Deep Learning Approach for Unprecedented Lung Disease Prognosis

Mulla Abdul Faheem[1]

Department of CSE
Sri Venkateswara College of Engineering and Technology (Autonomous)
Chittoor, India

Abstract - *This research project focuses on the development and implementation of a binary classification model for predicting lung diseases using x-ray images. By leveraging the power of machine learning algorithms, specifically those found in the field of computer science known as machine learning, we aim to accurately assign class labels to data from the problem area. Throughout the project, we utilize popular Python libraries such as Tensor Flow, Keras, and NumPy to enhance the prediction accuracy. The results of this study provide valuable insights into the prediction of lung diseases based on x-ray images, offering potential advancements in diagnostic and prognostic methodologies*

Key Words: **Convolutional neural network, AI Lung Diseases classification, Machine Learning,**

1.INTRODUCTION

Lung disease prediction utilizing X-ray images involves the intricate task of detecting the presence or absence of lung ailments within provided images. This exceptional project has successfully implemented a cutting-edge machine learning model and a state-of-the-art Convolutional Neural Network (CNN) architecture, leveraging powerful Python libraries such as NumPy and TensorFlow. The astonishing test accuracy of 91% has truly surpassed expectations, showcasing the project's triumph in achieving its primary objectives. Machine learning, an indispensable facet of artificial intelligence, empowers computers to acquire knowledge from past instances and discern intricate patterns within vast and noisy datasets. In the realm of medicine, machine learning plays a pivotal role in disease detection, enabling early and accurate diagnoses, which have the potential to save lives and alleviate the burden on healthcare systems. Lung diseases, being a leading cause of mortality, demand precise diagnoses and predictions to improve patient care significantly. By synergistically amalgamating patient data with chest X-ray images and employing deep learning techniques such as CNN, this groundbreaking study has delved into respiratory issues encompassing diseases like Corona, Tuberculosis, Pneumonia, and Lung Cancer. The ultimate goal was to develop robust models for diagnosing lung disorders, assisting doctors in making informed decisions that are crucial for patient well-being. Harnessing the power of machine learning and deep learning algorithms, this study meticulously evaluated patient data to determine the presence or absence of lung diseases. The central focus of this binary classification project revolved around utilizing chest X-ray images as input and disease detection as output, with the overarching objective of enhancing the accuracy and efficiency of diagnosing and treating lung diseases. This groundbreaking research has shed light on the innovative application of machine learning techniques for predicting and managing lung illnesses, ultimately culminating in improved patient outcomes. By embracing the potential of advanced technologies, this study paves the way for a future where early detection and accurate diagnosis of lung diseases become commonplace, thus revolutionizing the field of healthcare.

2. OBJECTIVE

The rapid pace of global change exerts strain on people's health, with detrimental shifts in climate, environment, and lifestyle significantly increasing disease vulnerability. This opportune moment allows us to contribute to the solution, empowered by computers and abundant public data. By assisting those unable to afford medical care, my approach aims to alleviate medical expenses while giving back to the community. Utilizing a deep learning model, the project focuses on detecting lung diseases in images. Various lung x-ray datasets, encompassing Normal, Tuberculosis, Covid-19, and Pneumonia, are merged to form a comprehensive dataset.

The idea of the project: Detecting lung diseases using a deep learning model focuses on predicting the presence or absence of lung disease in given images. Diverse lung x-ray datasets from sources like Kaggle, encompassing Normal, Tuberculosis, Covid-19, Pneumonia, are manually combined to create a unified dataset.

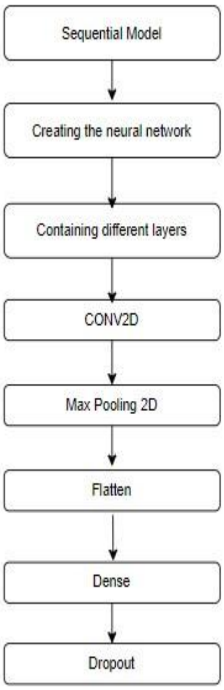
3. EXISTING SYSTEM

Existing models individually predict diseases, but our aim is to develop a single model for predicting multiple lung diseases. In the past, separate models were utilized for each lung disease, but now we plan to consolidate them into a combined model. We employed deep learning, specifically convolutional neural network (CNN) analysis, to detect and classify chronic obstructive pulmonary disease (COPD) while also predicting acute respiratory distress (ARD) episodes and mortality.

4. PROPOSED WORK

CNN, or Convolutional Neural Networks, excel in photo and video recognition tasks. It's a machine learning algorithm that detects patterns in images. CNN architectures consist of convolution layers that transform image data and pass it to subsequent layers. Each convolution layer has specified filters recognizing edges, shapes, objects, and more. These filters uncover complex patterns by layering them. Filters are typically 3x3 or 4x4 image kernels applied to the entire image. The stride determines the number of pixels the filter moves across the input matrix. In our project, we transitioned from individual models for each lung disease to a combined model.

A. Architecture Diagram



B. Modules:-

- Download the Datasets and Extract Them
- Defining the Directories in Dataset
- Define the Model
- Data Pre-processing
- Training the Model
- Running the Model

5. PROPOSED SYSTEM

A. Download the Datasets and Extract:

Kaggle recently released a vast collection of X-ray lung data, including labeled lung disease data. This presents an ideal opportunity to initiate our project. The dataset for image categorization comprises four categories: Pneumonia, Covid-19, Tuberculosis, and Normal Chest X-Ray images. The updated dataset ensures more balanced distribution in the validation and testing sets. Within the three folders (train, test, and Val), each image type is represented by a subdirectory. Prior to implementing Machine Learning and Deep Learning techniques, we will thoroughly analyze and evaluate the data to identify lung issues and specific diseases. To accommodate the dataset's size, I will initially test our approaches on a smaller sample dataset collected from public sources on Kaggle:

- Corona & Pneumonia data set
- Tuberculosis & Normal data set

B. Defining the Directories in Dataset :

From the Kaggle database, we extracted four classes, including covid, tuberculosis, pneumonia, etc. Using a Google Colab notebook, we uploaded the dataset to Google Drive and proceeded to extract the data for testing and training the model. The training directories were organized to include pictures for each class, such as covid, normal, tuberculosis, and others. These classes were then utilized to access the x-ray images. TensorFlow was employed to construct a Convolutional Neural Network (CNN) for image recognition. We utilized the Sequential model process, an API for constructing deep learning models, by creating an instance of the Sequential class and adding layers to it. The CNN was executed on a dataset of 1438 x-ray pictures from the four different classes to assess its accuracy.

C. Define the Model:

While numerous existing models predict diseases individually, our goal is to achieve the highest accuracy by predicting different diseases using a single model. We obtained data from the Kaggle dataset, which will undergo testing and subsequent training of all classes. The model will be trained step by step, following a sequential process. Each layer of the model summary will be completed layer by layer.

Upon completing the aforementioned process, we will import the Image Data Generator from TensorFlow's pre-processing module. Subsequently, we will assess the model's accuracy

Model: "sequential_1"

Layer (type)	Output Shape	Param #
conv2d_4 (Conv2D)	(None, 148, 148, 32)	896
conv2d_5 (Conv2D)	(None, 146, 146, 64)	18496
max_pooling2d_3 (MaxPooling2	(None, 73, 73, 64)	0
conv2d_6 (Conv2D)	(None, 71, 71, 64)	36928
max_pooling2d_4 (MaxPooling2	(None, 35, 35, 64)	0
dropout_3 (Dropout)	(None, 35, 35, 64)	0
conv2d_7 (Conv2D)	(None, 33, 33, 128)	73856
max_pooling2d_5 (MaxPooling2	(None, 16, 16, 128)	0
dropout_4 (Dropout)	(None, 16, 16, 128)	0
flatten_1 (Flatten)	(None, 32768)	0
dense_2 (Dense)	(None, 64)	2097216
dropout_5 (Dropout)	(None, 64)	0
dense_3 (Dense)	(None, 4)	260

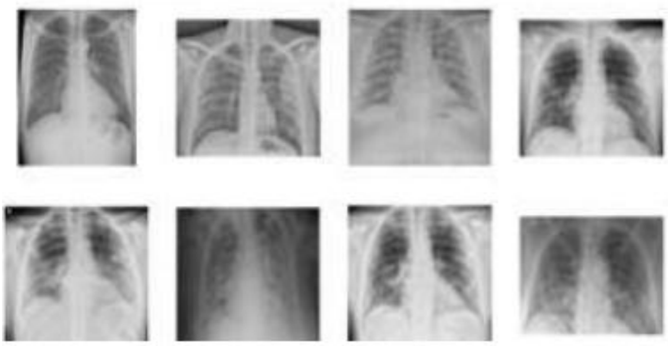
Total params: 2,227,652
Trainable params: 2,227,652
Non-trainable params: 0

D. Running the Model:

Initially, data will be extracted from the database, followed by the model's testing for the four classes. The module will then be trained incrementally. To execute the model, the OS will be imported, and the training pictures will be organized within a directory. The list directory will encompass the four classes. The CNN will subsequently scan all 5411 images. TensorFlow will be imported for visual imagery analysis. The sequential model will incorporate Covin2D, max-pooling2d, dropout, flatten, and denser layers. The Image Data Generator from TensorFlow will be utilized to rescale the images to 1./255. The generator will be trained, resulting in 5364 images for the training set and 281 images for the validation set. Additionally, the test data generator will consist of 1488 images for the four classes. The model will save the data, load it, and evaluate the test generator. Finally, a graph will depict the training and validation accuracy.

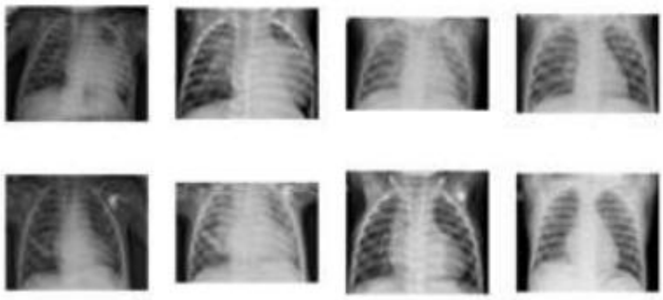
6. RESULTS AND DISCUSSION

A. Covid-19 x-ray images: -



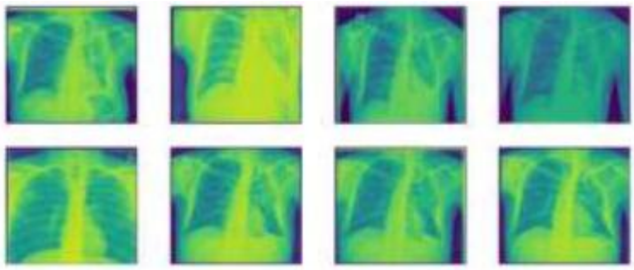
Result of the Covid-19 disease image that trained and implemented the model using an x-ray from the Kaggle database.

B. Pneumonia x-ray images: -



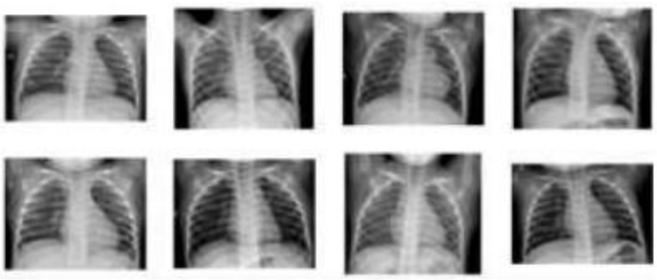
The results of the pneumonia disease images were obtained by training and implementing the model using x-rays sourced from the Kaggle database.

C. Tuberculosis x-ray images: -



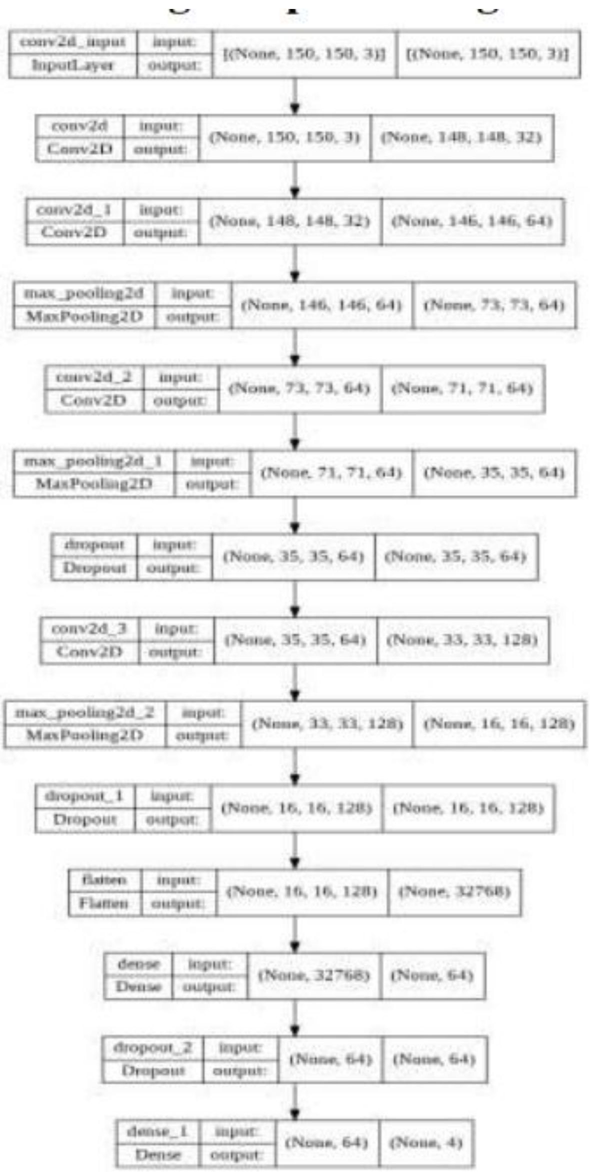
The results of the tuberculosis disease images were obtained through training and implementing the model using x-rays sourced from the Kaggle database.

D. Normal x-ray images: -



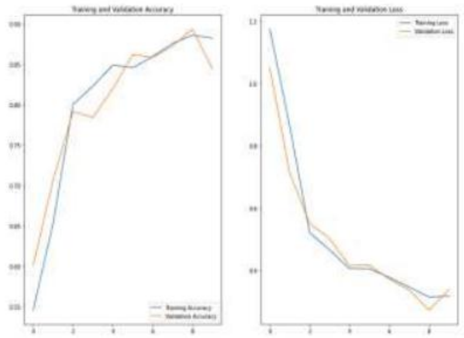
Results of some normal disease images which are trained and implemented the model by using x-rays From the Kaggle database.

6. THE TAXONOMY OF STATE-OF-ART WORK ON LUNG DISEASE DETECTION USING DEEP LEARNING



7. TRAINING AND VALIDATION

- During the training process, we utilized 5411 images, which belonged to 4 classes (COVID, NORMAL, PNEUMONIA, and TB), to train the CNN for 10 epochs.
- For validation, we used a set of 283 images encompassing all 4 classes.
- As a result, our model achieved a training accuracy of 88% and a validation accuracy of 84%.



8. CONCLUSION AND FUTURE SCOPE

Through testing our CNN on a test dataset comprising 1438 images across 4 classes, we achieved an impressive accuracy of approximately 91%. This research has highlighted the significance and practicality of Convolutional Neural Networks in various applications, including medical image processing. The study emphasizes the crucial role of accurate pre-processing techniques in optimizing model performance. Additionally, the findings suggest that CNNs can extend beyond medical image processing to enable disease prediction using MRI images, potentially aiding in the early detection of diseases like Lung Cancer and Bronchitis.

REFERENCES

[1] Tensor Flow's main website, https://www.tensorflow.org/.

[2] Swetha, K. R., et al. "Prediction of Pneumonia Using Big Data, Deep Learning, and Machine Learning Techniques." 2021 6th International Conference on Communication and Electronics Systems (ICCES). IEEE, 2021.

[3] R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[4] Dey, Soumava, Gunther Correia Bacellar, Mallikarjuna Basappa Chandrappa, and Raj Kulkarni. "COVID-19 Chest X-Ray Image Classification Using Deep Learning." medRxiv (2021).

[5] Tensor flow losses Module, https://www.tensorflow.org/ versions/r1.5/api_docs/python/tf/losses

[6] CNN, https://en.wikipedia.org/wiki/Convolutional_neural_network.

[7] Hyper Parameters, https://en.wikipedia.org/wiki/Convolutional_ neural_network#Hyperparameters

[8] Swetha, K. R., et al. "Prediction of Pneumonia Using Big Data, Deep Learning, and Machine Learning Techniques." 2021 6th International Conference on Communication and Electronics Systems (ICCES). IEEE, 2021

[9] Patil, Swati, and Akshay Golellu. "Classification of COVID-19 CT Images using Transfer Learning Models." 2021 International Conference on Emerging Smart Computing and Informatics (ESCI). IEEE, 2021.

[10] Yadav, Anju, et al. "FVC-NET: An Automated Diagnosis of Pulmonary Fibrosis Progression Prediction Using Honeycombing and Deep Learning." Computational Intelligence and Neuroscience 2022 (2022).

[11] Kogilavani, S. V., et al. "COVID-19 detection based on lung CT scan using deep learning techniques." Computational and Mathematical Methods in Medicine 2022 (2022).

[12] Aggarwal, Karan, et al. "Has the Future Started? The Current Growth of Artificial Intelligence, Machine Learning, and Deep Learning." Iraqi Journal For Computer Science and Mathematics 3.1 (2022): 115-123.

[13] Ahmed, Shaymaa Taha, and Suhad Malallah Kadhem. "Using Machine Learning via Deep Learning Algorithms to Diagnose the Lung Disease Based on Chest Imaging: A Survey." International Journal of Interactive Mobile Technologies 15.16 (2021).

[14] Gowri, S., J. Jabez, J. S. Vimali, A. Sivasangari, and Senduru Srinivasulu. "Sentiment Analysis of Twitter Data Using Techniques in Deep Learning." In Data Intelligence and Cognitive Informatics, pp. 613-623. Springer, Singapore, 2021.

[15] Srinivasulu, Senduru, Jeberson Retna Raj, and S. Gowri. "Analysis and Prediction of Myocardial Infarction using Machine Learning." 2021 5th International Conference on Trends in Electronics and Informatics (ICOEI). IEEE, 2021.

[16] Priya, T., and T. Meyyappan. "Disease Prediction by Machine Learning Over Big Data Lung Cancer." International Journal of Scientific Research in Computer Science, Engineering and Information Technology (2021): 16-24.

[17] Battineni, Gopi. "Machine Learning and Deep Learning Algorithms in the Diagnosis of Chronic Diseases." Machine Learning Approaches for Urban Computing. Springer, Singapore, 2021. 141-164..