**PNEUMONIA DETECTION SYSTEM**

 A

Report submitted in partial fulfilment of the requirement for the

degree of

B.Tech.

In

###### Information Technology

By

##### Ria Kalra (2101640130061)

##### Under the guidance of

##### (Ashish Tripathi Sir)

##### (Assistant Professor)

##### Project Id: I



Pranveer Singh Institute of Technology, Kanpur

Dr A P J A K Technical University

Lucknow

## DECLARATION

This is to certify that Report entitled “**PNEUMONIA DETECTION SYSTEM**….”which is submitted by **Ria Kalra** in partial fulfilment of the requirement for the award of degree B.Tech. in Information Technology to Pranveer Singh Institute of Technology, Kanpur Dr. A P J A K Technical University, Lucknow comprises only our own work and due acknowledgement has been made in the text to all other material used.

#### Date: 1 March 2023

Ria Kalra (2101640130061)

## 

## 

## Certificate

This is to certify that Report entitled “**PNEUMONIA DETECTION SYSYTEM...”** which is submitted by **Ria Kalra** in partial fulfilment of the requirement for the award of degree B.Tech. in Information Technology to Pranveer Singh Institute of Technology, Kanpur affiliated to Dr. A P J A K Technical University, Lucknow is a record of the candidate own work carried out by him under my supervision. The matter embodied in this thesis is original and has not been submitted for the award of any other degree.

| Signature:     Piyush Bhushan Singh  HoD IT Department,  PSIT, Kanpur |  | Signature:  Ashish Tripathi  Assistant Professor  IT Department,  PSIT, Kanpur |
| --- | --- | --- |
|  |  |  |

**ACKNOWLEDGEMENT**

*It gives us a great sense of pleasure to present the report of the B.Tech. Project undertaken during B.Tech. Second Year. We owe special debt of gratitude to our project supervisor Ashish Tripathi Sir.., Department of Information Technology, Pranveer Singh Institute of Technology, Kanpur for his constant support and guidance throughout the course of our work. His sincerely, thoroughness and perseverance have been a constant source of inspiration for us. It is only his cognizant efforts that our endeavours have seen light of the day.*

*We also take the opportunity to acknowledge the contribution of Mr. Piyush Bhushan Singh, HoD, Information Technology Department, Pranveer Singh Institute of Technology, Kanpur for his full support and assistance during the development of the project.*

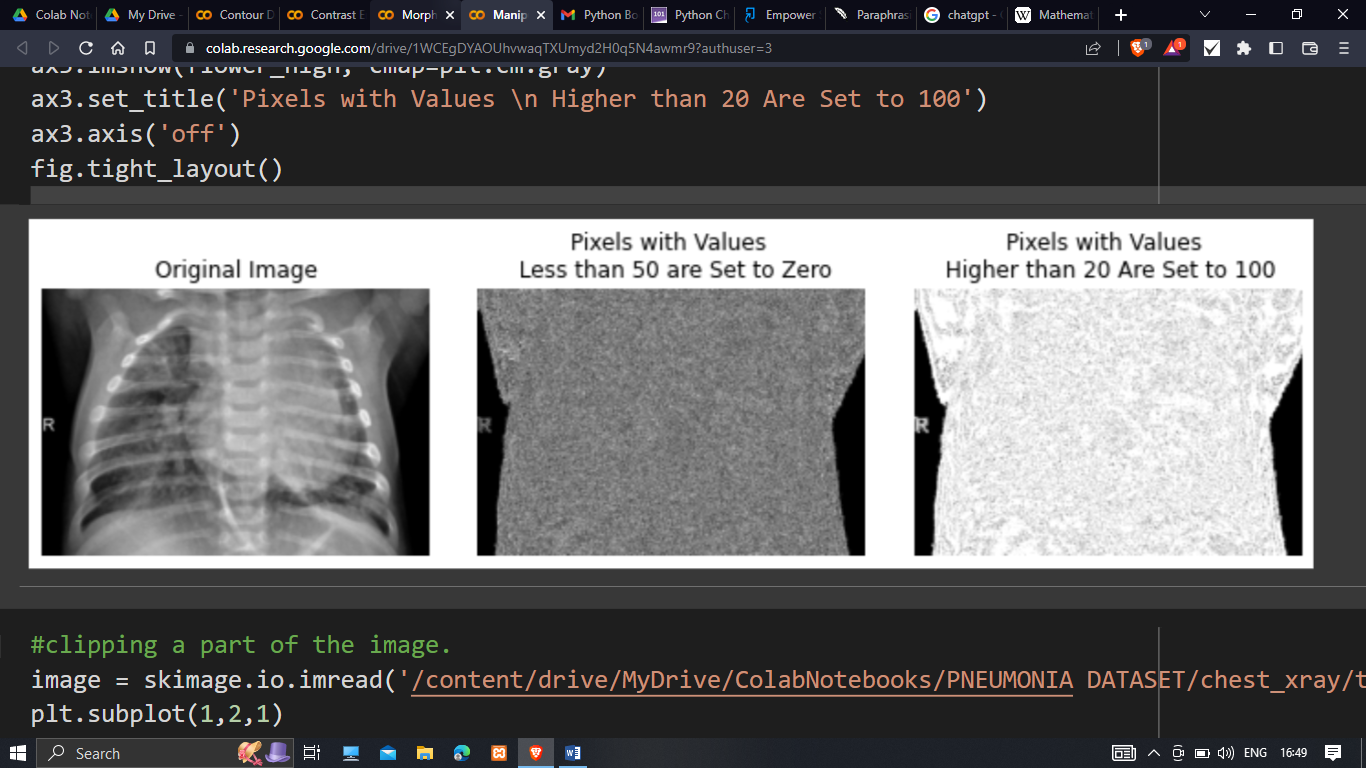
*We also do not like to miss the opportunity to acknowledge the contribution of all faculty members of the department for their kind assistance and cooperation during the development of our project. Last but not the least, we acknowledge our friends for their contribution in the completion of the project.*

| *Signature*  *Name:Ria Kalra*  *Roll No. 2101640130061* |  |
| --- | --- |
|  |  |

**ABSTRACT**

*.*

A Digital Image is an image composed of pixels, each with finite, discrete units for numeric representation for its intensity. Chest X-Rays which are used to diagnose pneumonia, need expert radiotherapists for evaluation. Thus, developing an automatic system for detecting pneumonia would be beneficial for treating the disease without any delay particularly in remote areas. Due to the success of deep learning algorithms in analysing medical images, CNNs) have gained much attention for disease classification. In addition, features learned by Pre-trained CNN models on large-scale datasets are much useful in image classification tasks. In this work, we appraise the functionality of pre-trained CNN models utilized as feature-extractors followed by different classifiers for the classification of abnormal and normal chest X-Rays. We analytically determine the optimal CNN model for the purpose. Statistical results obtained demonstrates that pretrained CNN models employed along with supervised classifier algorithms can be very beneficial in analysing chest X-ray images, specifically to detect Pneumonia. In this project Transfer learning and a CNN Model is used to detect whether the person has pneumonia or not using chest x-ray. Image enhancement is one of the most complex and important tasks in digital image processing. Image enhancement Pre-processing techniques **a**re used in improving the visual quality of images. The majority of applications, including radiography, MRI, ultrasound imaging, tomography, fundus imaging, and now use medical imaging. Contrast and Image quality are the major issues in medical imagery. Image Enhancement techniques makes the image more clear for human perception or machine analysis. The process of Image enhancement doesn't raise the inbuilt information content of the data, but can highlight the features of interest to detect the objects in a simple and efficient manner for accurate results.



**Fig 1 (**Chest X-ray image form our dataset)

**TABLE OF CONTENTS**

1. INTODUCTION
   1. MOTIVATION
   2. BACKGROUND OF PROBLEM
   3. PROBLEM STATEMENT
   4. PROPOSED WORK
2. DESIGN METHODOLOGY
3. IMPLEMENTATION
4. TESTING RESULT AND ANALYSIS
5. CONCLUSION AND FUTURE ENHANCEMENTS

REFERENCES

**CHAPTER – I INTRODUCTION**

Pneumonia is an acute pulmonary infection that can be caused by bacteria, viruses, or fungi

and infects the lungs, causing inflammation of the air sacs and pleural effusion, a condition in which the lung is filled with fluid. It accounts for more than 15% of deaths in children under the age of five years. l. Radiological examination of the lungs using computed tomography (CT), magnetic resonance imaging (MRI), or radiography (X-rays) is frequently used for diagnosis. X-ray imaging constitutes a non-invasive and relatively inexpensive examination of the lungs

* 1. **MOTIVATION**

Some of the most popular methods for Pneumonia detection, currently, include Sputum tests, Blood tests and CT scan. Methods like Blood tests take longer time to give out the results. Whereas, methods like CT scan are very expensive for a common man. These factors inspired us to consider this project where a patient can get a rough idea about his/her health , thus decreasing costs for common man and number of tests required and moreover help in early detection of disease.

Some of the most popular methods for Pneumonia detection, currently, include Sputum tests, Blood tests and CT Scan.

Methods like Blood tests take longer time to give out results. Whereas, methods like Sputum test is comparatively inaccurate and

lastly, methods like CT Scan are very expensive for a common man. These factors inspired us to consider this project where a

patient can get a rough idea about his health, thus decreasing costs for common man and number of tests required.

Some of the most popular methods for Pneumonia detection, currently, include Sputum tests, Blood tests and CT Scan.

Methods like Blood tests take longer time to give out results. Whereas, methods like Sputum test is comparatively inaccurate and

lastly, methods like CT Scan are very expensive for a common man. These factors inspired us to consider this project where a

patient can get a rough idea about his health, thus decreasing costs for common man and number of tests required

Some of the most popular methods for Pneumonia detection, currently, include Sputum tests, Blood tests and CT Scan.

Methods like Blood tests take longer time to give out results. Whereas, methods like Sputum test is comparatively inaccurate and

lastly, methods like CT Scan are very expensive for a common man. These factors inspired us to consider this project where a

patient can get a rough idea about his health, thus decreasing costs for common man and number of tests require

* 1. **BACKGROUND OF PROBLEM**

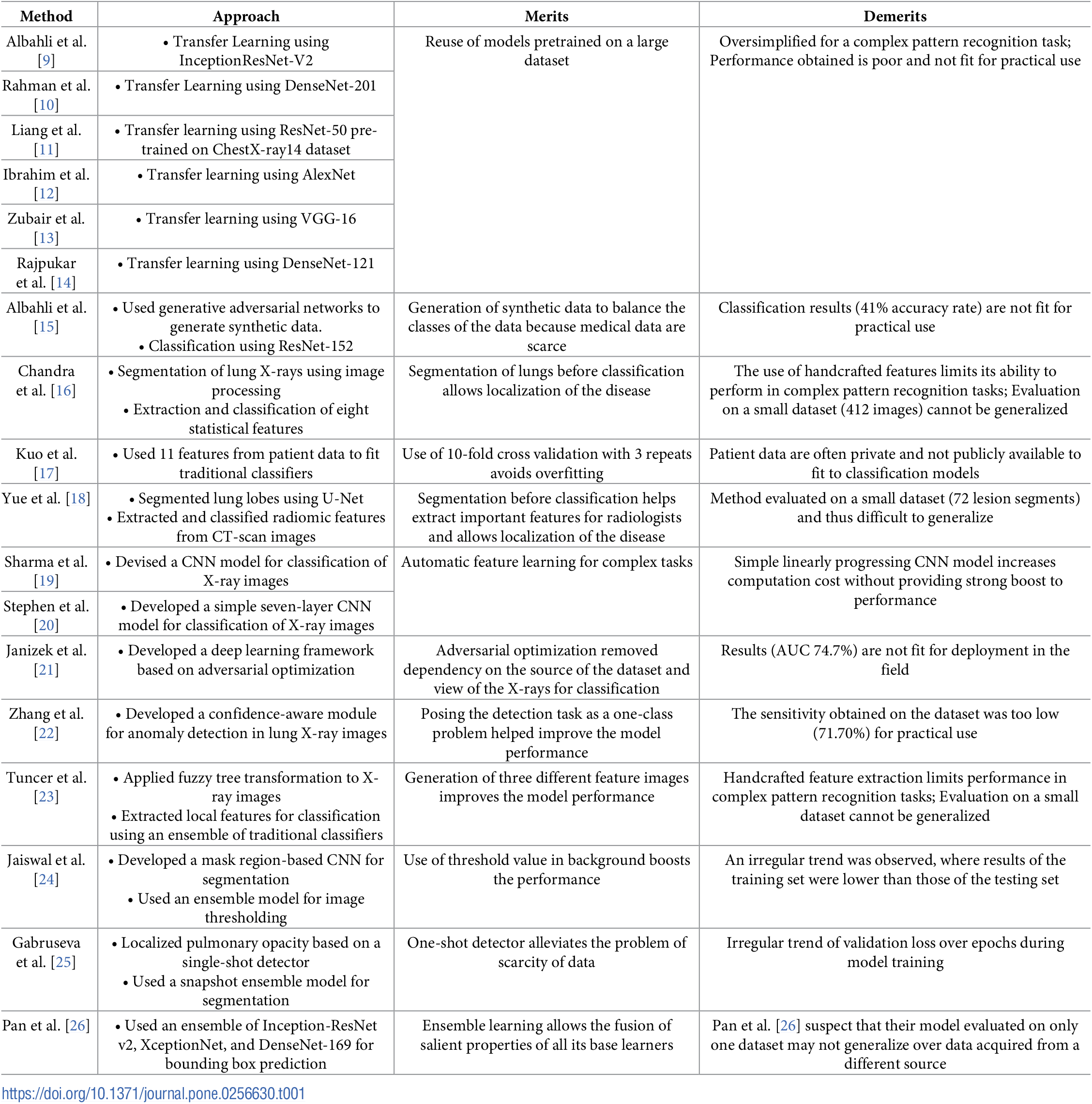
**1.2.1 CURRENT SYSYTEM**

Using the transfer learning paradigm, the well-known CheXNet model is capable of detecting the novel coronavirus pneumonia based on relevant and meaningful features with precise localization. COVID-CXNet is a step towards a fully automated and robust COVID-19 detection system.

To solve the data scarcity problem in biomedical image classification tasks, transfer learning, wherein knowledge gained from a large dataset is used to fine-tune the model on a current small dataset, is currently a frequently used approach. Recently, Ibrahim et al. [[12](https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0256630#pone.0256630.ref012)], applied purely transfer learning approaches in which different CNN models pre-trained on ImageNet] data are used for pneumonia classification. [Table 1](https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0256630#pone-0256630-t001) tabulates the development of the state of the art for the pneumonia detection problem.

**1.2.2 ISSUES IN CURRENT SYSTEM**

The following Table 1 shows the merits and issues faced in the current system for evaluation in Pneumonia Detection System.

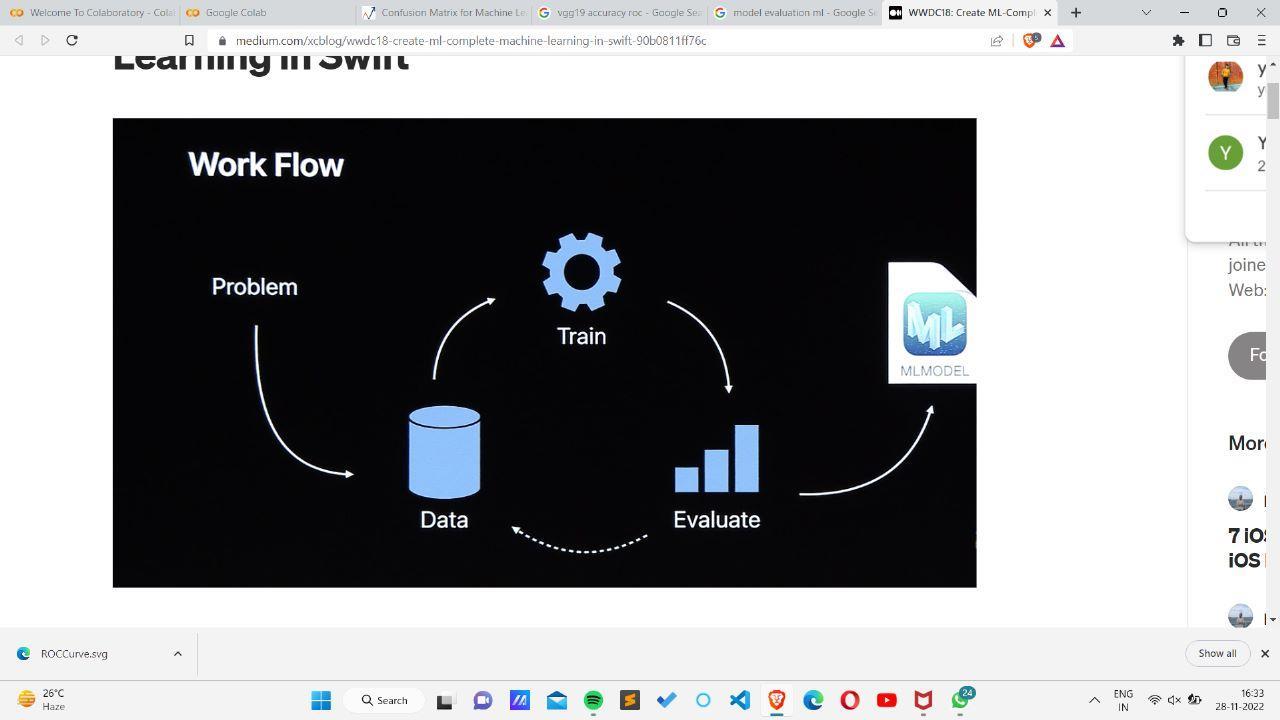


* 1. **PROBLEM STATEMENT**

By efficiently training through a relatively small image set, our fine-tuned models show high performance in the classification of COVID-19 pneumonia. Our conviction is that the proposed computer-aided diagnosis mechanism could outstandingly improve the diagnosis of COVID-19 cases. This is very helpful in a pandemic, especially when the available health resources do not match.

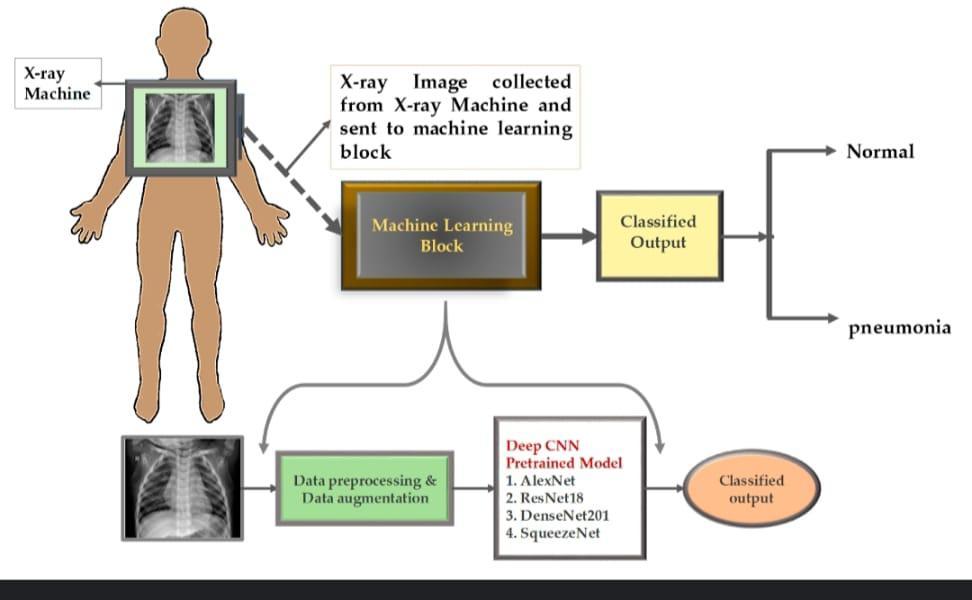
**1.4 PROPOSED WORK**

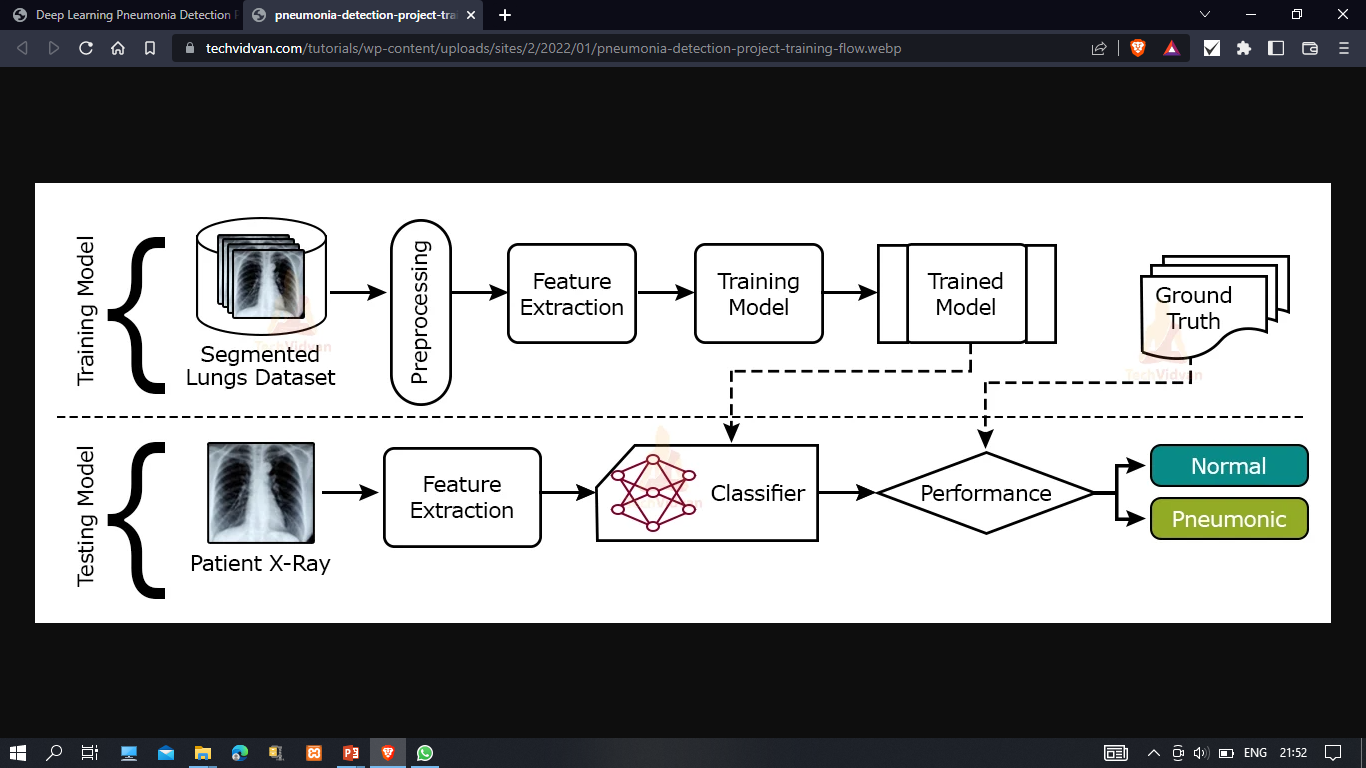
The proposed system COV-VGX extracts distinct features from chest X-ray images ,which takes an image and predicts whether the X-ray image is COVID-19 or pneumonia infected, normal cases, or not. Hence, the system proposes a binary classifier that decides between COVID-19 and pneumonia cases. After collecting the dataset from different sources, the dataset is pre-processed. Using transfer learning, pertained model VGG-16 is trained for the model evaluation. Few new layers are added with the base model to avoid the overfitting problem of training the model. Description of datasets Many available datasets of X-ray images from normal people and pneumonia infection cases are identified. Data augmentation helped the proposed model to perform better with exceptional features of the training images. Convolutional neural network (VGG-16) The features of the image are extracted through a series of convolutional layers. Using transfer learning Transfer learning is a method where a model trained for one classification problem is used in training for another classification problem. To evaluate the performance of COV-VGX, several performance metrics are used. Both the multiclass and binary classifiers are evaluated separately concerning accuracy, precision, recall, and F1-score. The metrics are evaluated separately for each class label. To evaluate these metrics, four basic terms are considered: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). Accuracy=TP + TNTP + TN + FP + FN 0.0<Accuracy

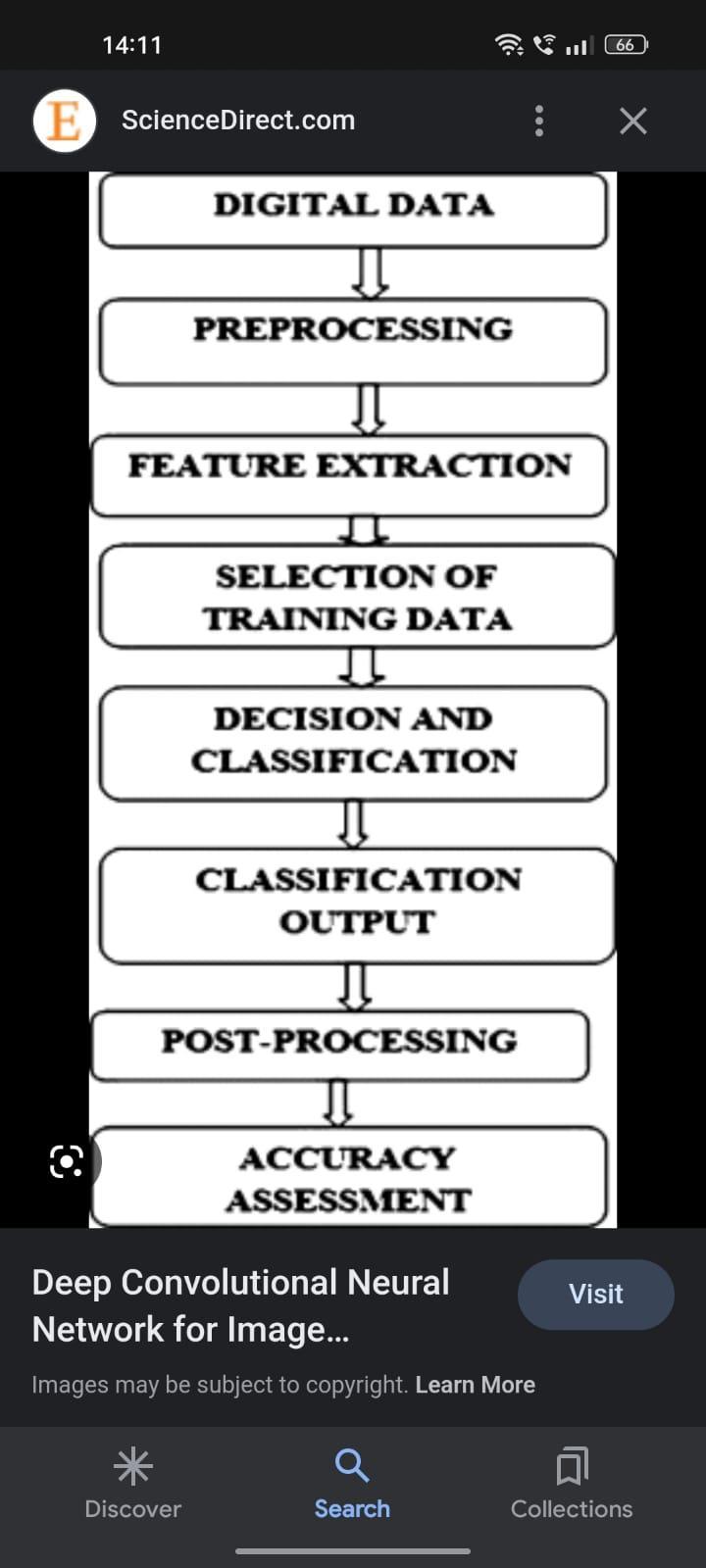
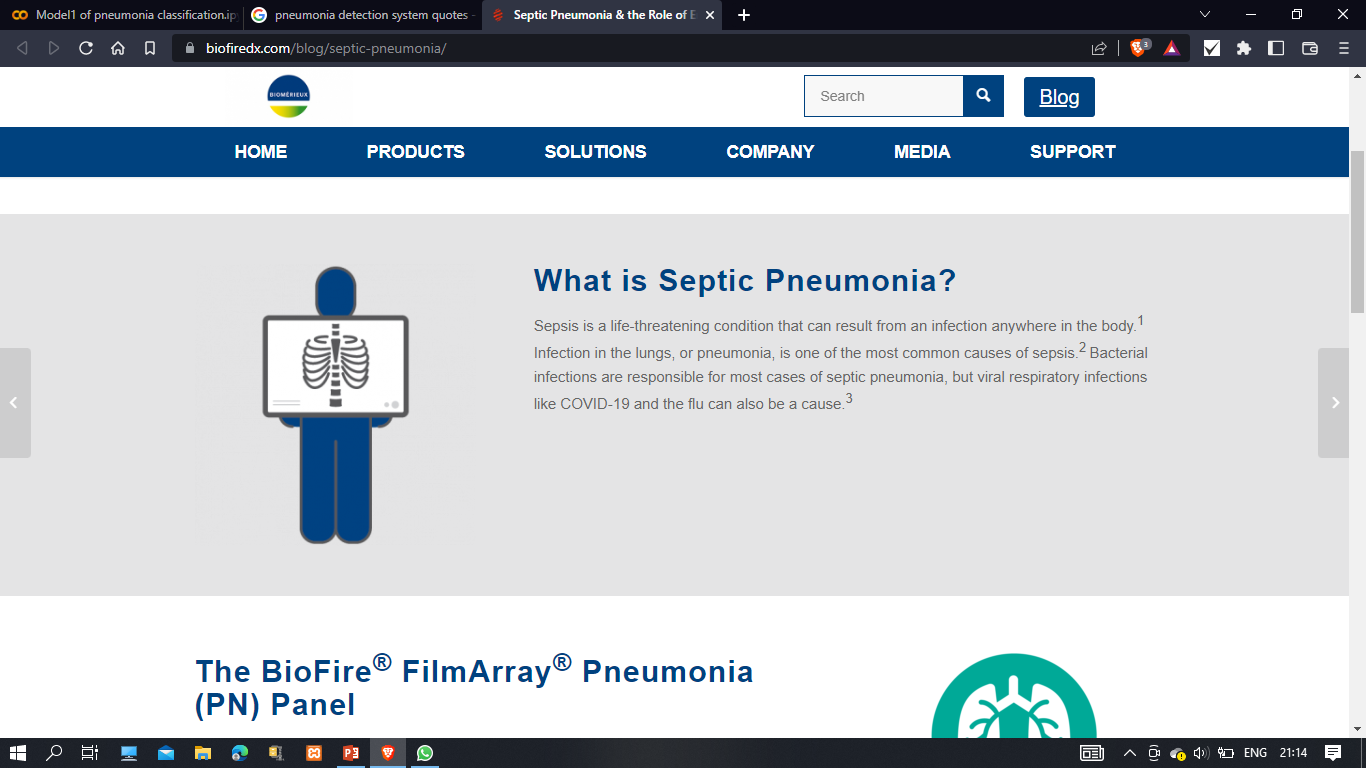


**CHAPTER 2 – DESIGN METHODOLOGY**

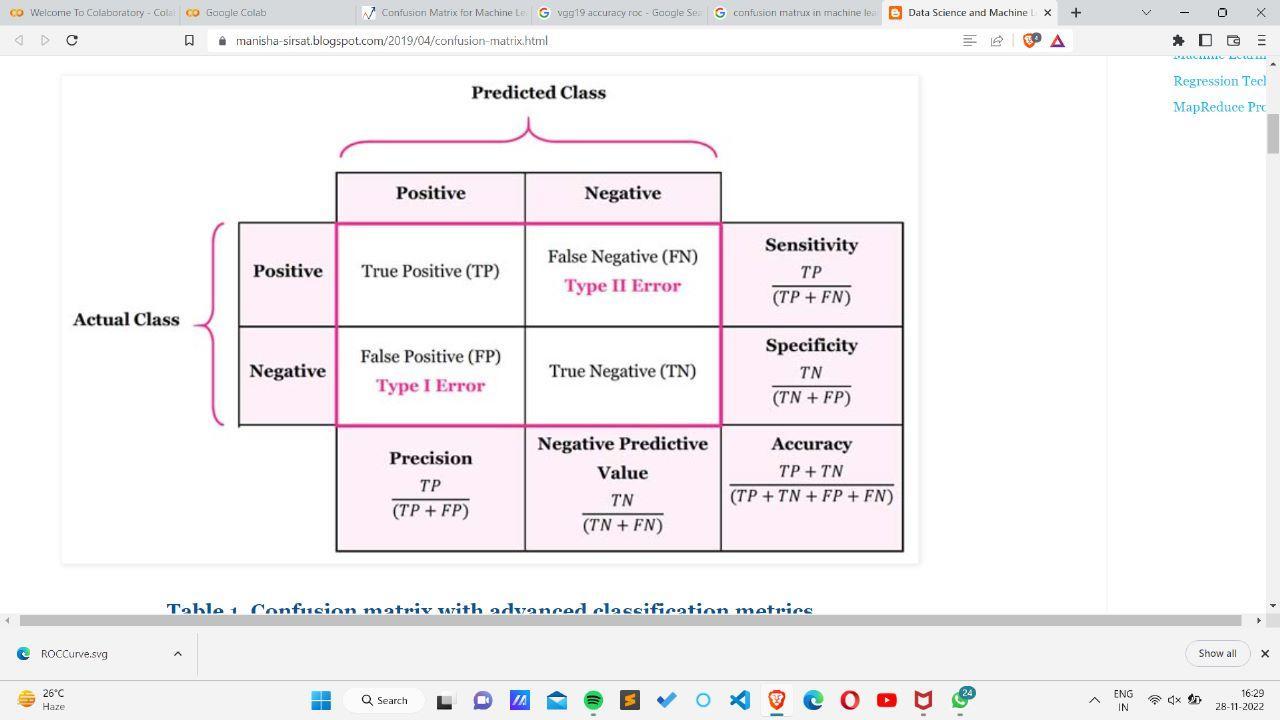
Due to the success of deep learning algorithms in analyzing medical images, Convolutional Neural Networks (CNNs) have gained much attention for disease classification.We analytically determine the optimal CNN model for the purpose. Statistical results obtained demonstrates that pre-trained CNN models employed along with supervised classifier algorithms can be very beneficial in analyzing chest X-ray images, specifically to detect Pneumonia.



****

** **

**MODEL EVALUATION**

****

**CHAPTER-3 IMPLEMENTATION**

import matplotlib.pyplot as plt

import numpy as np

import PIL

import tensorflow as tf

from tensorflow import keras

from tensorflow.keras import layers

from tensorflow.keras.models import Sequential

import numpy as np

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import tensorflow as tf

from keras import Sequential

from tensorflow.keras.layers import \*

from tensorflow.keras.models import \*

import pathlib

dataset\_url = "https://storage.googleapis.com/download.tensorflow.org/example\_images/flower\_photos.tgz"

data\_dir = tf.keras.utils.get\_file('flower\_photos', origin=dataset\_url, untar=True)

data\_dir = pathlib.Path(data\_dir)

import pathlib

data\_dir='/content/drive/MyDrive/ColabNotebooks/PNEUMONIA DATASET/chest\_xray/test'

data\_dir = pathlib.Path(data\_dir)

from google.colab import drive

drive.mount('/content/drive')

%cd drive/'My Drive/ColabNotebooks/PNEUMONIA DATASET'/

image\_count = len(list(data\_dir.glob('\*/\*.jpeg')))

print(image\_count)

### Create a dataset

Define some parameters for the loader:

batch\_size·=·32

img\_height·=·180

img\_width·=·180

test\_ds = tf.keras.utils.image\_dataset\_from\_directory(

  data\_dir,

  validation\_split=0.2,

  subset="training",

  seed=123,

  image\_size=(img\_height, img\_width),

  batch\_size=batch\_size)

val\_ds = tf.keras.utils.image\_dataset\_from\_directory(

  data\_dir,

  validation\_split=0.2,

  subset="validation",

  seed=123,

  image\_size=(img\_height, img\_width),

  batch\_size=batch\_size)

class\_names = test\_ds.class\_names

print(class\_names)

#visulaise the plot

import matplotlib.pyplot as plt

plt.figure(figsize=(10, 10))

for images, labels in test\_ds.take(1):

  for i in range(9):

    ax = plt.subplot(3, 3, i + 1)

    plt.imshow(images[i].numpy().astype("uint8"))

    plt.title(class\_names[labels[i]])

    plt.axis("off")

for image\_batch, labels\_batch in test\_ds:

  print(image\_batch.shape)

  print(labels\_batch.shape)

  break

## Configure the dataset for performance

AUTOTUNE = tf.data.AUTOTUNE

train\_ds = test\_ds.cache().shuffle(1000).prefetch(buffer\_size=AUTOTUNE)

val\_ds = val\_ds.cache().prefetch(buffer\_size=AUTOTUNE)

normalized\_ds = test\_ds.map(lambda x, y: (normalization\_layer(x), y))

image\_batch, labels\_batch = next(iter(normalized\_ds))

first\_image = image\_batch[0]

# Notice the pixel values are now in `[0,1]`.

print(np.min(first\_image), np.max(first\_image))

model\_1=base\_model3=tf.keras.applications.ResNet101(

    include\_top=False,

    weights="imagenet",

    input\_tensor=None,

    input\_shape=(180,180,3),

    pooling=None,

    classes=1000,

)

for layer in base\_model3.layers:

    layer.trainable = False

model = Sequential()

model.add(base\_model3)

model.add(GaussianNoise(0.25))

model.add(GlobalAveragePooling2D())

model.add(Dense(1024,activation='relu'))

model.add(BatchNormalization())

model.add(GaussianNoise(0.25))

model.add(Dropout(0.25))

model.add(Dense(4, activation='sigmoid'))

model.summary()

### Compile the model

model.compile(optimizer='adam',

              loss=tf.keras.losses.SparseCategoricalCrossentropy(from\_logits=True),

              metrics=['accuracy'])

### Train the model

acc = history.history['accuracy']

val\_acc = history.history['val\_accuracy']

loss = history.history['loss']

val\_loss = history.history['val\_loss']

epochs\_range = range(epochs)

plt.figure(figsize=(8, 8))

plt.subplot(1, 2, 1)

plt.plot(epochs\_range, acc, label='Training Accuracy')

plt.plot(epochs\_range, val\_acc, label='Validation Accuracy')

plt.legend(loc='lower right')

plt.title('Training and Validation Accuracy')

plt.subplot(1, 2, 2)

plt.plot(epochs\_range, loss, label='Training Loss')

plt.plot(epochs\_range, val\_loss, label='Validation Loss')

plt.legend(loc='upper right')

plt.title('Training and Validation Loss')

plt.show()

## Data augmentation

data\_augmentation = keras.Sequential(

  [

    layers.RandomFlip("horizontal",

                      input\_shape=(img\_height,

                                  img\_width,

                                  3)),

    layers.RandomRotation(0.1),

    layers.RandomZoom(0.1),

  ]

## Predict on new data

Use your model to classify an image that wasn't included in the training or validation sets.

#sunflower\_url = "https://drive.google.com/file/d/1g974fHANqFg6UYQZ5uAjWHqOWbS71nbu/uc?usp=share\_link"

sunflower\_path = "/content/drive/MyDrive/ColabNotebooks/PNEUMONIA DATASET/chest\_xray/test/NORMAL/IM-0001-0001.jpeg"

#tf.keras.utils.get\_file('Red\_sunflower', origin=sunflower\_url)

img = tf.keras.utils.load\_img(

    sunflower\_path, target\_size=(img\_height, img\_width)

)

img\_array = tf.keras.utils.img\_to\_array(img)

img\_array = tf.expand\_dims(img\_array, 0) # Create a batch

predictions = model.predict(img\_array)

score = tf.nn.softmax(predictions[0])

print(

    "This image most likely belongs to {} with a {:.2f} percent confidence."

    .format(class\_names[np.argmax(score)], 100 \* np.max(score))

)

base\_model3.save('PNEUMONIA DATASET\_Ria.h5')

new\_model = tf.keras.models.load\_model('PNEUMONIA DATASET\_Ria.h5')

## Use TensorFlow Lite

# Convert the model.

converter = tf.lite.TFLiteConverter.from\_keras\_model(model\_1)

tflite\_model = converter.convert()

# Save the model.

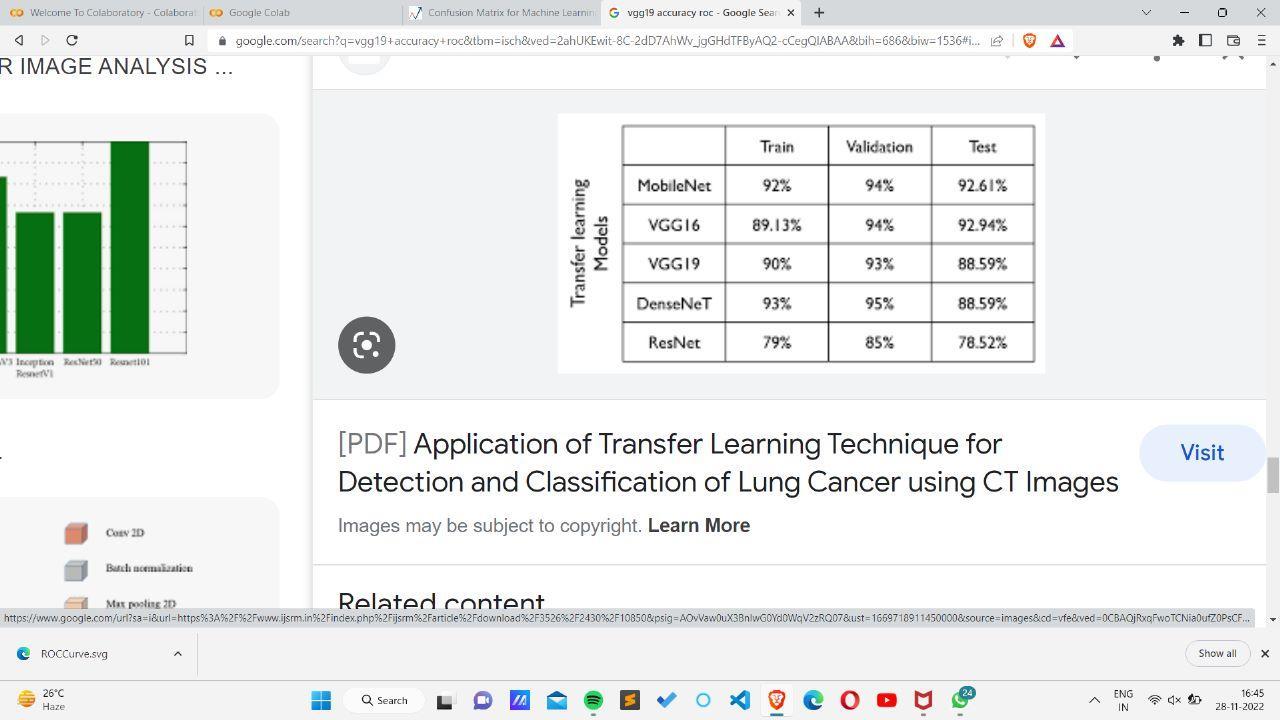
with open('model.tflite', 'wb') as f:

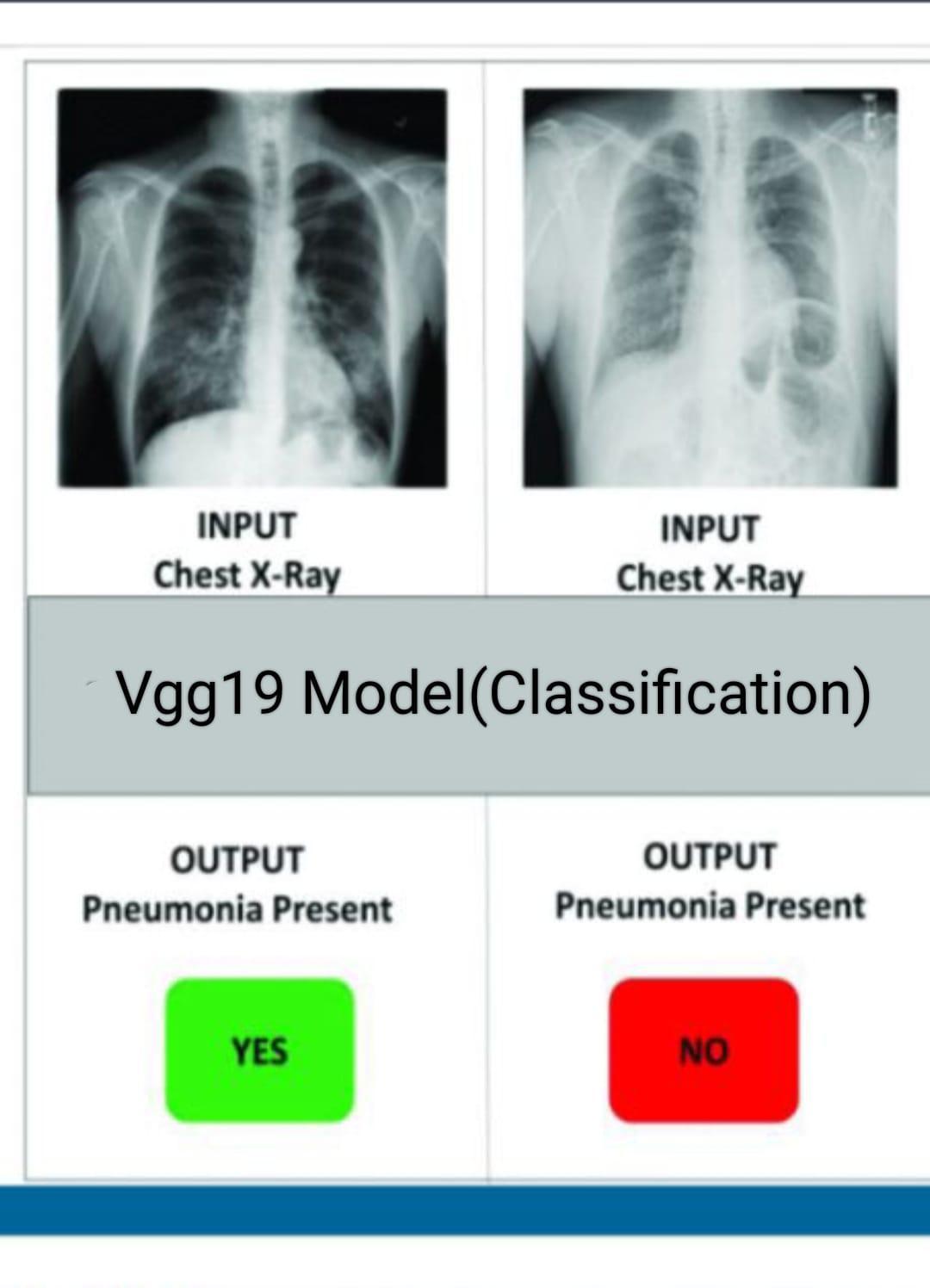
  f.write(tflite\_model)

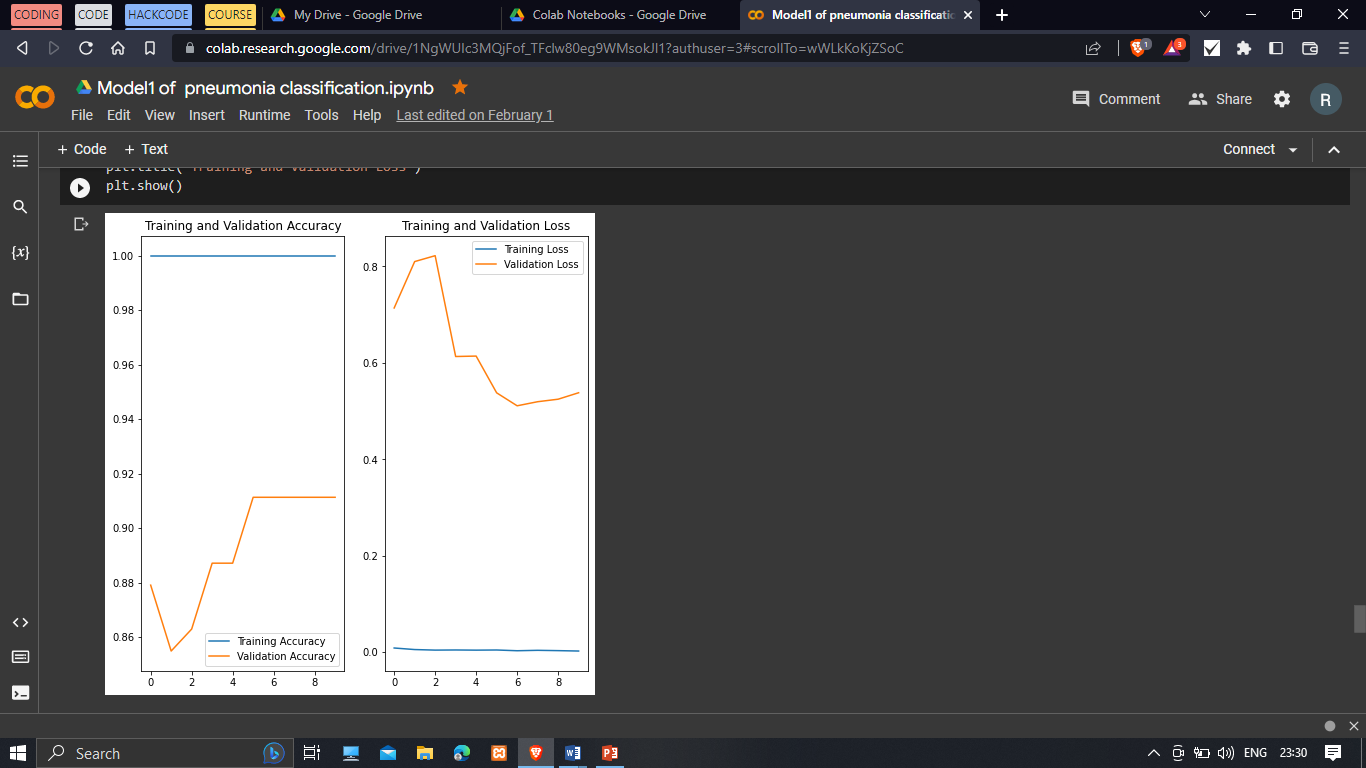
**CHAPTER -4 TESTING RESULT AND ANALYSIS**

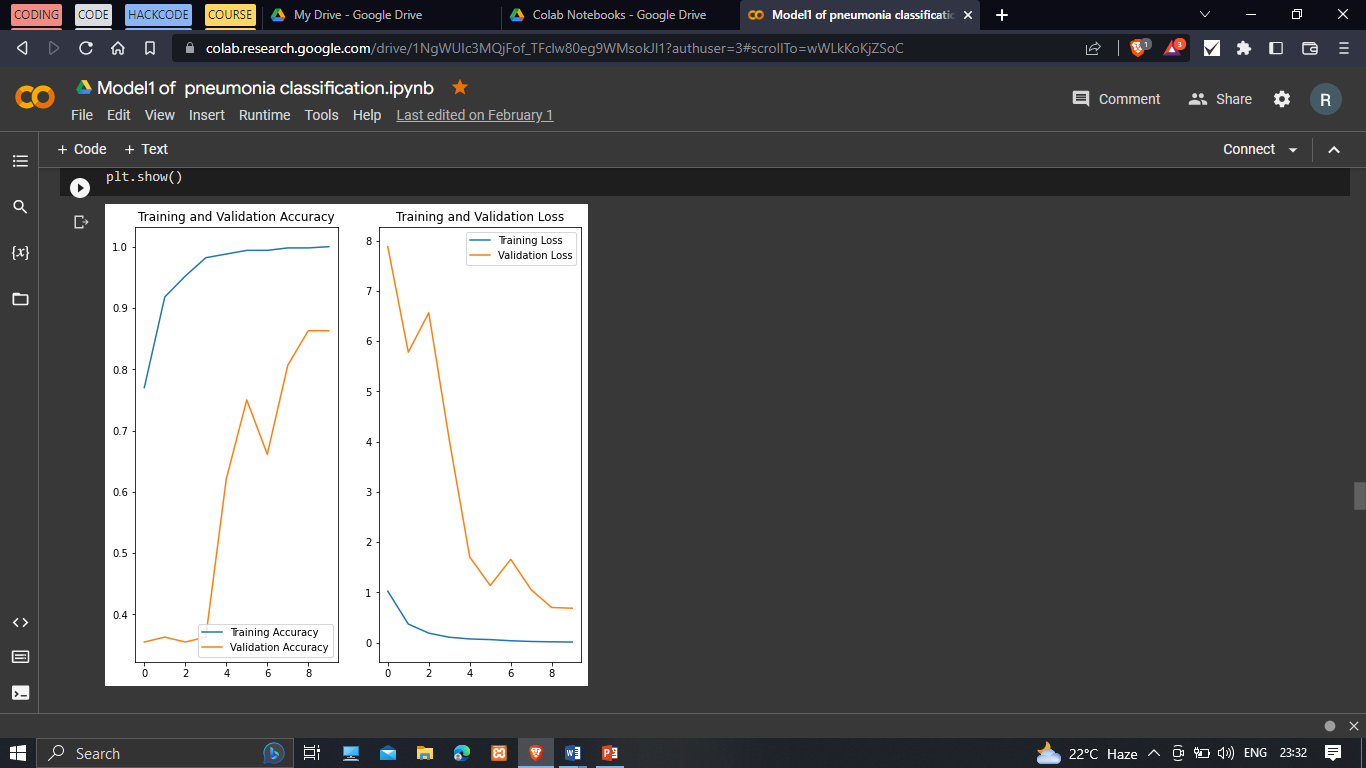
On the basis of comparison with other Models

Vgg19 found to be most efficient in predicting the result with 90% accuracy.

****

****

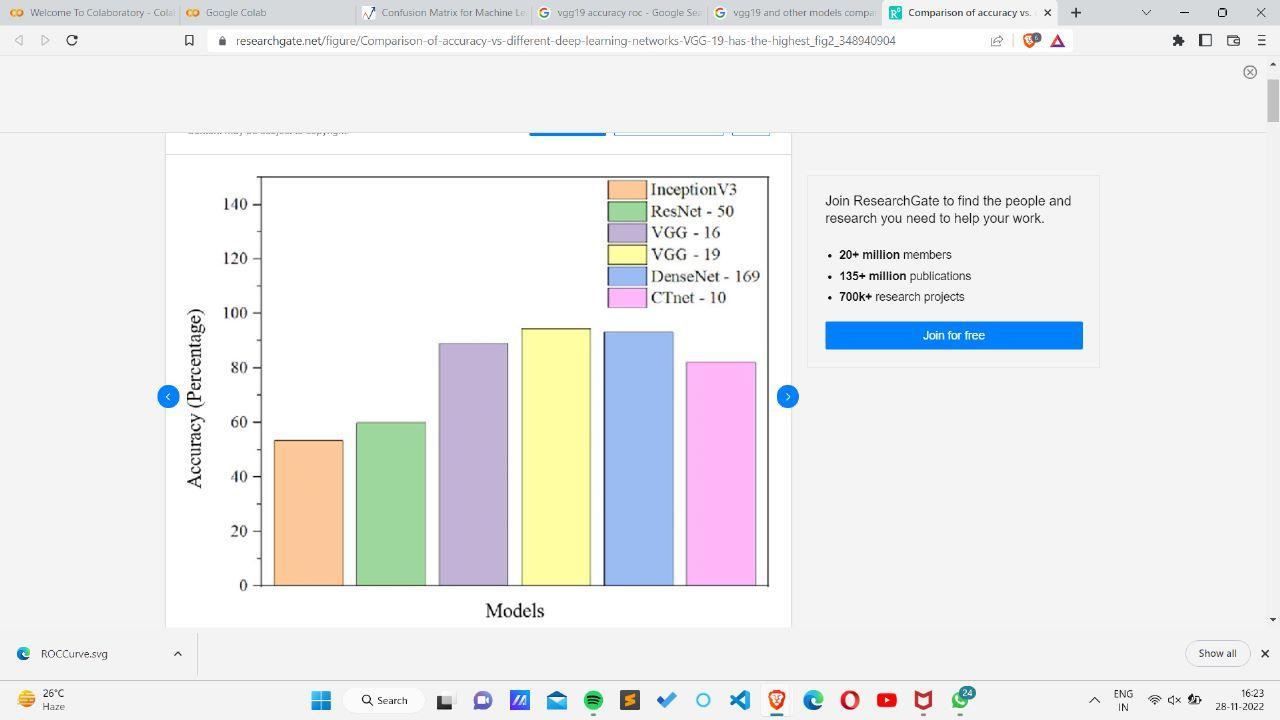
 **AFTER OVERFITTING**

 **BEFORE OVERFITTING**

**CHAPTER -4 CONCLUSION AND FUTURE ENHANCEMENTS**

The highest accuracy reported in the above-mentioned proposed model Vgg19 in classifying the normal vs. pneumonia patients using X-ray images using deep learning algorithms is 90.84%.

Successfully building Deep Learning diagnosis systems while maintaining people’s privacy is a frontier that the healthcare industry has already established.



A number of experiments were performed on a dataset of X ray images and evaluated by a set of metrics. All results confirmed the effectiveness and efficiency of the proposed approach.

**FUTURE SCOPE**

In future work, a large-sized dataset of different lung diseases will be collected to train the model to improve the healthcare sectors. Moreover, an energy evaluation measure will be used to compute a real result of the energy consumption of the proposed approach.

Early detection of pneumonia increases the survival rate of patients. It is one of the medical systems that help to train radiologists to improve their accuracy of diagnosing pneumonia.

A approach for the automatic detection of pneumonia in energy-efficient medical systems is important:

--To improve the quality

--Reduced cost

--Time response

Data augmentation and transfer learning have also been used to tackle the obstacle of the insufficient training dataset. Different scores, such as recall, precision and accuracy, were computed to prove the robustness of the model. The proposed model attained an accuracy of 98.14%, a high AUC score of 99.71 and an F1 score of 98.3. The future works involve developing an algorithm which can localize the parts of the lung affected by pneumonia.

**REFERENCES**

DATASET

[**https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia**](https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia)

MODEL

<https://keras.io/api/applications/>

PLATFORM

**colab.research.google.com**

IMAGES AND OTHER INFORMATION

**techvidvan.com**

**towardsdatascience.com**

[**www.researchgate.net**](http://www.researchgate.net/)

[**www.foreseemed.com**](http://www.foreseemed.com/)

**github.com**

<https://colab.research.google.com/drive/1T7t324thWAqpWkNLoaDsgCz2tjvlJRy3?authuser=3>

<https://en.wikipedia.org/wiki/Mathematical_morphology>

<https://www.ijert.org/research/the-importance-of-contrast-enhancement-in-medical-images-analysis-and-diagnosis-IJERTV7IS120006.pdf>

mi-research.net

[www.sciencedirect.com](http://www.sciencedirect.com)

[www.researchgate.net](http://www.researchgate.net)