Mini Project Synopsis Report

on

Improvisation of Chest-Radiograph Classification Using Dimensionality Reduction

By

Bishant Raaj Bhujel (202000224)

Mayal Punu Lepcha (202000283)

Rinchen Tempa Bhutia (202000117)

In partial fulfillment of requirements for the award of degree in Bachelor of Technology in Computer Science and Engineering (2023)

Under the Project Guidance of

Mr. Ashis Pradhan

Assistant Professor (SG)

Dept. of CSE, SMIT

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

SIKKIM MANIPAL INSTITUTE OF TECHNOLOGY

(A constituent college of Sikkim Manipal University)

MAJITAR, RANGPO, EAST SIKKIM – 737136

TABLE OF CONTENTS

ABSTRACT

The Covid-19 pandemic has exposed the limitations of our healthcare systems and the pressure such a situation puts on our medical workforce. The high volume of patients, the shortage of healthcare workers, and limited resources have highlighted the need for efficient and effective diagnostic tools to improve the overall efficiency and effectiveness of the diagnostic process.

One such diagnostic tool is a Chest Radiograph Classifier. Chest Radiographs are a valuable diagnostic tool as they are fast to obtain, easily accessible, and can be used to detect the presence of a variety of lung diseases. Although a variety of image classification models are available, they still suffer from a major problem: the problem of high dimensionality of data.

The high dimensionality of the Chest Radiograph images can make it difficult for traditional machine learning algorithms to classify them effectively. To tackle this problem, we aim to use various dimensionality reduction algorithms to reduce the dimensionality of the images and extract the most relevant features for effective disease classification for our model.

The goal of this project is to build a chest radiograph classifier that not only accurately classifies diseases based on an input x-ray image, but also addresses the problem of high dimensionality. By reducing the dimensionality of the images, the proposed classifier aims to not only improve the accuracy of disease classification but also improve the computational efficiency of the model making it more practical for real-world applications.

INTRODUCTION

Chest radiographs, also known as chest X-rays, are a valuable diagnostic tool in the medical field. They are widely used to detect the presence of a variety of lung diseases, such as pneumonia, tuberculosis, and lung cancer. Chest radiographs are fast to obtain, easily accessible, and relatively inexpensive, making them an ideal tool for screening and diagnosis. They can provide a wealth of information about a patient's lung function and overall health, including the size, shape, and position of the lungs and the heart, as well as the presence of fluid, masses, and other abnormalities.

Chest Radiograph Classifier (Image Classification Model) can be an effective diagnostic tool that can be used in detecting and classifying lung infections such as COVID-19, Pneumonia and Tuberculosis and hence can assist doctors in interpreting chest X-Rays. It can help to reduce the workload of radiologists by assisting in the process of medical image analysis which can improve the efficiency of the diagnostic process.

However, such Classifiers still have certain limitations that can affect their performance and accuracy. In our project, we plan to address the limitation of high dimensionality in chest x-ray images by using dimensionality reduction techniques. Chest X-rays are high-dimensional medical images, meaning that they contain many pixels and features. This high dimensionality can create several problems for learning algorithms when it comes to analyzing and classifying these images. In addition, High-dimensional images require more computational resources and time to process, which can make it difficult to develop and implement real-time diagnostic tools.

LITERATURE SURVEY

SL No.	Author(s)	Paper and Publication Details	Findings	Relevance to the Project
1.	*Tawsifur Rahman, Amith Khandakar, Yazan Qiblawey, Anas Tahir.*	"Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images." Computers in Biology and Medicine 132, Elsevier. [4th March 2021]	Study confirmed that deep learning models can be used to accurately detect COVID-19 from chest X-ray (CXR) images.	Dataset being used was produced by this research.
2.	*Thet Thet Khaing, Phyu Sin Nyein, Myint Soe Khyaing, Khaing Khaing Wai*	"Dimension Reduction of Images Using Principal Component Analysis Algorithm" Iconic Research and Engineering Journals [May 2020]	Study found that PCA effectively reduced the file size of the images and improved the transmission time for the compressed images.	Provides algorithm for applying PCA to images.
3.	*Ayesha, Shaeela; Hanif, Muhammad Kashif;*	"Overview and Comparative Study of Dimensionality Reduction Techniques for High Dimensional Data" Information Fusion [Jan 2020]	Study found that Linear techniques are less computationally intensive but nonlinear techniques can be useful for complex data.	Discusses and compares various dimensionality reduction techniques.
4.	*Taufit Rahmat, Azlan Ismail, Sharifah Aliman*	"Chest X-Rays Image Classification in Medical Image Analysis" Universiti Teknologi, Malaysia. [27th December 2018]	Study discusses approaches for classifying chest X-ray images, including the classification problem types, datasets used, splitting ratios, etc.	Discusses various approaches for CXR image classification.

PROBLEM DEFINITION

- There is a need for an accurate and efficient diagnostic tool to assist medical professionals and improve the diagnostic process and hence reduce the burden on healthcare workers. Chest radiographs are a widely used and accessible tool for detecting a variety of lung conditions, and their importance has been highlighted during the Covid-19 pandemic with the high volume of patients, shortage of healthcare workers, and limited resources. Developing an effective diagnostic tool such as a Chest X-Ray Classifier is crucial for addressing the pressing needs of our healthcare system.

- The problem of high dimensionality in chest radiograph images is a significant challenge in the classification of lung diseases and image classification models in general. High-dimensional data can make it difficult for traditional machine learning algorithms to process and classify the images effectively, resulting in lower accuracy and higher computational costs. Hence there is a need for an efficient computational system that can in turn be cost effective without compromising its performance or accuracy.

SOLUTION STRATEGY

Our project can be divided into mainly three parts:

1) Selection of Dataset and Preprocessing.
2) Using Dimensionality Reduction Algorithms on CXR Image Dataset for feature extraction.

   Algorithms: PCA (Principal Component Analysis)

   ICA (Independent Component Analysis)

3) Building Chest Radiograph Classifier on the dimensionality-reduced dataset.

   Algorithms: CNN (Convolutional Neural Networks)

In addition, on successful completion of the above two tasks, we will also deploy a web application created using ReactJs where the said model will be integrated through Flask (Python). Users will directly be able to upload CXR Images and get instant results from our model.

General Methodology:

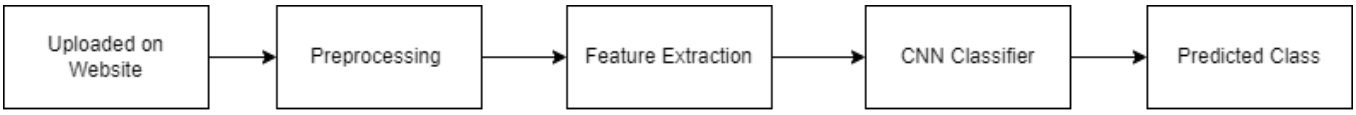
STEP 1: Download, Preprocess the dataset which involves tasks like resizing images, splitting data into train, validation and test sets.

STEP 2: Use dimensionality reduction algorithms like Principal Component Analysis (PCA)/ Independent Component Analysis (ICA) to reduce the dimensionality of the image data.

STEP 3: Train a machine learning model on the reduced-dimensionality image data. This model will be used to predict the labels of new chest X-ray images.

STEP 4: Evaluate the performance of the trained model on the test set and make any necessary adjustments to the model or the preprocessing steps.

STEP 5: Deploy the model on a web application (ReactJS) by creating a REST API using Flask (Python). The API will take in an image file and return the predicted label.



System Flow

- DATASET

  The dataset for our model is titled "COVID-19 Radiography Database" and it contains around 20,000 annotated CXR images.

  Classes = [Covid-19, Normal, Lung Opacity (Non Covid-19 infection), Viral Pneumonia]

  Hence, when given a CXR image our model will predict one of the above classes.

  Illustration below:

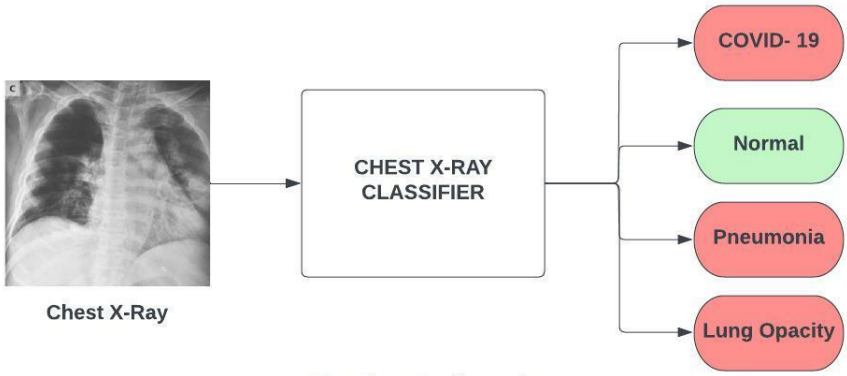


Fig: Chest Radiograph Classifier

The details of our dataset are as follows:

20,000 total CXR Images

1) 6000 X-rays showing lung opacity
2) 3600 X-rays with COVID-19
3) 10,000 X-rays of normal cases
4) 1345 X-rays showing pneumonia.

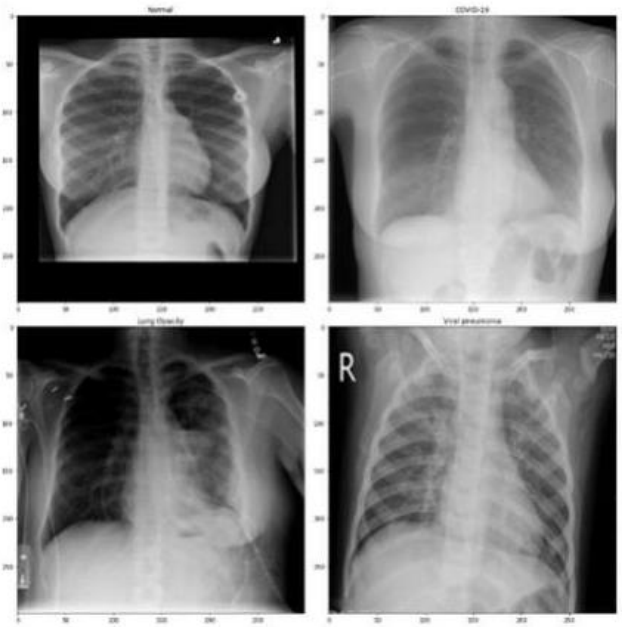
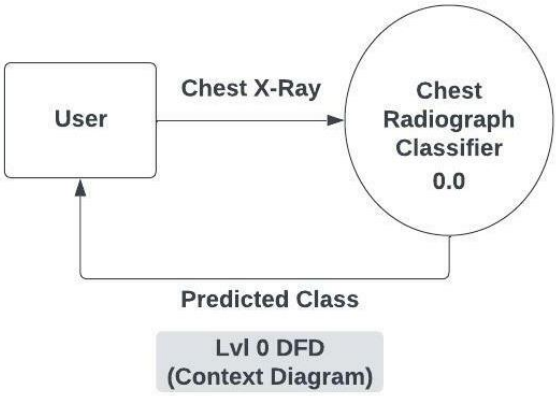


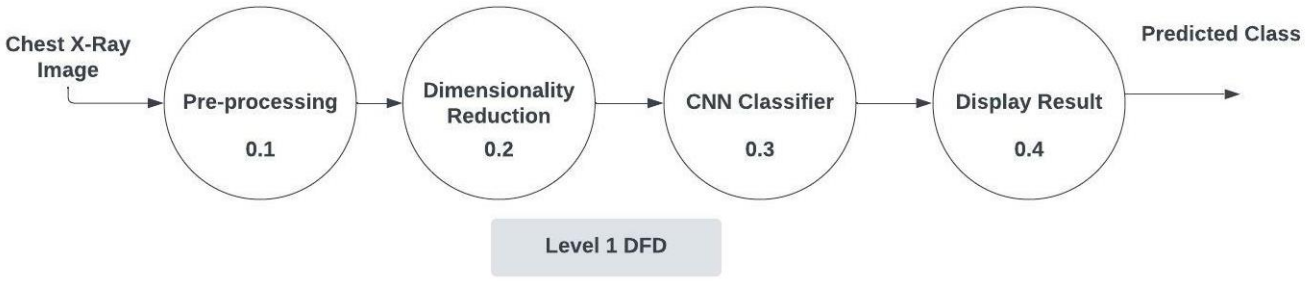
Fig 1: Samples from Dataset

DESIGN DIAGRAMS

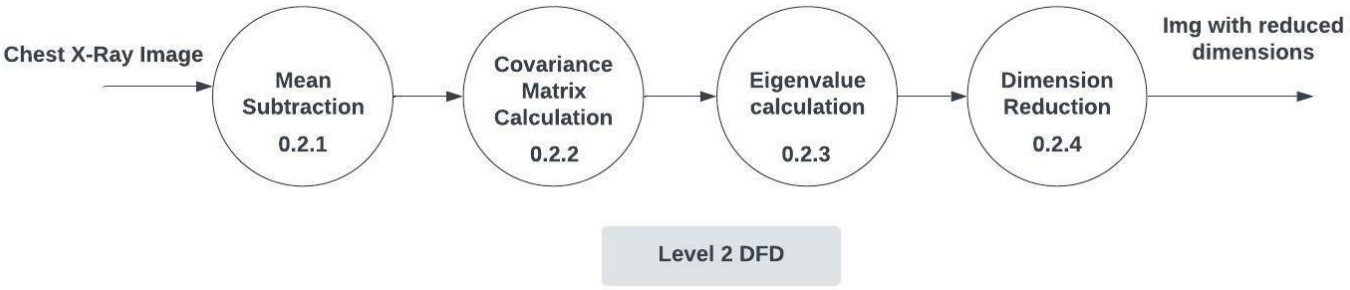
DATA FLOW DIAGRAMS



This is a level 0 DFD (context diagram) for our model, which is essentially a glance view of our system. Here, the user gives an input to our model, which is a CXR image, and the model gives the output as a predicted class.



This is a level 1 DFD for our model which shows the processes that our CXR image will typically go through. Firstly, we have basic preprocessing then dimensionality reduction after which the reduced image will be passed on to the CXR Classifier and the result will be displayed to the user.



Level 2 DFD

This is a level 2 DFD for our system. Here we have decided to expand the Dimensionality Reduction process from our Level 1 DFD. Here, after various processes the image with reduced dimensions is produced.

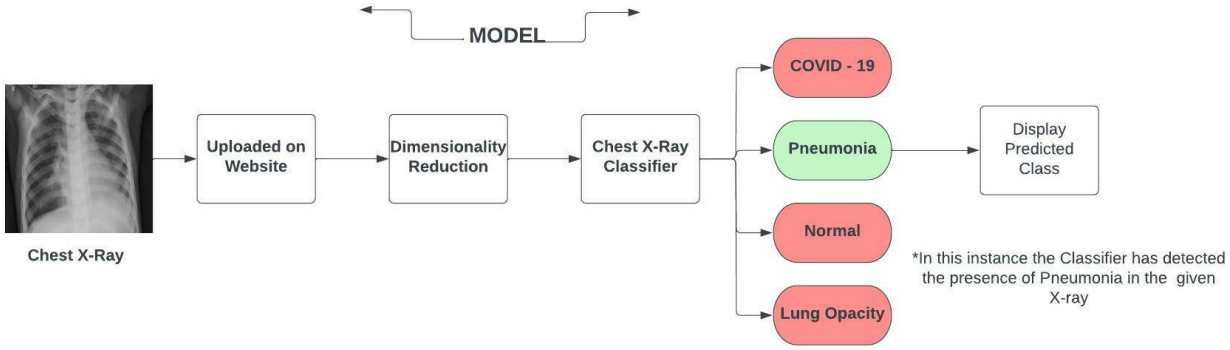


Fig : Glimpse of System

This figure depicts a glimpse of our proposed system. Here the user will upload a CXR Image through our website and it will be passed on to the model where the output (predicted class) will be returned to the user.

GANTT CHART

ACTIVITY	December 2022	January 2023	February 2023	March 2023	April 2023
Literature Survey					
Problem Definition					
Design and Development					
Testing and Validation					
Documentation					

	PROPOSED WORK
	ONGOING WORK
	COMPLETED WORK

REFERENCES

1. *Tawsifur Rahman, Amith Khandakar, Yazan Qiblawey, Anas Tahir*, "Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images", [March 2021]
   Link: https://www.sciencedirect.com/science/article/pii/S001048252100113X

2. *Thet Thet Khaing, Phyu Sin Nyein, Myint Soe Khyaing, Khaing Khaing Wai,* "Dimension Reduction of Images Using Principal Component Analysis Algorithm ", [May 2020]
   Link: https://www.irejournals.com/formatedpaper/1702269.pdf
3. *Ayesha, Shaeela; Hanif, Muhammad Kashif*, "Overview and Comparative Study of Dimensionality Reduction Techniques for High Dimensional Data", [January 2020]
   Link: https://www.sciencedirect.com/science/article/abs/pii/S156625351930377X
4. *Taufit Rahmat, Azlan Ismail, Sharifah Aliman,* "Chest X-Rays Image Classification in Medical Image Analysis" [December 2018]
   Link: https://www.researchgate.net/publication/330105218_CHEST_X-RAYS_IMAGE_CLASSIFICATION_IN_MEDICAL_IMAGE_ANALYSIS