RADIOGRAPHY DATASET COVID-19 PRE-PROCESSING,PCA,LINEAR SVM

Project Report

Submitted By

Name: Shubhanshu Singh Rathore
Roll No: 25CSM1S07
Course: M.Tech (Computer Science)
Semester: — 1st
Institute: —NIT Warangal

Submitted To

Name: _Prof. Kadambari ma'am
Department: Computer Science & Engineering

---

1. Abstract

This project focuses on implementing **Principal Component Analysis (PCA)** for image dimensionality reduction. In modern machine learning, high-dimensional image data leads to computational complexity and slower processing. PCA is used to reduce the number of features while preserving the most important information.

In this project, images are preprocessed, converted into feature vectors, and then transformed using PCA to reduce dimensionality. The reduced features can be used for further machine learning tasks like classification or clustering with improved efficiency.

---

2. Objectives

The main objectives of the project are:

1. To load and preprocess image datasets.
2. To implement PCA from scratch using NumPy.
3. To reduce image dimensionality while preserving maximum variance.
4. To save reduced features for future machine learning tasks.

---

3. Tools and Technologies Used

Component	Description
Language	Python
IDE	Jupyter Notebook
Libraries	NumPy, OS, Pathlib
Technique	Principal Component Analysis (PCA)
Dataset	Image Dataset from local directory

---

4. Methodology

The methodology of the project consists of the following steps:

4.1 Image Loading

Images are loaded from a structured directory. The dataset contains training and testing images stored inside different folders.

4.2 Preprocessing

- Images are resized to **128 × 128 pixels**.
- Each image is flattened into a one-dimensional vector.
- Data is normalized for PCA computation.

4.3 PCA Implementation

Steps followed for PCA:

1. Compute Mean Vector
2. Subtract Mean from Data
3. Compute Covariance Matrix
4. Calculate Eigenvalues & Eigenvectors
5. Sort Eigenvectors based on eigenvalues
6. Select top **100 principal components**

7. Project original data into reduced space

4.4 Feature Transformation

Both **training** and **testing** datasets are transformed using the computed PCA components.

5. System Architecture / Workflow

Image Dataset

↓

Image Preprocessing

↓

Flattening into Vectors

↓

Mean Centering

↓

Covariance Matrix

↓

Eigen Decomposition

↓

Principal Components Selection

↓

Dimension Reduction

↓

Saved Reduced Features

6. Implementation Details

Important parameters used:

Parameter	**Value**
Image Size	128 × 128

Parameter	Value
PCA Components (K)	100
Output Directory	/content/pca_output

The outputs generated:

- X_train_reduced.npz
- X_test_reduced.npz
- pca_components.npy
- pca_mean.npy
- explained_variance.npy

---

7. Results and Discussion

After applying PCA:

- The image dimensionality is reduced significantly.
- Storage size is reduced.
- Computation becomes faster.
- Major variance is still preserved using top 100 components.

This makes the dataset suitable for efficient training of machine learning models.

---

8. Conclusion

This project successfully demonstrates the use of PCA for dimensionality reduction in image datasets. By reducing large image features into meaningful components, we can improve both computation speed and memory efficiency while preserving key information.

This technique can be further extended to:

- Face Recognition
- Medical Image Analysis

- Object Detection Tasks

---

9. Future Scope

1. Apply PCA + Classifiers like SVM or KNN.
2. Compare PCA performance with Autoencoders.
3. Use larger and real-time datasets.

---

10. References

1. I.T. Goodfellow, Deep Learning, MIT Press
2. Scikit-learn Documentation – PCA
3. Online Research Papers on Dimensionality Reduction