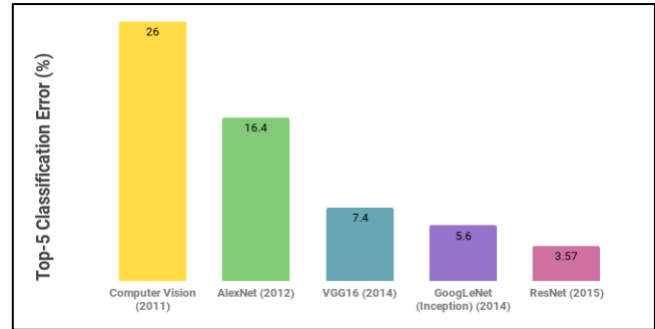Fig. 4: Source

Simple CNN approach: In subsequent stage, we developed our own convolutional neural network by testing more than 100 different architectures with different number of layers, number of neurons per layers, normalizations, pooling techniques and hyperparameters using techniques of grid and random research and adapting the model after each additional improvement of validation accuracy. The best validation accuracy achieved so far uses 3 convolutional layers followed by two fully connected linear layers. griand random research and adapting the model after each additional improvement of validation accuracy. The best validation accuracy achieved so far uses 3 convolutional layers followed by two fully connected linear layers. Fig. 5 manifests the CNN model architecture developed for this exercise. Rectified Linear Unit (ReLU) addressed the non-linearity whereas using batch normalization(BatchNorm2d) after each convolutional layer, and dropout between the connected linear layers were used to standardize the inputs to a neural network and stabilize the validation loss along the epochs, reducing the variance of validation accuracy. Batch normalization solves a major problem of internal covariate shift i.e., it standardizes the input images of different amplitude which helps neural network to understand the correlation between them.

To avoid overfitting and generalizations issues, we used the validation data to adjust our own model and calibrate the hyperparameters, and the test data to find the final accuracy of the model.

During forward propagation, batch normalization was implemented before ReLU activation layer so that data linearity can be achieved easily on normalized\centered data.

Brief overview of our CNN architecture:
The first convolutional layer self.conv1 has an input channel of dimension 3 since we are using RGB image of size 224 x 224. The kernel size is chosen to be of size 3x3 with stride of 1. Image padding of size 1 applied to keep uniformity between input and output image. The output dimension at this layer is **128 x 224 x 224.**
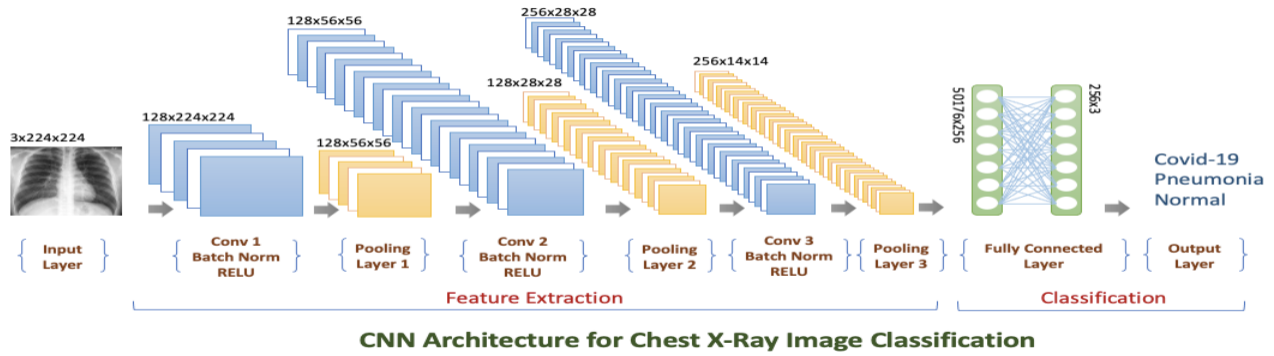
Fig. 5: Our CNN model architecture

Subsequently batch normalization function is applied to the network followed by ReLU activation function to introduce non-linearity. In order to reduce the number of training parameters and computation cost in accordance with controlling overfitting issue, a max pooling layer with kernel size 4x4 and stride 4 was introduced after RelU activation function. Max pooling layer down sample the output feature maps to **128 x 56 x 56.**

The second convolutional layer self.conv2 has an input channel of dimension 128 in order to keep it consistent with

the dimension of the feature map from previous layer. Like the previous convolutional layer, kernel size is chosen to be of size 3x3 with stride of 1 and image padding of size 1 applied to keep uniformity between input and output image. The output dimension at this layer is **128 x 56 x 56** so no such changes occurred in the transformation channel dimension. Like previous layers, both batch normalization and ReLU activation function are applied sequentially to stabilize the neural network once again. Then another max pooling layer with kernel size 2x2 and stride 2 is introduced to down sample the feature map led to increased efficiency in model training process. The down-sampled feature map dimension: **128 x 28 x 28.**

The third convolutional layer self.conv3 is introduced to upgrade the output channel before feeding into linear layers. Like previous convolutional layers, subsequent maxpool, ReLU activation and batch normalization are applied to the output images. The output dimension at this layer is **256 x 14 x 14.**

Finally, two connected linear layers self.linear1 and self.linear2 are used at the end and a dropout layer is added in between those two linear layers to reduce overfitting. We used recommended dropout value of p=0.5. A flattened version of the feature map is passed to the first fully connected layer. So, the input dimension size is **256 x 14 x 14 = 50176 nodes.** Then this layer is connected to the final linear layer of 128 nodes. The output dimension of final layer should match the total classes which is 3 for this project(Covid-19, Pneumonia and Normal). Hence, the final stage of

the CNN architecture comprised of two fully connected layers of size **50176 x 256** followed by **256 x 3.**

2.3 Analytical Infrastructure
We used open-source machine learning library PyTorch to build the predictive model for this image classification project. In order to achieve high accuracy and maximum throughput, we used both cloud platform Google Colab and machines with modern GPU configurations in order to harness better computing power to train and re-train the model.

2.4 Model Optimization
Neural network model can outperform others if the hyperparameters are optimized properly in order to get higher accuracy. As learning rate is a very imperative hyper-parameter while training the deep learning networks as it correlates with loss, we tried a few different rates in order to achieve the best optimal one. The model learning rate was in range between le-5 to le-2. In terms of hyperparameters, we observed that pre-trained models presented more stable validation accuracy when compared to our simple CNN model, which demanded more epochs with same learning rate in order to provide stable validation results. For some models we also tested the concept of dynamic learning rates using scheduler functionalities, starting with a value of 1e-2, and reducing it gradually until 1e-5 every time the validation loss didn't improve along the training epochs.

The optimizer played an integral part to minimize model error rate during training phase. We used both Adam and SGD as optimizer for model training purpose. SGD tend to perform well with learning rate le-2 as the model exhibited lower error rate\loss while training with these hyperparameters. Our initial model training with Adam optimizer took less time, but it did not produce higher accuracy as expected thus we selected SGD over Adam optimizer for our learning phase.

Different number epochs ranging from 5 to 25 were selected to identify the optimal number of iterations for pre-trained models which could produce best training result with lowest loss. However, we intended to use higher number of epochs

= 100 for training our CNN model which produced more stable and better accuracy so far.

2.5 Model Evaluation Metrics and Results

We obtained certain statistical measures such as true positive (TP), true negative (TN), false positive (FP), and false negative (FN) to validate accuracy of the pre-trained and our CNN models. **True positive (TP):** number of correctly identified disease X-ray images, **False Negative (FN):** number of incorrectly classified disease X-ray images, **True Negative (TN):** number of correctly identified healthy X-ray cases False Positive, **False Positive (FP):** incorrectly identified healthy X-ray cases.

Table 2 below depicts the preliminary evaluation metrics of different models acquired for this image classification task.

Model	Accuracy	Precision	Recall	F1 score
VGG16	0.92731	0.910803	0.928168	0.919067
VGG19	0.933022	0.930662	0.924649	0.927626
ResNet18	0.950338	0.953714	0.923171	0.937496
ResNet34	0.953488	0.956445	0.940649	0.948349
ResNet50	0.922481	0.929862	0.885737	0.906222
ResNet101	0.934884	0.943559	0.907175	0.924327
Inception	0.965732	0.968638	0.959973	0.963973
EfficientNet	0.939252	0.919735	0.961581	0.938566
Our model	0.967442	0.96544	0.957881	0.961060

Fig. 6 shows the confusion matrix representing three classes as predicted by our model.
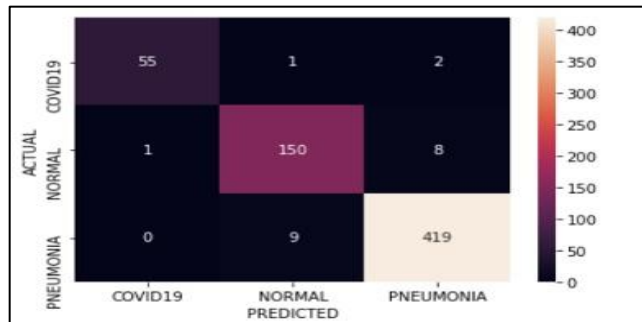

Fig. 6: Confusion matrix of our CNN model

2.6 Identifying the False Positives and False Negatives

Fig. 7 shows all 21 images classified incorrectly by the Simple CNN model from all 645 samples of test dataset. This analysis suggests that most of the misclassifications occur between pneumonia and normal images and were used to additional investigation of different image transformations.
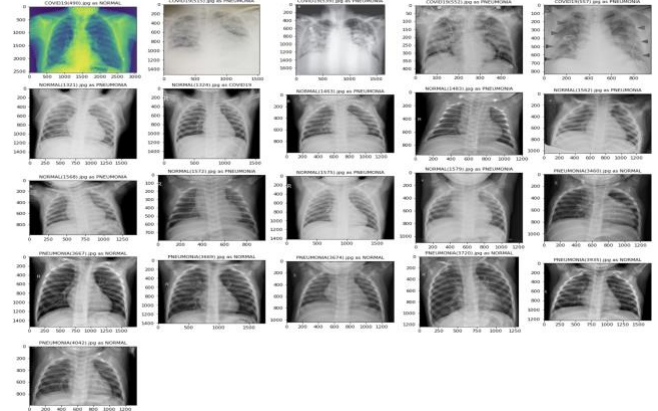

Fig. 7: False Positive and False Negative Images

3. Discussion and conclusion

The reasonable high accuracies obtained by several pre-trained methods along with our own CNN model suggest that deep learning networks are capable of extracting key features from the chest x-ray images. Eventually the efficient feature extraction process helps to distinguish the images accurately compared to other conventual machines learning methods. Initial literature survey revealed that, most of the previous research work pertain to binary classification (Covid vs Non-Covid) showed better accuracy results compared to multi-class chest x-ray image classification(See Table 1). We intended to develop a more robust multiclass CNN model that could outperform some of the pre-trained models specific to image classification task outlined in section 2.2 above. Our initial results seem to be promising in order to achieve our goal of using our own CNN architecture as final model for this project. However, we yet to complete full research on testing CNN architectures with different combination of layers and make necessary adjustments in hyperparameter selection as of this writing. Furthermore, given the limited test data sample for this project, we decided to incorporate k-fold cross validation technique to evaluate our models in our next validation step. This resampling procedure will have a single parameter called k that refers to the number of groups that our chest X-ray image dataset will be split into. For our model evaluation practice, we will choose k = 5 i.e., one unique fold will be chosen as a test set and the remaining folds will be selected as a training set. We will keep analyzing the images that were misclassified in the validation and test stage to understand how to improve accuracy of our model. Although 6K samples provided a good model accuracy so far, in the future, accuracy of these models could be improved with greater number of Covid-19 chest x-ray image samples used in the train stage. We aim to transform our final model into a publicly accessible web application platform which eventually would expedite the Covid-19 screening test, eliminate test errors incurred due to similar nature of Covid-19 and pneumonia affected chest x-ray images, and lower the overall diagnosis expenses by making the process autonomous and easy to operate by medical professionals.

4. REFERENCES

[1] Zoabi, Y., Deri-Rozov, S. & Shomron, N. Machine learning-based prediction of Covid-19 diagnosis based on symptoms. *NPJ Digit. Med.* **4,** 3 (2021).

[2] Jain, G., Mittal, D., Thakur D. & Mittal M. A deep learning approach to detect Covid-19 coronavirus with X-ray images. *Biocybern. Biomed. Eng.* **40,** 4 (2020).

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

[4] Ozturk, T., Talo, M., Yildirim, E.A., Baloglu, U.B., Yildirim, O. & Acharya, U.R. Automated detection of COVID-19 cases using deep neural networks with X-ray images. *Comput. Biol. Med*. **121** (2020).

[5] Majeed, T., Rashid, R., Ali D. & Asaad A. Covid-19 detection using CNN transfer learning from X-ray images. *MedRxiv* https://doi.org/10.1101/2020.05.12.20098954 (2020).

[6] Wang, L., Lin, Z.Q. & Wong, A. COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images. *Sci. Rep.* **10** (2020).

[7] Apostolopoulos, I.D. & Mpesiana, T.A. Covid-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks. *Phys. Eng. Sci. Med.* **43** (2020).

[8] 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.* **196** (2020).

[9] Oh, Y., Park, Sangjoon. & Ye, J.C. Deep learning COVID-19 features on CXR using limited training data Sets. *IEEE Trans. Med. Imaging*. **39**, 8 (2020).

[10] Chowdhury, M., Rahman, T., Khandakar, A., Mazhar, R., Kadir, M.A., Mahbub, Z.B., Islam, K.R., Khan, M.S., Iqbal, A., Al-Emadi, N., Reaz, M. & Islam, M.T. Can AI help in screening Viral and COVID-19 pneumonia? *IEEE Access*. **8** (2020).

[11] Pereira, R., Bertolini, D., Teixeira, L.O., Silla Jr, C.N. & Costa, Y.M. COVID-19 identification in chest X-ray images on flat and hierarchical classification scenarios. *Comput. Methods Programs Biomed.* **194** (2020).

[12] Ahuja, S., Abhimanyu, Reddy, Pasam, Vineet & Marques, Oge, Artificial intelligence and COVID-19: A multidisciplinary approach. *Integr Med Res.* **9,** 3(2020).