Turkish Online Journal of Qualitative Inquiry (TOJQI)
Volume 12, Issue 5, May 2021: 2192-2203

Early Screening Of Covid-19 Pneumonia Patients Using X-Rays By The Developed Artificial Intelligence System

M.S. Kavitha[a], K. Saranya[b], V. Priyadharshini[c], S. Maurya[d], R Preetha[e], M. Gayathri[f]

[a,b,c,d,e]Department of Electrical and Electronics Engineering, R. M. K. Engineering College, Chennai, Tamil Nadu, India
[f]E&Y LLP

Abstract

The global pandemic of Covid-19 has affected millions of lives . Any technical tool that allows for rapid screening of COVID-19 infection will help Healthcare practitioners to benefit greatly from precision of the results. The most commonly used clinical technique for the diagnosis of COVID-19 is the reverse transcription polymerase chain reaction (RT-PCR), which is expensive, less sensitive, and involves specialised medical staff. X- ray imaging is a simple tool and it's possible to use it as a good alternative for COVID-19 diagnosis. The research aims to prove that Artificial Intelligence (AI) could help detect COVID-19 from chest X-ray (CXR) images more quickly and accurately. The training and testing datasets were obtained from National Institute of Health(NIH) through kaggle.com. A Transfer learning technique was used with the help of image augmentation to train and verify many pretrained deep Convolutional Neural Network(CNN). The system is trained to classify the data into two different modules: i)Pneumonia affected and Normal ;(ii) Covid-19 Positive and Covid- 19 Negative with image augmentation. Furthermore the patients with non-pneumonia lungs who tested negative for Covid-19 will be given a second diagnosis based on their symptoms. Using Natural Language Processing(NLP) the probability of the patient suffering from Covid-19 can be predicted. For validation dataset random CXR images from the Veteran's Administration (VA) Picture archiving and communication (PAC) system were obtained .Our system's high accuracy can assist in increasing the speed and detection of COVID-19 diagnosis. This will be beneficial in this pandemic, where the risk of illness and preventive measures are at odds with available resources.

Keywords: Artificial Intelligence, Covid-19, Chest x- ray, Convolutional Neural Network, Deep Learning, Pneumonia

1. Introduction

The ongoing Covid-19 pandemic is caused by a new strain and has led to the death and hospitalization of millions globally. The Covid-19 infected people demonstrate moderate symptoms such as fever, breathlessness, sore throat, loss of taste and smell etc. The infected people may also show symptoms for severe respiratory ailments like pneumonia in both lungs. Covid-19 pneumonia is the most feared complications which is a life threatening disorder. The majority of Covid-19 pneumonia patients encounter abnormal chest x-ray findings; the most prominent of which are peripheral ground glass opacity (GGO) affecting lower lobes and linear opacity.

Reverse Transcription Polymerase Chain reaction uses respiratory specimens inorder to test and it appears to be the

most popular clinical screening technique for Covid-19 patients. Although RT-PCR is used as a reference method for diagnosing Covid-19 in patients, it is a time-consuming, manual, complex, and laborious procedure.. A delay in disease prevention is caused due to significant shortage in the supply of RT-PCR. This allowed to conduct a researchbased on AI for Covid-19 pneumonia diagnosis with high level of accuracy.

Artificial intelligence is a field that can identify meaningful associations in data sets and is widely used in healthcare settings to detect, cure and forecast outcomes in medical images such as MRIs, X-rays, and CT scans. AI-driven software can be programmed to recognise Covid-19 and pneumonia indications. Deep learning has risen to be the core technology of AI, and it has been discovered to diagnose lung disease accurately. Effective screening and immediate medical response are required to stop Covid-19 from spreading.

This paper investigates the potential for AI to diagnose patients with COVID-19 pneumonia using CXR's. This can be executed more accurately by developing a model to classify patients' CXR images through two levels and predicting the possibility of COVID-19 infection by analysing the symptoms.

2. Methodology

A. Image Acquisition:

Total image datasets include 1200 Covid-19 CXRs ,1345 pneumonia CXRs and 1341 Normal CXRs. As the clinical diagnosis of CXR images can be quite challenging, hence moved to deep learning work on Covid-19 pneumonia classification. Inorder to achieve clinically relevant Computer-Aided Detection (CAD) the Chest x-rays of unique patients from National Institute of Health datasets from Kaggle.com are accessed.

B. Data pre-processing:

Datasets of 1200 Covid-19 CXRs ,1345 pneumonia CXRs and 1341 Normal CXRs are pre-processed and filtered by the gaussian filter for removing salt and pepper grains. The module is trained with main three labels – Covid-19 which include CXRs of infected Covid-19 patients, Normal holds healthy CXR images, Viral pneumonia includes CXRs of pneumonia affected patients. It is trained and tested with balanced datasets so it can yield all the classes prediction with

high accuracy and be devoid of malfunction. The balanced datasets used in our classification model obtain a higher balanced accuracy and balanced detection rate.

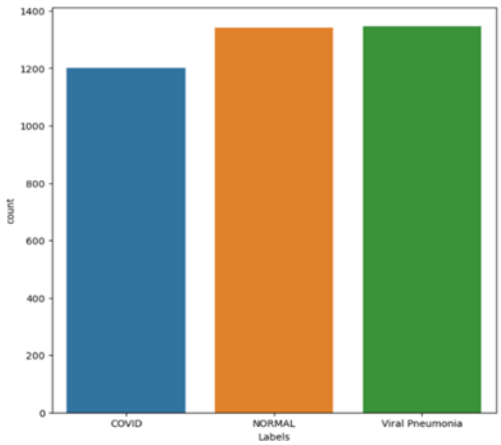


Figure 1: Labeled Balanced Dataset

3. Architecture:

Three 2D convolutional layers hold the entry flow framework with kernel dimensions (7,7),(1,1) and (3,3); two- dimensional max pooling layers with pool size (3,3) and (2,2) along with two batch normalization and inception blocks. Chest X-ray volumes of size (224,224,3) is given to input layer. This returns feature volumes of size (56,56,256) to the middle layer. Every associated layer is a combination of d*dense block and t*transition block with their corresponding values. The middle flow is

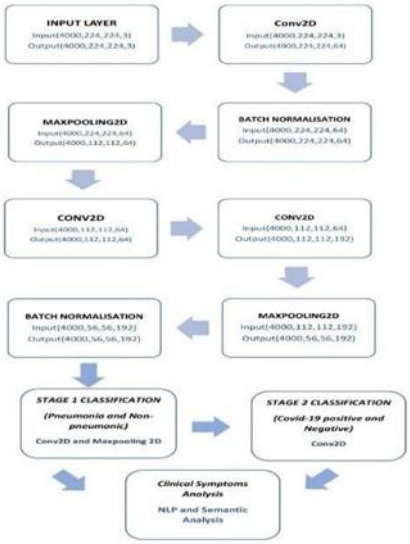


Figure 2 : Proposed frame work

made up of four interconnected layers. The output of the entry flow is processed by the middle flow, which returns 7x7x1024 feature volumes to the exit flow. Classification of feature volumes is done at the exit flow. It combines two- dimensional maximum pooling, two-dimensional average pooling, and fully-connected linked layers. 2D max pooling layers have a kernel size of 2x2x1 and a stride of 2x2x1. The

kernel size of the 2D average pooling layer is 7x7, and the stride is 1. Exit flow divides the feature volumes obtained by middle flow into categories such as: pneumonia affected, normal and Covid-19(+ve) and Covid-19(-ve) as shown below.

Further the non-pneumonia and Covid-19 negative undergo another processing through NLP by analysing the clinical symptoms. Here, the dataset contains information regarding the clinical symptoms including age and sex by means of a binary indication as to whether the tested individual is aged 60 years or above. The scikit-learn library function is used in this phase as it provides supervised and unsupervised learning algorithms. This developed model predicts possibility of Covid-19 infection on the Normal/Covid-19 negative CXRs in which eight binary features were used: sex, age 60 years or older, known contact with an infected person, and other five clinical symptoms – cough, fever, sore throat, breathlessness and headache.

Model: "sequential"

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 222, 222, 32)	896
max_pooling2d (MaxPooling2D)	(None, 111, 111, 32)	0
conv2d_1 (Conv2D)	(None, 109, 109, 64)	18496
max_pooling2d_1 (MaxPooling2	(None, 36, 36, 64)	0
conv2d_2 (Conv2D)	(None, 34, 34, 64)	36928
max_pooling2d_2 (MaxPooling2	(None, 17, 17, 64)	0
flatten (Flatten)	(None, 18496)	0
dense (Dense)	(None, 512)	9470464
dropout (Dropout)	(None, 512)	0
dense_1 (Dense)	(None, 256)	131328
dropout_1 (Dropout)	(None, 256)	0
dense_2 (Dense)	(None, 128)	32896
dropout_2 (Dropout)	(None, 128)	0
batch_normalization (BatchNo	(None, 128)	512
dense_3 (Dense)	(None, 64)	8256
dropout_3 (Dropout)	(None, 64)	0
dense_4 (Dense)	(None, 3)	195

Total params: 9,699,971
Trainable params: 9,699,715
Non-trainable params: 256

Figure 3: Sequential Model

4. Algorithm

The classification tool is developed using the CNN algorithm and each layer in the neural network is built in Keras using the sequential model. The conv 2D layer helps in building structured convolution network. In training module CNN takes two-dimensional CXRs as input. The input images from the datasets are broken down into pixels of size 224x224x3

.These CXRs are then resized to 56x56x224. Feature map is created by extracting the features from resized images using a Gaussian filter. The segmentation of CXRs is done using a Graph-cut algorithm. By the end of segmentation we get

trained weights extracted from chest x-rays. Max pooling and average pooling is applied in three layers to the feature map. The pooled images from the convolutional layers are converted into a single long feature vector by flattening. Every neuron in the neural network gets inputs from the other layers through the dense layer and it is fed to the droupout layer where overfitting is reduced. Batch Normalization works on every layer and also in between the layers of the structure. It affects each layer independently to do more of learning and also works on standardizing the inputs or outputs.

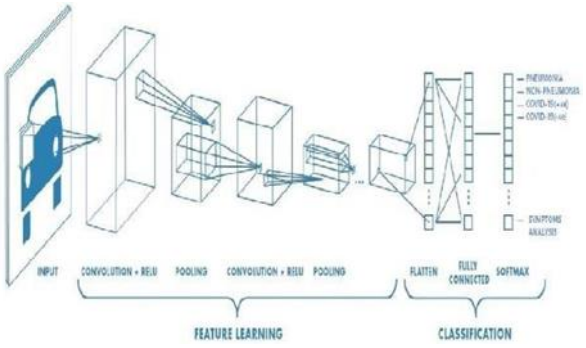


Figure 4: Algorithm of the proposed structure

The second diagnosis based on symptoms is analysed using the Support Vector Machine (SVM) algorithm. To analyse the symptoms for prediction the concept of maximal margin hyperplane is used. This outlines the decision boundaries among the eight symptoms. This proposed model is developed by Linear SVC, a faster implementation method of Support Vector Classification. At first libraries such as numpy, pandas are imported for better data analysis .Various Data based on eight symptoms are fed in the form of binary representation in a separate excel file and imported. The imported data file is read and scaled for optimisation of the data prior to training. The hyperplane in the linear SVM model can be expressed explicitly as:

$w \bullet x + b = 0,$

where w and b are linear SVM model parameters, and x is feature vector of the sample

After training the model with the sample dataset, the hyperplane with maximal margin is chosen to split positive symptoms from negative symptoms. Using joblib the model is saved and loaded to make predictions on the metadata set.

5. Implementation

The proposed framework is developed by 2D CNN which uses the CXRs of Covid-19, pneumonia affected and normal lungs for classification. In the entry flow the image is read and then converted from BGR to RGB scale and reduced to

224 pixels. The input image is the CXRs which is pre- processed and filtered with a gaussian filter for removing salt and pepper grains. By partitioning a digital image into various segments, the image segmentation optimises the representation of an image. The Graph-cut algorithm to the pre-processed CXRs to derive segmented CXRs.

The segmented region is extracted for feature collection using various techniques of first order and second order characteristics. The Gray Level Co-Occurrence Matrix (GLCM) is applied, as it is a statistical method commonly used to extract textural features from images. Following feature extraction the proposed architecture 2D CNN framework yields a two-level classification of a fully- connected model.

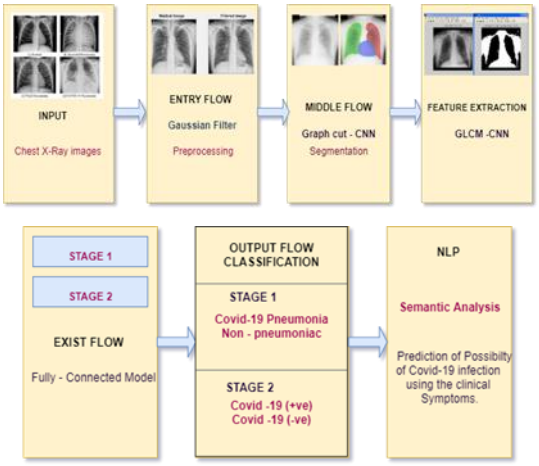


Figure 5: Workflow of the proposed module

At the exit flow the CXRs are classified as patients having pneumonia and non- pneumonia in the first level. The CXRs of patients diagnosed with pneumonia undergo another level of classification where the CXRs volumes are classified as Covid-19(+ve) or Covid-19(-ve). The patients with Normal CXRs(non-pneumonic and Covid-19 negative) are further questioned about their indications and are fed in the developed GUI to find the probability infection in them.

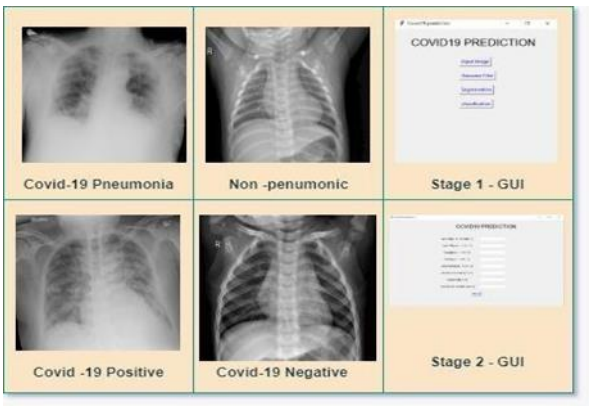
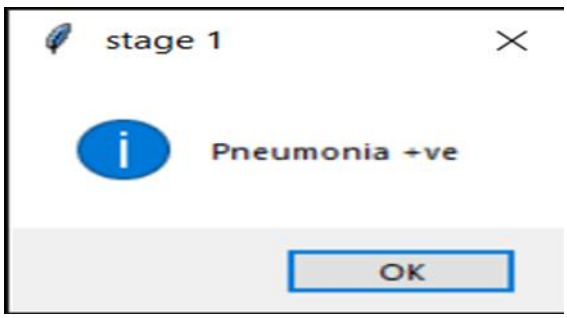


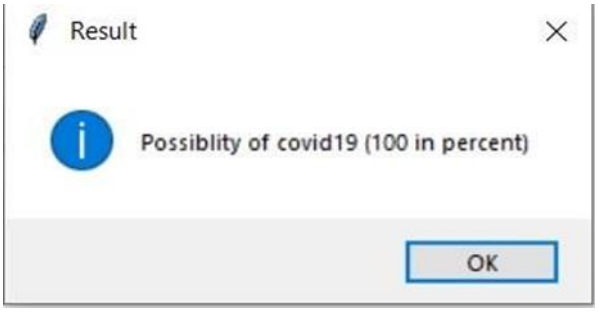
Figure 6: Classification framework

The labeled CXR images are provided as above. It is accessed and the output can be depicted using a Graphical User Interface.Level 1 and Level 2 GUIs are shown above. The sample GUI output for pneumonia type level 1 classification and level 2 Covid-19 testing classification is shown below. Finally the possibility rate of having Covid -19 infection on normal CXRs by analyzing the clinical symptoms is also obtained as given below.



(i). Level-1 classification output

(ii). Level-2 classification output



(iii). Final result – Possibility of having Covid- 19 infection in normal CXR's

Figure 7: Level wise output in the GUI

6.Results And Discussion

The input CXR images are divided in two levels: level 1 and level 2 with the help of the proposed framework. GUI is developed to display the output in two levels. Pneumonic and non-pneumonic lungs are categorized in level 1, and the pneumonic lungs are classified further into Covid-19(+ve) and Covid-19(-ve) in level 2.The project improves accuracy not only for Covid 19 but also for pneumonia by using the two-level procedure. The non-pneumonic and Covid 19 negative are further analysed after the two-level phase by providing symptoms using Natural Language Processing .

The preparation, verification, and evaluation analysis are summarised in the graphs below. The developed model was educated on the dataset of pneumonic and non-pneumonic CXRs in stage one classification. It's precision is shown in the graph below.

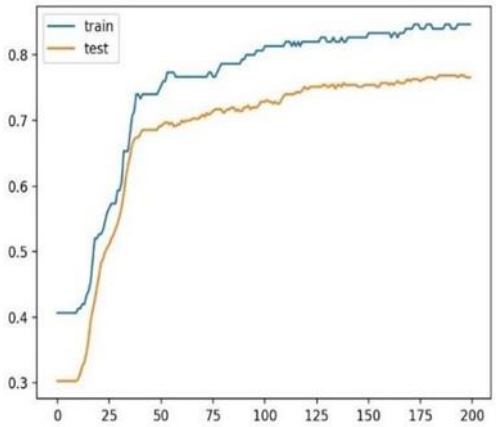


Figure 8: Level 1 accuracy graph

The model is further graded in level 2 as Covid 19(+ve) or Covid 19(-ve). Testing and validation of datasets indicates the accuracy as shown in the graph below.

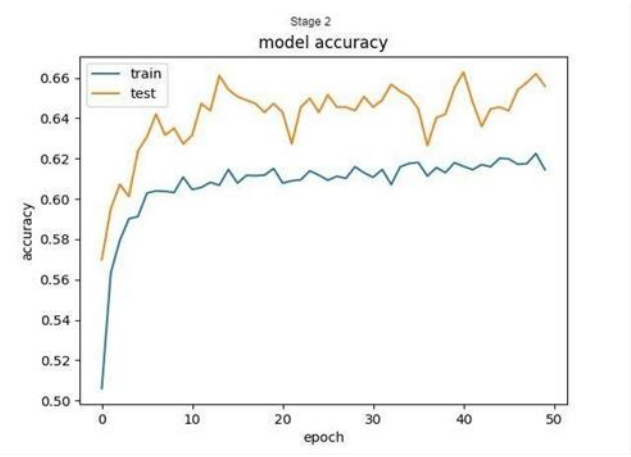


Figure 9: Level 2 accuracy graph

The accuracy of possibility of Covid-19 in non-pneumonic and Covid-19 negative patients is predicted by analysing the clinical symptoms through NLP is 96%, and is shown in the graph below.

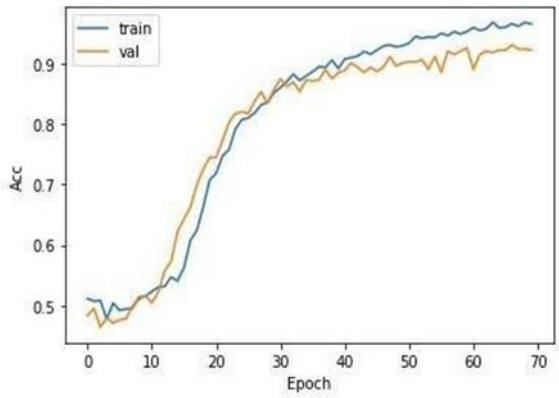


Figure 10: Symptoms accuracy graph

This confusion matrix machine learning approach assesses the system's precision, accuracy, and sensitivity. The

Confusion Matrix is a machine learning tool for calculating Precision, Accuracy, and the AUC-ROC curve.

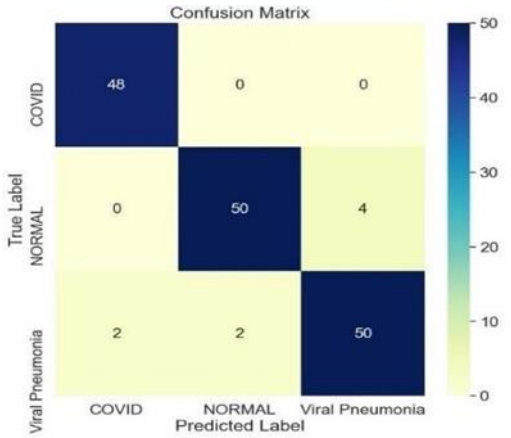


Figure 11: Confusion Matrix

tests. These findings will have a great impact on the screening and initial diagnosis of many diseases like Covid-

19. This research would open new doors in the field of medicine with the help of AI which will bring a new era in medical science.

7. Future Scope

With rapid increase in use of wireless patient monitoring devices lots of medical related information is available online through virtual assistants. This enormous data that is available will act as the key to process, diagnose and predict ailments. There has been a significant increase in the feedback obtained, this inherently results in higher accuracy and precision in the output. It promises a great future for the integration of AI with medical science. This research aims in eliminating the shortcomings of Covid-19 pneumonia diagnosis in the present system. The system would be of great help in remote areas where radiologists are scarce. Furthermore it can also be used to detect other chest or lung related diseases like tuberculosis. It is possible to make the model more robust by making use of more data on a much larger scale. It can further be developed for specific strains of the corona virus or other maladies on the basis of symptoms present in different geographical locations. Though the data available for training is humongous there is still a huge amount gap in the conversion of data from analog to the digital form. This transformation is a strenuous work but it will yield great outcomes in future.

Table I: Confusion Matrix of Level 1

Predicted/Actual	Non-pneumonic	Pneumonic
Non-pneumonic	48	2
Pneumonic	2	50

A. Filtration:

Annexure

Table II: Confusion Matrix of Level 2

Predicted/Actual	Covid-19(+ve)	Covid-19(-ve)
Covid-19 (+ve)	40	4
Covid-19 (-ve)	2	50

```
from tensorflow.keras.applications.vgg16 import preprocess_input
image1 = cv2.imread('Normal.png')
img = cv2.resize(image1, (512, 512))
blur = cv2.GaussianBlur(img,(5,5),0)
plt.imshow(blur)
plt.title('Gausian filter')
plt.show()
img = preprocess_input(img)
img = img[np.newaxis, :, :, :]
pred = model.predict(img)
print(pred[1])
```

As shown in the graphs above, accuracy of 96.04% is found for level 1 classification and 95.15% for level 2 classification during the testing of the CXRs. The symptoms calculation for the non-pneumonic and Covid-19 negative patients has yielded an accuracy of 96% . An aggregated precision and sensitivity of 95% for level 1 and 96 % for level 2 using the confusion matrix deep learning mechanism can be achieved. Thirty data sets from real-time unique patients for verification was used, and the project yielded 27 true values and 3 false values.

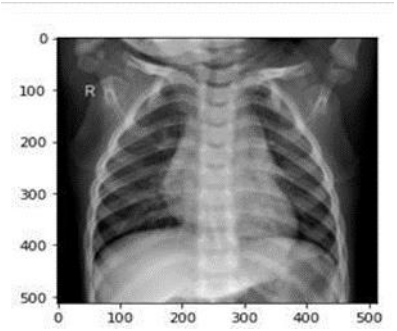


Figure 12: Gaussian Filter CXR

8 Conclusion

There has been a paradigm shift in healthcare owing to healthcare data availability and the rapid increase in the analytic techniques. Our research provides a glimpse into how artificial intelligence will transform medical practicesin future. AI will pave the way for better healthcare with fewer errors in diagnosing diseases. There is no need for manual feature extraction, this model is automated with an end-to- end structure making it easier for use. This automation saves time and also ,it provides an insight to early prognosis of COVID-19 pneumonic patients using CXRs and a prediction model is developed to estimate the probability of infection from Normal/negative Covid-19 CXRs. This has proven to be

a good alternative to the normally and widely used RT-PCR

The GLCM works on filtering the input CXRs to remove the salt and pepper grains. Gaussian filter is applied to each input image and the filtered image is sent for segmentation process for further data analysis.

B.Data Augmentation:

The image augmentation is done to training datasets as shown below:

```
img_width = img_height = 224

if K.image_data_format() == 'channels_first':
    input_shape = (3, img_width, img_height)
    X_train=X_train.reshape(X_train.shape[0],3,img_width,img_height)
    X_cv=X_cv.reshape(X_cv.shape[0],3,img_width,img_height)
    X_test=X_test.reshape(X_test.shape[0],3,img_width,img_height)

else:
    input_shape = (img_width, img_height, 3)
    X_train=X_train.reshape(X_train.shape[0],img_width,img_height,3)
    X_cv=X_cv.reshape(X_cv.shape[0],img_width,img_height,3)
    X_test=X_test.reshape(X_test.shape[0],img_width,img_height,3)
```

C.Segmentation

```
plt.figure(figsize=(12,12))
plt.subplot(1,3,1)
plt.imshow(cv2.resize(cv2.imread('Normal.png'), (512, 512)))
plt.title('Original Image')
plt.subplot(1,3,2)
plt.imshow(np.squeeze(pred[0]) > .5)
plt.title('Segmented image')
plt.show()
```

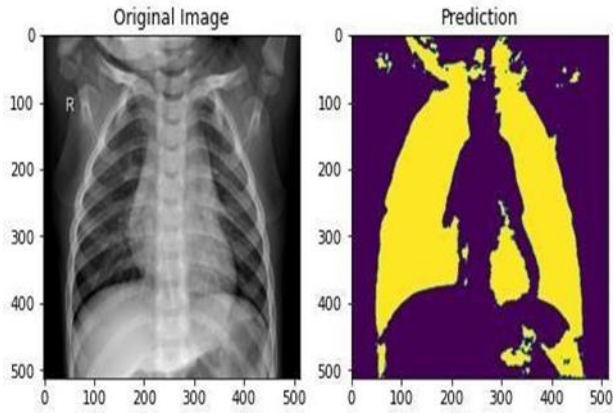


Figure 13: Segmentation of CXR

The original chest x-ray when processed the features are extracted and the trained weights are obtained .The predicted image above shows the segmented CXR of original one.Graph-cut algorithm is made use of in segmentation of the filtered images for feature maps.

D.Classification Tool

```
if round(classes[0][0]) == 1:
    print('Pneumonia')
    if round(classes[0][0])==1:
        print('Covid')
    else:
        print('Non covid')
elif round(classes[0][1]) == 1:
    print('Normal')
else:
    print('Pneumonia')
    if round(classes[0][0])==1:
        print('Covid')
    else:
        print('Non covid')
```

The output for the classified model is portrayed by making use of the following code:

E.Symptom Analysis

This second diagnosis involves analysing the symptoms using the SVM algorithm by implementing the linear SVC as shown below:

```
from sklearn.svm import LinearSVC
svc = LinearSVC()
svc.fit(X_train,y_train)
joblib.dump(svc,'svm.pkl')

# load svm model
svm_model = joblib.load('svm.pkl')
y_pred = svm_model.predict(X_test)
from sklearn.metrics import accuracy_score
print(accuracy_score(y_test, y_pred))
print(y_pred)
```

References

[1] "Covid-19 situation update worldwide, as of 16 may 2020," 2020. https://www.ecdc.europa.eu/en/geographical- distribution-2019-ncovcases.

[2] A. Jacobi, M. Chung, A. Bernheim, and C. Eber, "Portable chest x-ray in coronavirus disease-19 (covid-19): A pictorial review," Clinical Imaging, 2020.

[3] "Imaging the coronavirus disease covid-19," 2020. https://healthcare-ineurope.com/en/news/imaging-the- coronavirus-disease-covid-19.html.

[4] Selvan, R. et al. Lung segmentation from chest X-rays using variational data imputation. ICMLWorkshop on The Art of Learning with Missing Values (2020). ArXiv preprint arXiv:2020.2005.10052.

[5] A. Mitani, A. Huang, S. Venugopalan, G. S. Corrado, L. Peng, D. R. Webster, N. Hammel, Y. Liu, and A. V. Varadarajan, "Detection of anaemia from retinal fundus images via deep learning," Nature Biomedical Engineering, vol. 4, no. 1, pp. 18–27, 2020.

[6] J. P. Cohen, P. Morrison, and L. Dao, "Covid-19 image data collection," arXiv preprint arXiv:2003.11597, 2020.

[7] P. Afshar, S. Heidarian, F. Naderkhani, A. Oikonomou,K. N. Plataniotis, and A. Mohammadi, "Covid-caps: A capsule network-based framework for identification of covid-19 cases from x-ray images,"arXiv preprint arXiv:2004.02696, 2020.

[8] Vaishya, R., Javaid, M., Khan, I. H. & Haleem, A. Artificial intelligence (AI) applications for COVID-19 pandemic Diabetes & Metab. Syndr. Clin. Res. & Rev. (2020).

[9] Ozturk, T. et al. Automated detection of COVID-19 cases using deep neural networks with X-ray images. Comput. Biol. Medicine 103792 (2020).

[10] Yoo, S. H. et al. Deep learning-based decision-tree classifier for COVID-19 diagnosis from chest X-ray imaging. Front. Medicine 7, 427 (2020).

[11] 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. arXiv preprint arXiv:2004.09363 (2020).

[12] Abbas, A., Abdelsamea, M. M. & Gaber, M. M. Classification of COVID-19 in chest X-ray images using detrac deep convolutional neural network. arXiv preprint arXiv:2003.13815 (2020).

[13] C. Zheng, X. Deng, Q. Fu, Q. Zhou, J. Feng, H. Ma, W. Liu, and X. Wang, "Deep learning-based detection for covid- 19 from chest ct using weak label," medRxiv, 2020.

[14] 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 preprint arXiv:2003.11055, 2020.

[15] T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim, and U. R. Acharya, "Automated detection of covid- 19 cases using deep neural networks with x-ray images," Computers in Biology and Medicine, p. 103792, 2020.

[16] M. Ilyas, H. Rehman, and A. Na¨t-Ali, "Detection of covid-19 from chest x-ray images using artificial intelligence: An early review," arXiv preprint arXiv:2004.05436, 2020.

[17] A. Narin, C. Kaya, and Z. Pamuk, "Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks," arXiv preprint arXiv:2003.10849, 2020.

[18] Sethy, P. K. & Behera, S. K. Detection of coronavirus disease (COVID-19) based on deep features. Preprints (2020).

[19] Wang, L. & Wong, A. COVID-Net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images. arXiv preprint arXiv:2003.09871 (2020).

[20] Farooq, M. & Hafeez, A. COVID-ResNet: A deep learning framework for screening of covid19 from radiographs. arXiv preprint arXiv:2003.14395 (2020).

[21] Borkowski, A. A. et al. Using artificial intelligence for COVID-19 chest X-ray diagnosis. medRxiv (2020).

[22] Apostolopoulos, I. D. & Mpesiana, T. A. COVID-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks. Phys. Eng. Sci. Medicine 1 (2020).

[23] Wynants, L. et al. Prediction models for diagnosis and prognosis of COVID-19 infection: systematic review and critical appraisal. BMJ 369 (2020).

[24] Peng, Y. et al. COVID-19-CT-CXR: a freely accessible and weakly labeled chest X-ray and CT image collection on COVID-19 from biomedical literature. arXiv preprint arXiv:2006.06177(2020).

[25] Vayá, M. d. l. I. et al. BIMCV COVID-19+: a large annotated dataset of RX and CT images from COVID-19 patients. arXiv preprint arXiv:2006.01174 (2020).

[26] Pereira, R. M., Bertolini, D., Teixeira, L. O., Silla Jr, C.

[27] N. & Costa, Y. M. COVID-19 identification in chest X-ray images on flat and hierarchical classification scenarios. Comput. Methods Programs Biomed. 105532 (2020).

[28] Linda Wang, Z. Q. L. & Wong, A. COVID-Net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest radiography images. arXiv preprint arXiv:2003.09871 (2020).

[29] Zhang, K. et al. Clinically applicable AI system for accurate diagnosis, quantitative measurements, and prognosis of COVID-19 pneumonia using computed tomography. Cell 181, 1423–1433 (2020). [29] A. Bustos,

[30] Pertusa, J.-M. Salinas, and M. de la Iglesia-Vaya, "Padchest: A large chest x-ray image dataset with multi- label annotated reports," arXiv preprint arXiv:1901.07441, 2019.Paszke, A. et al. Pytorch: An imperative style, high- performance deep learning library. In Wallach, H. et al. (eds.) Advances in Neural Information Processing Systems 32, 8024– 8035 (Curran Associates, Inc., 2019)