ML PROJECT MINI REPORT

Thoracic Disease Detection

1. Introduction

Early and accurate detection of thoracic diseases such as pneumonia, atelectasis, and pleural effusion from Chest X-ray (CXR) images is vital for effective treatment. This project develops an automated diagnostic system that leverages **Deep Convolutional Neural Networks (CNNs)** for feature extraction and **shallow machine learning** for interpretable classification. The task is formulated as a **multi-label classification** problem, as a single X-ray can exhibit multiple diseases.

---

2. Methodology

Dataset: NIH Chest X-ray Dataset containing 14 thoracic pathologies.
Feature Extraction: A CNN is used (with transfer learning) to derive **Bag of Visual Words (BOVW)** feature vectors from CXR images.
Model Architecture:

- **Stage 1 – Deep Feature Extraction:** CNN compresses high-dimensional pixel data into a meaningful feature representation.
- **Stage 2 – Shallow Classification:** An **optimized Logistic Regression (LR)** model performs the final classification using the BOVW vectors.

Disease	ROC-AUC SCORES {Avg score:- 0.7301}
Cardiomegaly	0.6295
Edema	**0.7726**
Emphysema	0.7583
Hernia	0.6752
Pneumonia	0.4882
Fibrosis	**0.4674**
Pneumothorax	0.7301