International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

A Study on Lung Disease Detection System

G. Arunalatha

Submitted: 15/05/2024 **Revised:** 27/06/2024 **Accepted:** 10/07/2024

Abstract: In order to avoid complications, including death, early diagnosis and treatment of lung diseases are essential. The best diagnostic technique currently available is a chest X-ray, which is essential to clinical care. For a person with a lung disease, using deep learning to predict the disease from chest x-rays and chest CT scan may save their life. This is made possible by the instantaneous, high-predictability of the results. In this paper, the various methods for expert lung disease diagnosis is presented. Its main goal is to develop a system that will help radiologists to identify the lung conditions.

***Keywords**: Lung disease, Deep learning, CNN*

1. Introduction:

The human system relies heavily on the expansion and relaxation of the lungs to take in oxygen and expel carbon dioxide. Lung diseases are respiratory conditions that impact the different breathing-related organs and tissues, resulting in conditions affecting the airways, lung tissue, and lung circulation. While some respiratory illnesses, such as the flu and the common cold, only cause minor discomfort and inconvenience, others, such as lung cancer, pneumonia, and tuberculosis, can be fatal and cause severe acute respiratory issues. In India, many people suffer from a variety of lung conditions. These illnesses are most likely caused by smoking, infections, and genetics. In order to breathe oxygen and expel carbon dioxide, the lungs, which are essential organs, expand and contract numerous times throughout the day.

According to epidemiological statistics on respiratory disorders published by the World Health Organization (WHO), there are 210 million cases of chronic obstructive pulmonary disease (COPD) and 30 million cases of asthma worldwide. Studies show that between 15 and 25 million Indians suffer from asthma. Physicians commonly use the non-invasive and reasonably priced lung auscultatory technique to assess the state of the cardiopulmonary system. Early detection of the aforementioned diseases can significantly improve survival rates and reduce the number of human casualties. Common tests that identify the existence of these conditions include CT scans and chest X-ray images.

Assistant Professor, Department of Computer Science and Engineering,
Perunthalaivar Kamarajar Institute of Engineering and Technology (PKIET), Karaikal.
Puducherry.
Email: vigneshgayu1121@gmail.com

As a subset of artificial intelligence with representation learning, deep learning is a subfield of machine learning that offers cutting-edge accuracy. The ability of this tool to read image data, process it, and produce results based on previously trained data has garnered attention recently. In order to classify new test images that the model hasn't previously visualized, deep learning models can utilize the features and patterns they have learned from dataset images. model.

2. Lung Diseases:

I. Tuberculosis (TB)

Bacteria cause tuberculosis (TB), an infectious disease that primarily affects the lungs. When people who have tuberculosis cough, sneeze, or spit, it spreads through the air. Additionally, it may spread to the spine and brain, among other areas of our body. It is caused by a particular kind of bacteria known as Mycobacterium tuberculosis. An estimated 25% of people worldwide are thought to have contracted tuberculosis. Of those infected with TB, 5–10% will eventually experience symptoms and develop TB disease. In 2023, TB killed 1 in 25 million people, including 161,000 HIV-positive individuals. After three years when coronavirus disease (COVID-19) surpassed tuberculosis (TB) as the world's leading infectious agent-related cause of death, TB has most likely returned to its previous position. It was also the main cause of deaths from antimicrobial resistance and the leading cause of death for HIV-positive individuals.

II. Pneumonia

A lung infection called pneumonia can be severe that occurs when an infection causes fluid or pus to fill the alveoli, which are air sacs in our lungs, and the bronchioles, which are tubes in the airways that connect to them. It is difficult to take in enough oxygen as a result. It results in the accumulation of fluid or pus in the lungs' alveoli, or air sacs. Anyone can contract this lung infection from bacteria, viruses, or fungi. However, adults 50 years of age and older and children under the age of two are more vulnerable. This is because they may not have robust immune systems to combat it. Lifestyle choices such as excessive alcohol consumption and cigarette smoking can also increase the risk of contracting pneumonia. In 2019, pneumonia killed 740,180 children under the age of five, making up 14% of all pediatric deaths.

III. Lung cancer

One type of cancer that begins as lung cell growth is lung cancer. The two spongy organs in the chest that regulate breathing are called lungs. The most common cause of cancer-related deaths globally is lung cancer. Non-small cell lung cancer and small cell lung cancer are the two primary varieties. Both of these varieties develop and are handled differently. The more prevalent kind of lung cancer is non-small cell lung cancer. The risk of lung cancer is highest among smokers. The longer and more cigarettes smoked, the higher the risk of lung cancer. Giving up smoking dramatically reduces the likelihood of smoking, even if a person smoked for many years.

IV. Pleural effusion

The pleura is a delicate membrane that encases the lungs and coats the inner surface of the chest cavity. A small amount of fluid facilitates the movement of the pleura on the lung's exterior against the chest wall during respiration. This fluid accumulates in the area between the lung and the chest wall, often as a result of conditions such as pneumonia or heart failure. Significant pleural effusions can lead to breathing difficulties and may require medical intervention to be drained.

V. COVID-19:

Coronavirus disease (COVID-19) is an infectious illness resulting from the SARS-CoV-2 virus. The majority of individuals infected with this virus will encounter mild to moderate respiratory symptoms and will recover without the need for specialized medical intervention. Nevertheless, a subset of patients may experience severe illness and necessitate medical care. Individuals who are older or who have pre-existing health conditions, such as cardiovascular disease, diabetes, chronic respiratory conditions, or cancer, are at a heightened risk of developing severe complications. It is important to note that anyone, regardless of age, can contract COVID-19 and may experience serious health outcomes or even death.

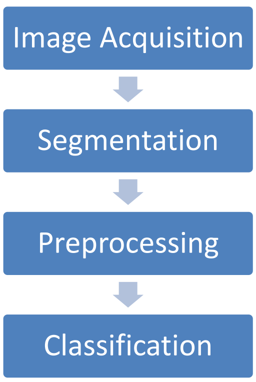

Fig.1 Lung Disease Detection Model

3. Literature Survey:

For predicting lung diseases like pneumonia and COVID-19 from patient chest X-ray images, a novel framework [1] has been developed. Adaptive and accurate region of interest (ROI) estimation, image quality enhancement, dataset acquisition, and normalization are the components of the framework. For classification, soft computing techniques like ensemble classifiers, K-nearest neighbor (KNN), support vector machines (SVM), artificial neural networks (ANN), and deep learning classifiers are employed. disease prediction and feature extraction. The characteristics of appearance, form, texture, and intensity

Fibrotic lung disease was automatically classified from high-resolution CT scans using a deep learning algorithm [2]. Two tertiary referral centers for interstitial lung disease provided a database of 1157 anonymized high-resolution CT scans demonstrating evidence of diffuse fibrotic lung disease. TensorFlow was used in this study to develop deep learning algorithms. The Google Brain Team created the open-source TensorFlow machine learning library, which makes it easier to express and use machine learning algorithms.

VDSNet [3] is a novel hybrid deep learning framework that combines CNN, VGG, data augmentation, and spatial transformer network (STN) to detect lung diseases from X-ray images. Tensorflow, Keras, and Jupyter Notebook are used as implementation tools. The NIH chest X-ray image dataset, which was gathered from the Kaggle repository, is used to test the new model. Important records for the set of data constructed as follows are enclosed in the data: view position, gender, age, snapshot data, and lung X-ray images. Advanced data normalization and augmentation techniques were used in the development of the Lung Disease Classification (LDC) system [4]. In this work, a convolution neural network was utilized to extract the spectrogram features.

Using a multiclass cross-entropy loss function on a compound scaling framework with EfficientNet as a baseline, a CX-Ultranet (Chest X-ray Ultranet)[5] is proposed to classify and identify thirteen thoracic lung diseases from chest X-rays. On the NIH Chest X-ray Dataset, the CX-Ultra net achieves an average prediction accuracy of 88%. Compared to existing state-of-the-art models, it takes about 30% less time. The suggested CX-Ultra net effectively addresses the problem of class imbalance and provides higher average accuracy.

To categorize lung CT image patches into seven classes, including six distinct ILD patterns and healthy tissue, a deep CNN was proposed [6]. In order to capture the low-level textural features of the lung tissue, a novel network architecture was created. Five convolutional layers with two kernels and LeakyReLU activations make up the suggested network [6]. Three dense layers and average pooling of size equal to the final feature map size come next. The final dense layer has seven outputs, which correspond to the classes that were taken into consideration: micronodules, consolidation, reticulation, honeycombing, ground glass opacity (GGO), healthy, and a combination of GGO and reticulation. A dataset of 14696 image patches derived from 120 CT scans were used from various scanners and hospitals—to train and assess the CNN.

4. Convolutional Neural Network

A Convolutional Neural Network (CNN) is structured with multiple layers that can be classified into three primary categories: convolutional layers, pooling layers, and fully connected layers. As data traverses through these layers, the CNN's complexity escalates, enabling it to progressively recognize larger segments of an image and more abstract characteristics.

I. Convolutional Layer

The convolutional layer serves as the core component of a CNN, where the majority of the computational tasks are executed. This layer employs a filter or kernel—a compact matrix of weights—that traverses the receptive field of the input image to identify specific features. The process initiates with the kernel sliding across the image's dimensions, systematically covering the entire image through numerous iterations. At each position, a dot product is computed between the kernel's weights and the pixel values beneath it. This operation results in the transformation of the input image into a series of feature maps or convolved features, each indicating the presence and intensity of particular features at various locations within the image.

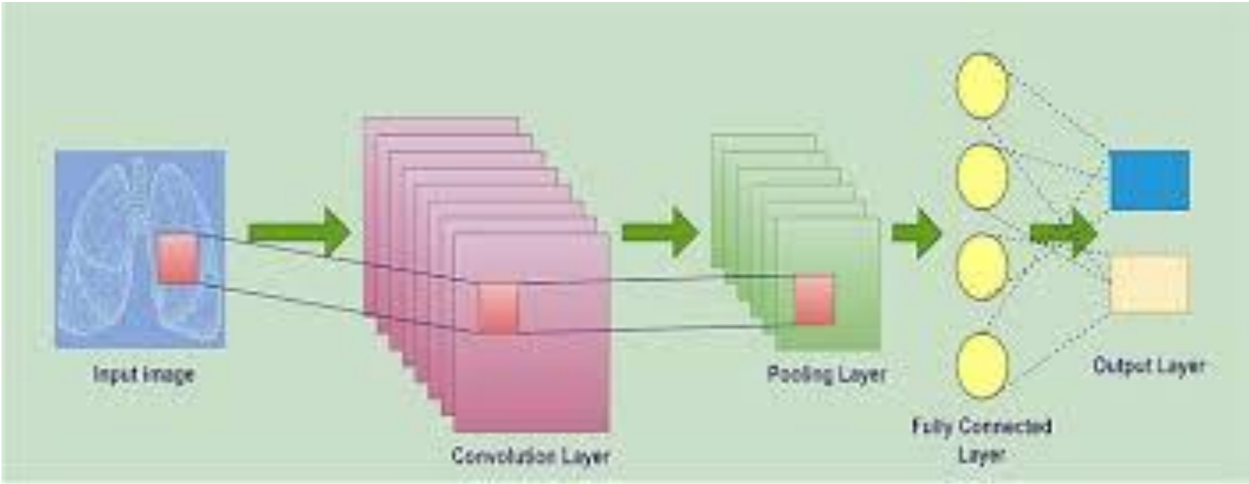

Fig.2 CNN architecture

II. Pooling Layer

Following the convolutional layer, the pooling layer constitutes an essential element of a CNN. While it also involves a sweeping mechanism across the input image, its primary function diverges from that of the convolutional layer. The pooling layer's objective is to diminish the dimensionality of the input data while preserving vital information, thereby enhancing the overall efficiency of the network. This is generally accomplished through downsampling, which reduces the number of data points in the input.

III. Fully Connected Layer

The fully connected layer is pivotal in the concluding phases of a CNN, tasked with classifying images based on the features that have been extracted in the preceding layers. The term "fully connected" indicates that each neuron in one layer is linked to every neuron in the subsequent layer. This layer synthesizes the diverse features obtained from the earlier convolutional and pooling layers, facilitating the final classification process.

5. Dataset:

Kaggle Dataset

The collective dataset is a combination of 16 datasets and contains 85,105 frontal chest X-ray images. The collection consists of 10 classes of X-ray images that include COVID-19 (15,660 images), Effusion (13,501 images), Lung Opacity (7179 images), Mass (5603 images), Nodule (6201 images), Pulmonary Fibrosis (3357 images), Pneumonia (9878 images), Pneumothorax (6870 images), and Tuberculosis (3184 images), in addition to the Controlled (13,672 images) class.

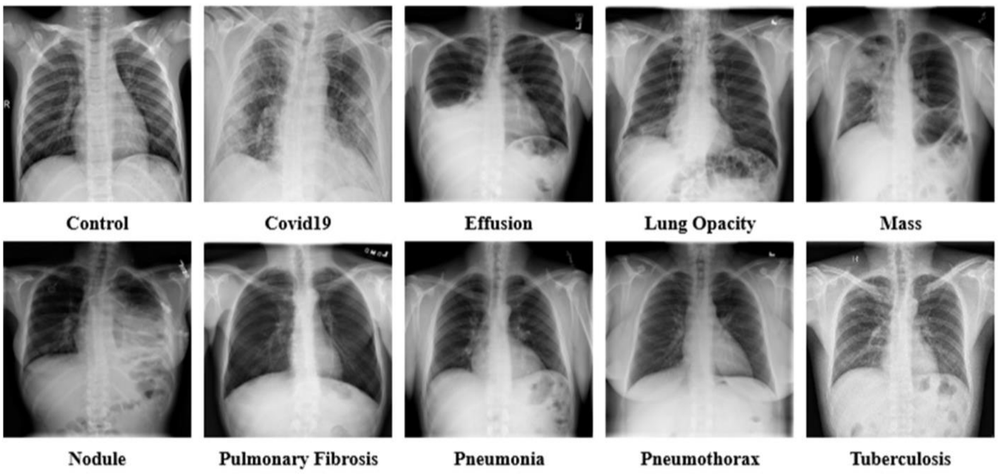

Fig.3 Lung Disease X-ray images Data set

CXIP dataset

Chest X-Ray Images for Pneumonia (CXIP) is the second lung disease dataset collected from Kaggle (https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia). CXIP dataset consists of 5856 samples in three categories- normal (1583), bacterial pneumonia (2790), and viral pneumonia (1483).

Shenzhen dataset

The Shenzhen Hospital dataset (SH) containing CXR images was created by the People's Hospital in Shenzhen, China. It includes both abnormal (containing traces of tuberculosis) and standard CXR images.

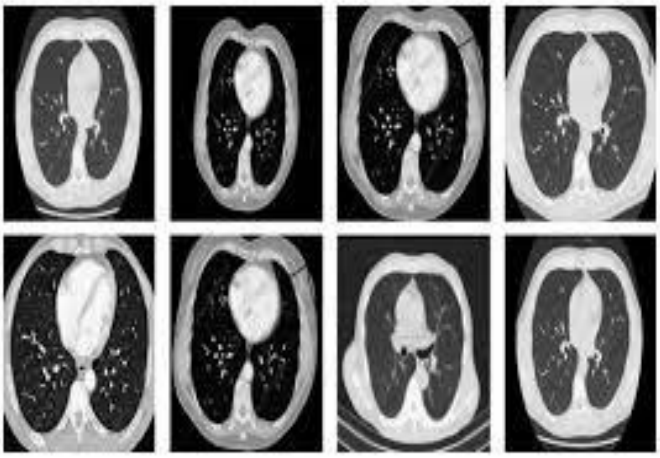

Fig.4 Lung Disease CT scan images Data set

Indiana dataset

The dataset was collected from various hospitals affiliated to the Indiana University School of Medicine. It consists of 7470 chest radiographs including the frontal and lateral images of disease annotations, such as cardiac hypertrophy, pulmonary edema, opacity, or pleural effusion.

KIT dataset

The dataset consists of 10,848 DICOM cases from the Korea Tuberculosis Institute under the Korea Association of Tuberculosis, including 7020 cases of normal and 3828 cases of abnormalities (tuberculosis).

MC dataset

The dataset was collected from the Department of Health and Human Services in partnership with Montgomery County, Maryland in the United States. The group consisted of 138 frontal chest radiographs from the Montgomery County Tuberculosis Screening Program, of which 80 were normal and 58 were tuberculosis with the image sizes as 4020×4892 or 4892×4020 pixels.

JSRT dataset

The dataset was compiled by the Japanese Society of Radiological Technology (JSRT) and includes 247 chest radiographs, of which 154 have pulmonary nodules (100 malignant and 54 benign), and 93 have no nodules. All of the X-ray images are 2048×2048 pixels in size, while the color depth of the grayscale is 12 bits.

6. Evaluation Metrics:

i) Accuracy

Accuracy is a fundamental evaluation metric for assessing the overall performance of a classification model. It is the ratio of the correctly predicted instances to the total instances in the dataset

Accuracy=TP+TN/TP+FP+TN+FN

ii) Precision

Precision evaluates the accuracy of the positive prediction made by the classifier. In precision specifies how many were actually positive in the total number of instances.

Precision = (TP) / (TP) + (FP)

iii) Recall

The recall is also known as sensitivity or true positive rate. It is the ratio of the number of true positive predictions to the total number of actual positive instances in the dataset. Recall measures the ability of a model to identify all relevant instances.

Recall = (TP) / (TP) + (FN)

iv) F1-Score

F1 score is the harmonic mean of precision and recall. It provide a single metric that balances the trade-off between precision and recall. It is especially useful when the class distribution is imbalanced.

F1 Score = 2 x [(Precision x Recall)/(Precision + Recall)]

7. Conclusion:

Pneumonia, tuberculosis, and coronavirus disease are examples of lung diseases that affect the airways and other lung structures. Lung disease classification is a difficult task that requires some specific methods. It is obvious that one of the main causes of death and disability worldwide is lung disease. Improving long-term survival rates and improving the likelihood of recovery depend heavily on early detection. A branch of machine learning called deep learning deals with algorithms that draw inspiration from the structure and operation of the brain. This paper presents a variety of lung disease detection methods.

References:

[1]. Goyal, Shimpy, and Rajiv Singh. "Detection and classification of lung diseases for pneumonia and Covid-19 using machine and deep learning techniques." Journal of Ambient Intelligence and Humanized Computing 14.4 (2023): 3239-3259.

[2]. Walsh, Simon LF, et al. "Deep learning for classifying fibrotic lung disease on high-resolution computed tomography: a case-cohort study." The Lancet Respiratory Medicine 6.11 (2018): 837-845.

[3]. Bharati S, Podder P, Mondal MRH. Hybrid deep learning for detecting lung diseases from X-ray images. Inform Med Unlocked. 2020;20:100391. doi: 10.1016/j.imu.2020.100391. Epub 2020 Jul 4. PMID: 32835077; PMCID: PMC7341954.

[4]. Tariq, Zeenat, Sayed Khushal Shah, and Yugyung Lee. "Lung disease classification using deep convolutional neural network." 2019 IEEE international conference on bioinformatics and biomedicine (BIBM). IEEE, 2019.

[5] Kabiraj, Anwesh, et al. "Detection and classification of lung disease using deep learning architecture from x-ray images." International Symposium on visual computing. Cham: Springer International Publishing, 2022.

[6] Anthimopoulos, Marios, et al. "Lung pattern classification for interstitial lung diseases using a deep convolutional neural network." IEEE transactions on medical imaging 35.5 (2016): 1207-1216.