

COMPUTER VISION PROJECT REPORT
LUNG DISEASE DETECTION

NAME: MOHANA KRISHNAN M

ROLL NO: 2023510032

BRANCH: AI AND DS

DEPARTMENT: IT

SUBJECT CODE: AD23402

INTRODUCTION:

In this project, we're building a smart system that looks at chest X-rays and checks for lung diseases like Bacteria pneumonia, Viral pneumonia, COVID-19, or TB. Using computer vision and machine learning, it learns to find patterns in the images, just like a doctor. This can help doctors get faster, more accurate results especially in places with fewer medical experts.

REASON FOR CHOOSING THIS PROJECT

Lung diseases like pneumonia, COVID-19, and TB are still common and dangerous if not detected early. In many places, especially rural areas, there aren't enough doctors to read chest X-rays. Even in hospitals, reading X-rays takes time and skill. So, we decided to build a system using computer vision and machine learning to help detect lung problems faster and more accurately.

METHODOLOGY

The system was developed using image processing and machine learning techniques to classify chest X-ray images into different lung disease categories. The steps involved are as follows:

1. Importing Libraries

Essential Python libraries such as OpenCV, NumPy, scikit-learn, matplotlib, and tqdm were imported to handle image processing, data handling, model building, and visualization.

2. Image Loading and Preprocessing

Chest X-ray images were loaded in grayscale, histogram equalization was applied for better contrast, and all images were resized to a uniform size to ensure consistency. Preprocessing helped reduce noise and standardize the input for the model.

3. Feature Extraction

Histogram of Oriented Gradients (HOG) was used to extract important shape and texture features from the lung images. These features capture the structure of the lungs, which helps in distinguishing between different diseases.

4. Model Training

A Support Vector Machine (SVM) with a linear kernel was used as the classifier. The extracted HOG features were used to train the model to recognize patterns associated with normal lungs and various lung diseases.

5. Model Evaluation

The trained model was evaluated using metrics like accuracy, confusion matrix, and classification report to understand how well it performs on unseen data.

6. Prediction

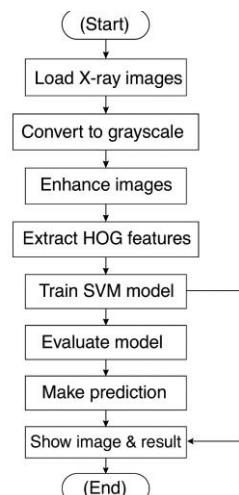
The final model takes a test X-ray image, processes it through the same steps, and predicts the type of lung disease. The prediction is displayed along with the model's accuracy.

Streamlit Web Application

To make our lung disease detection system easy to use, we created a simple web app using **Streamlit**. The interface is clean and user-friendly, so anyone can try it—even without any technical knowledge.

Users can upload a **chest X-ray image**, and then choose which model they want to use for prediction, like **SVM (RBF)**, **SVM (Linear)**, **Random Forest**, or **KNN**. Once the image is processed, the app shows the predicted result—such as whether the lungs are normal or affected by pneumonia, COVID-19, tuberculosis, etc.—along with a **confidence score**.

Architecture Diagram:



EXPLANATION:

The process begins with loading X-ray images, followed by converting them to grayscale for simplification. Next, image enhancement improves clarity, and HOG features are extracted to highlight important patterns. The extracted features are then used to train an SVM model, which is later evaluated for accuracy. Once trained, the model is applied to make predictions on new images, and finally, the image and its result are displayed. If the model performs poorly, adjustments are made, and the training cycle repeats.

DATASET AND ITS DESCRIPTION

The dataset we have chosen is lung disease detection it includes labeled chest X-ray images for **Normal**, **Bacterial Pneumonia**, **Viral Pneumonia**, **COVID-19**, and **Tuberculosis** conditions. These images are used to train the model to classify and detect lung diseases based on visual patterns.

- Dataset Name: Lung Disease Chest X-ray Dataset
- Number of Classes: 5
 - Bacterial Pneumonia
 - Corona Virus Disease
 - Normal (No Disease)
 - Tuberculosis
 - Viral Pneumonia
- Image Types: PNG and JPG formats
- Image Size: All images are resized to 128x128 pixels and converted to grayscale
- Folder Structure:
 - train/ – used to train the model
 - test/ – used to evaluate how well the model performs

WHY THIS DATASET?

- Focuses on realistic and diverse chest X-ray images representing **Normal**, **Bacterial Pneumonia**, **Viral Pneumonia**, **COVID-19**, and **Tuberculosis** conditions.
- Allows for multi-class classification, enabling detection of multiple lung diseases.
- Balanced dataset: Includes both healthy and diseased lung samples for robust training.
- Ideal for training and validating machine learning models using meaningful visual patterns in medical imaging

RESULT:

The model was trained using three different classifiers: **Support Vector Machine (SVM)** with a linear, rbf , **Random Forest Classifier (RFC)**, and **K-Nearest Neighbors (KNN)**. The **SVM (rbf)** model achieved the highest classification accuracy of **79.90%** on the test set of 2025 images. The **RFC** and **KNN** models showed lower accuracy, with results being useful for comparison in understanding model performance.

MODEL ACCURACY TABLE	
Model	Accuracy (%)
SVM (RBF)	79.90
SVM (Linear)	75.51
Random Forest	72.49
K-Nearest Neighbors	63.80

Justification for Model Performance

The **SVM (RBF)** model achieved the highest accuracy of **79.90%** because the RBF kernel captures complex, non-linear patterns in the data, making it well-suited for classifying different lung conditions in X-rays.

The **SVM (Linear)** model performed well with an accuracy of **75.51%**, but it's a bit simpler and may not capture the intricate features in the images as effectively as the RBF kernel.

The **Random Forest Classifier (RFC)** scored **72.49%**, which is decent, but it struggled to detect subtle features necessary for lung disease classification, likely due to the complexity of the X-ray images.

The **K-Nearest Neighbors (KNN)** model had the lowest accuracy of **63.80%**. KNN works well for simpler datasets but tends to struggle with high-dimensional data like X-rays, where relationships are harder to define based on proximity alone

Conclusion

In this project, we built a machine learning system to detect lung diseases from chest X-ray images. Using **SVM with an RBF kernel**, we achieved an accuracy of **79.90%**, showing that the system can support doctors in making faster, more accurate diagnoses. While other models like **SVM (Linear)**, **RFC**, and **KNN** had lower accuracy, they still offered valuable comparisons. With more data and better models, the system's accuracy can improve further. In the future, it could be expanded for real-time use or in areas with limited access to expert doctors.

OUTPUT:



Chest X-ray Disease Classification

Select a model for prediction

SVM (Linear Kernel)

SVM (Linear Kernel)

SVM (RBF Kernel)


K-Nearest Neighbors

Random Forest


Select a model for prediction

SVM (Linear Kernel)

Upload a Chest X-ray Image

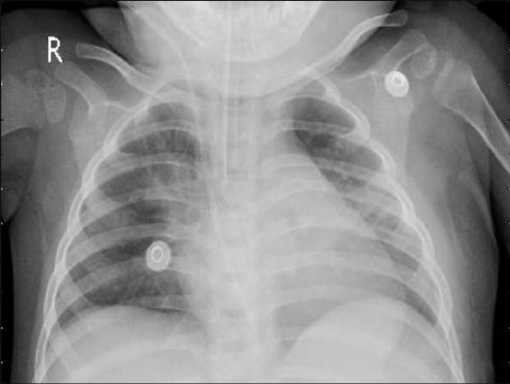
 Drag and drop file here
Limit 200MB per file • JPG, JPEG, PNG

Browse files

 bp.jpeg 13.1KB

×

The use_column_width parameter has been deprecated and will be removed in a future release. Please utilize the use_container_width parameter instead.




Uploaded Chest X-ray

Predict

Predict

✓ Predicted Lung Condition: Bacterial Pneumonia

☑ Confidence: 60.29%

 View Debug Info

Model: SVM (Linear Kernel)

Raw class index: 0

Class probabilities:

- Bacterial Pneumonia: 60.29%
- Corona Virus Disease: 17.77%
- Normal: 1.84%
- Tuberculosis: 0.12%
- Viral Pneumonia: 19.98%

