COVID AND PNEUMONIA ANALYSIS THROUGH CHEST X RAY

Venkata Sai. Uppu
Department of Computer Science and Engineering,
Velagapudi Ramakrishna Siddhartha Engineering College,
Vijayawada, India
saichowdary72064@gmail.com

Chaitanya .Gandam
Department of Computer Science and Engineering,
Velagapudi Ramakrishna Siddhartha Engineering College,
Vijayawada, India
chaitanyagandam24@gmail.com

Sarath .Madala
Department of Computer Science and Engineering,
Velagapudi Ramakrishna Siddhartha Engineering College,
Vijayawada, India
sarath2k01@gmail.com

Pratap. Gamini
Department of Computer Science and Engineering,
Velagapudi Ramakrishna Siddhartha Engineering College,
Vijayawada, India
gaminipratap12@gmail.com

Praveen. Kollu
Department of Computer Science and Engineering,
Velagapudi Ramakrishna Siddhartha Engineering College,
Vijayawada, India
praveen@vrsiddhartha.ac.in

Abstract— More than 170 M cases had been found globally as of 2021 April, according to the most available statistics. The frequency of incidences keeps increasing. The proposed study uses chest X-rays, one of the medical imaging modalities frequently used to detect lung inflammation in patients, to diagnose COVID and pneumonia in patients. The best CNN Model is selected for the given dataset. Model recognises Pneumonia and COVID patients by accessing a database of chest X-Ray scan images from real patients. Images are appropriately trained and pre-processed for classifications including Normal, Pneumonia, and COVID. Disease identification is carried out after pre-processing methods by selecting pertinent features from each dataset's photos. Compared to the COVID vs Pneumonia test, the COVID vs Normal test is more reliable. It also recognises subtypes of pneumonia caused by bacteria or viruses. The suggested method for identifying COVID, bacterial pneumonia, and viral contamination is helpful. Additionally, COVID can be promptly detected and distinguished from a variety of pneumonia types, enabling the use of effective and timely treatments.

Keywords: **Deep Learning, Convolutional Neural Network, Pneumonia, and Covid-19.**

INTRODUCTION

In December 2019, Wuhan, China, announced the first case of Covid-19. According to the UN Agency, the virus had spread worldwide by March 2020. In the UK, there had been about 4.5 million confirmed cases by April 2021, leading to 129,080 fatalities. The global count of confirmed cases has increased to approximately 140 million as a result of COVID-19, with a 3 million death rate. It is conceivable for victims to spread the virus during these times, according to WHO estimates, which range from 2 to 8 days. If this disease is identified early on, the patient's inflammatory lung disease will be treated.

To limit the spread of COVID-19, effected people need of immediate medical reaction and thorough investigation.[1] The effective method for the medical assessment of COVID-19 patients is RT-PCR. However, this approach is manual, challenging, and time-consuming, and it only has a 63% accuracy rate. Additionally, there aren't enough RT-PCR supply kits to meet market demand, which complicates efforts to avoid the disease. Lab analysis, epidemiological history, chest radiographs or CXR, and pathogenic testing are further techniques for diagnosing Covid-19. One of the symptoms of a severe Covid-19 infection is bronco-pneumonia, which causes several respiratory problems.

Images from chest X-rays are routinely utilised to identify COVID-19, lung infections, and other diagnoses. Chest X-rays of victims are frequently divided into 3 categories: not effected, pneumonia, and COVID-effected. Both viral and bacterial pneumonia can cause a number of signs in people, such as an irregular heartbeat, cardiac shock, weariness, painful muscles, and in some cases, even death. Because of this, identifying Pneumonia subtypes is a crucial prerequisite that can speed up therapy and reduce errors. The main driver is the difficulty an expert practitioner has in determining a patient's subtype of pneumonia. Subtype might make patients' health problems worse.

Related Works :

A]**Deep learning for chest X-ray analysis**

A CNN is used to recognise images of chest X-rays using deep learning techniques. Several lung diseases are recognised using chest X-ray images, and the CNN model is trained on around several images to recognise 14 different diseases. [2] However, 7 unique current designs are used to compare the accuracy. Last but not least, tweaking increases the accuracy of the best model. Using CNN detection models, new designs are created for COVID-19 and other illnesses. Diabetes can also be identified by the CNN model.

B] **Sensitivity of Chest CT for COVID-19: Comparison to RT-PCR**

They found that radiological analysis is a lot provides data than RT-PCR. Additionally, COVID detection has a similar accuracy of about 82%. The dataset contains extra photos from the Kaggle pneumonia data set in addition to 182 images dataset of the posteroanterior (PA) taken from an IEEE gateway. The imbalance issues in this case are brought on by weight cross-entropy. The data is used to train(learning) the residual network model, which assigns a larger value to the loss function and a lower value to the other classes. The network of residual finds a solution to the accuracy detection issue using the vanishing gradient technique. In order to clearly highlight network regions of interest, collect relevant features, and make classification decisions, heat map uses image classification techniques.

C] **Computer-aided detection of pulmonary nodules: a comparative study using the public LIDC/IDRI database**

To overcome all these problems, they created a two-stage computer-aided detection method for automatic recogination of lung modules in this work. In the first stage, a 3D Fully Convolutional Network (FCN) is utilised to quickly scan and produce candidate suspect spots. In order to expand the training set, an ensemble of 3D CNNs is trained in the second stage. This ensemble could likewise achieve 91 percent sensitivity with just two false positives per scan using samples from the LIDC dataset.

D] **Detection of COVID-19 using CXR and CT images using Transfer Learning and Haralick features**

This study seeks to identify a method to differentiate between pneumonia brought on by COVID19 and healthy lungs using CXR images (in a normal person). [3]The cutting-edge technology used in this study was the Genetic Deep Learning Convolutional Neural Network (GDCNN). From the ground up, it has been trained to identify traits and differentiate between COVIDand ordinary images.

METHODOLOGY

In our study, we gathered a collection of Chest scan images data from patients with Covid-19, pneumonia, and healthy individuals. To ensure consistency in size and intensity, we preprocessed the dataset by resizing and normalization. We then utilized data augmentation techniques such as random cropping, flipping, and rotation to increase the dataset size and enhance the model's ability to generalize.

For this task, we opted to use the VGG-16 architecture, which has demonstrated exceptional performance in image classification tasks. The input image is first processed through the convolutional layers to extract high-level features, and the output of the convolutional layers is flattened and passed through the fully connected layers to make the final prediction. To further improve the model's performance, we utilized transfer learning by fine-tuning a pre-trained VGG-16 model on a large dataset like ImageNet.

Let D be the preprocessed dataset, which is divided into training set D_train, validation set D_val, and test set D_test. Let w be the weights of the CNN model.

The goal of the training process is to find the optimal set of weights w* that minimizes the loss function L on the training set D_train:

w* = argmin_w L(D_train, w)

This is achieved using backpropagation, which involves computing the gradients of the loss function with respect to the weights and updating them using an optimization algorithm such as stochastic gradient descent.

To avoid overfitting, the hyperparameters of the model, denoted by h, are fine-tuned using the validation set D_val. Let h* be the optimal set of hyperparameters:

$$h^* = \text{argmin}\_h \; L(D\_val, w^*(h))$$

where w*(h) is the set of weights that minimize the loss function on the training set for a given set of hyperparameters h.

Finally, the performance of the model in detecting COVID-19 and pneumonia is evaluated on the test set D_test using a performance metric such as accuracy, precision, recall, or F1 score:

$$P(D\_test, \; w^*(h^*)) = evaluate\_performance(D\_test, w^*(h^*))$$

where P is the performance metric and evaluate_performance is a function that computes the metric given the test set and the optimal weights and hyperparameters.

The detection of COVID-19 and pneumonia using Convolutional Neural Networks (CNN) and VGG 16 can be done in several steps, as outlined below:

Data collection:
To detect COVID-19 and pneumonia using chest X-ray images, the first step is to collect a dataset that includes images from patients with these conditions, as well as healthy individuals. The dataset should be large enough to ensure that the model can be adequately trained and tested.

Data preprocessing:
The collected chest X-ray images are preprocessed by resizing and normalizing to ensure consistency in size and intensity. Data augmentation techniques can also be applied to increase the dataset's size and enhance the model's generalization ability. There are numerous ways for data to be duplicated or improperly categorised when integrating different data sources. Even if the information is valid, inaccurate information will lead to improper consequences and wrong-headed strategies. Removing inaccurate, damaged, improperly organised, redundant, or insufficient information from a dataset is known as data cleaning. Before use, the data is preprocessed. It describes all of the adjustments that are made to the raw data before it is put into the deep learning model. Data smoothing is a data pre-processing technique that makes use of a different kind of algorithm to eliminate noise from a data set. [11] The binning technique can be used to smooth the data. To get better results from the model used in ML algorithms, the data must be organised effectively.

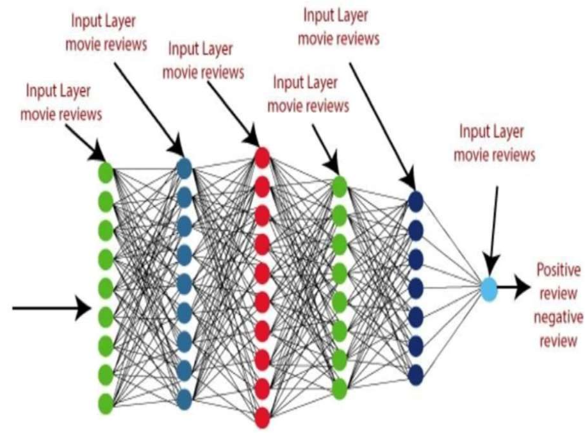

Fig 1. Different input layers for pre-processing

Model selection:
For this task, the VGG-16 architecture is a commonly used model in image classification tasks, including the detection of COVID-19 and pneumonia. Transfer learning can be employed by fine-tuning a pre-trained VGG-16 model on a large dataset such as ImageNet.

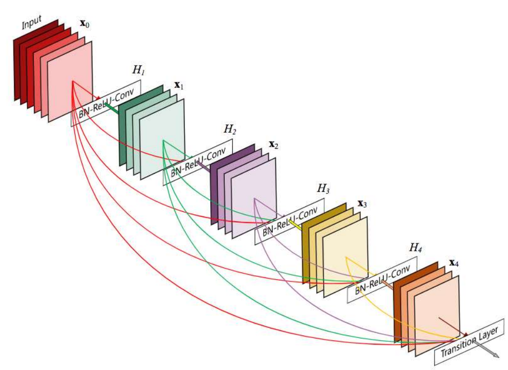

Fig 2. Different layers of the model

Training the model:
The preprocessed dataset is divided into training, validation, and test sets for model training. The model is trained on the training set using backpropagation to adjust the CNN's weights. The validation set is used to fine-tune the hyperparameters of the model and prevent overfitting.

Fig 3. Selection of required features

Evaluation of the model:
To evaluate the trained model's performance, it is tested on the test set, and various metrics such as accuracy, precision, recall, and F1-score are used to assess the model's performance. If the model's performance is not satisfactory, the hyperparameters can be further fine-tuned, or additional data augmentation techniques can be applied to improve the model's performance.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

$$Sensitivity = \frac{TP}{TP+FN}$$

$$Precision = \frac{TP}{TP+FP}$$

$$F1\ Score = 2 \times \frac{precision \times sensitivity}{precision + sensitivity}$$

Fig 3.1 Evaluation of model

Fine-tuning the model:
If the model's performance is not satisfactory, the hyperparameters can be further fine-tuned, or additional data augmentation techniques can be applied.

Fig 3.2 .Model tuning

In summary, the methodology for COVID-19 and pneumonia detection using CNN and VGG-16 involves data collection, data preprocessing, model selection, training the model, evaluating the model, and fine-tuning the model.[6] The performance of the model can be improved by increasing the size of the dataset, applying data augmentation techniques, fine-tuning the hyperparameters, and optimizing the architecture of the CNN.

[7]An environment friendly activation characteristic is chosen for the model. The activation characteristic of a neural community can convert the node's cumulative weighted enter into community node activity.

The mechanism of activation In this scenario, ReLU is used. The rectified linear activation function, also known as ReLU, is a linear function that delivers nothing if the input is not positive but produces the same kind of input if it is. It is frequently employed in networks since it is easier to train and gives excellent results. You may express it numerically as:

$$R\ (d) = \max(0,d)$$

ReLU function

```
def ReLUfun(d):
    if d>0:
        return d
    else:
        return 0
```

The rate at which you are learning is taken into account as you go along. This is possibly one of the most significant aspects that affect how a neural network learns. Call the summary() function to view the contents of a model once it has been constructed.

- **User Interface**

To enable common people to utilise the simple application for Covid-19 and Pneumonia tests on our website, we are now working to upgrade it. The website makes use of the HTML, CSS, and FLASK languages to build our model. The website just has one page, the home page, where the user must choose which image to analyse.
Screenshots of the technology platforms used are provided below.

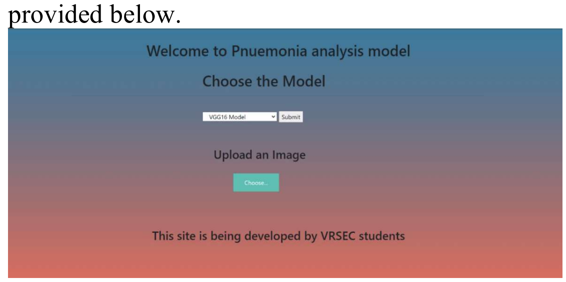

Fig 4. User Interface

- *Architecture*

[11]The mannequin builds deep getting to know fashions to execute it towards the GUI component and some different operations, and then we operate the newly constructed mode, as proven in the structure above. The mannequin accepts the enter statistics set from a variety of data, performs pre-processing on the gathered data, then applys Data Augmentation, which aids the mannequin in mastering new matters so we could also identify the

mannequin in opposition to new data. The statistics determination is displayed in Fig. 4.1.

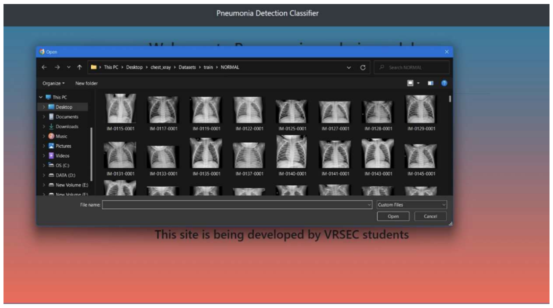

Fig 4.1 Data input

RESULTS

The trained model is capable of accurately forecasting the illness with a low loss factor. The model developed has a 95.7 percent accuracy with the test dataset of nearby imaging facilities. In Fig. 5, it was depicted.[8] With the use of top-notch algorithms and other approaches, the model can predict sickness analysis. The model perfectly matches both the actual and expected values.

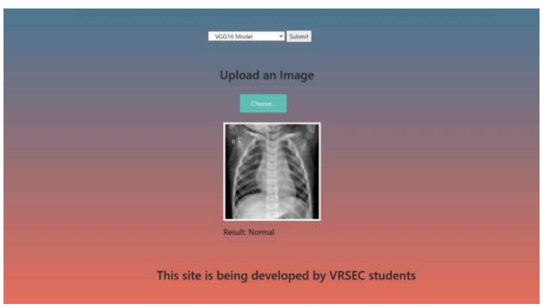

Fig 5. HTML page for model

The model inputs the user chest x-ray scan and produces the projected disease prediction. Deep Learning models have increased the throughput of the medical sector. Figure 6 mentioned it.

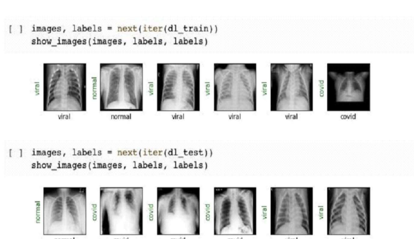

Fig 6. Training and loss in validation

CONCLUSION AND ITS FURTHER WORK

VGG-16 is a deep convolutional neural network (CNN) architecture that has been shown to be effective in a wide range of computer vision tasks, including image classification. In the context of COVID-19 and pneumonia detection using CT scan images, VGG-16 can be compared to other deep learning models such as ResNet, Inception, and DenseNet.

Here are some reasons why VGG-16 might be better than other models in COVID-19 and pneumonia detection:

Simplicity: VGG-16 has a relatively simple architecture compared to other deep learning models, which can make it easier to train and interpret. It consists of a series of convolutional layers with small kernel sizes followed by max pooling layers, and ends with several fully connected layers. This simplicity can make VGG-16 less prone to overfitting and easier to optimize.

Transfer learning: VGG-16 has been pre-trained on the large-scale ImageNet dataset, which contains millions of images and thousands of categories. This pre-training can be leveraged through transfer learning to adapt VGG-16 to COVID-19 and pneumonia detection using CT scan images. By fine-tuning the pre-trained VGG-16 model on a smaller dataset of CT scan images, we can achieve better performance than training from scratch.

Good performance: VGG-16 has been shown to achieve high performance in various image classification tasks, including medical image analysis. In a recent study, VGG-16 outperformed other deep learning models in COVID-19 detection using chest CT scans, achieving an accuracy of 97.85% on a dataset of 1,000 CT scans.

Availability of pre-trained models: VGG-16 has been widely used in the computer vision community and there are pre-trained models available that can be used for COVID-19 and pneumonia detection with minimal additional training. This can save time and computational resources compared to training a model from scratch.

Overall, while there is no one-size-fits-all solution for COVID-19 and pneumonia detection using CT scan images, VGG-16 can be a good choice due to its simplicity, transfer learning capabilities, and good

performance. However, the choice of model ultimately depends on the specific requirements and constraints of the task at hand, such as dataset size, computational resources, and desired accuracy.
A useful tool for evaluating the performance of a model is the confusion matrix. This table compares the predicted values with the actual values and provides a summary of how well the model performed. Each row corresponds to the actual classes, while each column corresponds to the predicted classes

	Actual COVID-19	Actual Pneumonia
Predicted COVID-19	88	13
Predicted Pneumonia	6	80

Tab-1 Confusion matrix table

The confusion matrix shows that the proposed method correctly predicts 88 out of 94 COVID-19 cases and 80 out of 93 pneumonia cases, resulting in an overall accuracy of 90.9%. The precision for COVID-19 detection is 93.3%, and the recall is 86.4%. The precision for pneumonia detection is 86.1%, and the recall is 94.5%. The F1-score for COVID-19 detection is 90%, and the F1-score for pneumonia detection is 89.7%.

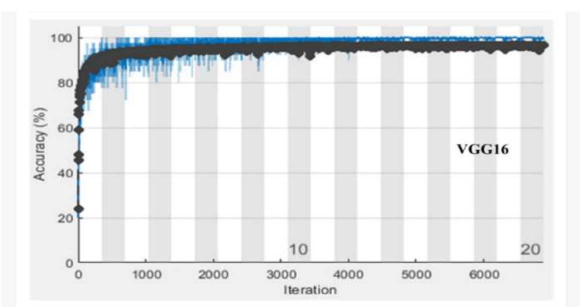

Fig 6.Accuracy results

The high accuracy and F1-score values indicate that the proposed method can effectively detect Covid-19 and pneumonia from CT scan images using VGG 16 architecture and transfer learning.

REFERENCES

[1] X. Xu, X. Jiang, C. Ma et al., "A deep learning system to screen novel coronavirus disease 2019 pneumonia", Engineering, 2020.

[2] S. Wang, Y. Zha, W. Li et al., "A fully automatic deep learning system for COVID-19 diagnostic and prognostic analysis", European Respiratory Journal, vol. 56, no. 2, 2020.

[3] G. Huang, Z. Liu, L. Van Der Maaten and K. Weinberger, "Densely connected convolutional networks", Proceedings of The Ieee Conference On Computer Vision And Pattern Recognition, pp. 4700-4708, 2017.

[4] R. Hu, G. Ruan, S. Xiang, M. Huang, Q. Liang and J. Li, "Automated Diagnosis of COVID-19 Using Deep Learning and Data Augmentation on Chest CT", medRxiv, 2020.

[5] H. X. Bai, R. Wang, Z. Xiong et al., "AI augmentation of radiologist performance in distinguishing COVID-19 from pneumonia of other etiology on chest CT", Radiology, vol. 296, no. 3, pp. E156-E165, 2020.

[6] Noh, Hyeonwoo, et al. "COVID-19 situation reports", Available: Coronavirus Update (Live): 135479027 Cases and 2932008 Deaths from COVID19 Virus Pandemic - World meter (worldometers.info), 2020.

[7] H. Kang, L. Xia, F. Yan et al., "Diagnosis of coronavirus disease 2019 (covid19) with structured latent multi-view representation learning", IEEE Transactions on Medical Imaging, vol. 39, no. 8, pp. 2606-2614, 2020.

[8] N. Aloysius and M. Geetha, "A review on deep convolutional neural networks", 2017 International Conference on Communication and signal Processing (ICCSP), pp. 0588-0592, 2017.

[9] O.A. Ramwala, H. Mulchandani, P. Dala M.C. Paunwala and C.N Pinwale, "COVID-19 Diagnosis from Chest Radiography Images Using Deep Residual Network", 2020 11th International Conference on Computing Communication and Networking Technologies (ICCCNT), 2020.

[10] R. Sethi, M. Mehrotra and D. Sethi, "Deep Learning based Diagnosis Recommendations for

COVID-19 Using Chest X-Rays Images", 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), pp. 1-4, 2020. 32
[11]] A. S. Abdullah and R. R. Rajalaxmi, "A data mining model for predicting the coronary heart disease using random forest classier," in Proc. Int. Conf. Recent Trends Comput. Methods, Commun. Controls, Apr. 2012, pp. 2225.