

**CSA203 Artificial Intelligence and Machine Learning**

**Final Report On Covid Detection**

Group members:

Pema Chozom, 12210024

Pelden Wangchuk, 12210023

Parjeet Mongar, 12210022

Jigme Tenzin, 12210011

Jigme Wangyel Wangchuk, 12210012

**Tutor :** Mrs. Nima Dema

[**Abstract 3**](#_1cel0ddjoixt)

[**Introduction 3**](#_ekycu637s9h4)

[**Aim 3**](#_bzjc71ekyt9)

[**Goal 3**](#_i4ba6v8gdnsg)

[**Problem Statement 4**](#_jsuel6ysn712)

[**Scope 4**](#_psxtadf5g1fa)

[a. User scope 4](#_ckz7acozom89)

[b. System scope 4](#_twu72b7f5umy)

[**Literature review(related work) 4**](#_g1o726i0nvbz)

[a. Automatic detection of COVID-19 from chest radiographs using deep learning by Pandit M. et al.(2022) 4](#_w0esigc1ft18)

[b. Artificial Intelligence Applied to Chest X-Ray Images for the Automatic Detection of COVID-19 4](#_i44ot7ejudt8)

[c. Automatic detection from X-ray images utilizing transfer learning with convolutional neural networks 4](#_v1magu9zjnzo)

[**Methodology 5**](#_emlm98r0cnk2)

[a. Data Collection and Preprocessing 5](#_2qxzancrv5el)

[b. Data Split 6](#_iwaqx64vaw2r)

[c. Model training 7](#_gj4677pk2dbz)

[I. Logistic regression 7](#_vgc4sj7jc4pu)

[II. Naive Bayes 8](#_cp1cspx76g2o)

[III. Decision Tree 8](#_2v1rkidizh3e)

[IV. SVM 8](#_qjqpd7pj004s)

[GuassianNB pipeline 8](#_gtr9tfc1sct3)

[I. GradientBoostingClassifier 9](#_cyl0djmevsdu)

[II. BaggingClassifier 10](#_nszmmvcbp0nw)

[**Conclusion 10**](#_8ahapthrr4r0)

# **Abstract**

The COVID-19 pandemic poses global health risks; developing precise diagnostic techniques is crucial for early detection and containment. A machine learning method is proposed for X-ray image analysis. By training a large dataset of labeled X-ray images (supervised), we aim to create a predictive model capable of accurately classifying X-ray scans as either COVID-19 positive or negative.To do this, we first got a wide range of data, including X-ray pictures of COVID-19 patients as well as healthy individuals. Our machine learning effort offers a reliable method for X-ray image-based COVID-19 infection prediction. The created model demonstrates its potential as an efficient diagnostic tool that can help medical professionals quickly identify infected individuals. To improve the model's resilience and usability in actual clinical situations, additional improvements and validation studies are required.

# **Introduction**

Global health, economy, and communities have all seen a record-breaking effect as a result of the COVID-19 pandemic outbreak. There is an urgent need for precise and effective diagnostic techniques to locate affected people and launch necessary therapies since the virus spreads quickly still. In this situation, medical imaging methods including X-ray imaging have shown possibilities for assisting in the diagnosis of COVID-19. The state of a patient's lungs, which are frequently impacted by respiratory disorders like COVID-19, can be understood quite well using X-ray imaging. When it comes to the presence or absence of COVID-19 infection, the visual patterns shown in X-ray pictures can provide valuable information. Medical experts must manually evaluate these photos, which can be time-consuming and prone to inaccuracy. Therefore, using machine learning techniques to automate and improve the accuracy of COVID-19 diagnosis based on X-ray pictures has drawn a lot of attention.

# **Aim**

The purpose of this project is to make the viral detection process more accessible and effective, particularly in areas with limited resources where RT-PCR testing might not be readily available.

# 

# **Goal**

Develop and deploy a low-cost, rapid, and accurate viral detection system that can be easily utilized in resource-constrained areas to facilitate early identification and containment of viral outbreaks.

# **Problem Statement**

The outbreak of COVID-19 pandemic has caused a global health crisis and led to a significant increase in demand for the detection of the virus. The current testing methods such as RT-PCR and antigen tests are limited by their availability, cost, and time. Therefore, there is a need for an efficient and reliable method for COVID-19 detection that can be easily accessible and scalable.

# **Scope**

## User scope

Targeted users are the people seeking medical checkup for covid

## 

## System scope

Using the X-ray image of chest, it can predict whether the user is suffering from covid or

not

# **Literature review(related work)**

## 

## Automatic detection of COVID-19 from chest radiographs using deep learning by Pandit M. et al.(2022)

Pandit M. et al. (2022) studied automatic detection of COVID-19 from chest radiographs using deep learning. They used a dataset of 1428 chest radiographs from healthy individuals and patients with common bacterial pneumonia. The pre-trained VGG16 model successfully trained the network on small chest radiographs, achieving 96% accuracy for COVID and non-COVID classes and 92.5% accuracy for COVID, non-COVID pneumonia, and normal classes.

## 

## Artificial Intelligence Applied to Chest X-Ray Images for the Automatic Detection of COVID-19

The article discusses the training of a Grad-CAM Convolutional Neural Network on 79,500 X-ray images from various sources, including 8500 COVID-19 examples. Three experiments were conducted using three preprocessing methods, achieving a classification accuracy of 91.5%. However, preprocessing is needed for lung segmentation.

## Automatic detection from X-ray images utilizing transfer learning with convolutional neural networks

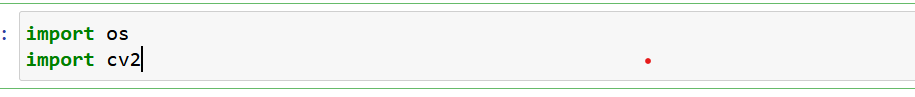
The study uses transfer learning with convolutional neural networks to automatically detect COVID-19 from X-ray images. The images are normalized to extract improved features, which are then used in deep learning-based image categorization algorithms. Five CNN systems were used, including VGG19, MobileNetV2, Inception, Xception, and InceptionResNetV. The MobileNetV2 net achieved the best results, with 96.78% and 94.72% accuracy in two- and three-class classifications, respectively.

# **Methodology**

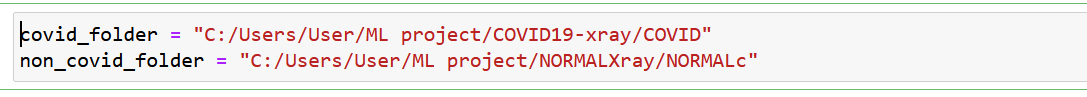
## Data Collection and Preprocessing

We collected a dataset from Kaggle which was collected by Anas AAbo, a teaching assistant from Alexandria, Alexandria Governorate, Egypt. This dataset was drawn from the sources like github and mendeley and contains images of covid and non-covid (normal) x-ray. The dataset contains 1161 files for covid and 3836 files. Clearly, we will be working on a binary-class classification problem. We used Grayscale pixel values as features.

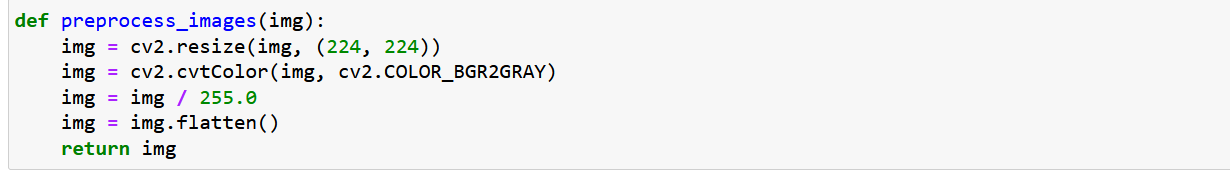
There were two folders, one for covid images and one for non-covid images. So, to load this dataset from our system we imported os, which is used for accessing the file system. And to perform image processing and computer vision tasks we imported cv2 (Computer vision uses image processing to recognize and categorize image data.).



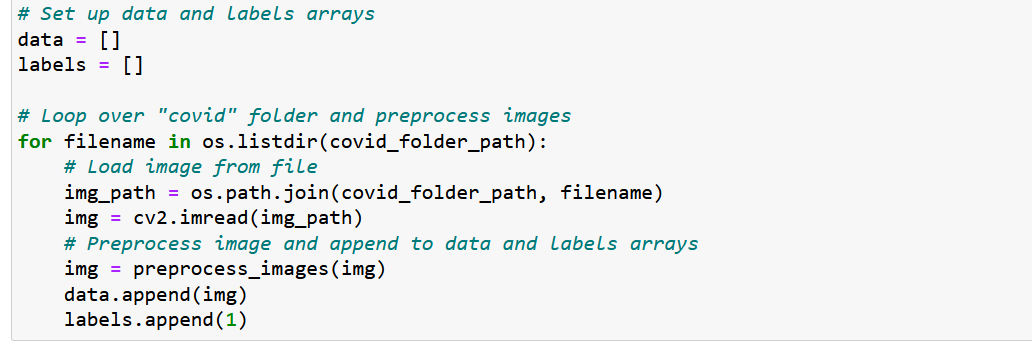
For the image processing we needed to load the files from the folder. So, we defined path to the Covid-19 X-ray folder and non-covid X-ray folder



For easier comparison of the results we defined the size of the images after preprocessing to standard size, that is, (244,244) pixel value. The image is converted to grayscale which normalizes the pixel value between 0 and 1. And image is standardized by dividing by 255, which is standard pixel value. Then the image is converted into 1D array .

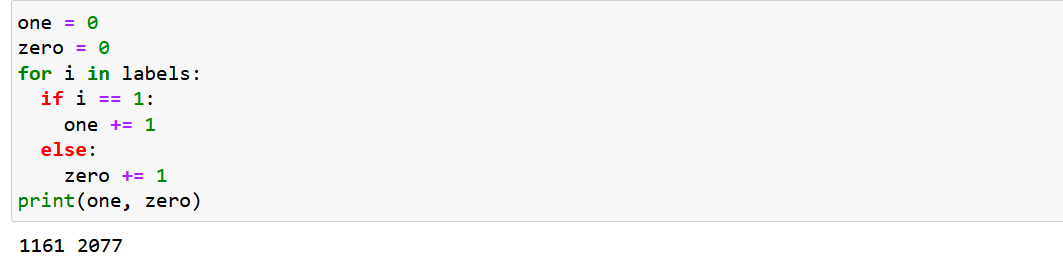


To load images from a folder i.e, covid and non-covid folder, and to preprocess the images, first we have created a list with the respective names. To get the list of all the files in the directory, we have used the os.listdir() function using a for loop to iterate each file in a directory. . And we appended the images to the data created earlier. And appended covid which is 1 and non-covid 0 in labels(target)

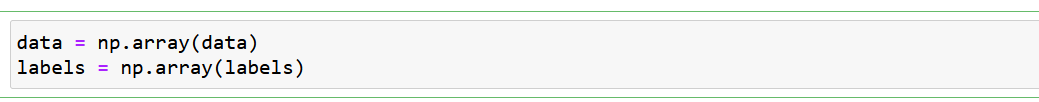




We counted the number of labels where there were found of label 0 and of label 1.

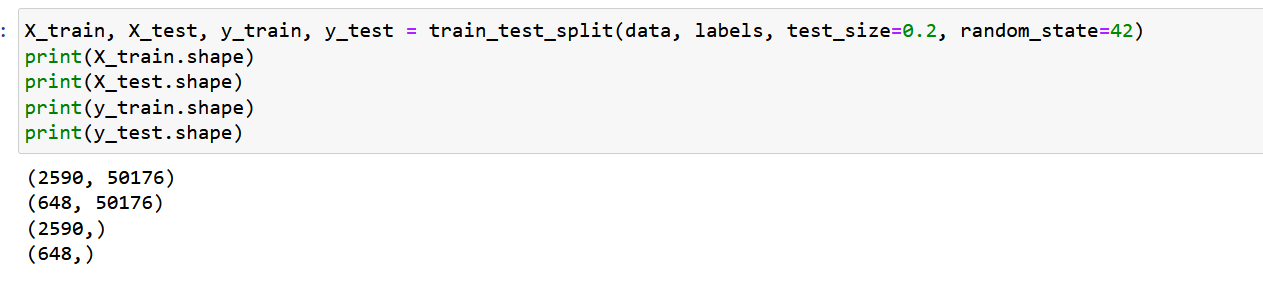


We converted the data into numpy array.



## Data Split

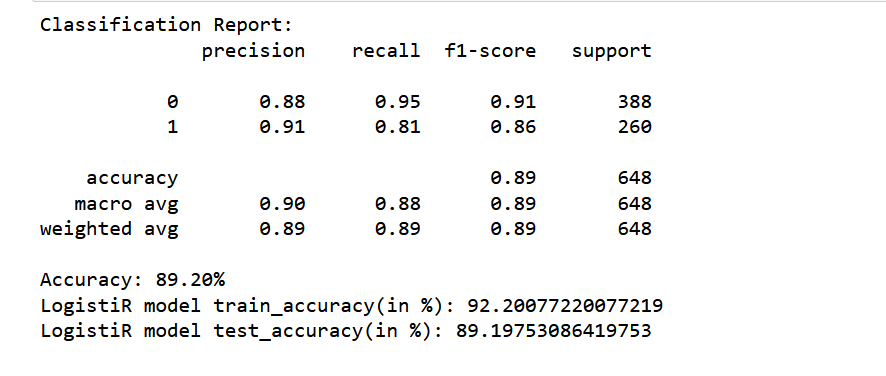
We split the data into train and test sets keeping “data” as features and “labels” as target. We kept 20% of the original data as a test set and the rest 80% as the training set which includes x\_train having 2590 rows and 50176 columns and y\_train having 2590 rows, and x\_test having 648 rows and 50176 columns and y\_test having 648 rows.



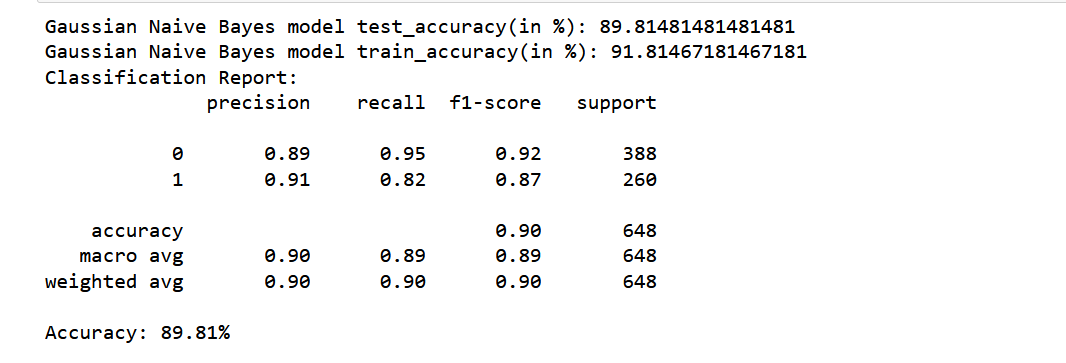
## Model training

We implemented total of 4 model and got the following score:

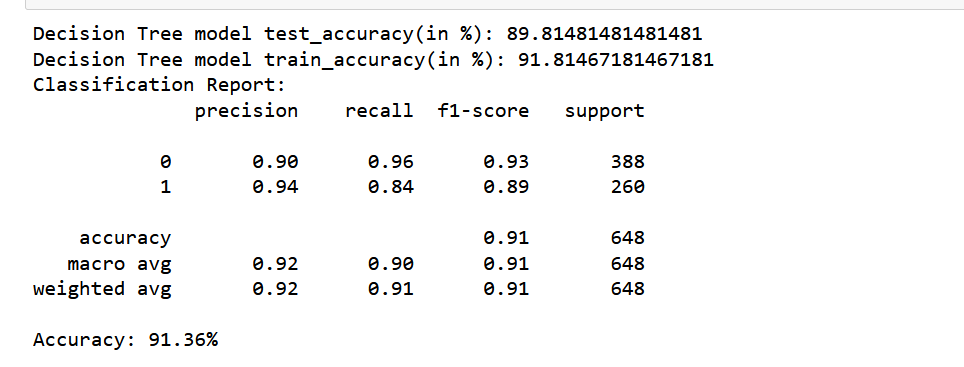
### I. Logistic regression



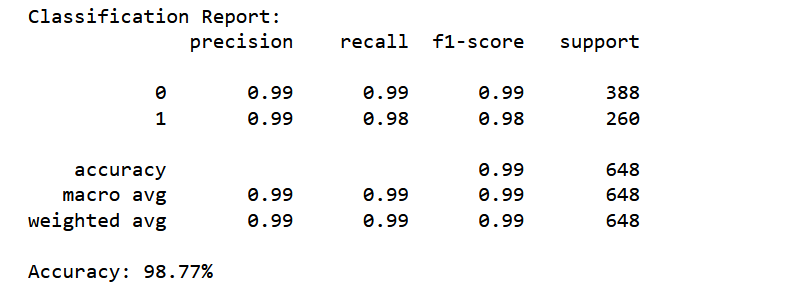
### II. Naive Bayes



### III. Decision Tree

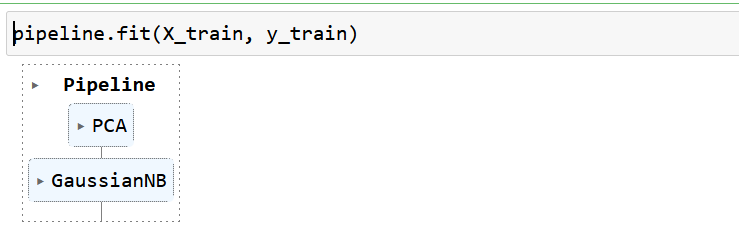


### IV. SVM

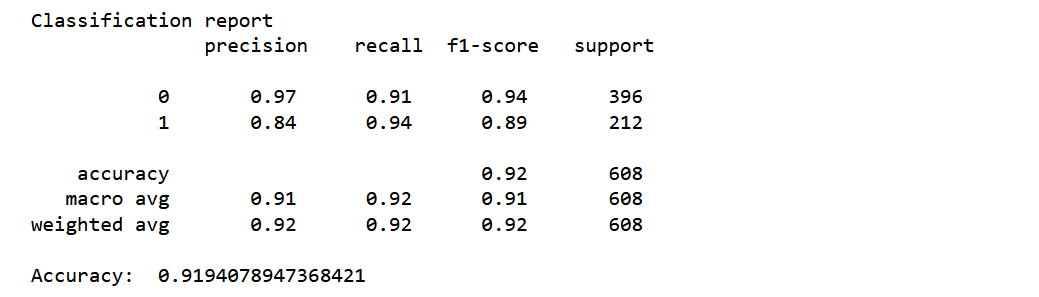


### GuassianNB pipeline

After that we created pipeline where GuassionNB was selected as estimator and PCA to reduce dimensionality:

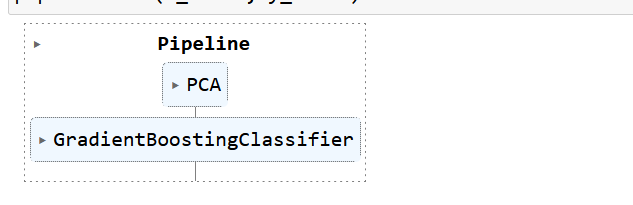


The performance of the pipeline model is shown below:

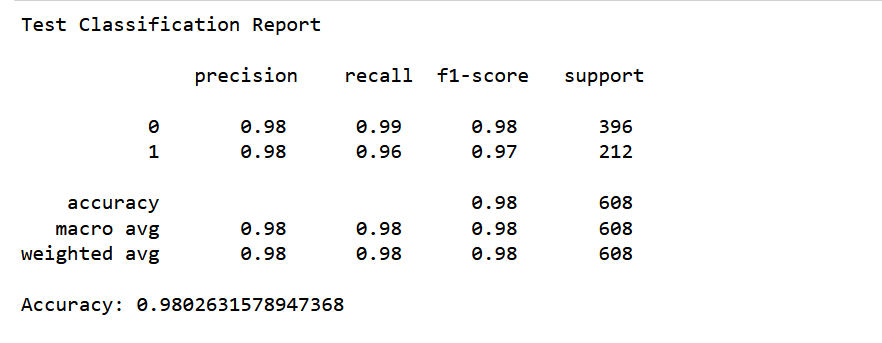


Since the model was not classifying non-covid images, it was only able to classify normal x-ray images. We created model using ensemble learning algorithm (GradientBoostingClassifier and BaggingClassifier). And created pipeline for these.

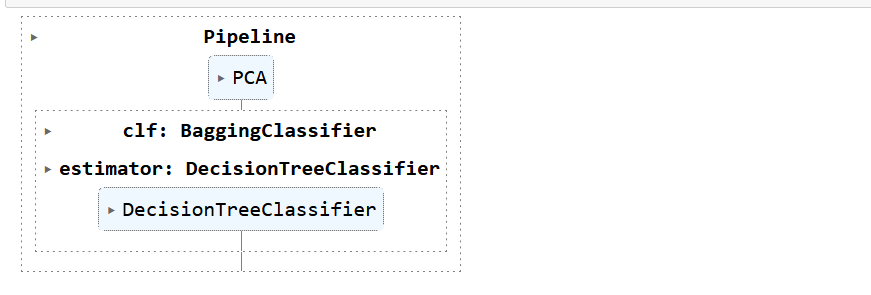
### GradientBoostingClassifier



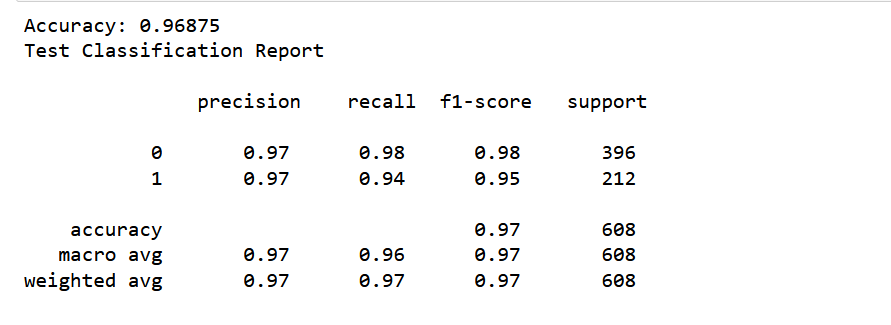
The performance is give below:



### BaggingClassifier



The performance for the above pipeline model



We chose BaggingClassifier since GradientBoostingClassifier doesn’t train our estimator. Created pickel model using the above pipeline model:



# 

# **Conclusion**

For our project, our final dataset dimension was reduced by using PCA to 300.and the labels include 1 for covid and 0 for non-covid x-ray images. We set out to evaluate and get insights from a significant amount of data in this dataset study (where data contains 3037 rows and 50176 columns). We used advanced machine learning techniques, exploratory data analysis, and data cleansing in our analysis. Among four estimator we chose Decision Tree as our final estimator, where ensemble BaggingClassifier is used and the accuracy score of 97% was achieved. To integrate our model with front-end we created pickel file and used flaskenv, where all necessary packages were installed and used accordingly.