**Lung Disease Prediction Using ML**

**Models in the Orange Tool**

A PROJECT REPORT

**21CSC305P – Digital Health Eco System and Application Deployment Techniques**

**(2021 Regulation)**

**III Year/ V Semester**

**Academic Year: 2024 -2025**

Submitted by

ROHAAN S S [RA2211026010228]

THARUN K [RA2211026010241]

RICHARD RAJESH J[RA2211026010263]

#### Under the Guidance of

## Dr. T.R Saravanan

Associate Professor

Department of Computational Intelligence

### *in partial fulfillment of the requirements* *for the degree of*

## BACHELOR OF TECHNOLOGY

## in

## COMPUTER SCIENCE ENGINEERING

## with specialization in

## ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING



## SCHOOL OF COMPUTING

## COLLEGE OF ENGINEERING AND TECHNOLOGY

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

## KATTANKULATHUR- 603 203

### NOVEMBER 2024



# SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

# KATTANKULATHUR – 603 203

## BONAFIDE CERTIFICATE

Certified that **21CSC305P -** **Digital Health Eco System and Application Deployment Techniques** project reporttitled “**LUNG DISEASE PREDICTION USING MACHINE LEARNING MODELS IN THE ORANGE TOOL**” is the bonafide work of “**ROHAAN S S [RA2211026010228] ,THARUN K [RA2211026010241] RICHARD RAJESH J [RA2211026010263]”** who carried out the task of completing the project within the allotted time.

|  |  |
| --- | --- |
| **SIGNATURE**  Dr. T.R Saravanan  **Course Faculty**  Associate Professor  Department of Computational Intelligence  SRM Institute of Science and Technology  Kattankulathur | **SIGNATURE**  Dr. R. Annie Uthra  **Head of the Department**  Professor  Department of Computational Intelligence  SRM Institute of Science and Technology Kattankulathur1 |

**ABSTRACT**

Lung diseases are among the leading causes of mortality and morbidity globally, with conditions like bacterial pneumonia, tuberculosis, viral pneumonia, and COVID-19 imposing significant health challenges. Timely and accurate diagnosis is crucial to improving patient outcomes and mitigating long-term health effects. Traditional diagnostic approaches, such as chest X-rays and CT scans, rely on the expertise of radiologists to interpret imaging, which is time-intensive and may vary across practitioners. The emergence of machine learning (ML) and artificial intelligence (AI) provides a promising solution to streamline the diagnostic process by automating disease detection with high accuracy and speed.

In this study, we explore the application of various machine learning models for lung disease prediction using the Orange data mining tool. The dataset includes medical images representing five distinct classes of lung conditions: Bacterial Pneumonia, COVID-19, Normal (healthy lungs), Tuberculosis, and Viral Pneumonia. We utilized Orange’s drag-and-drop interface to construct a workflow that imports, preprocesses, and analyzes image data, allowing for a simplified model-building process without extensive programming. Five ML models were evaluated in this study: Support Vector Machines (SVM), k-Nearest Neighbors (kNN), Random Forest, Neural Networks, and Gradient Boosting. Feature extraction was conducted using the “Image Embedding” widget, which converts high-dimensional images into manageable feature vectors suitable for model training.

The performance of each model was assessed based on accuracy, Area Under the Curve (AUC), F1 Score, precision, and recall. Among the models tested, SVM achieved the highest classification accuracy at 91.36%, closely followed by the Neural Network with an accuracy of 91.27%. The confusion matrix analysis revealed that the models were particularly effective in distinguishing COVID-19 and normal lung images but encountered some misclassifications between classes with overlapping features, such as Bacterial and Viral Pneumonia.

This research demonstrates the effectiveness of using Orange to deploy ML workflows in healthcare, showing that the tool’s user-friendly interface can simplify the adoption of machine learning in clinical diagnostics. The SVM model’s high accuracy underscores its potential as a robust choice for lung disease classification in medical imaging. The study highlights the benefits of integrating ML with accessible tools like Orange to bridge the gap between complex algorithms and practical clinical applications, paving the way for more accessible and automated diagnostics in healthcare.