PROJECT REPORT

CovidVision: Advanced COVID-19 Detection from Lung X-rays with Machine Learning or Deep Learnings

Team Id: NM2023TMID17493

Project Report

1. INTRODUCTION

1.1 Project Overview

CovidVision is an innovative project that aims to develop a state-of-the-art system for advanced COVID-19 detection from lung X-rays using machine learning or deep learning techniques. The project seeks to leverage the power of artificial intelligence to assist healthcare professionals in accurately identifying COVID-19 infection patterns in lung X-ray images. By providing an automated and reliable solution, CovidVision aims to contribute to the early detection and efficient management of COVID-19 cases.

1.2 Purpose

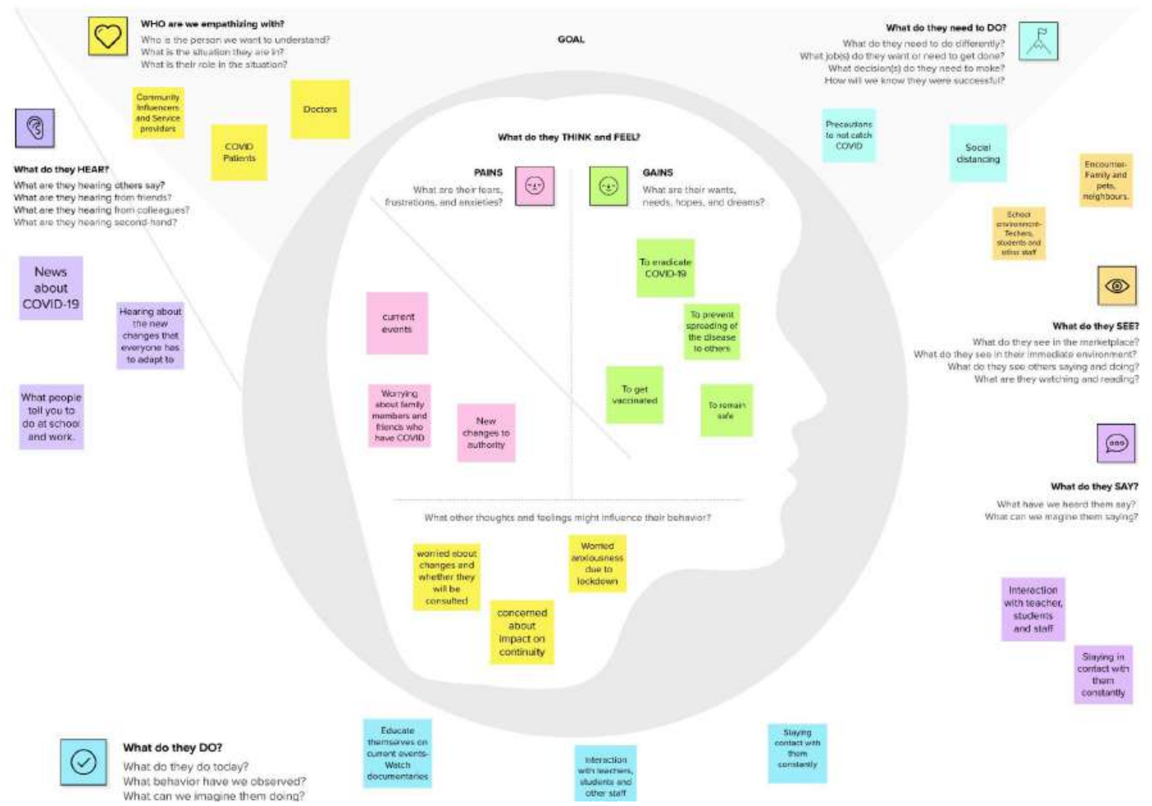
The purpose of CovidVision is to leverage advanced machine learning or deep learning techniques to improve the detection of COVID-19 from lung X-rays. By providing accurate and efficient diagnosis, it aims to support healthcare professionals, optimize resources, aid in public health measures, and drive advancements in the field of COVID-19 detection and management.

2. IDEATION & PROPOSED SOLUTION

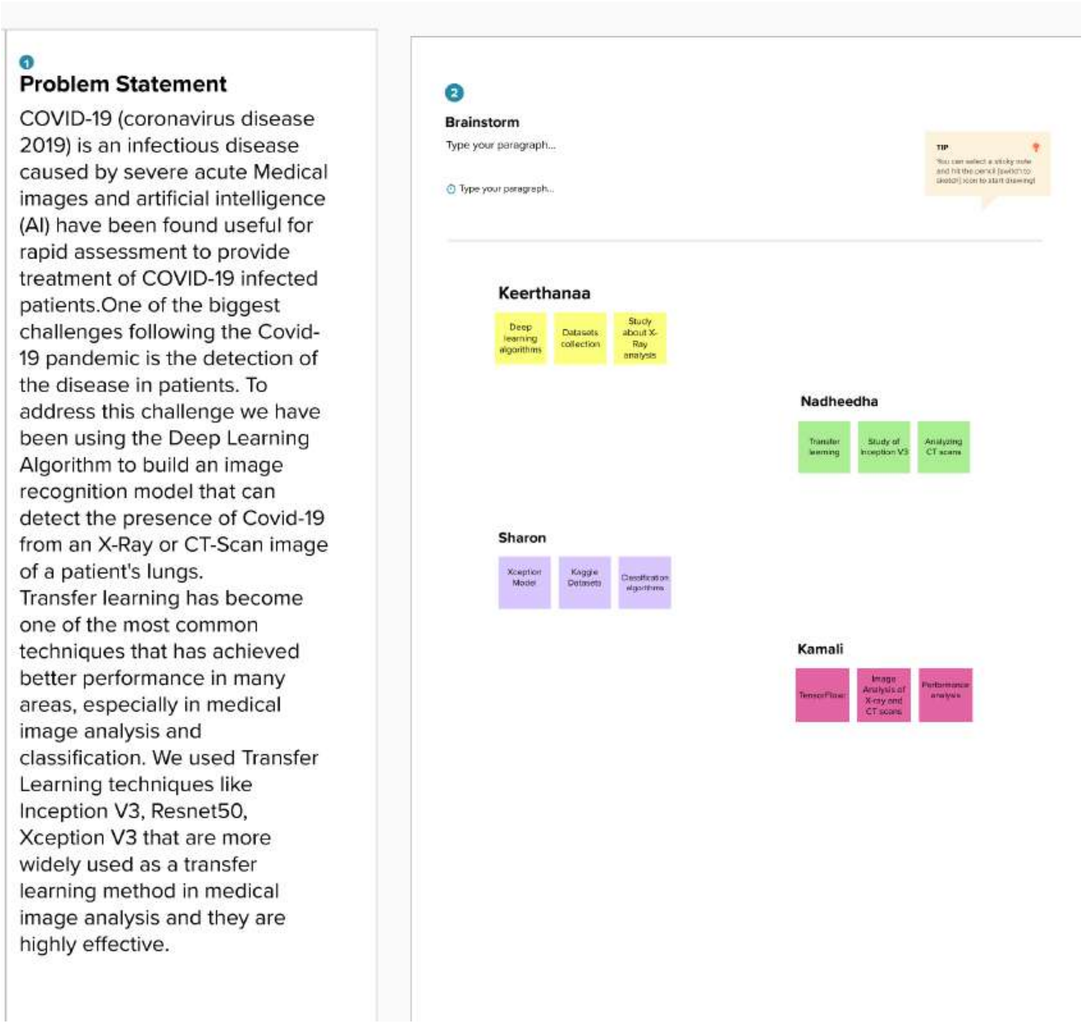
2.1 Problem Statement Definition

There is a critical need for a reliable and accurate automated system to detect and identify COVID-19 infection patterns from lung X-rays using machine learning or deep learning techniques. The current manual process of examining lung X-rays for COVID-19 diagnosis is time-consuming and prone to human error, leading to delays in treatment and resource allocation. Additionally, the increasing number of COVID-19 cases overwhelms healthcare systems, making it challenging for healthcare professionals to efficiently diagnose and manage patients.

2.2 Empathy Map Canvas



2.3 Ideation & Brainstorming



3

Group ideas

Take turns sharing your ideas while clustering similar or related notes as you go. Once all sticky notes have been grouped, give each cluster a sentence-like label. If a cluster is bigger than six sticky notes, try and see if you and break it up into smaller sub-groups.

20 minutes

TIP

Add customizable tags to sticky notes to make it easier to find, browse, organize, and categorize important ideas as themes within your mural.

The user interacts with the UI (User Interface) to choose the image. → The chosen image analyzed by the model which is integrated with flask application.The Xception Model analyzes the image, then the prediction is showcased on the Flask UI. → To accomplish this, we have to complete all the activities and tasks listed below → Data Collection.

Create a Train and Test path. → Data Pre-processing. → Import the required library → Configure ImageDataGenerator class

ApplyImageDataGenerator functionality to Trainset and Testset → Model Building → Pre-trained CNN model as a Feature Extractor → Adding Dense Layer

Configure the Learning Process → Train the model → Save the Model → Test the model

Application Building → Create an HTML file → Build Python Code

The main idea here is to analyze the image using Xception model and then to predict it on Flask UP. Then after collecting training and testing the data is pre-processed and the model is built. For feature Extraction, we use a pre-trained CNN model. Then we add a dense layer and configure the learning process then to train and save the model. Finally, the model is tested and the application is built and the python code is built

2.4 Proposed Solution

CovidVision proposes the development of an advanced COVID-19 detection system that leverages machine learning or deep learning techniques to accurately detect COVID-19 cases from lung X-rays. The proposed solution encompasses the following key components:

-

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	There is a critical need for a reliable and accurate automated system to detect and identify COVID-19 infection patterns from lung X-rays using machine learning or deep learning techniques. The current manual process of examining lung X-rays for COVID-19 diagnosis is time-consuming and prone to human error, leading to delays in treatment and resource allocation. Additionally, the increasing number of COVID-19 cases overwhelms healthcare systems, making it challenging for healthcare professionals to efficiently diagnose and manage patients.
2.	Idea / Solution description	CovidVision is an innovative project that aims to develop a state-of-the-art system for advanced COVID-19 detection from lung X-rays using machine learning or deep learning techniques. The project seeks to leverage the power of artificial intelligence to assist healthcare professionals in accurately identifying COVID-19 infection patterns in lung X-ray images. By providing an automated and reliable solution, CovidVision aims to contribute to the early detection and efficient management of COVID-19 cases.
3.	Novelty / Uniqueness	Advanced Detection Capability Interpretability and Explainability Integration with Healthcare Systems CovidVision addresses ethical considerations such as fairness and privacy. CovidVision emphasizes continuous improvement by facilitating periodic model updates and retraining. By incorporating new labeled data and adapting to evolving COVID-19 patterns

4.	Social Impact / Customer Satisfaction	CovidVision can assist healthcare professionals in making informed decisions regarding diagnosis, treatment, and patient management. Enabling faster and more accurate detection of COVID-19 cases. Accurately distinguishing between COVID-19 positive cases, non-COVID lung diseases, and healthy individuals. The early detection and management of COVID-19 cases facilitated by CovidVision can lead to potential cost savings in the healthcare system. CovidVision can have a global impact by extending its reach to regions with limited access to expert radiologists or medical facilities. The development and deployment of CovidVision can enhance preparedness for future outbreaks or similar health crises
5.	Business Model (Revenue Model)	The primary revenue streams include subscription or licensing models, where healthcare institutions and technology providers pay a fee to access and use the CovidVision software. Additionally, usage-based models can be employed, charging customers based on the number of lung X-ray images processed or COVID-19 detection requests made. Key activities involve continuous research and development, data management, customer support, and training. Channels include online platforms, direct sales, and partnerships with healthcare institutions and technology providers. The business model aims to deliver accurate detection, efficiency, integration, and interpretability while generating revenue through customer subscriptions, licensing, and usage-based fees.
6.	Scalability of the Solution	The scalability of CovidVision enables it to effectively handle the growing volume of lung X-ray images and meet the increasing demand for COVID-19 detection. By incorporating scalable architecture, optimizing performance, and implementing monitoring and scaling mechanisms, CovidVision delivers a scalable solution that supports the healthcare industry in the battle against COVID-19

3. REQUIREMENT ANALYSIS

3.1 Functional requirement

Following are the functional requirements of the proposed solution.

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	Image Input and Processing	a. Accept lung X-ray images as input in a standard format. b. Preprocess the input images to normalize intensity, remove noise, and enhance features. c. Handle variations in image sizes, orientations, and resolutions.
FR-2	Model Training and Testing	a. Train a machine learning or deep learning model using a diverse dataset of labeled lung X-ray images. b. Implement appropriate algorithms, such as CNNs, for COVID-19 detection. c. Conduct model testing and evaluation using independent test datasets to measure accuracy, sensitivity, specificity, and other performance metrics. d. Perform model optimization techniques, such as hyperparameter tuning and regularization, to improve the model's performance.
FR-3	COVID-19 Classification	a. Utilize the trained model to classify lung X-ray images into COVID-19 positive, non-COVID lung diseases, or healthy lung categories. b. Assign probability scores or confidence levels to the classification outputs. c. Handle the classification of multiple lung X-ray images in a batch or real-time streaming scenario.
FR-4	User Interface	a. Develop a user-friendly interface for healthcare professionals to interact with the system. b. Enable easy uploading and retrieval of lung X-ray images for analysis.
FR-5	Integration	Integrate the CovidVision system with existing healthcare systems or diagnostic tools, allowing seamless incorporation into the clinical workflow.
FR-6	Security and Privacy	Implement appropriate security measures to protect patient data and ensure compliance with privacy regulations. b. Safeguard the confidentiality and integrity of the processed lung X-ray images and associated metadata. c. Implement user authentication and access control mechanisms to restrict system usage to authorized personnel.

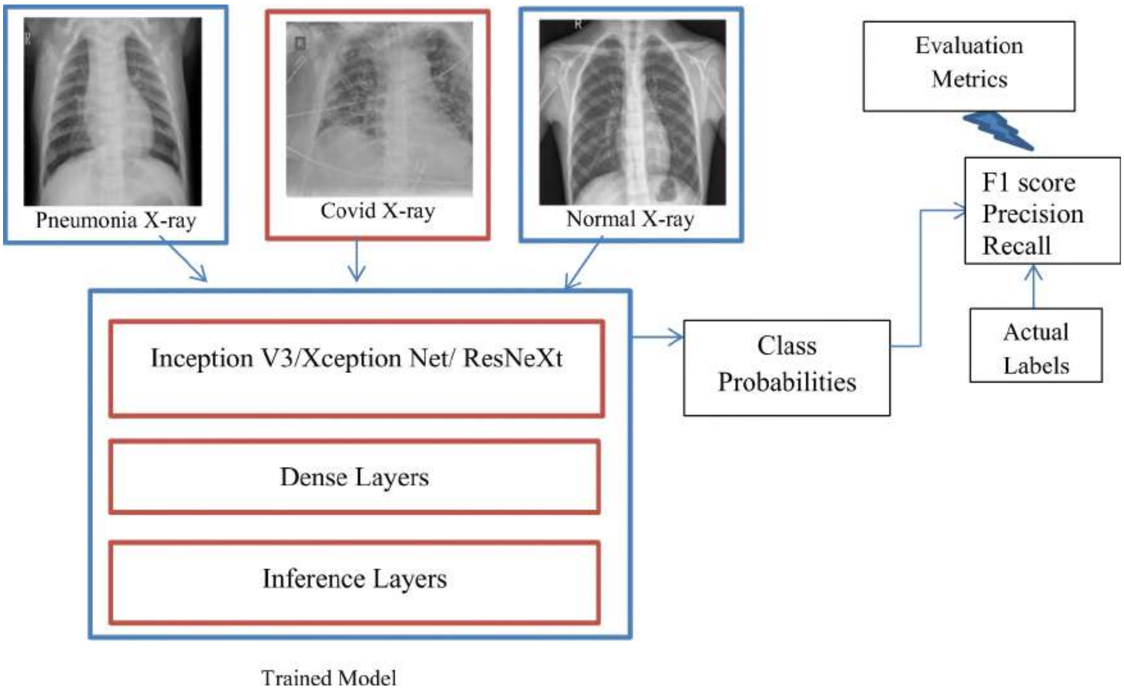
3.2 Non-Functional requirements

Following are the non-functional requirements of the proposed solution.

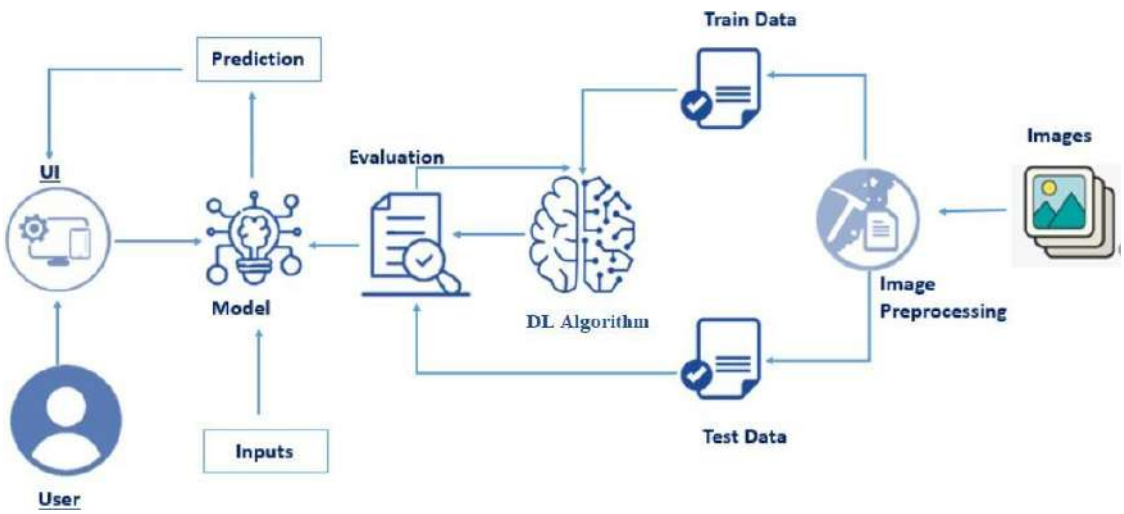
FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	The system has an intuitive and easy-to-use interface for healthcare professionals, requiring minimal training and technical expertise.
NFR-2	**Security**	The system employs secure transmission protocols when exchanging data between the user interface, backend, and external systems.
NFR-3	**Reliability**	The system has a high degree of reliability, ensuring consistent and dependable performance under varying conditions.
NFR-4	**Transparency**	The system scales to accommodate increased data volume and user demand without significant performance degradation.
NFR-5	**Access Control**	The system enforces appropriate access controls to restrict system usage to authorized personnel and protect against unauthorized use.
NFR-6	**Scalability**	The system scales to accommodate increased data volume and user demand without significant performance degradation.
NFR-7	**Data Privacy**	The system adheres to privacy regulations and protects patient data from unauthorized access, ensuring confidentiality and data integrity.
NFR-8	**Transparency**	The system provides clear explanations and visualizations to enable healthcare professionals to understand the basis for the classification decisions.
NFR-9	**Explainability**	The system offers insights into the features, patterns, or regions within the lung X-ray images that contribute to the COVID-19 detection, fostering trust and collaboration.
NFR-10	**Accuracy**	The system achieves high accuracy in classifying COVID-19 cases from lung X-ray images, minimizing false positives and false negatives.
NFR-11	**Reliability**	The system has a high degree of reliability, ensuring consistent and dependable performance under varying conditions.

4. PROJECT DESIGN

4.1 Data Flow Diagrams



4.2 Solution & Technical Architecture



4.3 User Stories

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Team Member
Hospital Staffs	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	I can access my account / dashboard	High	keerthanaa
		USN-2	As a user, I will receive confirmation email once I have registered for the application	I can receive confirmation email & click confirm	High	sharon
	Login	USN-5	As a user, I can log into the application by entering email & password		High	Kamali
Healthcare administrator	Integration		Integrate CovidVision into our existing picture archiving and communication system (PACS), enabling seamless transfer of lung X-ray images for COVID-19 analysis and improving workflow efficiency.	I can integrate CovidVision into our existing pictures of lung X-Rays	High	Nadheedha
Radiology department manager	Monitoring		Monitor the performance and usage statistics of CovidVision, including the number of processed images, accuracy rates, and resource utilization, to ensure optimal system performance and cost-effective usage.	I can monitor the performance and usage statistics of CovidVision	Medium	Sharon
Healthcare IT professional	Security		Ensure the secure and encrypted transmission of lung X-ray images to CovidVision, protecting patient confidentiality and complying with data protection regulations.	Security of the patient data is achieved successfully	High	Kamali
Patient	Usage		I want CovidVision to assist healthcare professionals in accurately detecting COVID-19 from my lung X-ray, enabling early detection, appropriate treatment, and better overall outcomes.	The outcome of the project is accurate	Low	Nadheedha
Healthcare technology provider			I want to integrate CovidVision into our diagnostic software platform, expanding our offering with advanced COVID-19 detection capabilities and providing added value to our customers.	Integration is done successfully	Medium	Sharon

5. CODING & SOLUTIONING (Explain the features added in the project along with code)

5.1 Feature 1 (Image Preprocessing)

The image preprocessing feature in CovidVision plays a crucial role in enhancing the quality and consistency of lung X-ray images before they are processed for COVID-19 detection. Here are some key aspects of the image preprocessing feature:

Image Enhancement: CovidVision applies various enhancement techniques to improve the quality and visibility of lung X-ray images. These techniques may include contrast adjustment, histogram equalization, and sharpening filters to enhance important features and improve the overall image quality.

Noise Reduction: Lung X-ray images can often contain noise, which can interfere with the accuracy of COVID-19 detection. CovidVision employs noise reduction algorithms, such as filters or denoising techniques, to reduce unwanted noise and artifacts while preserving important image details.

Standardization and Normalization: To ensure consistency across different lung X-ray images, CovidVision applies standardization and normalization techniques. This involves scaling the pixel intensities to a standardized range and adjusting the image dimensions or resolutions to a consistent format for uniform processing.

Region of Interest (ROI) Extraction: CovidVision identifies and extracts the relevant regions of interest within the lung X-ray images. This involves segmenting and isolating the lung area from the rest of the image, allowing the algorithms to focus on the specific regions that contain critical information for COVID-19 detection.

Artifact Removal: Certain artifacts or inconsistencies may be present in lung X-ray images, such as motion blur, artifacts from medical devices, or image artifacts introduced during the acquisition process. CovidVision employs algorithms to detect and remove such artifacts, ensuring cleaner and more accurate input for COVID-19 detection.

5.2 Feature 2 (Performance Monitoring and Analytics)

Performance monitoring and analytics in CovidVision provide valuable insights into the system's performance, resource utilization, and accuracy rates. Here are some key aspects of the performance monitoring and analytics feature:

Processing Time and Throughput: CovidVision tracks the processing time required for analyzing each lung X-ray image. It measures the time taken from image upload to the generation of COVID-19 detection results. Monitoring the processing time allows healthcare professionals to assess system efficiency and identify any bottlenecks that may impact turnaround time.

Resource Utilization: The system monitors resource utilization, including CPU usage, memory consumption, and storage capacity. It ensures efficient allocation of resources, identifies potential performance issues, and helps optimize the infrastructure to handle varying workloads effectively.

Accuracy Rates: CovidVision captures and analyzes accuracy rates, measuring the system's performance in correctly identifying COVID-19 cases from lung X-ray images. These rates are typically evaluated against ground truth data or expert annotations, providing insights into the model's precision, recall, and overall diagnostic performance.

Error Analysis: The performance monitoring and analytics feature in CovidVision includes error analysis capabilities. It helps identify common patterns or types of errors made by the system, allowing for continuous improvement and fine-tuning of the algorithms. Understanding the nature of errors can lead to enhancements in model training, feature extraction, or data preprocessing techniques.

Reporting and Visualization: CovidVision generates comprehensive reports and visualizations to present performance metrics and analytics in a clear and accessible manner. These reports can include statistical summaries, charts, and graphs, enabling healthcare professionals, administrators, and researchers to monitor and evaluate the system's performance over time.

Alerting and Notification: The performance monitoring feature may include alerting and notification mechanisms. It can alert system administrators or support teams when certain performance thresholds or error rates are exceeded, allowing for proactive intervention and prompt resolution of issues.

Data Insights and Trends: By analyzing performance data over an extended period, CovidVision can identify trends, patterns, and potential areas for improvement. These insights can guide further enhancements in model training, algorithm optimization, or infrastructure scaling, ensuring the system evolves to meet changing needs.

5.3 Database COVID-19 Detection:

User-Friendly Interface: COVIDVision offers a user-friendly interface that allows healthcare professionals to easily upload and view lung X-ray images. The interface provides a seamless user experience, enabling intuitive interactions and quick access to detection results.

Interpretability and Explainability: The system incorporates features to provide interpretability and explainability of the COVID-19 detection results. It highlights the key regions or features in the lung X-ray images that contribute to the classification decision, aiding healthcare professionals in understanding and trusting the results.

Integration with Existing Systems: COVIDVision can integrate with existing healthcare systems, such as picture archiving and communication systems (PACS) or electronic health record (EHR) systems. This allows for seamless data exchange, streamlined workflows, and integration into the existing diagnostic infrastructure.

Scalable Infrastructure: The solution is designed to handle large volumes of lung X-ray images with scalable infrastructure. It leverages cloud computing or distributed processing techniques to ensure efficient processing and storage of images, allowing for scalability as the demand increases.

Performance Monitoring and Analytics: COVIDVision includes monitoring and analytics capabilities to track system performance, resource utilization, and accuracy rates. It provides insights and reports on the processing time, throughput, and other metrics to ensure optimal performance and identify areas for improvement.

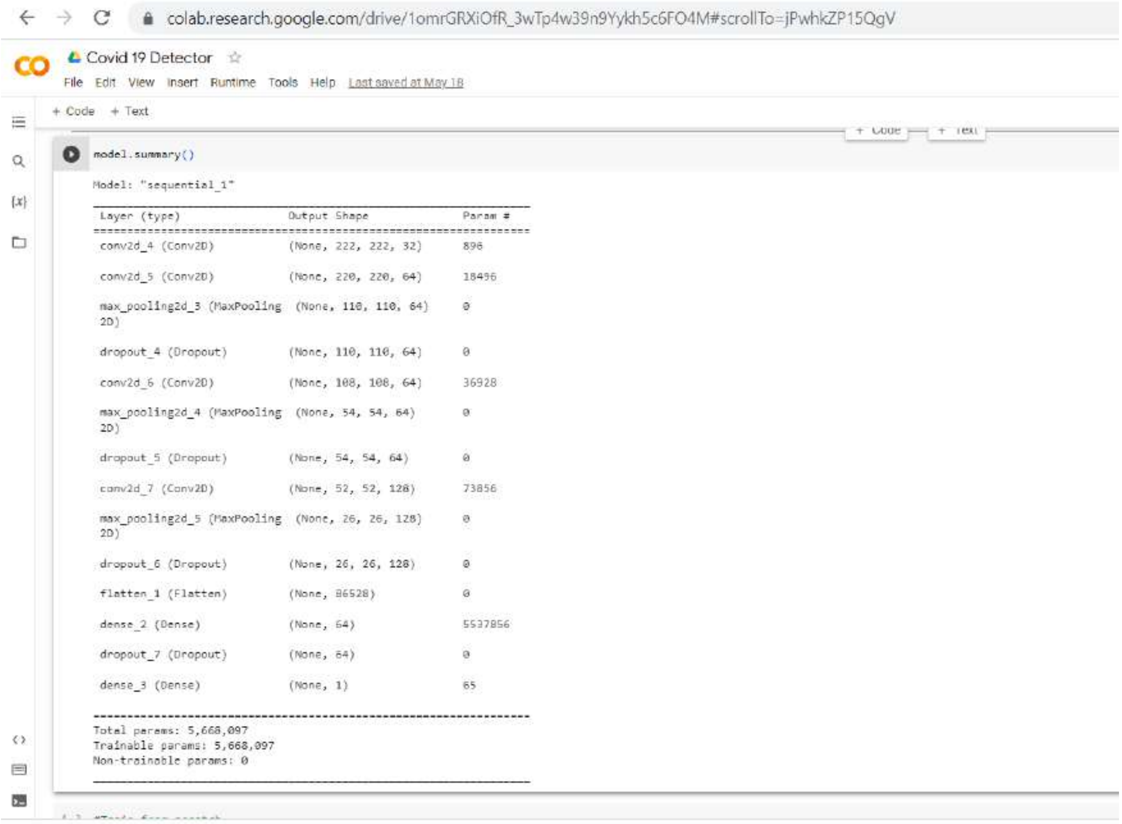
Security and Privacy: The solution prioritizes data security and privacy by implementing robust encryption, access controls, and anonymization techniques. It adheres to relevant data protection regulations to protect patient confidentiality and ensure secure data transmission.

Continuous Improvement and Updates: COVIDVision incorporates a feedback loop to continuously improve its performance. It can be updated with new datasets, algorithm enhancements, and technological advancements to ensure the system remains up-to-date and provides state-of-the-art COVID-19 detection capabilities
Schema (if Applicable)

6. RESULTS

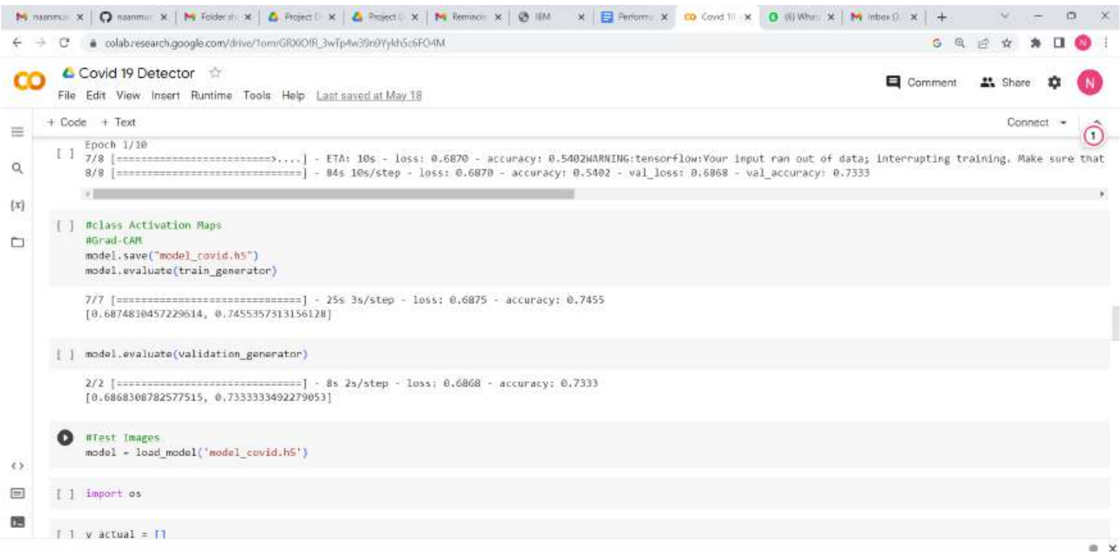
6.1 Performance Metrics

1)Model Summary:



```
model.summary()

Model: "sequential_1"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d_4 (Conv2D)           (None, 222, 222, 32)      896

 conv2d_5 (Conv2D)           (None, 220, 220, 64)      18496

 max_pooling2d_3 (MaxPooling  (None, 110, 110, 64)     0
 2D)

 dropout_4 (Dropout)         (None, 110, 110, 64)      0

 conv2d_6 (Conv2D)           (None, 108, 108, 64)      36928

 max_pooling2d_4 (MaxPooling  (None, 54, 54, 64)       0
 2D)

 dropout_5 (Dropout)         (None, 54, 54, 64)        0

 conv2d_7 (Conv2D)           (None, 52, 52, 128)       73856

 max_pooling2d_5 (MaxPooling  (None, 26, 26, 128)      0
 2D)

 dropout_6 (Dropout)         (None, 26, 26, 128)       0

 flatten_1 (Flatten)         (None, 86528)             0

 dense_2 (Dense)             (None, 64)                5537856

 dropout_7 (Dropout)         (None, 64)                0

 dense_3 (Dense)             (None, 1)                 65

=================================================================
Total params: 5,668,097
Trainable params: 5,668,097
Non-trainable params: 0
_________________________________________________________________
```

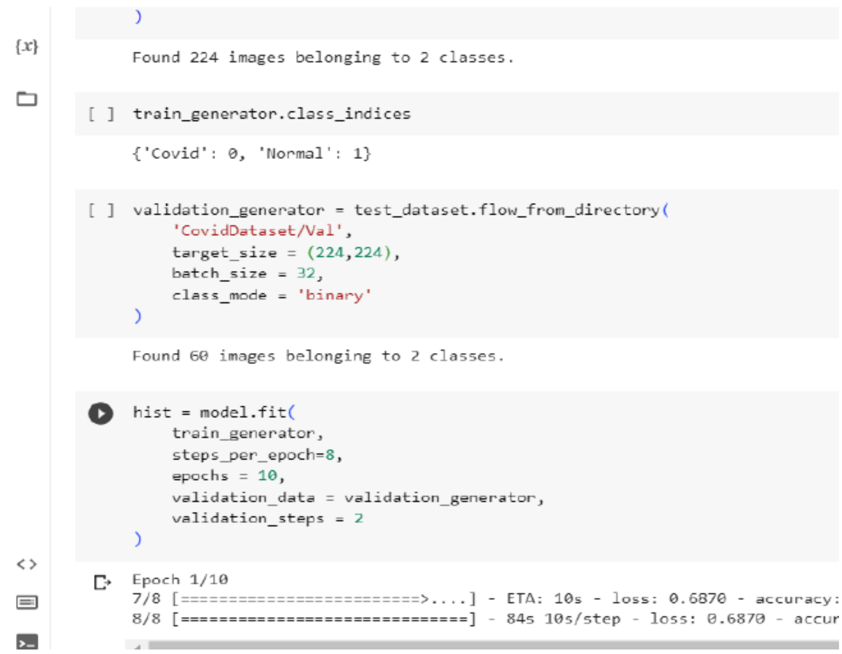
2)Accuracy:



```
Epoch 1/10
7/8 [=========================>....] - ETA: 10s - loss: 0.6870 - accuracy: 0.5402WARNING:tensorflow:Your input ran out of data; interrupting training. Make sure that
8/8 [==============================] - 84s 10s/step - loss: 0.6870 - accuracy: 0.5402 - val_loss: 0.6868 - val_accuracy: 0.7333

#class Activation Maps
#Grad-CAM
model.save("model_covid.h5")
model.evaluate(train_generator)

7/7 [==============================] - 25s 3s/step - loss: 0.6875 - accuracy: 0.7455
[0.6874810457229614, 0.7455357313156128]

model.evaluate(validation_generator)

2/2 [==============================] - 8s 2s/step - loss: 0.6868 - accuracy: 0.7333
[0.6868308782577515, 0.7333333492279053]

#Test Images
model = load_model('model_covid.h5')

import os

y_actual = []
```

I) Training Accuracy - accuracy: 0.7455

II) Validation Accuracy -accuracy: 0.7333



Hence we have created a model to predict whether a person having COVID or not based on the lung X-Ray taken from the patients

This will be useful for helping the doctors and patients to identify the presence of COVID.

7. ADVANTAGES & DISADVANTAGES

Advantages of COVID-19 detection using X-rays:

1. Quick results: X-ray imaging is a relatively quick method for COVID-19 detection. The images can be obtained in a matter of minutes, allowing for a rapid assessment of lung involvement.

2. Availability: X-ray machines are widely available in healthcare settings, making it accessible for COVID-19 screening in various medical facilities, including hospitals, clinics, and emergency rooms.

3. Cost-effective: Compared to some other diagnostic tests like PCR (polymerase chain reaction) or CT scans, X-rays are generally less expensive, making them a cost-effective option for mass screening or resource-limited settings.

Disadvantages of COVID-19 detection using X-rays:

1. Limited sensitivity: X-ray imaging may not always detect early-stage COVID-19 infections or mild cases where lung involvement is minimal. False-negative results are possible, especially in the early days of infection.

2. Lack of specificity: X-ray images alone cannot definitively confirm a COVID-19 infection as they do not provide a specific identification of the virus. Other respiratory illnesses or conditions with similar radiographic findings can potentially be misdiagnosed as COVID-19.

3. Radiation exposure: Although the radiation dose from a single chest X-ray is relatively low, repeated or unnecessary X-rays may increase the cumulative radiation dose, which can be a concern, especially for vulnerable populations or frequent testing.

4. Operator dependency: Interpreting X-ray images requires expertise and experience. There can be variations in interpretation among radiologists or healthcare providers, leading to potential errors or inconsistencies in COVID-19 diagnosis.

8. CONCLUSION

In conclusion, X-ray imaging can be a useful tool for COVID-19 detection, offering advantages such as quick results, availability in healthcare settings, and cost-effectiveness. However, it also has several limitations, including limited sensitivity in detecting early-stage or mild cases, lack of specificity in confirming COVID-19 infection, potential radiation exposure, and reliance on operator expertise for interpretation. X-rays alone should not be considered a definitive diagnostic method for COVID-19, and confirmatory testing through PCR or other techniques is typically necessary for accurate diagnosis. It is essential to use X-ray imaging in conjunction with clinical evaluation and other diagnostic methods to improve the accuracy of COVID-19 detection and management.

9. FUTURE SCOPE

The future scope for COVID-19 detection using X-rays holds potential for advancements and improvements. Here are some possible areas of development:

1. Artificial Intelligence (AI) and Machine Learning: Integration of AI algorithms and machine learning techniques can enhance the accuracy and efficiency of COVID-19 detection from X-ray images. These technologies can assist in automating the analysis process, reducing human error, and improving the sensitivity and specificity of diagnosis.

2. Quantitative Assessment: Future developments may focus on developing quantitative assessment tools based on X-ray imaging. This involves the use of image analysis algorithms to measure specific radiographic features or patterns associated with COVID-19 infection, providing a more objective and standardized approach to diagnosis.

3. Combined Imaging Modalities: Combining X-ray imaging with other imaging modalities, such as computed tomography (CT) or ultrasound, may offer a more comprehensive and accurate assessment of COVID-19. These multimodal approaches can provide complementary information, enabling a more precise diagnosis and evaluation of lung involvement.

4. Portable and Point-of-Care Devices: Advancements in technology may lead to the development of portable X-ray devices or point-of-care systems specifically designed for COVID-19 detection. These devices would enable rapid screening in various settings, including remote areas, emergency situations, or resource-limited environments.

5. Long-term Monitoring and Follow-up: X-ray imaging may also play a role in long-term monitoring and follow-up of COVID-19 patients, tracking the progression or resolution of lung abnormalities, and evaluating treatment outcomes. Serial X-ray imaging can provide valuable information for assessing recovery and determining the need for further intervention.

It's important to note that these potential future developments would require rigorous research, validation, and collaboration between medical professionals, radiologists, engineers, and data scientists to ensure their effectiveness, safety, and widespread implementation.

10. APPENDIX

Source Code

```
# Dataset : http://cb.lk/covid_19

!wget http://cb.lk/covid_19
```

```
!unzip covid_19
```

```
TRAIN_PATH = "CovidDataset/Train"
VAL_PATH = "CovidDataset/Test"
```

```
import numpy as np
import matplotlib.pyplot as plt
import keras
from keras.layers import *
from keras.models import *
from keras.preprocessing import image
```

```
from keras.layers.attention.multi_head_attention import activation
#CNN Based Model in Keras
model = Sequential()
model.add(Conv2D(32,kernel_size=(3,3),activation='relu',input_shape=(224,224,3)))
model.add(Conv2D(64,(3,3),activation='relu'))
model.add(MaxPooling2D(pool_size=(2,2)))
model.add(Dropout(0.25))

model.add(Conv2D(64,(3,3),activation='relu'))
model.add(MaxPooling2D(pool_size=(2,2)))
model.add(Dropout(0.25))

model.add(Conv2D(128,(3,3),activation='relu'))
model.add(MaxPooling2D(pool_size=(2,2)))
model.add(Dropout(0.25))

model.add(Flatten())
model.add(Dense(64,activation='relu'))
model.add(Dropout(0.5))
model.add(Dense(1,activation="sigmoid"))

model.compile(loss=keras.losses.binary_crossentropy,optimizer='adam',metrics=['accuracy'])
```

Covid_19_Detector (1) (1).ipynb ×

C: > Users > Keerthana > Downloads > Covid_19_Detector (1) (1).ipynb > # Dataset : http://cb.lk/covid_19

+ Code + Markdown | Run All Clear All Outputs | Outline ···

```python
model.add(MaxPooling2D(pool_size=(2,2)))
model.add(Dropout(0.25))

model.add(Flatten())
model.add(Dense(64,activation='relu'))
model.add(Dropout(0.5))
model.add(Dense(1,activation="sigmoid"))

model.compile(loss=keras.losses.binary_crossentropy,optimizer='adam',metrics=['accuracy'])
```

[44]

```python
model.summary()
```

[45]

```
Model: "sequential_1"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d_4 (Conv2D)           (None, 222, 222, 32)      896

 conv2d_5 (Conv2D)           (None, 220, 220, 64)      18496

 max_pooling2d_3 (MaxPooling  (None, 110, 110, 64)     0
 2D)

 dropout_4 (Dropout)         (None, 110, 110, 64)      0
```

DatasetCreator.ipynb

Covid_19_Detector (1) (1).ipynb DataSet Creator.ipynb ×

C: > Users > Keerthana > Downloads > DataSet Creator.ipynb > import pandas as pd

+ Code + Markdown | Run All Clear All Outputs | Outline ···

```python
import pandas as pd
import os
import shutil
```

[5]

```python
#create the data for positive samples

FILE_PATH = "covid-chestxray-dataset-master/metadata.csv"
IMAGES_PATH = "covid-chestxray-dataset-master/images"
```

[2]

```python
df = pd.read_csv(FILE_PATH)
print(df.shape)
```

[3]

```
(950, 30)
```

```python
df.head()
```

Covid_19_Detector (1) (1).ipynb DataSet Creator.ipynb

C: > Users > Keerthana > Downloads > DataSet Creator.ipynb > import pandas as pd

Python 3.9.13

	patientid	offset	sex	age	finding	RT_PCR_positive	survival	intubated	intubation_present	went_icu	...	date	location	folder	
0	2	0.0	M	65.0	Pneumonia/Viral/COVID-19	Y	Y	N	N	N	...	January 22, 2020	Cho Ray Hospital, Ho Chi Minh City, Vietnam	images	2020_(
1	2	3.0	M	65.0	Pneumonia/Viral/COVID-19	Y	Y	N	N	N	...	January 25, 2020	Cho Ray Hospital, Ho Chi Minh City, Vietnam	images	2020_(
2	2	5.0	M	65.0	Pneumonia/Viral/COVID-19	Y	Y	N	N	N	...	January 27, 2020	Cho Ray Hospital, Ho Chi Minh City, Vietnam	images	2020_(
3	2	6.0	M	65.0	Pneumonia/Viral/COVID-19	Y	Y	N	N	N	...	January 28, 2020	Cho Ray Hospital, Ho Chi Minh City, Vietnam	images	2020_(
4	4	0.0	F	52.0	Pneumonia/Viral/COVID-19	Y	NaN	N	N	N	...	January 25, 2020	Changhua Christian Hospital, Changhua	images	

Cell 1 of 11

Covid_19_Detector (1) (1).ipynb DataSet Creator.ipynb

C: > Users > Keerthana > Downloads > DataSet Creator.ipynb > import pandas as pd

```python
TARGET_DIR = "Dataset/Covid"

if not os.path.exists(TARGET_DIR):
    os.mkdir(TARGET_DIR)
    print("Covid folder created")
```

[6]

```
Covid folder created
```

```python
cnt = 0

for (i,row) in df.iterrows():
    if row["finding"]=="Pneumonia/Viral/COVID-19" and row["view"]=="PA":
        filename = row["filename"]
        image_path = os.path.join(IMAGES_PATH, filename)
        image_copy_path = os.path.join(TARGET_DIR,filename)
        shutil.copy2(image_path,image_copy_path)
        #print("Moving image", cnt)
        cnt+=1
```

[23]

```python
#Sampling of Images from Kaggle

import random
```

Covid_19_Detector (1) (1).ipynb   DataSet Creator.ipynb

C: > Users > Keerthana > Downloads > DataSet Creator.ipynb > import pandas as pd

+ Code + Markdown | Run All Clear All Outputs | Outline ...

```python
#Sampling of Images from Kaggle

import random
KAGGLE_FILE_PATH = "chest_xray/train/NORMAL"
TARGET_NORMAL_DIR = "Dataset/Normal"
```
[24]

```python
image_names = os.listdir(KAGGLE_FILE_PATH)
```
[25]

```python
random.shuffle(image_names)
```
[26]

```python
for i in range(196):

    image_name = image_names[i]
    image_path = os.path.join(KAGGLE_FILE_PATH,image_name)

    target_path = os.path.join(TARGET_NORMAL_DIR, image_name)

    shutil.copy2(image_path, target_path)
    print("Copying image ",i)
```


Covid_19_Detector (1) (1).ipynb   DataSet Creator.ipynb

C: > Users > Keerthana > Downloads > DataSet Creator.ipynb > import pandas as pd

+ Code + Markdown | Run All Clear All Outputs | Outline ...

```python
    image_path = os.path.join(KAGGLE_FILE_PATH,image_name)

    target_path = os.path.join(TARGET_NORMAL_DIR, image_name)

    shutil.copy2(image_path, target_path)
    print("Copying image ",i)
```

[28]

```
Copying image  0
Copying image  1
Copying image  2
Copying image  3
Copying image  4
Copying image  5
Copying image  6
Copying image  7
Copying image  8
Copying image  9
Copying image  10
```

GitHub Link:
https://github.com/naanmudhalvan-SI/PBL-NT-GP-14476-1683893793/tree/main/Assignments

Project Video Demo Link:
Click here