Pneumonia Detection from Chest X-ray Images

-Vikas Tunge(2022OCTVUGP0031)

Problem Statement

To detect pneumonia in chest X-ray images using both traditional ML methods and deep learning models, with and without preprocessing techniques, and compare their performances.

Dataset

- **Source**: Chest X-ray dataset (data/chest_xray/train)
- **Classes**: Pneumonia and Normal
- **Total Images Used**: 5,216 (approx.)
- **Image Size**: Resized to 128×128

Approaches

1. **Preprocessing Techniques Applied**:
    - Grayscale conversion
    - Gaussian blur
    - Rotation
    - Histogram equalization
    - Fourier transform
    - Pyramid downsampling
2. **Dimensionality Reduction** (for Random Forest):
    - PCA and Truncated SVD (100 components)
3. **Models Used**:
    - **CNN**: Basic 1-layer ConvNet
    - **Random Forest**: With PCA/SVD or raw pixels

Results

With Preprocessing:

Model	Accuracy	Precision	Recall	F1-Score
CNN	98%	0.98	0.98	0.98
RandomForest	91%	0.92	0.91	0.91

Without Preprocessing:

Model	Accuracy	Precision	Recall	F1-Score
CNN	**99%**	0.99	0.99	0.99
RandomForest	**96%**	0.96	0.96	0.96

Insights & Justification

- CNN consistently outperformed Random Forest, with or without preprocessing.
- Surprisingly, the **raw (non-preprocessed)** images gave **better results**, suggesting the CNN was effective at feature learning without manual preprocessing.
- Preprocessing helped Random Forest slightly, but CNN benefited more from raw, unaltered data.
- Dimensionality reduction (PCA/SVD) was essential for traditional models due to high-dimensional input.

Conclusion

Deep learning (CNN) is highly effective for X-ray based pneumonia detection and performs best on raw data. Traditional ML methods like Random Forest can work reasonably well with preprocessing and dimensionality reduction, but are outperformed by CNNs in both accuracy and robustness.