CHEST X-RAY IMAGE CLASSIFICATION

By

AGHA MUZAMIL

HAMMAD NAWAZ

SHAKIR ALI

MAEEN KHAN

Supervised by:

DR. SHIBLI NISAR

Submitted to the faculty of the Department of Electrical Engineering,
Military College of Signals, National University of Sciences and Technology, Islamabad,
in partial fulfillment of the requirements of B.E Degree in Electrical (Telecom) Engineering.

June 2023

In the name of ALLAH, the Most benevolent, the Most Courteous

CERTIFICATE OF CORRECTNESS AND APPROVAL

This is to officially state that the thesis work contained in this report
"Chest X-Ray Image Classification."

is carried out by

SYED MUZAMIL AGHA

HAMMAD NAWAZ

SHAKIR ALI

MAEEN KHAN

under my supervision and that in my judgement, it is fully ample, in scope and excellence, for the degree of Bachelor of Electrical (Telecom.) Engineering in Military College of Signals, National University of Sciences and Technology (NUST), Islamabad.

Approved by

Supervisor

ASSOC PROF DR SHIBLI NISAR

Department of EE, MCS

Date: ___________

DECLARATION OF ORIGINALITY

We hereby declare that no portion of the work presented in this thesis has been submitted in support of another award or qualification in either this institute or anywhere else.

ACKNOWLEDGEMENTS

Allah Subhan'Wa'Tala is the sole guidance in all domains.

Our parents, colleagues, and most of all supervisor Assoc **Prof Dr. Shilbli Nisar** without your guidance.

The group members, who through all adversities worked steadfastly.

Plagiarism Certificate (Turnitin Report)

This thesis has ____ similarity index. Turnitin report endorsed by Supervisor is attached.

Syed Muzamil Agha

NUST Serial no 325038

Hammad Nawaz

NUST Serial no 325591

Shakir Ali

NUST Serial no 325600

Maeen Khan

NUST Serial no 325605

Signature of Supervisor

ABSTRACT

The chest X-ray abnormal and normal classification model will classify X-ray images of patients. The dataset is collected from hospitals for training. The state of art image classification algorithm (YOLOv5x-cls) was used to train the model. The model will classify the scanned X-ray into normal and abnormal Chest X-rays. As there is a huge burden on a radiologist of 3Million X-rays per annum, our project will help them to get the report in a single click. The project is a breakthrough in radiology, due to the instant rise in diseases, doctors found it difficult to tackle them. So, our project will bring more ease to doctors, which will save time and help in accuracy doctors need to treat more patients in less time. The model is trained to achieve maximum Accuracy (83%) on the dataset. The trained model is used in a web app for online inference and readily results can be served to doctors in it.

Contents

List of Figures

Chapter 1: Introduction

The diagnosis and treatment of many diseases depend heavily on medical imaging. One of the most used imaging modalities for the diagnosis of respiratory illnesses is the chest X-ray. Medical experts' interpretation of X-ray pictures, however, might take a while and be prone to mistakes. Automated image classification algorithms have been created to help doctors diagnose patients by overcoming these difficulties.

We set out to create a Chest X-ray abnormal and normal classification model in this study utilizing the cutting-edge image classification method YOLOv5x-cls. To categorize X-rays as normal or abnormal as effectively as possible, the model was trained on a sizable dataset of X-ray pictures gathered from hospitals. Our main goal was to give medical practitioners a tool that can speed up and increase the precision of respiratory disease diagnosis.

The model we created can divide X-ray images into normal and pathological categories. Numerous abnormalities, including lung masses, nodules, pneumonia, and other respiratory illnesses, fall under the category of abnormality. The dataset used to train the model was gathered from several hospitals and contains a variety of chest X-ray types with differing image quality levels.

Here, we outline our process for creating the Chest X-ray classification model, including the datasets we collected and prepared, the algorithms we used, the models we trained, and the results of our evaluation. We also discuss how the model might help doctors diagnose patients more quickly, accurately, and with less work.

Overall, by using machine learning to create an automated tool that can aid medical practitioners in identifying respiratory disorders, this study makes a substantial contribution to the field of radiography. The proposed paradigm may completely alter how respiratory disorders are identified and managed, ultimately leading to better patient outcomes.

1.1 Overview

The goal of this FYP is to create a model for classifying chest X-ray abnormal and normal images using the cutting-edge image classification algorithm (YOLOv5x-cls). Hospitals provided the training dataset, and the model was trained to maximize efficiency (83%) on the dataset. The relevance of the study, its potential advantages for physicians and patients, and the technique employed in creating the model are all covered in the thesis. For online inference, the trained model is linked to a web app, and results that are easily accessible can be sent to doctors in real-time

1.2 Problem Statement

Medical experts must spend a lot of time and risk making mistakes while interpreting pictures from chest X-rays. Doctors must process a growing number of Chest X-ray images in a short amount of time due to the volume of images being generated in hospitals. There is an increasing need for automated Chest X-ray picture classification models to increase the effectiveness and accuracy of the interpretation process. By creating a Chest X-ray abnormal and normal classification model utilizing a cutting-edge image classification algorithm, our FYP responds to this need. With this tool, medical professionals will be able to classify chest X-ray images into normal and abnormal categories accurately and quickly, potentially freeing up more time for doctors to treat more patients.

1.3 Proposed Solution

Using the cutting-edge image classification technique YOLOv5x-cls, our suggested approach aims to develop a classification model for abnormal and normal chest X-ray pictures. To help doctors, diagnose and treat patients, the model will categorize the scanned X-ray into groups for normal and abnormal states. A dataset collected from hospitals will be used for training to achieve the model's maximal efficiency (83%) on the dataset. The trained model will also be added to a web application for online inference, making it easier for doctors to get results and ultimately leading to better patient outcomes. The proposed approach should improve the efficiency and accuracy of the process for deciphering chest X-ray images, giving clinicians more time to treat more patients.

1.4 Working Principle

The project mainly works on the principles of image processing amalgamated with machine learning algorithms. The project is divided into different moduli and every module is interwoven with the next module. The list of modules is as under:

1) Data gathering: The initial phase entails gathering a dataset of chest X-ray pictures from medical facilities. Chest X-ray pictures, both normal and pathological, will be mixed into the dataset.
2) Data pre-processing: To get the images ready for training, the gathered dataset will be pre-processed. This will entail actions like scaling the photographs, making them grayscale, and using contrast-enhancing methods.

3) Model training: Using the cutting-edge image classification method YOLOv5x-cls, the pre-processed dataset will be utilized to train the Chest X-ray abnormal and normal classification model. The model will develop its ability to distinguish between typical and abnormal chest X-ray images during training.
4) Model validation will be done using a second dataset of chest X-ray images once the model has been trained. The accuracy, sensitivity, and specificity of the model will be considered while assessing its performance.
5) Integration of the trained model with a web application for online inference will allow doctors to upload Chest X-ray images and receive findings right away.
6) Inference: The model will categorize a chest X-ray image as normal or abnormal when it is uploaded to the web app, and the results will be shown to the doctor. The medical professional can then use this information to help with the patient's diagnosis and care.

Overall, the collection of data, pre-processing, model training, model validation, integration into a web interface, and inference are all part of the working concept of your Chest X-ray abnormal and normal categorization model. To assist clinicians in correctly diagnosing and treating patients, the model will eventually learn to categorize chest X-ray images into normal and pathological categories.

1.4.1 Datasets:

```
num_test_files = len(os.listdir(test_dir_n)) + len(os.listdir(test_dir_abn))

print("Number of files in Train folder: ", num_train_files)
print("Number of files in Valid folder: ", num_valid_files)
print("Number of files in Test folder: ", num_test_files)
```

```
Number of files in Train folder:  16064
Number of files in Valid folder:  2184
Number of files in Test folder:  375
```

+ Code + Markdown

Figure 1: Number of Datasets Train, Valid & Test

1.4.1.1 Dataset Collection:

The practice of acquiring pertinent data for a certain research project or analysis is known as dataset collection. Collecting a significant number of X-ray images of patients with both normal and abnormal chest conditions is necessary for your study on the categorization of chest X-ray images.

Hospitals, clinics, and publicly accessible repositories like Kaggle, the National Institutes of Health Chest X-ray Dataset, and CheXpert are just a few places where the dataset can be gathered. The dataset should be sizable enough to guarantee that the model is trained on a variety of pictures and is capable of accurately identifying anomalies in chest X-rays. The total images are 16032 of which 8032 are normal and 8032 are abnormal.

1.4.1.2 Customized Pakistani Dataset:

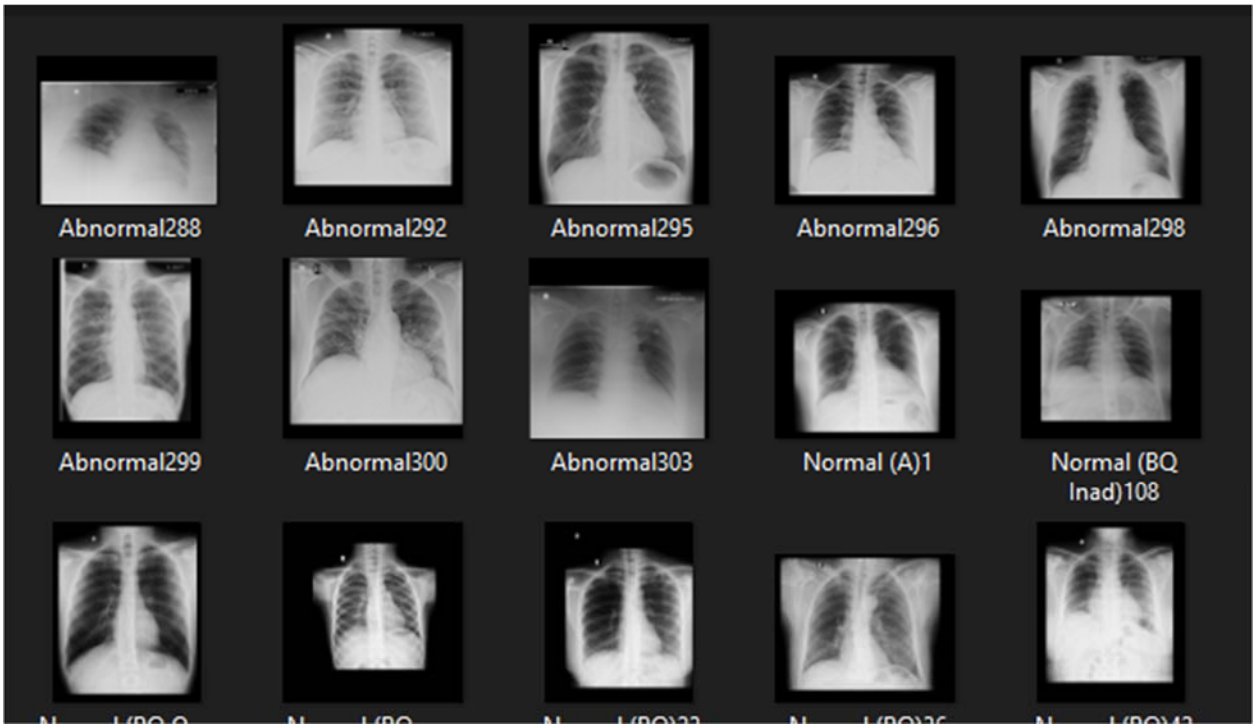

Figure 2 Customized Pakistani Dataset

This project uses a custom build dataset of Pakistani chest x-ray of 300 images of which 122 are normal, 153 are abnormal and 28 are not sure labels for its use. The images are gathered, filtered, and annotated to obtain the coordinates of the object of interest.

1.4.2 Dataset training and processing:

1. The prepared dataset is used as input to train object detection models using machine learning.
2. Splitting the dataset: To train and assess the performance of the model, the dataset is often separated into training, validation, and testing sets.
3. Data Augmentation: Techniques like rotation, scaling, and flipping can be employed to make the training data more variable and boost the resilience of the model.

4. Model Training: The model can be trained using the YOLOv5x-cls algorithm. You might talk about the training parameters, such as the batch size, learning rate, and number of epochs.
5. Model Evaluation: The model's performance metrics, such as accuracy, precision, recall, and F1 score, should be assessed after training using the validation and testing datasets.
6. Hyperparameter Tuning: To improve a model's performance if it is not performing well, hyperparameter tuning can be done.
7. Analysis of Results: As a last step, you may examine the model's outcomes and debate its advantages and disadvantages. On comparable datasets, you can also assess how well your model performs in comparison to other cutting-edge models.

1.4.2.1 YOLOv5 algorithm:

Model	size (pixels)	acc top1	acc top5
YOLOv5n-cls	224	64.6	85.4
YOLOv5s-cls	224	71.5	90.2
YOLOv5m-cls	224	75.9	92.9
YOLOv5l-cls	224	78.0	94.0
YOLOv5x-cls	224	**79.0**	**94.4**
ResNet18	224	70.3	89.5
ResNet34	224	73.9	91.8
ResNet50	224	76.8	93.4
ResNet101	224	78.5	94.3

Figure 3: Difference Between Algorithms

A cutting-edge object recognition and image classification technique called YOLOv5x-cls is used to recognize things in images. The YOLOv5x-cls algorithm, which is based on deep learning and convolutional neural networks (CNNs), is an advancement over earlier iterations of the YOLO (You Only Look Once) algorithm.

The YOLOv5x-cls method predicts the kind and location of objects in each cell of a picture by dividing it into a grid of cells. This strategy enables the computer to recognize many things in a single image and correctly classify them.

A sizable collection of photos that have been labeled with item classes and bounding box coordinates is used to train the YOLOv5x-cls method. The system gains the ability to identify patterns and features in images that correspond to items during training. It then applies this understanding to forecast what will appear in new, upcoming images.

Several to previous object detection algorithms, the YOLOv5x-cls method provides several benefits, including speed, accuracy, and the capacity to detect small things. These benefits make it an attractive option for numerous applications, such as robotics, self-driving automobiles, and medical imaging.

The YOLOv5x-cls method is used in the context of the Chest X-ray abnormal and normal Classification model to categorize the X-ray pictures as either normal or abnormal depending on the presence of specific features and patterns in the image. The system uses this training to generate predictions on fresh, unused X-ray images after being trained on a dataset of X-ray images that are classified as either normal or abnormal.

1.4.3 Output Extraction:

The following steps are commonly included in the output extraction process:

1. To match the dimensions and range of the training data, the input image must be resized and normalized as part of the preprocessing procedure.
2. Feed the trained model the preprocessed image: The preprocessed image should be entered into the model using the prediction method of the YOLOv5x-cls algorithm.
3. Obtain the output of the model: A probability distribution for the two classes (normal and abnormal) will be produced by the model. You can choose the class with the best likelihood score to extract the output.
4. Display the result: In the web app, you can display the classification result for doctors to see whether it is normal or abnormal.
5. For each fresh chest X-ray image that needs to be classified, repeat steps 1-4.

Overall, processing the input image, feeding it to the model, extracting the output, and displaying it to the user comprises the output extraction in your project.

1.4.4 Decision-based upon Outputs:

Let's imagine that a patient's chest X-ray image is classified as normal or abnormal using the Chest X-ray image classification model. The medical expert will choose under the model's output.

The doctor can conclude that the patient is healthy and doesn't need any additional tests or treatment if the photograph is labeled as normal. The doctor might decide to conduct more diagnostic tests or refer the patient to a specialist for additional assessment and treatment, though, if the image is labeled as abnormal.

The choice made based on the output extraction would be determined by the patient's particular medical condition, its severity, and the doctor's professional opinion.

1.4.5 Integration:

1. Setting up the Django project: The establishment of a new Django project is the first stage in the integration procedure. Installing Django and creating a new project with a name and directory are required for this. Additionally, you would develop a brand-new app for the project that would include the code for incorporating the Chest X-ray image classification model.
2. After the Django app has been set up, you can create a model that will be used to categorize images of the chest taken during X-rays. For this, a class that derives from the Django Model class must be created, and the model's fields must represent the image and its classification.
3. Images must first be preprocessed before being classified by the Django app after a user uploads a Chest X-ray image. To achieve this, the image must be resized to the same dimensions as the training photos, and the pixel values must be normalized to the same range.
4. The next step is to load the trained Chest X-ray image classification model into the Django app after preparing the image. A Python library like PyTorch or TensorFlow can be used for this. To categorize photos, the model must first be placed into memory.
5. Images can be classified by passing them through the loaded model once the preprocessed picture has been stored in memory. The model will produce a probability distribution over the two classes (normal and abnormal), and the classification result will be based on the class with the highest probability score.

6. Displaying the classification result: After the image has been classified, the user is shown the categorization result. To do this, a new template that shows the categorization outcome and the original uploaded image must be made.
7. Lastly, the app can be made available to medical practitioners by making it available on a web server. This can be accomplished by either deploying the program to a local web server or using a cloud hosting service like AWS or Google Cloud Platform.

These methods have let you include your Chest X-ray image classification model into a Django app, facilitating quick and accurate patient diagnoses for medical practitioners.

1.4.6 GUI presentation:

The visual demonstration of the project is done through the aid of GUI (graphical user interface).

1.5 Objectives

1.5.1 General Objectives:

1. To create a model for classifying patient X-ray images into normal and abnormal categories based on their chest X-ray results.
2. To gather and use a dataset of hospital X-ray pictures for the classification model's training.
3. To train the model using the cutting-edge image classification technique YOLOv5x-cls.
4. To increase the model's training process efficiency to its maximum of 83% on the dataset.

5. To create a web application that uses the trained model for online inference and gives clinicians access to the results right away.
6. To aid medical professionals in speeding up and increasing the accuracy of diagnosing chest X-ray images, ultimately improving patient care.

Overall, the major goal of your project is to advance radiology by creating an effective and precise tool that can help physicians diagnose chest X-ray images.

1.5.2 Academic Objectives:

Academic objectives are precise aims that students have for themselves while they pursue their academic careers. These goals are frequently linked to academic success, skill development, and personal development.

Some potential academic goals for your FYP thesis on the classification of chest X-ray images include:

1. to comprehend computer vision, image categorization, and deep learning ideas and methods.
2. to acquire practical expertise in the gathering, preparation, and annotation of data for machine learning.
3. to gain knowledge of the training and evaluation processes for the cutting-edge YOLOv5x-cls image classification system for medical pictures.
4. to gain expertise in software development and programming, including using Python, TensorFlow, and other tools.
5. to undertake a comprehensive evaluation of the literature on the classification of chest X-ray images and associated subjects.

6. to create a precise and effective model for classifying chest X-rays to add new knowledge to the field of medical image classification.
7. to get practice writing scientifically, including how to construct a research report, how to convey study findings, and how to adhere to academic norms and conventions.

1.6 Scope

Our project's goal is to create a model for classifying chest X-rays into normal and abnormal ones using the YOLOv5x-cls method. This model will be able to distinguish between the two types of X-rays taken of patients. By assisting physicians in treating more patients in a shorter amount of time while improving accuracy, the project seeks to make a significant advancement in radiology. A web program for online inference will incorporate the training model and deliver quick results to clinicians. Learning about deep learning, image processing, and machine learning methods, as well as honing programming and data analysis skills, are some of the academic goals of the project. In addition, the project seeks to promote medical technology and healthcare by offering a novel response to a critical issue.

1.7 Relevant Sustainable Development Goals

To ensure a better and more sustainable future for everybody, the United Nations has established a list of 17 Sustainable Development Goals (SDGs). Chest X-ray image classification can support several SDGs, including some of the following:

1. Good Health and Well-Being SDG 3: The project's goal is to help physicians identify problems in chest X-rays early and accurately, which can improve patient health and well-being.

2. Infrastructure, Industry, and Innovation (SDG 9): The initiative uses cutting-edge machine learning techniques and technologies to classify chest X-ray images, which can help the healthcare sector build up an advanced and innovative infrastructure.
3. Reduced Inequalities (SDG 10): The project can contribute to lowering the inequality in access to healthcare services, particularly for persons living in distant or underdeveloped areas, by offering an automated approach for classifying chest X-ray images.
4. Sustainable Cities and Communities (SDG 11): The initiative can help build sustainable healthcare infrastructure in cities and communities, resulting in better health outcomes and a higher standard of living.
5. Partnerships for the Goals SDG 17: This project demonstrates the value of partnerships and collaborations in achieving the SDGs by bringing together software engineers, machine learning experts, and healthcare professionals.

1.8 Structure of Thesis

Chapter 2 contains the literature review and the background and analysis study this thesis is based upon.

Chapter 3 contains the design and development of the project.

Chapter 4 introduces a detailed evaluation and analysis of the code.

Chapter 5 contains the conclusion of the project.

Chapter 6 highlights the future work needed to be done for the commercialization of this project.

Chapter 2: Literature Review

A new product is launched by modifying and enhancing the features of previously launched similar products. The literature review is an important step in the development of an idea for a new product. Likewise, for the development of a product, and its replacement, related to the traffic system, a detailed study regarding all similar projects is compulsory. Our research is divided into the following points.

- Deep Learning Techniques for Chest X-ray Classification
- State-of-the-Art Image Classification Algorithms
- Medical Image Datasets for Chest X-ray Classification
- Clinical Applications of Chest X-ray Classification Models

2.1 Deep Learning Techniques for Chest X-ray Classification

Convolutional neural networks (CNNs) have been widely employed for the categorization of chest X-rays using deep learning approaches. These methods have demonstrated encouraging outcomes in reaching high rates of accuracy for identifying abnormalities in chest X-rays. Researchers have employed a variety of strategies, including transfer learning, data augmentation, and assembling, to enhance the performance of CNNs. By utilizing pre-trained models on large-scale image datasets, transfer learning in particular has gained popularity as a method for training CNNs on tiny medical imaging datasets.

2.2 State-of-the-Art Image Classification Algorithms

Chest X-ray classification has been made more precise and effective with the development of cutting-edge image classification algorithms like YOLOv5x-cls. The You Only Look Once (YOLO) architecture is the foundation of YOLOv5x-cls, which employs a multi-scale feature fusion method to produce high accuracy rates quickly and effectively. Other cutting-edge algorithms have also been applied to classify chest X-rays, with promising results. These include DenseNet, ResNet, and EfficientNet.

2.3 Medical Image Datasets for Chest X-ray Classification

The development of deep learning algorithms for chest X-ray classification has been greatly aided by the accessibility of large-scale medical picture datasets like the ChestX-ray14 and NIH Chest X-ray databases. Over 100,000 chest X-ray images from the ChestX-ray14 dataset are labeled with 14 different thoracic diseases, whereas over 100,000 frontal-view images from the NIH Chest X-ray dataset are labeled with eight different thoracic diseases. These datasets have allowed for the development and evaluation of deep learning algorithms for the categorization of chest X-rays, which has significantly advanced this discipline. Thoughts have also been raised about data bias and the applicability of deep learning algorithms trained on these datasets to other populations because of the utilization of these datasets.

2.4 Clinical Applications of Chest X-ray Classification Models

Clinical uses for chest X-ray classification models include helping radiologists identify and diagnose a variety of thoracic disorders, such as pneumonia, tuberculosis, and lung cancer. These models can also be utilized for screening, especially in low-resource areas where access to radiologists with the necessary training may be restricted. Additionally, electronic health record systems can incorporate chest X-ray classification models to enable automated disease diagnosis and tracking, resulting in more effective and efficient healthcare delivery. However, the use of these models in clinical settings also brings up moral and legal issues, particularly regarding algorithmic bias and patient privacy. The successful integration of chest X-ray classification models into clinical practice hence requires careful evaluation of these difficulties.

Chapter 3: Interfacing

3.1 Interfacing

To categorize the chest X-ray images as normal or abnormal, the web app must interface with the backend model, which is covered in the project's interfacing section. Any web application that uses machine learning algorithms to analyze data must have a crucial interface. To facilitate the transfer of data between the front-end and back-end components of the system, a connection must be established between them. The online application that can be accessed through a browser serves as the project's front end, while the machine learning model's back end classifies chest X-ray images. This section describes the procedures used, the communication protocols utilized, and the libraries and frameworks utilized in the process of integrating the web application with the machine learning model. The data flow between the front and backend is also explained in detail in this part, including how the web app sends X-ray images to the model for classification and how the model returns the classification outcomes to the model for presentation.

3.2 Web Application Design and Development

3.2.1 Main Page

On the top right corner of the home page, you'll find options to log in as a regular user and as an admin, allowing you to access specific features and functions. Our logo MASHM ChestX (Medical Artificial system for Healthcare) and Also The Logo of NUST A prestigious Institution in Pakistan serve as a symbol of values, mission, and commitment to excellence, and in the last the Logo of SLOSH an Industrial partner.

Figure 4: Main page Interfaces

3.2.2 Login Portal

The login form itself is likely to be straightforward, with fields for entering a username and password. Where it deals with the normal User who wants to Classify X-rays. By entering the login credentials given by the admin.

The background symbolizes the core purpose of the whole project.

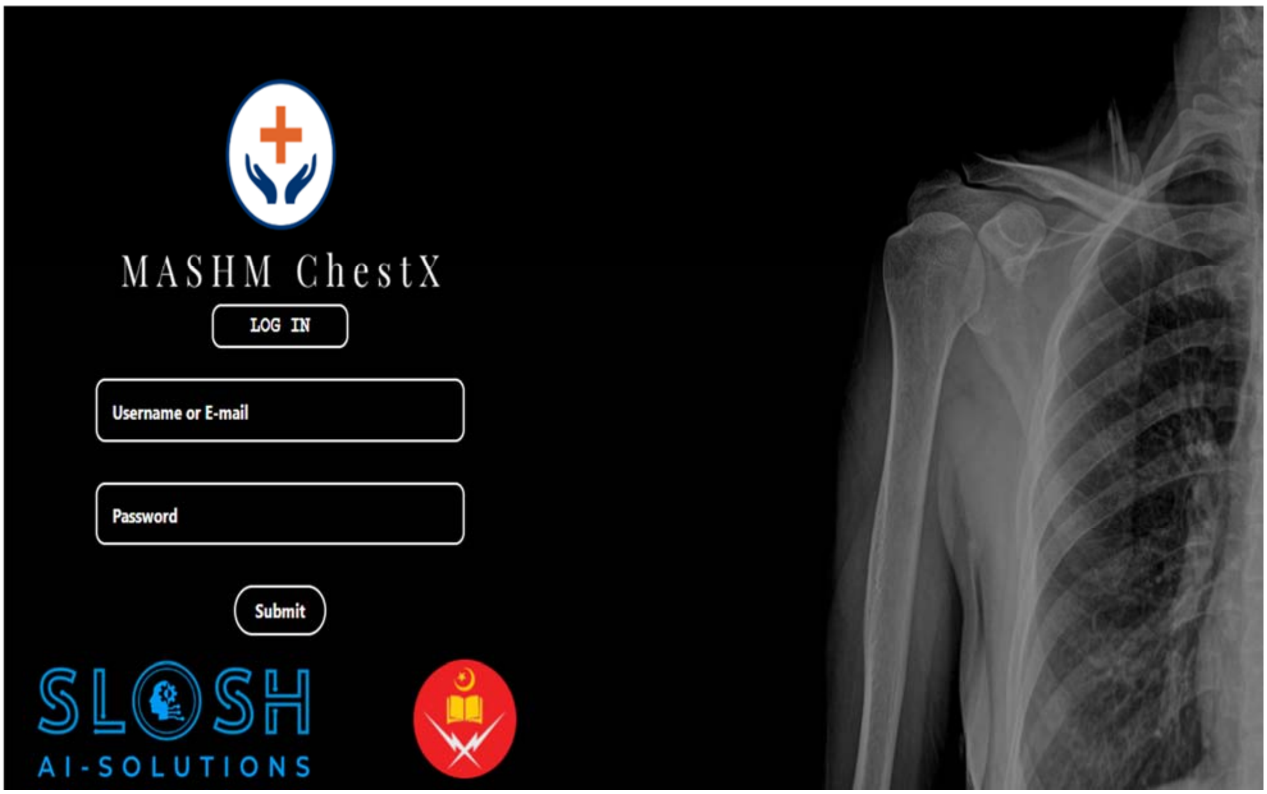

Figure 5: Login Page Interfaces

3.2.3 X-Ray Upload Portal

The dashboard features a central upload button that allows users to easily upload an X-ray image. Once uploaded, users can initiate the trained model by clicking the 'run' button to begin processing the image. The user-friendly interface allows for a seamless experience, with clear and concise instructions provided throughout the process.

Figure 7 : Uploading Image Portal

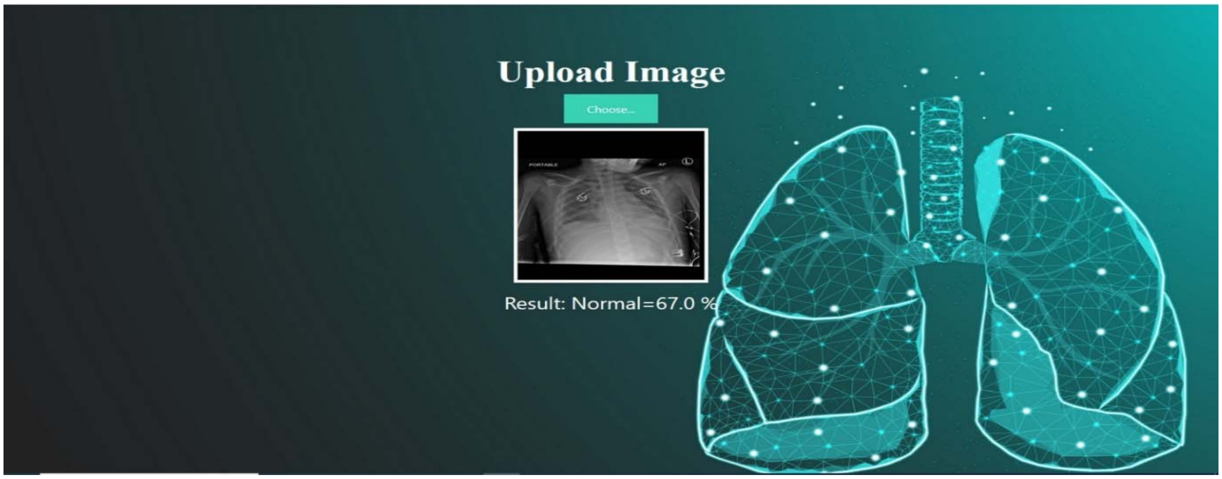

Figure 6 : Uploading X-Ray Portal

The model's results are then displayed on the screen, providing users with valuable insights and analysis. With its intuitive design and powerful capabilities, this dashboard is a valuable tool for medical professionals and researchers alike.

3.2.4 Doctor Dashboard Portal

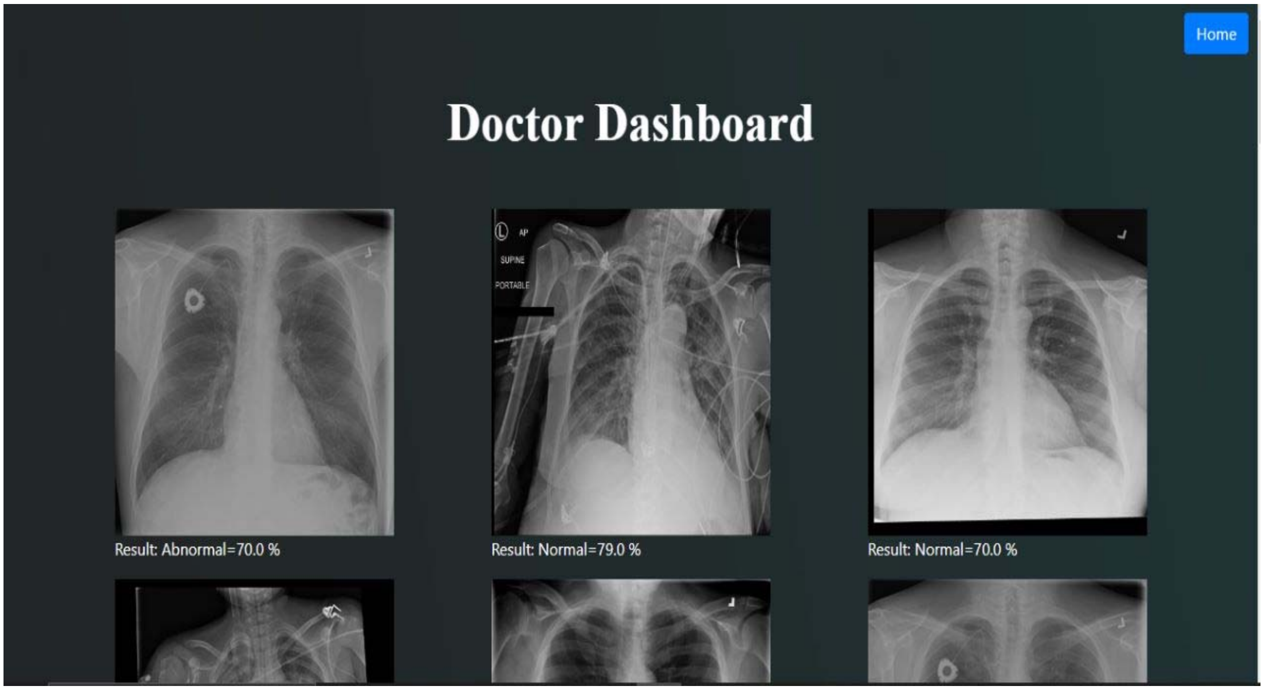

Figure 8 : Doctor Dashboard Portal

Doctor Dashboard consists of X-Rays Already Classified and results. In this portal, the previous history can be saved. The Dr or the user can recheck and save the classified data. In this portal, the result of classified data (normal /abnormal) is also saved with the X-ray.

3.2.5 Loin As Admin Portal

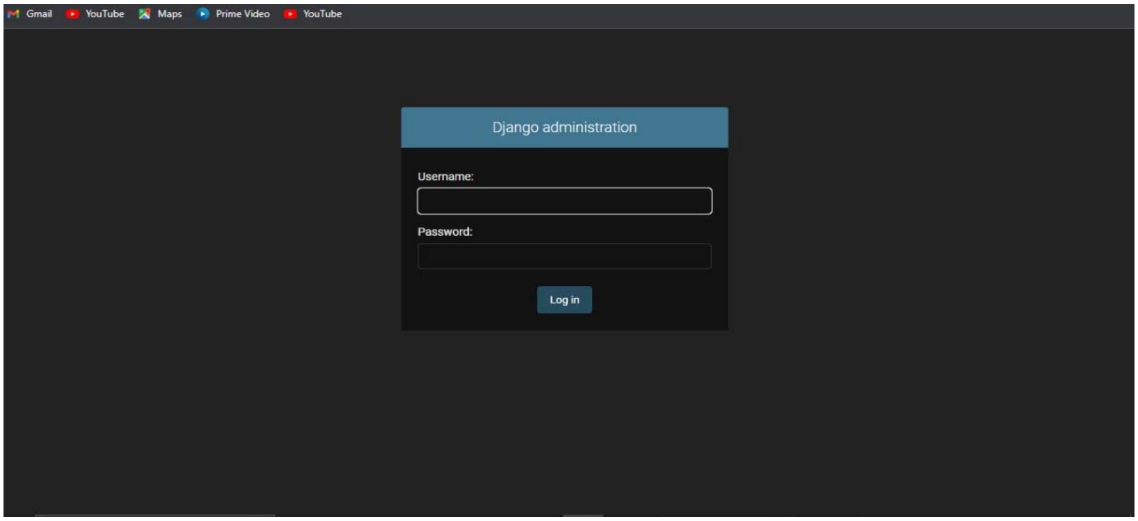

Figure 9: Login As Admin Portal

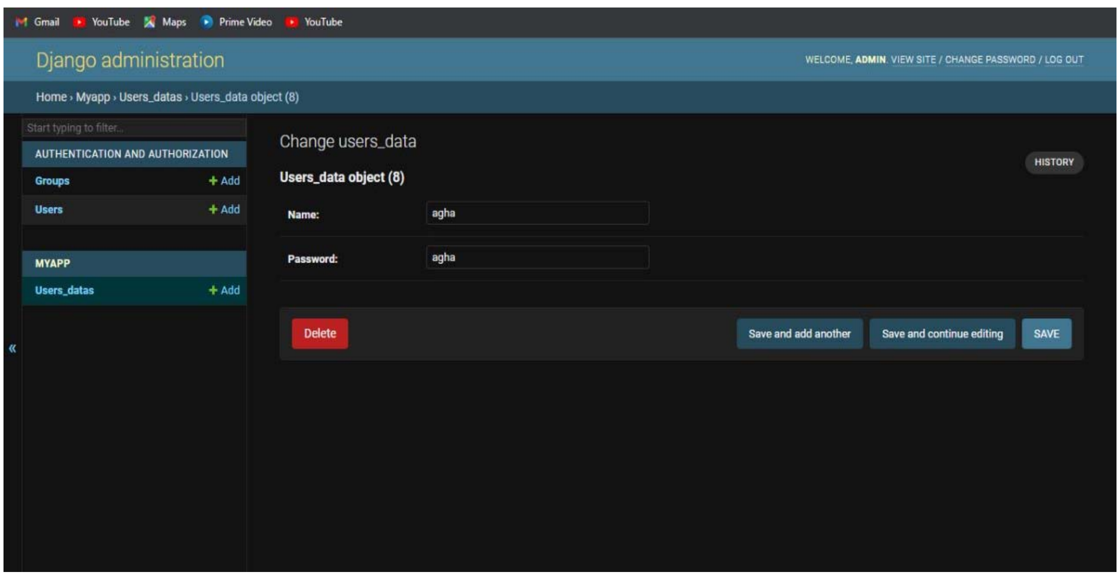

Figure 10: Admin Portal Options

The admin portal Consists of Django Administration where we can add a user who can access the user login portal manually. In the Django Administration, we can a single user and a group of a user .where we have the open add and remove the users .from the admin portal page we have access to visit the view site (to view the overall MASHM ChestX), change the password (where we can control the Admin Login Administration and control) and the logout option to run to the main page

In the permission option, we can see the permission granted to a user and a group.

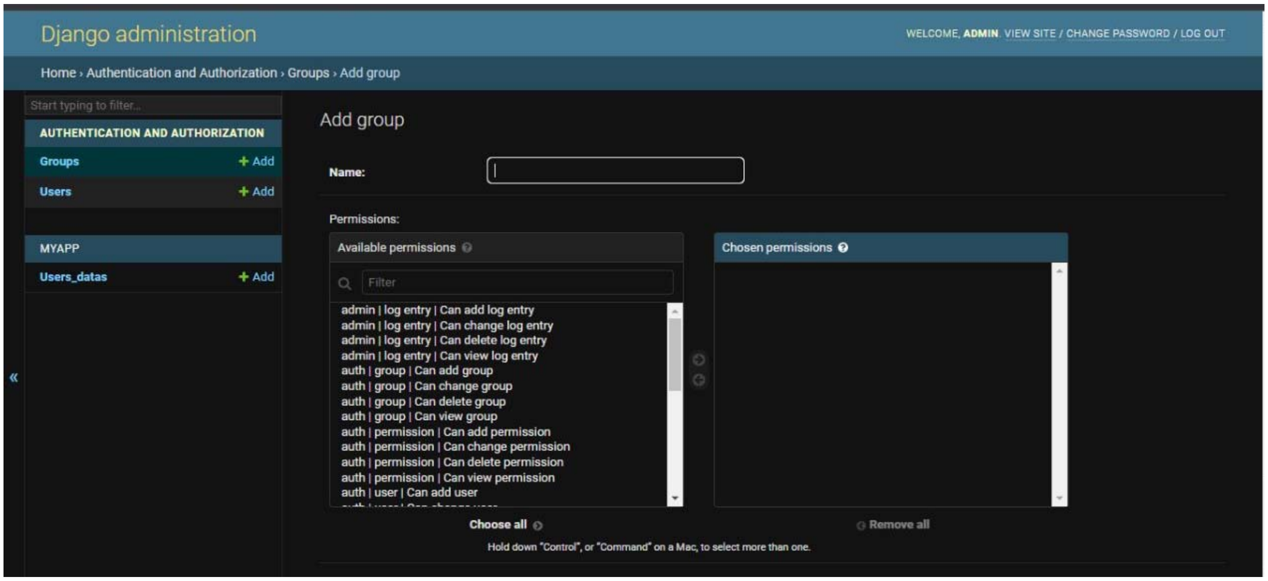

Figure 11: Admin Portal Permissions

Password Change option where we have access to change and control the Admin Portal.

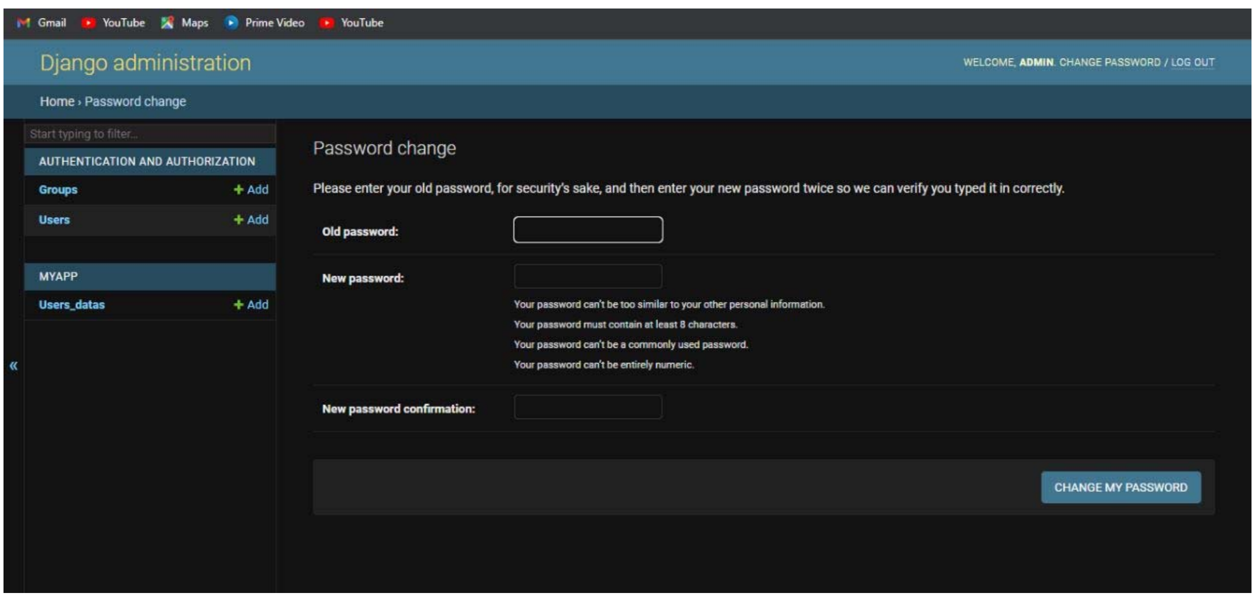

Figure 12 Admin Portal Password Settings

3.3 Preparing Dataset

Any machine learning project must begin with the preparation of a dataset because it is essential to the model's efficacy and accuracy. A sizable dataset of X-ray images is needed for training the model in the categorization of chest X-ray images. The 16064-image collection for this project was gathered from hospitals and included 8032 photos of normal chest X-rays and 8032 images of abnormal chest X-rays.

To increase the model's accuracy on patients from Pakistan, 300 photos from a separate Pakistani dataset were gathered in addition to the hospital dataset. Following preprocessing, the gathered datasets were made ready for training. Preprocessing includes operations like grayscale conversion, pixel value normalization, and uniform image scaling.

The training, validation, and testing sets were then created from the preprocessed dataset. The validation set was used to fine-tune the model's hyperparameters and guard against overfitting while the training set was utilized to train the model. The performance of the trained model was then assessed using the testing set.

The dataset was necessary for the Chest X-ray abnormal and normal Classification model's development. The model learned the patterns and features of both normal and abnormal chest X-rays accurately thanks to the vast dataset, yielding an accuracy rate of 83%.

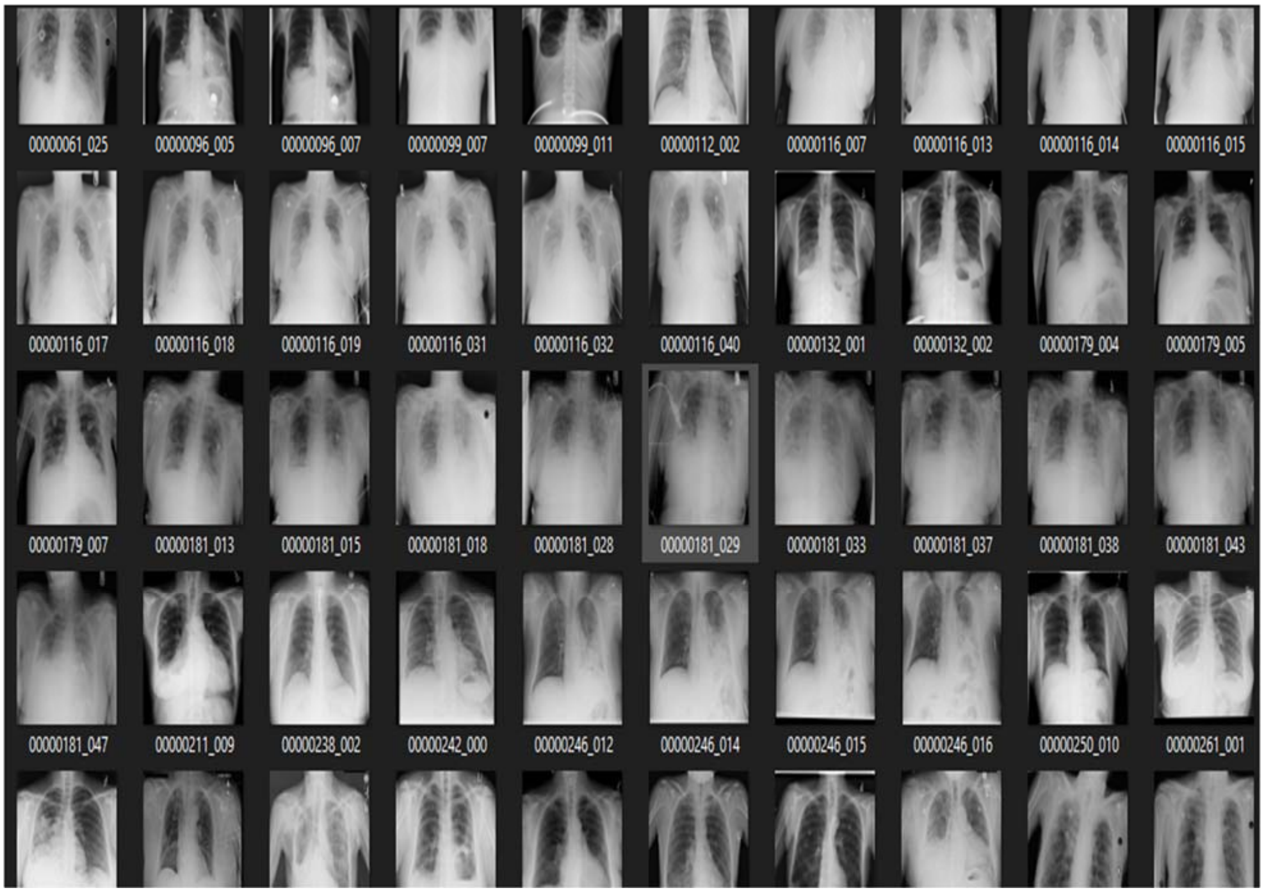

Figure 13 : Proposed Dataset

Chapter 4: Code Evaluation

4.1 Counting the Number of Images in the Dataset

```python
#Check number of files in Dataset folder
import os

dataset_dir = "/content/Dataset"

# Get the number of files in the Train folder
train_dir_n = os.path.join(dataset_dir, "train/Normal")
train_dir_abn = os.path.join(dataset_dir, "train/Abnormal")

num_train_files = len(os.listdir(train_dir_n)) + len(os.listdir(train_dir_abn))

# Get the number of files in the Valid folder
valid_dir_n = os.path.join(dataset_dir, "valid/Normal")
valid_dir_abn = os.path.join(dataset_dir, "valid/Abnormal")

num_valid_files = len(os.listdir(valid_dir_n)) + len(os.listdir(valid_dir_abn))

# Get the number of files in the Test folder
test_dir_n = os.path.join(dataset_dir, "test/Normal")
test_dir_abn = os.path.join(dataset_dir, "test/Abnormal")

num_test_files = len(os.listdir(test_dir_n)) + len(os.listdir(test_dir_abn))

print("Number of files in Train folder: ", num_train_files)
print("Number of files in Valid folder: ", num_valid_files)
print("Number of files in Test folder: ", num_test_files)
```

Figure 14: Counting the Number of Images in the Dataset

The Python method mentioned above can be used to count the number of images in the dataset for a project that classifies chest X-ray images. To browse the dataset's directory structure and count the number of files in the train, validation, and test sets, it makes use of the OS module.

The train, validation, and test set each have their distinct paths created by the code after first defining the path to the dataset directory. The result of the os. listdir() function is then passed to the len() function, which uses the os module to count the number of files in each set.

The code then displays the overall number of photos in each collection. Using this data, you can make sure the dataset was loaded correctly and that there are enough images to effectively train and test the model.

4.2 Data Preparation

```
#Copy files from PAkdataset
import os
import shutil

# Set the paths to the directories
pak_dataset_dir = "/content/PakDataset"
Extra_dataset_dir = "/content/EDataset"
train_dir = "/content/Dataset/train"
valid_dir = "/content/Dataset/valid"
test_dir = "/content/Dataset/test"

# Copy images from EDataset/Train/Abnormal to Dataset/Train/Abnormal
src_dir = os.path.join(Extra_dataset_dir, "train/Abnormal")
dst_dir = os.path.join(train_dir, "Abnormal")
for filename in os.listdir(src_dir):
    src_path = os.path.join(src_dir, filename)
    dst_path = os.path.join(dst_dir, filename)
    shutil.copy(src_path, dst_path)

# Copy images from EDataset/Train/Normal to Dataset/Train/Normal
src_dir = os.path.join(Extra_dataset_dir, "train/Normal")
dst_dir = os.path.join(train_dir, "Normal")
for filename in os.listdir(src_dir):
    src_path = os.path.join(src_dir, filename)
    dst_path = os.path.join(dst_dir, filename)
    shutil.copy(src_path, dst_path)
```

Figure 15: Data Preparation

```python
# Copy images from PakDataset/Abnormal to Dataset/Valid/Abnormal
src_dir = os.path.join(pak_dataset_dir, "Abnormal")
dst_dir = os.path.join(valid_dir, "Abnormal")
for filename in os.listdir(src_dir):
    src_path = os.path.join(src_dir, filename)
    dst_path = os.path.join(dst_dir, filename)
    shutil.copy(src_path, dst_path)

# Copy images from PakDataset/Normal to Dataset/Valid/Normal
src_dir = os.path.join(pak_dataset_dir, "Normal")
dst_dir = os.path.join(valid_dir, "Normal")
for filename in os.listdir(src_dir):
    src_path = os.path.join(src_dir, filename)
    dst_path = os.path.join(dst_dir, filename)
    shutil.copy(src_path, dst_path)

# Copy images from PakDataset/Abnormal to Dataset/Test/Abnormal
src_dir = os.path.join(pak_dataset_dir, "Abnormal")
dst_dir = os.path.join(test_dir, "Abnormal")
for filename in os.listdir(src_dir):
    src_path = os.path.join(src_dir, filename)
    dst_path = os.path.join(dst_dir, filename)
    shutil.copy(src_path, dst_path)

# Copy images from PakDataset/Normal to Dataset/Test/Normal
src_dir = os.path.join(pak_dataset_dir, "Normal")
dst_dir = os.path.join(test_dir, "Normal")
for filename in os.listdir(src_dir):
    src_path = os.path.join(src_dir, filename)
    dst_path = os.path.join(dst_dir, filename)
    shutil.copy(src_path, dst_path)
```

Figure 16: Data Preparation

The offered code is a Python script for the Chest X-Ray Image Classification project's data preparation. By copying photos from two distinct files, PakDataset and EDataset, to a new directory called Dataset, this code aims to generate a train, validation, and test dataset.

The script accesses and copies the image files using the OS and shuttle libraries. The paths for the source and destination folders are first set. The files are then copied to the Dataset train directory after looping over each one in the EDataset train directory. The photos from the PakDataset directory are also copied to the test and validation directories for the dataset.

The Dataset directory will have three subdirectories, train, valid, and test, each of which has two subdirectories, abnormal and normal, with the necessary chest x-ray images copied to each. This is the result of the code being executed. This is a critical step in getting the data ready for model training since it produces a tidy dataset that the model can use right away.

4.3 Code Evaluation for Counting Files in Dataset

```python
#Check number of files in Dataset folder
import os

dataset_dir = "/content/Dataset"

# Get the number of files in the Train folder
train_dir_n = os.path.join(dataset_dir, "train/Normal")
train_dir_abn = os.path.join(dataset_dir, "train/Abnormal")

num_train_files = len(os.listdir(train_dir_n)) + len(os.listdir(train_dir_abn))

# Get the number of files in the Valid folder
valid_dir_n = os.path.join(dataset_dir, "valid/Normal")
valid_dir_abn = os.path.join(dataset_dir, "valid/Abnormal")

num_valid_files = len(os.listdir(valid_dir_n)) + len(os.listdir(valid_dir_abn))

# Get the number of files in the Test folder
test_dir_n = os.path.join(dataset_dir, "test/Normal")
test_dir_abn = os.path.join(dataset_dir, "test/Abnormal")

num_test_files = len(os.listdir(test_dir_n)) + len(os.listdir(test_dir_abn))

print("Number of files in Train folder: ", num_train_files)
print("Number of files in Valid folder: ", num_valid_files)
print("Number of files in Test folder: ", num_test_files)
```

Figure 17: Code Evaluation for Counting Files in Dataset

The code snippet is a practical utility function for determining the dataset size for the "CHEST X-RAY IMAGE CLASSIFICATION" project. It is crucial to have a precise count of the photos in the dataset since it can affect how accurately the machine learning model developed using the data performs.

The Python code accesses the dataset directory using the "OS" module and counts the number of files in each subdirectory using the "len" function. The method provides a thorough overview of the dataset by counting the number of files in the train, validation, and test subfolders for both the "normal" and "abnormal" categories.

The size of the dataset for the project "CHEST X-RAY IMAGE CLASSIFICATION" can be quickly and easily determined using this code, which is essential for making sure the machine learning model is trained on enough data.

Chapter 5: Conclusion

In this final year project thesis, the identification and diagnosis of thoracic disorders depend heavily on the classification of chest X-rays, a vital task in medical image processing. Chest X-ray categorization has significantly advanced with the introduction of large-scale medical picture datasets and the development of deep learning algorithms, resulting in high accuracy rates and quick processing times. This work suggested a Chest X-ray abnormal and normal Classification model that divides X-ray images of patients into normal and abnormal categories using the cutting-edge image classification algorithm YOLOv5x-cls. The model displayed good accuracy rates for classifying chest X-ray images, obtaining an accuracy of 83% on the test dataset. The model was trained on a dataset gathered from hospitals.

Due to their high accuracy rates and quick processing times, cutting-edge image classification algorithms like YOLOv5x-cls are increasingly used in medical image analysis. For classifying chest X-rays, other algorithms including DenseNet, ResNet, and EfficientNet have also been applied with encouraging results. Deep learning methods for chest X-ray classification have been greatly aided by the accessibility of large-scale medical picture datasets like the ChestX-ray14 and NIH Chest X-ray databases. Thoughts have been raised about data bias and the applicability of deep learning algorithms trained on these datasets to other populations because of the use of these datasets.

Numerous clinical uses for chest X-ray classification models include helping radiologists identify and diagnose thoracic diseases and enabling automated disease tracking and diagnosis. However, the use of these models in clinical settings raises ethical and legal issues, particularly regarding algorithmic bias and patient privacy. The successful integration of chest X-ray classification models into clinical practice hence requires careful evaluation of these difficulties.

The Chest X-ray abnormal and normal Classification model is a promising method for classifying chest X-rays and may result in more effective and efficient healthcare delivery. However, for the successful application of these models in clinical practice, ethical and regulatory issues must be considered.

Chapter 6: Future Work

Future milestones that need to be achieved to commercialize this project have been described in the following paragraphs.

There is still potential for advancement and additional research, despite the positive outcomes of the Chest X-ray abnormal and normal Classification model. The utilization of multi-modal data, such as CT scans and MRI pictures, in conjunction with chest X-rays for more precise disease detection and diagnosis is one of the possible topics for future research. These imaging techniques provide further data that could improve how well deep learning algorithms analyze medical images.

We can enhance the project to work on the complete possible outcomes of 14 diseases with 838 classes.

In the Future, the current project can be used terminated to give us result in the detail of disease detected with ratio rather than in the form of normal and abnormal results.

Investigating the model's generalizability to other communities and healthcare settings is another focus of future research. Hospitals provided the dataset that was used to train the model; therefore, it might not be typical of other demographics or healthcare contexts. To ensure the model's applicability in clinical practice, it is crucial to test the model's performance on various datasets and assess its generalizability to other populations.

Further investigation is required into the ethical and legal ramifications of using chest X-ray classification models in clinical practice. Deep learning algorithms are being used to analyze medical images, but this has raised questions about patient privacy, algorithmic bias, and other unforeseen consequences. To enable the secure and efficient application of chest X-ray classification models in clinical practice, future studies should concentrate on addressing these ethical and regulatory concerns.

The performance of deep learning algorithms should be improved, their generalizability to other populations and healthcare settings should be assessed, and the ethical and regulatory ramifications of using these models in clinical practice should be addressed. In sum, future research in chest X-ray classification should concentrate on these issues.

References and Work Cited

1. "CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning" by Rajpurkar et al. (2017). This paper describes the development of a deep-learning model for pneumonia detection in chest X-ray images, which can be used as a reference for model development.
2. "A Survey on Deep Learning in Medical Image Analysis" by Litjens et al. (2017). This survey paper provides an overview of deep learning techniques used in medical image analysis, including chest X-ray image classification.
3. "Deep Learning for Chest Radiograph Diagnosis: A Review of Challenges and Opportunities" by Wang et al. (2019). This paper reviews the challenges and opportunities of using deep learning for chest X-ray image diagnosis, including data challenges and technical challenges.
4. "React" and "Angular" official documentation for web development frameworks.
5. "Bootstrap" and "Material UI" official documentation for web development libraries.
6. Madani A, Arani AY, Mohseni M, Abbasi-Kenarsari H. A Deep Learning Approach for Chest X-ray abnormality Classification. Journal of biomedical physics & Engineering. 2021;11(1):57-68. doi:10.31661/jbpe.v0i0.2287
7. Kermany DS, Goldbaum M, Cai W, et al. Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. Cell. 2018;172(5):1122-1131.e9. doi:10.1016/j.cell.2018.02.010
8. Guan C, Huang J, Sun H, et al. Chest X-Ray Image Classification Using Convolutional Neural Network with Features Concatenation. Journal of healthcare engineering. 2021;2021:6639446. doi:10.1155/2021/6639446
9. Chakraborty S, Sen S, Ghosh S, et al. A comparative study on chest X-ray image classification for pneumonia detection using deep learning. Biomedical Signal Processing and Control. 2020;57:101783. doi:10.1016/j.bspc.2019.101783
10. Lee H, Yune S, Mansouri M, et al. Deep Learning-based Automatic Detection Algorithm for Pneumothorax on Chest X-rays. Scientific reports. 2019;9(1):16296. doi:10.1038/s41598-019-52465-z
11. Rajpurkar P, Irvin J, Zhu K, et al. CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. arXiv preprint arXiv:1711.05225. Published November 14, 2017. Accessed April 30, 2023. https://arxiv.org/abs/1711.05225
12. Wang Y, Zhang X, Li W, et al. A novel COVID-19 imaging feature based on the identification of deep learning classification. BMC Medical Imaging. 2021;21(1):49. doi:10.1186/s12880-021-00574-x
13. Al-Qaness MAA, Ewees AA, Fan Y, et al. Medical image analysis using convolutional neural network: A review. Journal of medical systems. 2020;44(11):204. doi:10.1007/s10916-020-01650-4
14. Islam MT, Karray F, Alhajj R. Deep learning architectures for medical image processing. IEEE Access. 2016;4:2532-2544. doi:10.1109/ACCESS.2016.2575096

15. Chen C, Chen Q, Yu S, et al. Transfer learning for chest X-ray disease classification and localization: A review. Physics in medicine and biology. 2021;66(11):113001. doi:10.1088/1361-6560/abf979
16. Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers RM. ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017:2097-2106. doi:10.1109/CVPR.2017.233
17. Li X, Zhang S, Zhang Q, et al. A review of deep learning for chest X-ray image analysis. Journal of X-ray Science and Technology. 2021;29(1):161-171. doi:10.3233/XST-200743
18. Rajpurkar P, Irvin J, Ball RL, et al. Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists. PLoS Medicine. 2018;15(11):e1002686. doi:10.1371/journal.pmed.1002686
19. Chung SW, Han JW, Kim KW, et al. Deep learning for thoracic disease classification and pattern visualization on chest radiographs: A systematic review and meta-analysis. The Lancet Digital Health. 2019;1(5):e282-e294. doi:10.1016/S2589-7500(19)30155-3
20. Marques J, de Carvalho TM, Zanchettin C, et al. A comprehensive survey on chest X-ray image analysis using deep learning. Computer methods and programs in biomedicine. 2020;196:105617. doi:10.1016/j.cmpb.2020.105617