**Lung Infection Detection**

**COURSE PROJECT REPORT**

**18CSE398J -Machine Learning - Core Concepts with Applications**

**(2018 Regulation)**

**III Year/ VI Semester**

**Academic Year: 2022 -2023 (EVEN)**

By

**Akash Prasad – RA2011003010336**

**Ritik Raj – RA2011003010334**

Under the guidance of

**Vijayalakshmi V**

**Assistant Professor**

**Department of Data Science and Business Systems**



**DEPARTMENT OF DATA SCIENCE AND BUSINESS SYSTEMS**

**FACULTY OF ENGINEERING AND TECHNOLOGY**

**SRM INSTITUTE OF SCIENCE AND TECHNOLOGY**

**Kattankulathur, Kancheepuram**

**MAY 2023**

# **Abstract**

Lung Infection like pneumonia are very common in the covid 19 pandemic and post pandemic. Many people suffered with pneumonia but didn’t knew until it was to late. It is really difficult to detect this lung infection. This project intends to make the detection of pneumonia via X-ray using deep learning concepts. We bring CNN (Convolutional Neural Network) to classify by extracting features from the X-ray which includes 3 major difficult challenges: Feature extraction, class imbalance and detection. Coupled with hyper-parameter tuning and transfer learning in which adopt EfficientNet. We have used Google Colab and Kaggle for training these models. The best classification accuracy we got 99.43% using EfficientNet as a base model.

## **Introduction**

Pneumonia is a major public health problem affecting millions of people worldwide. It is a serious respiratory disease that affects the lungs and can be caused by many diseases such as bacteria, viruses or fungi. Accurate and timely diagnosis is essential for effective treatment, as early detection can improve patient outcomes.  
  
Recent advances in artificial intelligence and deep learning have yielded great results in medical image analysis and diagnosis. Deep learning techniques such as convolutional neural networks (CNNs) have been used successfully for image classification, including medical treatment.  
Analyzing chest X-ray and CT scans, the method has shown the correct diagnosis of many diseases, including pneumonia. Transfer learning also works well for image classification models.  
  
In this project, we are planning a deep study to detect lung diseases by lung x-ray. Our method uses a CNN architecture trained on large datasets of X-ray images. The training model can classify the output image as positive or negative for pneumonia, providing doctors with an accurate and effective diagnostic tool.  
We use several models before EfficientNet for better performance.  
  
The plan has the potential to improve the accuracy and speed of lung disease diagnosis, especially in limited areas. This work adds to the growing body of research using deep learning techniques for medical image analysis and demonstrating the potential of these techniques to improve patient outcomes in the field of respiratory diseases.

## **Dataset**

The dataset is organized into 3 folders (train, test, val) and contains subfolders for each image category (Pneumonia/Normal). There are 5,863 X-Ray images (JPEG) and 2 categories (Pneumonia/Normal).

Chest X-ray images (anterior-posterior) were selected from retrospective cohorts of pediatric patients of one to five years old from Guangzhou Women and Children’s Medical Center, Guangzhou. All chest X-ray imaging was performed as part of patients’ routine clinical care.

For the analysis of chest x-ray images, all chest radiographs were initially screened for quality control by removing all low quality or unreadable scans. The diagnoses for the images were then graded by two expert physicians before being cleared for training the AI system. In order to account for any grading errors, the evaluation set was also checked by a third expert.

## **Methods**

The methodology for the lung infection detection project typically involves the following:

steps:

1. Data collection: The first step is to collect a large and diverse dataset of chest X-ray

images that include both normal and infected lungs. The dataset should be

preprocessed to remove duplicates, and low-quality, irrelevant images.

We got the dataset from Kaggle.

2. Data preprocessing: This step involves preparing the dataset for training by

performing various operations such as image resizing, normalization, and data

augmentation to ensure that the model receives high-quality and diverse data. We used Keras and TensorFlow for data augmentation and preprocessing.

3. Model architecture: The next step is to select an appropriate deep-learning model

architecture. For lung infection detection, a Convolutional Neural Network (CNN) is a

common choice, which can learn to extract meaningful features from chest X-ray

images. Transfer learning with will also be used EfficentNet.

4. Training the model: In this step, the prepared dataset is used to train the selected

deep-learning model. The training process involves feeding the model with a large

number of chest X-ray images and iteratively adjusting its parameters to minimize the

loss function and maximize accuracy. Hyper-parameter tuning used to select best parameters for training like epochs, batch-size, etc

5. Model evaluation: Once the model is trained, it is evaluated on a separate

validation dataset to test its performance. The model is then fine-tuned based

on the validation results. Model is evaluated on the validation dataset.

6. Model testing: The final step involves testing the model on a test dataset to evaluate

its generalization ability and accuracy. The model's performance is compared to

existing methods to determine its effectiveness in detecting pneumonia. Model was tested on the testing dataset.

7. Deployment: The final step involves deploying the model for practical use. The model is deployed using Streamlit as a web application.

In summary, the methodology for the lung infection detection project involves collecting and preprocessing the dataset, selecting an appropriate deep learning model architecture, training the model, evaluating its performance, testing its generalization ability, and finally deploying the model for practical use

## **Experiment**

EfficientNet is a family of convolutional neural network models developed by Google that achieve state-of-the-art performance on image classification tasks while using fewer parameters and less computational resources than other comparable models.

In the context of lung infection detection, EfficientNet could be a promising choice as it could potentially improve the accuracy and efficiency of the detection model, which is especially important for medical applications. EfficientNet models are scalable in terms of their depth, width, and resolution, which could allow for fine-tuning to suit the specific requirements of the lung infection detection task.

To evaluate the performance of an EfficientNet model for lung infection detection, a suitable dataset of chest X-rays with labeled ground truth for lung infection would be required. The dataset should be split into training, validation, and test sets, with the majority of the data used for training and validation, and a smaller portion held out for final evaluation.

The model would need to be trained using appropriate hyperparameters and optimization techniques, and its performance evaluated on the test set using metrics such as accuracy, precision, recall, and F1 score. Additional experiments could be conducted to assess the robustness and generalization ability of the model, such as cross-validation or testing on an external dataset.

Overall, EfficientNet is a promising architecture for lung infection detection using deep learning, and could potentially yield high accuracy results with less computational resources compared to other architectures. Hyperparameter tuning and the Data augmentation ensured to have the best parameters and best quality of data to be used while building the model.

After the best working model is found it is saved and deployed using streamlit as a web application. Weights are also saved in order to use them for further development.

## **Result**

In general, deep learning models can achieve high accuracy for detecting lung infections, with reported accuracies ranging from 80% to 95% on different datasets. Other relevant metrics include precision, recall, and F1 score, which measure the model's ability to correctly identify positive cases while minimizing false positives and false negatives. Overall, the performance of a deep learning model for lung infection detection will depend on factors such as the quality and size of the dataset, the choice of architecture and hyperparameters, and the validation methods used to evaluate the effectiveness of the model. Achieving an accuracy of 99.43% on a lung infection detection project would be considered a very impressive result, but it is important to consider other metrics and validation methods to fully assess the reliability and effectiveness of the model. The classification report in Performance Analytics can show all the above metrics.

## **Conclusion**

In conclusion, deep learning models showed great promise for pneumonia detection, achieving high accuracy, sensitivity, and specificity in different datasets EfficientNet is a promising framework that can improve detection model accuracy and performance time using fewer parameters and computer resources An accuracy of 99.43% in the lung infection detection task is considered a highly impressive result.

Furthermore, the performance of the model should be evaluated on separate test data, validated by expert review, and compared with other existing methods for lung disease detection.

Overall, deep learning models have the potential to significantly improve the accuracy and efficiency of pneumonia diagnosis, leading to more accurate diagnosis and improved patient outcomes but continued research and development is needed to further improve the performance of these models and ensure their safety and reliability in clinical settings.

Moreover, the model is deployed with the help or streamlit as a web application.