DAYANANDA SAGAR UNIVERSITY

KUDLU GATE, BANGALORE – 560068

Bachelor of Technology
in
COMPUTER SCIENCE AND ENGINEERING

Major Project Phase-II Report

AN AUTOMATED FRAMEWORK FOR DIAGNOSING LUNGS RELATED ISSUES USING ML AND DATA ANALYTICS

By

MOHAMMED MAAZ AHMED KHAN	**ENG18CS0170**
MOHAMMED SIDDIQ S	**ENG18CS0171**
MYTHRI J L	**ENG18CS0179**
NAVEEN A	**ENG18CS0185**

Under the supervision of

Dr. RAJESH T M
Assistant Professor

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING,
SCHOOL OF ENGINEERING
DAYANANDA SAGAR UNIVERSITY,
BANGALORE

(2021-2022)

DAYANANDA SAGAR UNIVERSITY

School of Engineering
Department of Computer Science & Engineering
Kudlu Gate, Bangalore – 560068
Karnataka, India

CERTIFICATE

This is to certify that the Phase-II project work titled **"An Automated Framework For Diagnosing Lungs Related Issues Using Ml And Data Analytics"** is carried out by **Mohammed Maaz Ahmed Khan (ENG18CS0170), Mohammed Siddiq S (ENG18CS0171), Mythri J L (ENG18CS0179), Naveen A (ENG18CS0185)** bonafide students of Bachelor of Technology in Computer Science and Engineering at the School of Engineering, Dayananda Sagar University, Bangalore in partial fulfillment for the award of degree in Bachelor of Technology in Computer Science and Engineering, during the year **2021-2022**.

Dr. Rajesh T M	**Dr. Girisha G S**	**Dr. A Srinivas**
Assistant Professor Dept. of CS&E, School of Engineering Dayananda Sagar University Date: 20 / 06 / 2022	Chairman CSE School of Engineering Dayananda Sagar University Date: 20 / 06 / 2022	Dean School of Engineering Dayananda Sagar University Date: 20 / 06 / 2022

Name of the Examiner **Signature of Examiner**

1.

2.

DECLARATION

We, **Mohammed Maaz Ahmed Khan (ENG18CS0170), Mohammed Siddiq S (ENG18CS0171), Mythri J L (ENG18CS0179), Naveen A (ENG18CS0185),** are students of the seventh semester B.Tech in **Computer Science and Engineering**, at School of Engineering, **Dayananda Sagar University**, hereby declare that the phase-II project titled **"An Automated Framework For Diagnosing Lungs Related Issues Using Ml And Data Analytics"** has been carried out by us and submitted in partial fulfillment for the award of degree in **Bachelor of Technology in Computer Science and Engineering** during the academic year **2021-2022**.

STUDENTS DETAILS		SIGNATURE
Name1: **USN:**	**MOHAMMED MAAZ AHMED KHAN** **ENG18CS0170**	
Name2: **USN:**	**MOHAMMED SIDDIQ S** **ENG18CS0171**	
Name3: **USN:**	**MYTHRI J L** **ENG18CS0179**	
Name4: **USN:**	**NAVEEN A** **ENG18CS0185**	

Place: Bangalore

Date: 20 / 06 / 2022

ACKNOWLEDGEMENT

It is a great pleasure for us to acknowledge the assistance and support of many individuals who have been responsible for the successful completion of this project work.

First, we take this opportunity to express our sincere gratitude to the School of Engineering & Technology, Dayananda Sagar University for providing us with a great opportunity to pursue our Bachelor's degree in this institution.

We would like to thank **Dr. A Srinivas. Dean**, **School of Engineering & Technology**, **Dayananda Sagar University** for his constant encouragement and expert advice. It is a matter of immense pleasure to express our sincere thanks to **Dr. Girisha G S, Department Chairman**, **Computer Science, and Engineering**, **Dayananda Sagar University,** for providing the right academic guidance that made our task possible.

We would like to thank our guide **Dr. Rajesh T M**, **Associate Professor**, **Dept. of Computer Science and Engineering**, **Dayananda Sagar University**, for sparing his valuable time to extend help in every step of our project work, which paved the way for smooth progress and the fruitful culmination of the project.

We would like to thank our Project Coordinators **Dr. Meenakshi Malhotra** and **Dr. Bharanidharan N**, and all the staff members of Computer Science and Engineering for their support.

We are also grateful to our family and friends who provided us with every requirement throughout the course. We would like to thank one and all who directly or indirectly helped us in the Project work.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

SMLA:	Supervised Machine learning Algorithm
TL:	Transfer Learning

LIST OF FIGURES

LIST OF TABLES

Abstract

The recent developments of deep learning support the identification and classification of lung diseases in medical images. Hence, numerous works on the detection of lung disease using deep learning can be found in the literature. This project presents a survey of deep learning for lung disease detection in medical images. There has only been one survey paper published in the last five years regarding deep learning directed at lung disease detection. However, their survey is lacking in the presentation of taxonomy and analysis of the trend of recent work. The objectives of this project are to present a taxonomy of the state-of-the-art deep learning-based lung disease detection systems, visualize the trends of recent work on the domain and identify the remaining issues and potential future directions in this domain. The taxonomy consists of seven attributes that are common in the surveyed articles: image types, features, data augmentation, types of deep learning algorithms, transfer learning, the ensemble of classifiers, and types of lung diseases. The potential future direction suggested could further improve the efficiency and increase the number of deep learning-aided lung disease detection applications.

CHAPTER 1

INTRODUCTION

CHAPTER 1 INTRODUCTION

Lung diseases, also known as respiratory diseases, are diseases of the airways and the other structures of the lungs. Examples of lung diseases are pneumonia, tuberculosis, and Coronavirus Disease 2019 (COVID-19). According to the Forum of International Respiratory Societies, about 334 million people suffer from asthma, and, each year, tuberculosis kills 1.4 million people, 1.6 million people die from lung cancer, while pneumonia also kills millions of people. The COVID-19 pandemic impacted the whole world, infecting millions of people and burdening healthcare systems. It is clear that lung diseases are one of the leading causes of death and disability in this world. Early detection plays a key role in increasing the chances of recovery and improving long-term survival rates. Traditionally, lung disease can be detected via skin test, blood test, sputum sample test, chest X-ray examination, and computed tomography (CT) scan examination. Recently, deep learning has shown great potential when applied to medical images for disease detection, including lung disease.

The risk of lung diseases is enormous, especially in developing and low-middle-income countries, where millions of people are facing poverty and air pollution. According to the estimation of WHO, over 4 million premature deaths occur annually from household air pollution-related diseases, including asthma, and pneumonia. Hence, it is necessary to take necessary steps to reduce air pollution and carbon emission. It is also essential to implement efficient diagnostic systems which can assist in detecting lung diseases. Since late December 2019, a novel coronavirus disease 2019 (COVID-19) has been causing serious lung damage and breathing problems. In addition, pneumonia, a form of lung disease, can be due to the causative virus of COVID-19 or may be caused by other viral or bacterial infections. Hence, early detection of lung diseases has become more important than ever. Recently, digital technology has become more important worldwide. This project can provide doctors and other researchers with a direction for detecting lung disease with the help of a deep learning methodology. A large number of lung X-ray images are used as a dataset.

It is clear that lung diseases are one of the leading causes of death and disability in this world. Early detection plays a key role in increasing the chances of recovery and improving long-term survival rates. Traditionally, lung disease can be detected via skin test, blood test, sputum sample test, chest X-ray examination, and computed tomography (CT) scan examination.

1.1. SCOPE:

The risk of lung diseases is enormous, especially in developing and low-middle-income countries, where millions of people are facing poverty and air pollution. According to the estimation of WHO, over 4 million premature deaths occur annually from household air pollution-related diseases, including asthma, and pneumonia. Hence, it is necessary to take necessary steps to reduce air pollution and carbon emission. It is also essential to implement efficient diagnostic systems which can assist in detecting lung diseases. Since late December 2019, a novel coronavirus disease 2019 (COVID-19) has been causing serious lung damage and breathing problems. In addition, pneumonia, a form of lung disease, can be due to the causative virus of COVID-19 or may be caused by other viral or bacterial infections. Hence, early detection of lung diseases has become more important than ever. Recently, digital technology has become more important worldwide. This project can provide doctors and other researchers with a direction for detecting lung disease with the help of a deep learning methodology. A large number of lung X-ray images are used as a dataset. According to the Forum of International Respiratory Societies, about 334 million people suffer from asthma, and, each year, tuberculosis kills 1.4 million people, 1.6 million people die from lung cancer, while pneumonia also kills millions of people. The COVID-19 pandemic impacted the whole world, infecting millions of people and burdening healthcare systems. It is clear that lung diseases are one of the leading causes of death and disability in this world. Early detection plays a key role in increasing the chances of recovery and improving long-term survival rates. Traditionally, lung disease can be detected via skin test, blood test, sputum sample test, chest X-ray examination, and computed tomography (CT) scan examination. Hence, early detection of lung diseases has become more important than ever. Recently, digital technology has become more important worldwide. This project can provide doctors and other researchers with a direction for detecting lung disease with the help of a deep learning methodology. A large number of lung X-ray images are used as a dataset.

CHAPTER 2

PROBLEM DEFINITION

CHAPTER 2 PROBLEM DEFINITION

Machine Learning plays an important role in medical systems. Lung diseases are one of the leading causes of death. The early identification and prediction of lung diseases have become a necessity in the research, as it can facilitate the subsequent clinical management of patients. Machine Learning-based decision support systems provide the contribution to the doctors in their diagnosis decisions. The project considered is the classification of lung diseases like Pneumonia, Tuberculosis, Lung cancer, and Covid 19. Machine Learning and Deep Learning are used to process data as well as create models for diagnosing patients. Combining the processing of patient information with data from chest X-rays and CT scans, using CNN with the well-known pre-trained model, These Neural networks for data of this form are the methods used for this project to identify lung diseases. The risk of lung diseases is enormous, especially in developing and low-middle-income countries, where millions of people are facing poverty and air pollution. According to the estimation of WHO, over 4 million premature deaths occur annually from household air pollution-related diseases, including asthma, and pneumonia. Hence, it is necessary to take necessary steps to reduce air pollution and carbon emission. It is also essential to implement efficient diagnostic systems which can assist in detecting lung diseases. The risk of lung diseases is enormous, especially in developing and low-middle-income countries, where millions of people are facing poverty and air pollution. According to the estimation of WHO, over 4 million premature deaths occur annually from household air pollution-related diseases, including asthma, and pneumonia. Hence, it is necessary to take necessary steps to reduce air pollution and carbon emission. It is also essential to implement efficient diagnostic systems which can assist in detecting lung diseases. The risk of lung diseases is enormous, especially in developing and low-middle-income countries, where millions of people are facing poverty and air pollution. According to the estimation of WHO, over 4 million premature deaths occur annually from household air pollution-related diseases, including asthma, and pneumonia. Hence, it is necessary to take necessary steps to reduce air pollution and carbon emission. It is also essential to implement efficient diagnostic systems which can assist in detecting lung diseases.

CHAPTER 3

LITERATURE REVIEW

CHAPTER 3 LITERATURE REVIEW

In spite of launching the first CAD system for detecting lung nodules or affected lung cells in the late 1980s, those efforts were not enough. This is because there were many inadequate computational resources for the implementation of advanced image processing techniques at that time. Lung disease detection using basic image processing techniques is also time-consuming. After the successful invention of GPU and CNN, the performance of CAD (for lung disease diagnosing) and decision support arrangement got a high boost.

A full CNN is proposed in Ref. [6] to reduce the false-positive rate in classifying the lung nodules. This method can only analyze the nature of the CT scan images in order to reduce the probability of a wrong diagnosis.

In Ref. [7], a framework for deep learning is proposed to predict lung cancer and pneumonia offering two deep learning methods. Initially, they used modified AlexNet for the diagnosis of chest X-rays. Moreover, in the modified AlexNet, SVM is implemented for the purpose of classification

Moreover, deep learning methods are also proposed in Ref. [8] where several transfer learning methods such as DenseNet121, AlexNet, Inception V3, etc., are used for pneumonia diagnoses. However, the parameter tuning for their implemented methods is very complex.

CHAPTER 4

REQUIREMENTS

CHAPTER 4 REQUIREMENTS

4.1 Software Requirements

- Transfer learning
- React framework
- Tensorflow
- Python3
- OpenCV
- Keras
- Numpy
- Matplotlib
- Scikit-learn
- Tkinter

4.2 Hardware Requirements

- CPU: Intel i3+
- RAM: 2GB+
- Storage Space: At least 2GB
- Mobile phone (Android & iPhone)

4.3 System Features

- Constructing dataset for Tuberculosis, pneumonia, lung cancer, and COVID 19
- Model portable on IOS and Android platform
- Recognition of lung-related diseases

CHAPTER 5
METHODOLOGY

CHAPTER 5 METHODOLOGY

5.1. PROPOSED DESIGN:

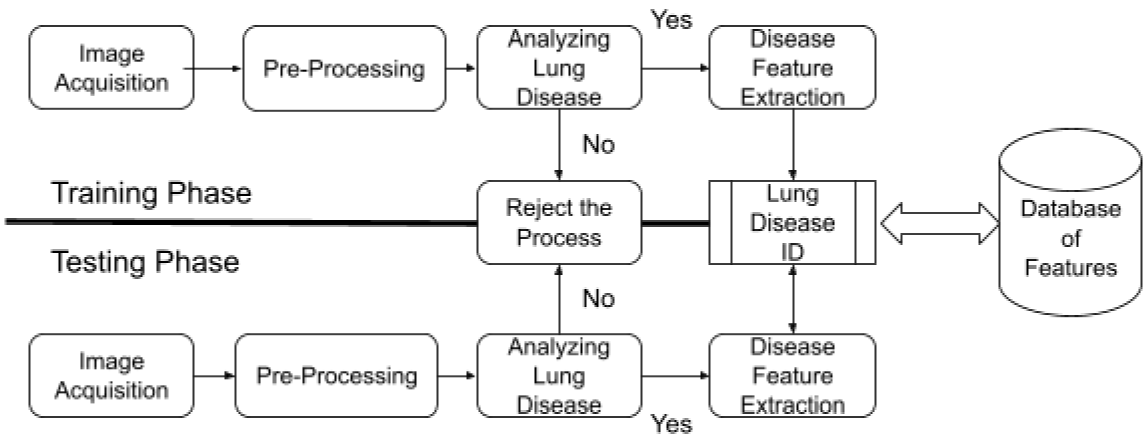


Figure 5.1 Overview of the project

5.1.1. **Image Acquisition Phase:**

The first step is to acquire images. To produce a classification model, the computer needs to learn by example. The computer needs to view many images to recognize an object. Other types of data, such as time-series data and voice data, can also be used to train deep learning models. In the context of the work used in this project, the relevant data required to detect lung disease will be images. Images that could be used include chest X-ray, and CT scan. The output of this step is images that will later be used to train the model.

5.1.2. **Preprocessing Phase:**

The second step is preprocessing. Here, the image could be enhanced or modified to improve image quality. Image modification such as lung segmentation and bone elimination could be used to identify the region of interest (ROI), whereby the detection of the lung disease can then be performed on the ROI. Edge detection could also be used to provide an alternate data representation. Data augmentation could be applied to the images to increase the amount of available data. Feature extraction could also be conducted so that the deep learning model could identify important features to identify a certain object or class. The output of this step is a set of images whereby the quality of the images is enhanced, or unwanted objects have been removed. The output of this step is images that were enhanced or modified that will later be used in training.

5.1.3. **Training Phase:**

In the third step, namely training, three aspects could be considered. These aspects are the selection of deep learning algorithms, usage of transfer learning, and usage of an ensemble. There are numerous deep learning algorithms, for example, multilayer perceptron neural network (MPNN), recurrent neural network (RNN), and the aforementioned CNN. Different algorithms have different learning styles. CNN works particularly well with images. A deep learning algorithm should be chosen based on the nature of the data at hand. Transfer learning refers to the transfer of knowledge from one model to another. Ensemble refers to the usage of more than one model during classification. Transfer learning and ensemble are techniques used to reduce training time, improve classification accuracy and reduce overfitting. The output of this step is models generated from the data learn.

5.1.4. **Classification Phase:**

In the fourth and final step, which is classification, the trained model will predict which class an image belongs to. For example, if a model was trained to differentiate X-ray images of healthy lungs and tuberculosis-infected lungs, it should be able to correctly classify new images (images that are never seen by the model before) into healthy lungs or tuberculosis-infected lungs. The model will give a probability score for the image. The probability score represents how likely an image belongs to a certain class. At the end of this step, the image will be classified based on the probability score given to it by the model.

CHAPTER 6
EXPERIMENTATION

CHAPTER 6 EXPERIMENTATION

6.1. FIGURES AND TABLES

Original Image Histogram Equalization Images CLAHE Image

Figure 6.1 Enhanced Images

Original Image Predicted Lung Mask Segmented Image

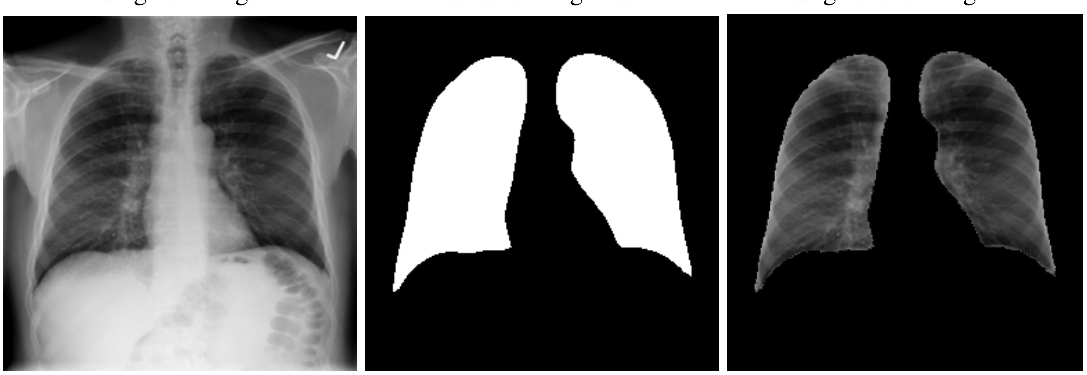


Figure 6.2 Segmented Images

	Mean	Standard Deviation	Entropy	Root Mean Square	Variance	Smoothness	Kurtosis	Skewness	Contrast	Co-relation	Energy	Homogeneity	Disease
0	94.763307	56.759613	7.150496	110.461478	3221.653650	33.536496	-1.030312	-0.234947	56.759613	0.996959	9.036597	0.448394	PNEUMONIA
1	78.442951	40.744653	7.285012	88.393571	1660.126779	20.283737	-0.429221	0.279648	40.744653	0.993575	12.966873	0.499646	PNEUMONIA
2	142.541590	46.531230	7.217841	149.944190	2165.155364	15.551852	-0.159196	-0.853114	46.531230	0.996846	7.564334	0.415658	PNEUMONIA
3	128.992903	53.869477	7.354550	139.789448	2901.920562	26.109848	-0.427046	-0.741057	53.869477	0.996619	10.623842	0.531515	PNEUMONIA
4	154.931903	62.773383	6.949651	167.165763	3940.497645	11.628231	0.943004	-1.394711	62.773383	0.994432	19.421998	0.528539	PNEUMONIA
5	138.402001	46.089991	7.205093	145.874607	2124.287230	19.181818	0.598208	-1.087847	46.089991	0.996991	7.900782	0.534019	PNEUMONIA
6	125.474252	51.981676	7.437894	135.815619	2702.094656	29.412371	-0.639158	-0.590980	51.981676	0.997084	8.477351	0.504962	PNEUMONIA
7	109.163294	47.310573	7.409364	118.974430	2238.290300	28.142857	-1.062299	-0.076061	47.310573	0.996169	10.549122	0.508046	PNEUMONIA
8	111.621449	65.720846	7.511337	129.532149	4319.229654	23.758065	-1.117018	-0.390145	65.720846	0.993462	28.551465	0.511194	PNEUMONIA
9	117.563693	52.621198	7.393003	128.802998	2768.990465	17.116208	-0.792609	-0.491352	52.621198	0.996232	11.828002	0.448739	PNEUMONIA
10	130.409016	49.367538	7.254372	139.440544	2437.153835	17.538793	-0.182829	-0.906614	49.367538	0.996740	9.721029	0.488947	PNEUMONIA

Figure 6.3 Extracted features

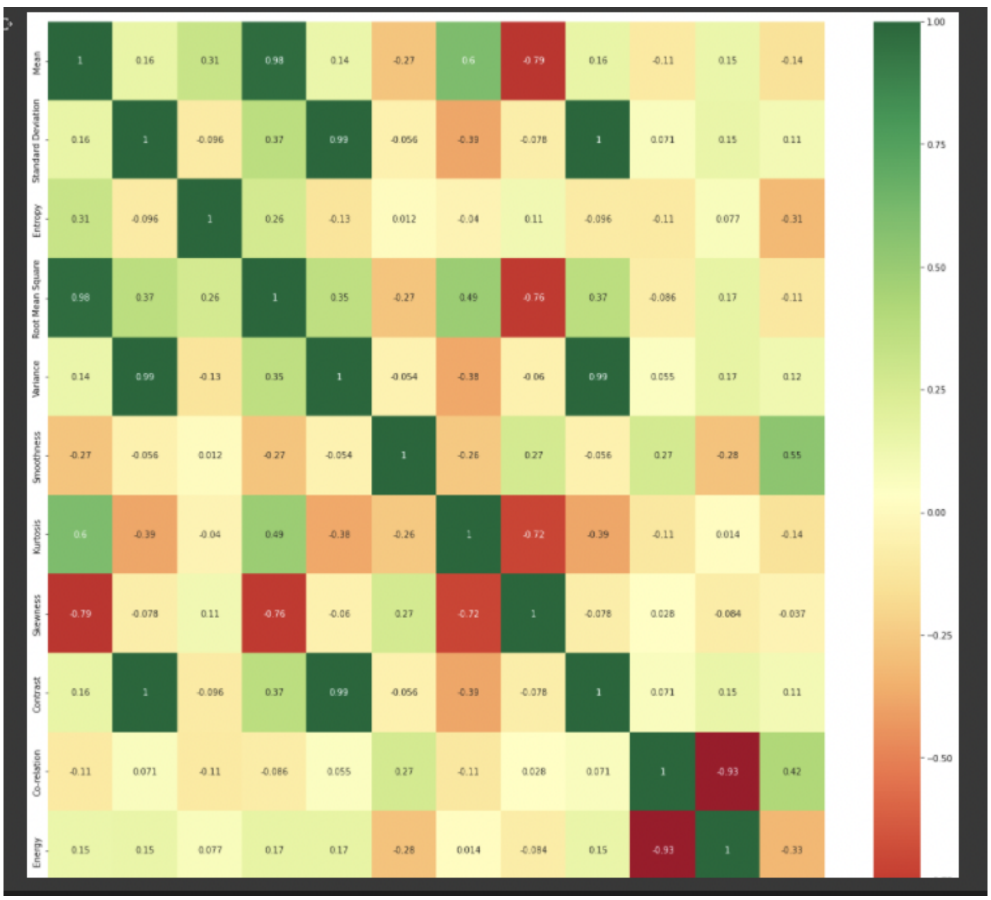


Figure 6.4 Correlations of each feature in dataset

6.2. ALGORITHMS

```
[ ] random_forest_modelh5 = loadModel("random_forest_model.h5")
    predict_train_data = random_forest_modelh5.predict(X_test)

    from sklearn import metrics

    print("Accuracy = {0:.3f}".format(metrics.accuracy_score(y_test, predict_train_data)))

    Accuracy = 0.840
```

Figure 6.5 Random Forest Classifier Algorithm

```
[ ] #Fitting Decision Tree classifier to the training set
    from sklearn.tree import DecisionTreeClassifier
    classifier= DecisionTreeClassifier(criterion='entropy', random_state=0)
    classifier.fit(X_train, y_train)

    DecisionTreeClassifier(criterion='entropy', random_state=0)

[ ] saveModel(classifier,"decisonTreeClassifier.h5")

    classifier_model = loadModel("decisonTreeClassifier.h5")
    y_pred= classifier_model.predict(X_test)
    from sklearn import metrics

    print("Accuracy = {0:.3f}".format(metrics.accuracy_score(y_test, y_pred)))
```

Figure 6.6 Decision Tree Classifier Algorithm

```
Logistic Regression

[ ] #Fitting Logistic Regression to the training set
    from sklearn.linear_model import LogisticRegression
    classifier= LogisticRegression(random_state=0)
    classifier.fit(X_train, y_train)

    /usr/local/lib/python3.7/dist-packages/sklearn/utils/validation.py:993: DataConversionWarnin
      y = column_or_1d(y, warn=True)
    /usr/local/lib/python3.7/dist-packages/sklearn/linear_model/_logistic.py:818: ConvergenceWar
    STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

    Increase the number of iterations (max_iter) or scale the data as shown in:
        https://scikit-learn.org/stable/modules/preprocessing.html
    Please also refer to the documentation for alternative solver options:
        https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression
      extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
    LogisticRegression(random_state=0)

[ ] saveModel(classifier,"LogisticRegression.h5")

    #Predicting the test set result
    classifier_model_lr = loadModel("LogisticRegression.h5")
    y_pred2= classifier_model_lr.predict(X_test)
    from sklearn import metrics

    print("Accuracy = {0:.3f}".format(metrics.accuracy_score(y_test, y_pred2)))

    Accuracy = 0.819
```

Figure 6.7 Logistic Regression Algorithm

Support vector Machines

```
[ ] from sklearn.svm import SVC # "Support vector classifier"
    classifier = SVC(kernel='linear', random_state=0)
    classifier.fit(X_train, y_train)

    /usr/local/lib/python3.7/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning:
      y = column_or_1d(y, warn=True)
    SVC(kernel='linear', random_state=0)

[ ] saveModel(classifier,"SVM.h5")

[ ] #Predicting the test set result
    svm_model = loadModel("SVM.h5")
    y_pred3= svm_model.predict(X_test)
    from sklearn import metrics

    print("Accuracy = {0:.3f}".format(metrics.accuracy_score(y_test, y_pred3)))

    Accuracy = 0.803
```

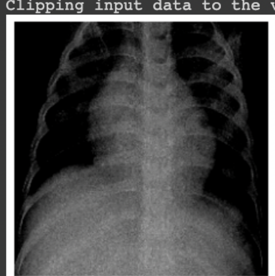
Figure 6.8 Support Vector Machines Algorithm

```
def get_class_string_from_index(index):
   for class_string, class_index in validation_generator.class_indices.items():
      if class_index == index:
         return class_string

x, y = next(validation_generator)
image = x[0, :, :, :]
true_index = np.argmax(y[0])
plt.imshow(image)
plt.axis('off')
plt.show()

# Expand the validation image to (1, 224, 224, 3) before predicting the label
prediction_scores = model.predict(np.expand_dims(image, axis=0))
predicted_index = np.argmax(prediction_scores)
print("True label: " + get_class_string_from_index(true_index))
print("Predicted label: " + get_class_string_from_index(predicted_index))
```

```
Clipping input data to the valid range for imshow with RGB data ([0..1] for floats or [0..255] for integers).
```



```
True label: PNEUMONIA
Predicted label: PNEUMONIA
```

Figure 6.9 Mobile Net Algorithm

```
[ ] import cv2
    import tensorflow as tf

    CATEGORIES= ['NORMAL','PNEUMONIA']
    def prepare(filepath):
        IMG_SIZE=224
        img_array=cv2.imread(filepath)
        img_array = img_array / 255.0
        new_array=cv2.resize(img_array,(IMG_SIZE,IMG_SIZE))
        return new_array.reshape(-1,IMG_SIZE,IMG_SIZE,3)


    prediction = model.predict(prepare('/content/drive/MyDrive/Final year Project/Pneumonia/val/NORMAL/NORMAL2-IM-1440-0001.jpeg'))
    prediction = np.argmax(prediction,-1)
    print(prediction)
    print(CATEGORIES[int(prediction[0])])

    [0]
    NORMAL
```

Figure 6.10 VGG Classifier Algorithm

CHAPTER 8

TESTING AND RESULTS

CHAPTER 7 TESTING AND RESULTS

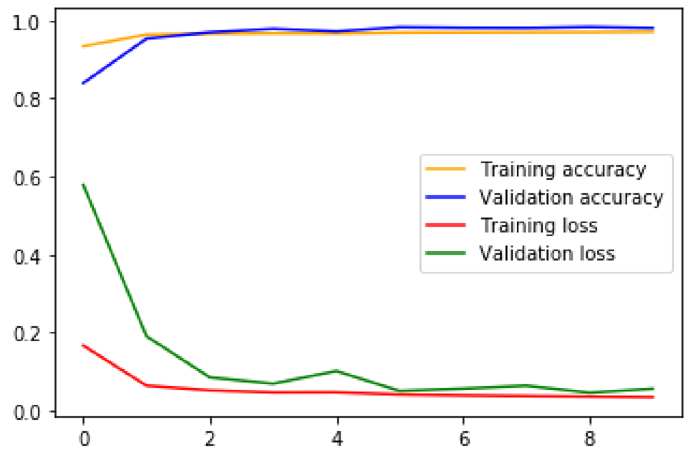


Figure 7.1 Summary of training evaluation & loss of Image Segmentation

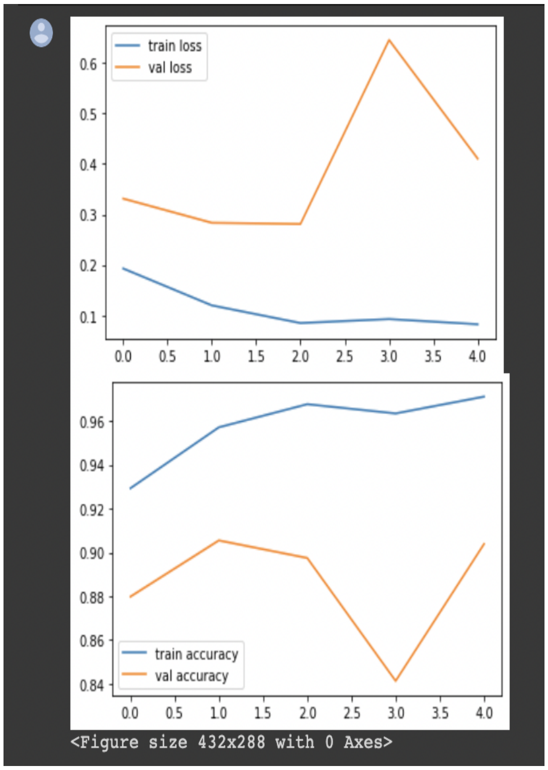
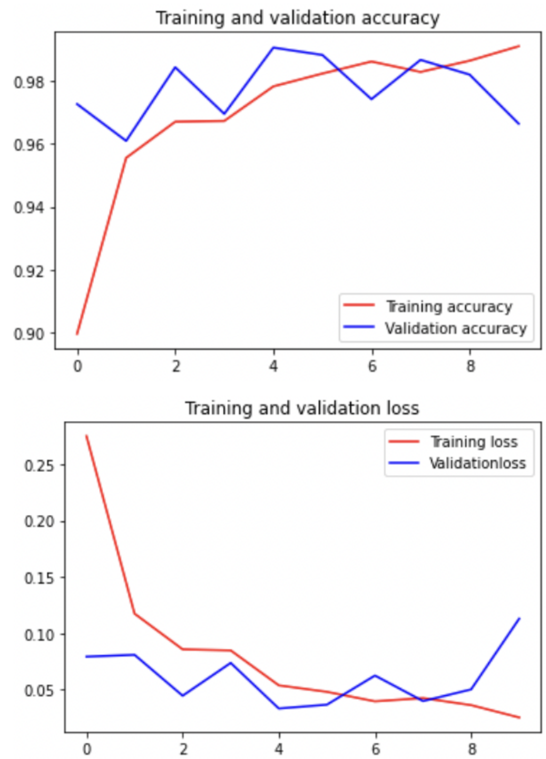
Table 7.1: Summary of Dataset

Category	Training dataset	Testing Dataset
Pneumonia	6119	2623
Normal	4965	2128

Table 7.2: Summary of Supervised and Transfer learning experimentation results

Method	Classifier	Training Accuracy	Testing Accuracy
SMLA	Random forest	86.3%	84.1%
SMLA	Logistic Regression	80.2%	75.1%
SMLA	Decision Trees	80.1%	79.1%
SMLA	SVM	79.22%	67.88%
TL	Mobile Net	88.11%	81.0%
TL	VGG	76.34%	61%

Table 7.3: Summary of training evaluation & loss of transfer learning

VGG Classifier	MobileNet training & Testing plot
	

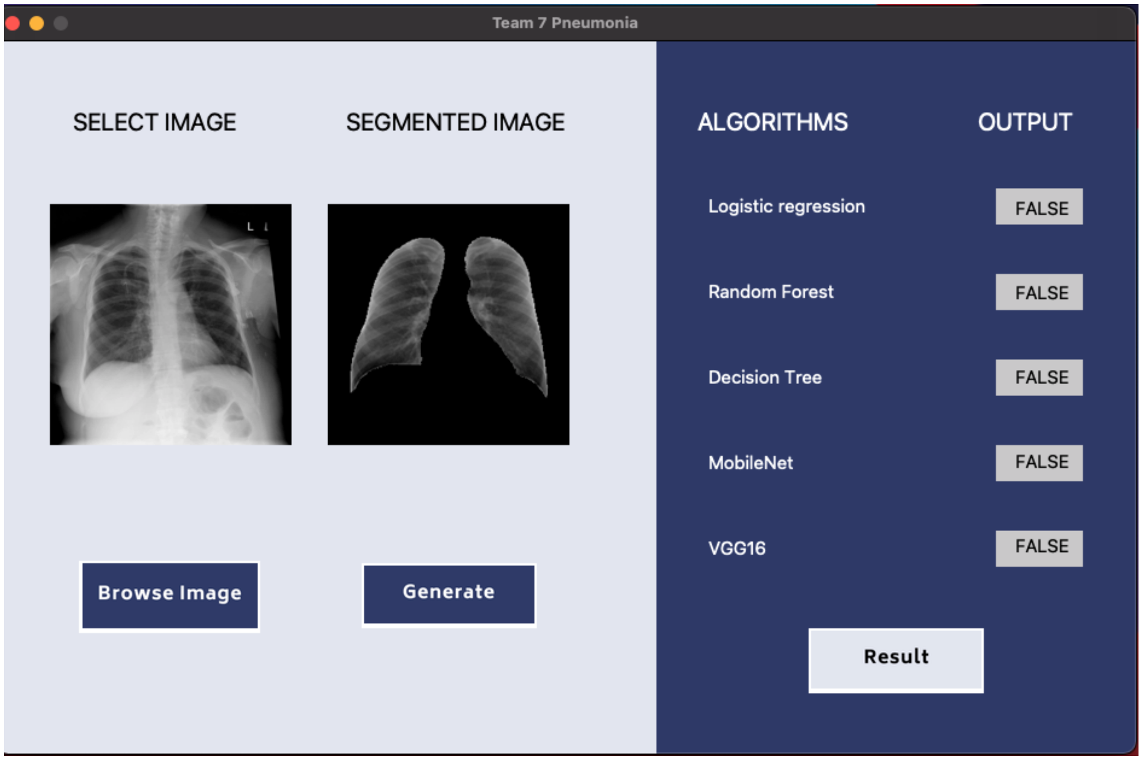


Figure 7.2 Integrated Application displaying negative for Pneumonia Disease

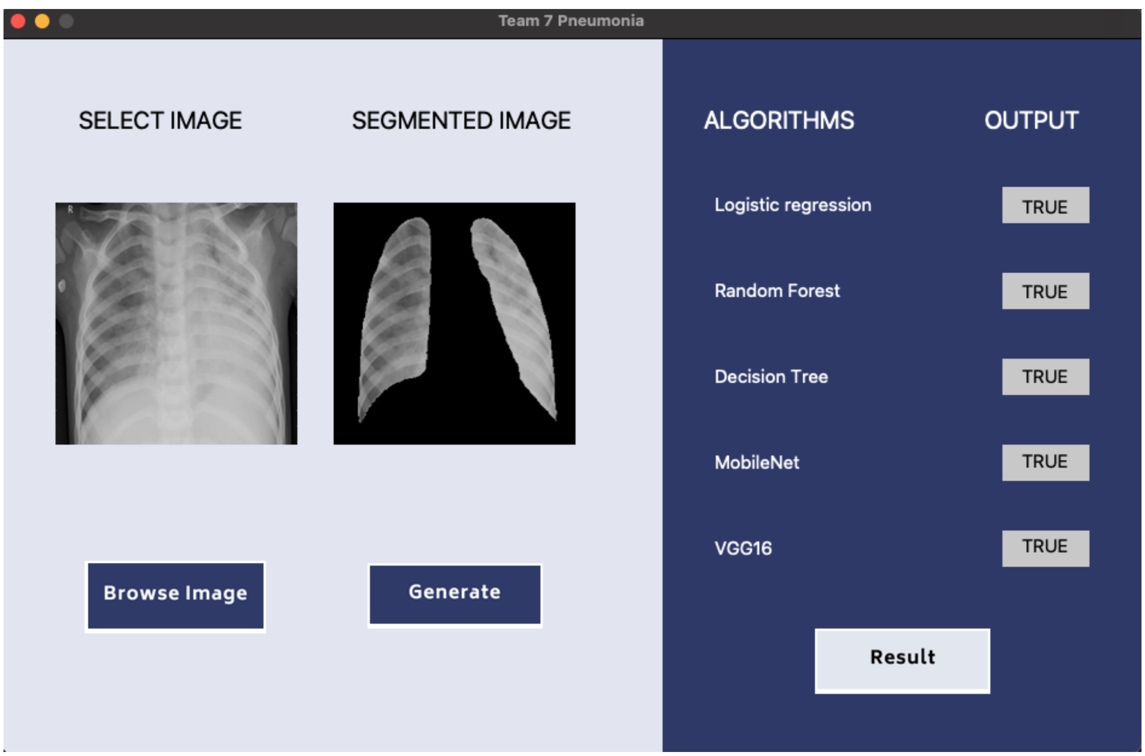


Figure 7.3 Integrated Application displaying Positive for Pneumonia Disease

REFERENCES

[1] Qin, Z. Z., Ahmed, S., Sarker, M. S., Paul, K., Adel, A. S. S., Naheyan, T., … Creswell, J. (2021). Tuberculosis detection from chest x-rays for triaging in a high tuberculosis-burden setting: an evaluation of five artificial intelligence algorithms. The Lancet Digital Health, 3(9),e543–e554. doi:10.1016/s2589-7500(21)00116-3

[2] Pasa, F., Golkov, V., Pfeiffer, F. et al. Efficient Deep Network Architectures for Fast Chest X-Ray Tuberculosis Screening and Visualization. Sci Rep 9, 6268 (2019). https://doi.org/10.1038/s41598-019-42557-4

[3] Yoo, Seung Hoon; Geng, Hui; Chiu, Tin Lok; Yu, Siu Ki; Cho, Dae Chul; Heo, Jin; Choi, Min Sung; Choi, Il Hyun; Cung Van, Cong; Nhung, Nguen Viet; Min, Byung Jun; Lee, Ho
(2020). Deep Learning-Based Decision-Tree Classifier for COVID-19 Diagnosis From Chest X-ray Imaging. Frontiers in Medicine, 7(), 427–. doi:10.3389/fmed.2020.00427

[4] Chandra T.B., Verma K. (2020) Pneumonia Detection on Chest X-Ray Using Machine Learning Paradigm. In: Chaudhuri B., Nakagawa M., Khanna P., Kumar S. (eds) Proceedings of 3rd International Conference on Computer Vision and Image Processing. Advances in Intelligent Systems and Computing, vol 1022. Springer, Singapore.
https://doi.org/10.1007/978-981-32-9088-4_3

[5] van Cleeff, M., Kivihya-Ndugga, L., Meme, H. et al. The role and performance of chest X-ray for the diagnosis of tuberculosis: A cost-effectiveness analysis in Nairobi, Kenya. BMC Infect Dis 5, 111 (2005). https://doi.org/10.1186/1471-2334-5-111

[6] Setio AAA, Traverso A, de Bel T, Berens MSN, van den Bogaard C, Cerello P, Chen H, Dou Q, Fantacci ME, Geurts B, et al. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the LUNA16 challenge. Med Image Anal 2017;42: 1–13.

[7] Bhandary Abhir, et al. Deep-learning framework to detect lung abnormality – a study with chest X-Ray and lung CT scan images. Pattern Recogn Lett January 2020;129:271–8 https://doi.org/10.1016/j.patrec.2019.11.013.

[8] Chouhan V, et al. A novel transfer learning-based approach for pneumonia detection in chest X-ray images. Appl Sci 2020;10(2):559. https://doi.org/10.3390/app10020559.

APPENDIX A

e-ISSN: 2395-0056 p-ISSN: 2395-0072

International Research Journal of Engineering and Technology (IRJET)

(An ISO 9001 : 2008 Certified Journal)

Is hereby awarding this certificate to

Mythri J L

In recognition the publication of the manuscript entitled

An Automated Framework for Diagnosing Lungs Related Issues using ML and Data Analytics

published in our Journal Volume 9 Issue 5 May 2022

Editor in Chief

E-mail : editor@irjet.net

Impact Factor : 7.529

www.irjet.net

e-ISSN: 2395-0056 p-ISSN: 2395-0072

International Research Journal of Engineering and Technology (IRJET)

(An ISO 9001 : 2008 Certified Journal)

Is hereby awarding this certificate to

Naveen A

In recognition the publication of the manuscript entitled

An Automated Framework for Diagnosing Lungs Related Issues using ML and Data Analytics

published in our Journal Volume 9 Issue 5 May 2022

Editor in Chief

E-mail : editor@irjet.net

Impact Factor : 7.529

www.irjet.net

e-ISSN: 2395-0056 p-ISSN: 2395-0072

International Research Journal of Engineering and Technology (IRJET)

(An ISO 9001 : 2008 Certified Journal)

Is hereby awarding this certificate to

Mohammed Maaz Ahmed Khan

In recognition the publication of the manuscript entitled

An Automated Framework for Diagnosing Lungs Related Issues using ML and Data Analytics

published in our Journal Volume 9 Issue 5 May 2022

Editor in Chief

Impact Factor : 7.529

www.irjet.net

E-mail : editor@irjet.net

e-ISSN: 2395-0056 p-ISSN: 2395-0072

International Research Journal of Engineering and Technology (IRJET)

(An ISO 9001 : 2008 Certified Journal)

Is hereby awarding this certificate to

Mohammed Siddiq S

In recognition the publication of the manuscript entitled

An Automated Framework for Diagnosing Lungs Related Issues using ML and Data Analytics

published in our Journal Volume 9 Issue 5 May 2022

Editor in Chief

Impact Factor : 7.529

www.irjet.net

E-mail : editor@irjet.net

e-ISSN: 2395-0056 p-ISSN: 2395-0072

International Research Journal of Engineering and Technology (IRJET)

(An ISO 9001 : 2008 Certified Journal)

Is hereby awarding this certificate to

Dr. Rajesh T M

In recognition the publication of the manuscript entitled

An Automated Framework for Diagnosing Lungs Related Issues using ML and Data Analytics

published in our Journal Volume 9 Issue 5 May 2022

Editor in Chief

E-mail : editor@irjet.net

Impact Factor : 7.529

www.irjet.net

International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 05 | May 2022 www.irjet.net p-ISSN: 2395-0072

AN AUTOMATED FRAMEWORK FOR DIAGNOSING LUNGS RELATED ISSUES USING ML AND DATA ANALYTICS

Mohammed Maaz Ahmed Khan[1], Mohammed Siddiq S[2], Mythri J L[3], Naveen A[4], Dr. Rajesh T M[5]

[1,2,3,4,5] *Department of Computer Science and Engineering, Dayananda Sagar University, Bangalore, Karnataka, India*

Abstract - *Compared to most other tissues, lungs are directly exposed to oxygen concentrations. Lung diseases are one of the leading causes of death. There are many different lung diseases, some of which are caused by viral, bacterial, or fungal infections. Other lung diseases are associated with environmental factors, including COVID19, tuberculosis, bronchitis, pneumonia, etc. Deep learning has shown great potential when applied to medical images for disease detection including lung disease. We build and compare two pre-trained models, MobileNet and VGG16 architectures using the Transfer learning approach. We have also used Supervised Machine Learning algorithms like Random forest, Decision Trees, Support Vector Machines, and Logistic Regression. This paper provides the analysis which we have performed on the different algorithms.*

***Key Words*: Fungal infections, Pneumonia, MobileNet, VGG16, Transfer Learning, Supervised Machine Learning algorithms**

1. INTRODUCTION

Lung disorders often called respiratory diseases, are illnesses that affect the lungs' airways and other components. Pneumonia, TB, and Coronavirus Disease are all examples of lung diseases (COVID-19). According to the Forum of International Respiratory Societies, around 334 million people have asthma, tuberculosis kills 1.4 million people each year, lung cancer kills 1.6 million people, and pneumonia kills millions. COVID-19 was a global pandemic that infected millions of individuals and put a strain on healthcare services. Lung illnesses are without a doubt one of the world's leading causes of death and disability. Early identification is crucial for boosting long-term survival rates and enhancing the possibilities of recovery. Lung disease is usually discovered by a physical exam & skin tests, blood tests, sputum sample tests, chest X-ray exams, and computed tomography (CT) scan exams are all used to identify cancer. Deep learning has recently shown considerable promise in disease identification using medical pictures, particularly lung disease.

The threat of lung illnesses is enormous, in particular in growing and low-middle-income countries, wherein tens of thousands and thousands of humans are dealing with poverty and air pollutants. According to the estimation of WHO, over four million untimely deaths arise yearly from household air pollutants-associated illnesses, together with asthma, and pneumonia. Hence, it's far more important to take important steps to lessen air pollutants and carbon emissions. It is likewise vital to enforce green diagnostic structures that can help in detecting lung illnesses. Since December 2019, a singular coronavirus sickness 2019 (COVID-19) has been inflicting critical lung harm and respiration problems. In addition, pneumonia, a form of lung sickness, may be because of the causative virus of COVID-19 or can be because of different viral or bacterial infections. Hence, early detection of lung illnesses has turned out to be more essential than ever. Recently, the virtual era has turned out to be a greater essential worldwide. This challenge can offer medical doctors and different researchers a course for detecting lung sickness with the assistance of a deep studying methodology. A huge quantity of lung X-ray pics are used as a dataset

1.1 Problem Definition

Machine Learning performs an essential position in clinical systems. Lung illnesses are one of the main reasons for death. The early identity and prediction of lung illnesses have emerged as a need withinside the research, as it may facilitate the subsequent scientific control of patients. Machine Learning-primarily based totally selection help systems offers the contribution to the medical doctors of their analysis decisions. The mission taken into consideration is the class of lung illnesses like Pneumonia, Tuberculosis, Lung cancer, and Covid 19. Machine Learning and Deep Learning are used to manage information in addition to creating fashions for diagnosing patients. Combining the processing of affected person statistics with information from chest X-rays and CT scans, using CNN with the famous pre-skilled model, These Neural networks for information of this shape are the techniques used for this mission to pick out lung illnesses.

2. RELATED WORKS

A fully CNN has been proposed in Ref. [6] to reduce the false-positive rate in classifying the lung nodules. This method can only analyze the nature of the CT scan images in order to reduce the probability of a wrong diagnosis.

In Ref. [7], a framework for deep learning is proposed to predict lung cancer and pneumonia offering two deep learning methods. Initially, they used modified AlexNet for the diagnosis of chest X-rays. Moreover, in the modified AlexNet, SVM is implemented for the purpose of classification.

Deep learning methods are also proposed in Ref. [8] where several transfer learning methods such as DenseNet121, AlexNet, Inception V3, etc., are used for pneumonia diagnosis. However, the parameter tuning for their implemented methods is very complex.

3. REQUIREMENTS

The Software requirements used for performing the experiments are as follows, Transfer learning, React framework, Tensorflow, Python3, OpenCV, Keras, Numpy, Matplotlib, Scikit-learn, and Tkinter

In order to successfully run the project the following hardware requirements are needed CPU Intel i3+, RAM 2GB+, Storage Space At least 2GB, Mobile phone (Android & iPhone)

3.1 System Features

Constructing dataset for Tuberculosis, pneumonia, lung cancer, and COVID 19, Model portable on IOS and Android platform, Recognition of lung-related diseases

4. METHODOLOGY

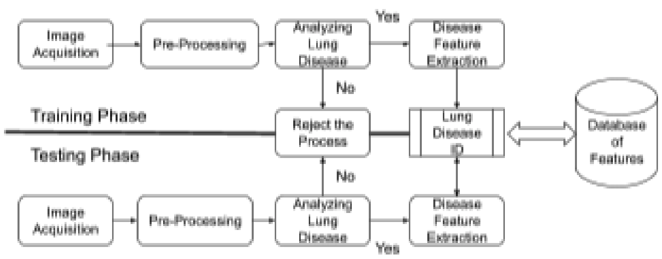


Fig -1: Proposed Design

4.1 Image Acquisition Phase

The first step is to acquire images. To produce a classification model, the computer needs to learn by example. The computer needs to view many images to recognize an object. Other types of data, such as time-series data and voice data, can also be used to train deep learning models. In the context of the work used in this project, the relevant data required to detect lung disease will be images. Images that could be used include chest X-ray, and CT scan. The output of this step is images that will later be used to train the model.

4.2 Preprocessing Phase

The second step is preprocessing. Here, the image could be enhanced or modified to improve image quality. Image modification such as lung segmentation and bone elimination could be used to identify the region of interest (ROI), whereby the detection of the lung disease can then be performed on the ROI. Edge detection could also be used to provide an alternate data representation. Data augmentation could be applied to the images to increase the amount of available data. Feature extraction could also be conducted so that the deep learning model could identify important features to identify a certain object or class. The output of this step is a set of images whereby the quality of the images is enhanced, or unwanted objects have been removed. The output of this step is images that were enhanced or modified that will later be used in training.

4.3 Training Phase

In the third step, namely training, three aspects could be considered. These aspects are the selection of deep learning algorithms, usage of transfer learning, and usage of an ensemble. There are numerous deep learning algorithms, for example, multilayer perceptron neural network (MPNN), recurrent neural network (RNN), and the aforementioned CNN. Different algorithms have different learning styles. CNN works particularly well with images. A deep learning algorithm should be chosen based on the nature of the data at hand. Transfer learning refers to the transfer of knowledge from one model to another. Ensemble refers to the usage of more than one model during classification. Transfer learning and ensemble are techniques used to reduce training time, improve classification accuracy and reduce overfitting. The output of this step is models generated from the data learned.

4.4 Classification Phase

In the fourth and final step, which is classification, the trained model will predict which class an image belongs to. For example, if a model was trained to differentiate X-ray images of healthy lungs and tuberculosis-infected lungs, it should be able to correctly classify new images (images that are never seen by the model before) into healthy lungs or tuberculosis-infected lungs. The model will give a probability score for the image. The probability score represents how likely an image belongs to a certain class. At the end of this step, the image will be classified based on the probability score given to it by model

5. TESTS AND RESULTS

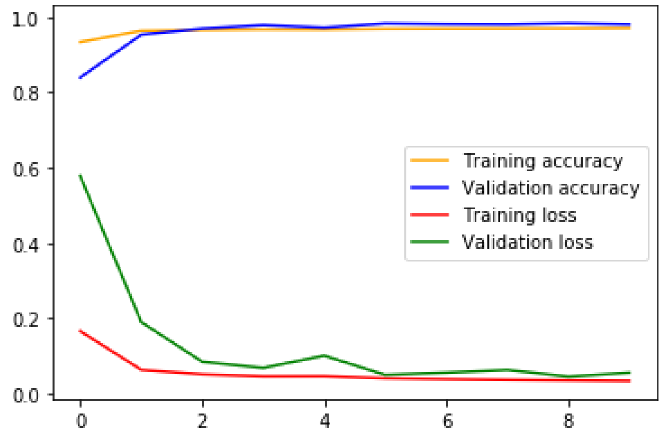


Chart -1: Summary of training evaluation & loss of Image Segmentation

Table -1: Summary of Dataset

Summary of Dataset		
Category	**Training dataset**	**Testing Dataset**
Pneumonia	6119	2623
Normal	4965	2128

Table -1: Summary of Supervised and Transfer learning experimentation results

Summary of Supervised and Transfer learning experimentation results		
Algorithm	Training Accuracy	Testing Accuracy
Random forest	86.3%	84.1%
Logistic Regression	80.2%	75.1%
Decision Trees	80.1%	79.1%
SVM	79.22%	67.88%
Mobile Net	99.83%	99.83%
VGG16	99%	98.24%

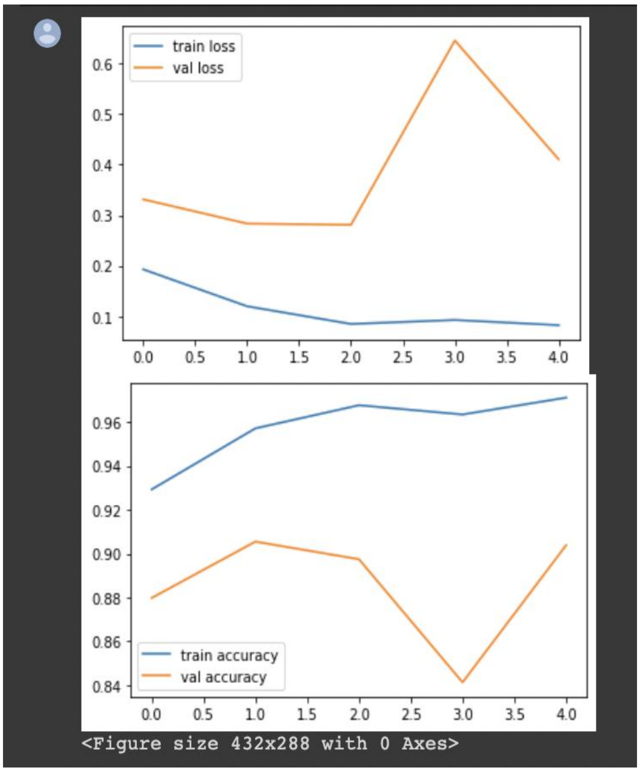


Chart -2: Summary of training evaluation & loss of VGG Classifier

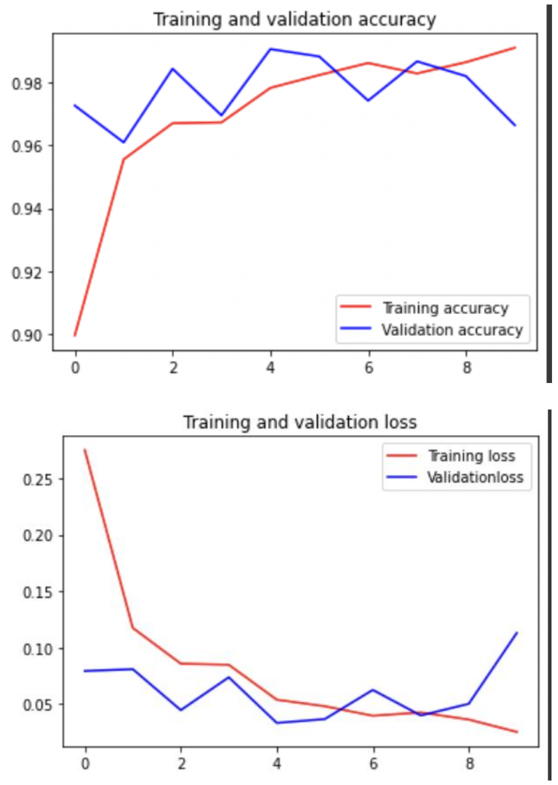


Chart -3: Summary of training evaluation & loss of MobileNet

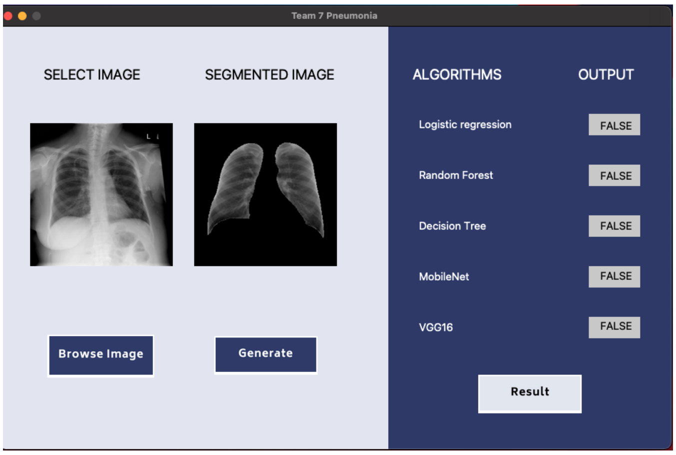


Fig -2: Integrated Application displaying negative for Pneumonia Disease

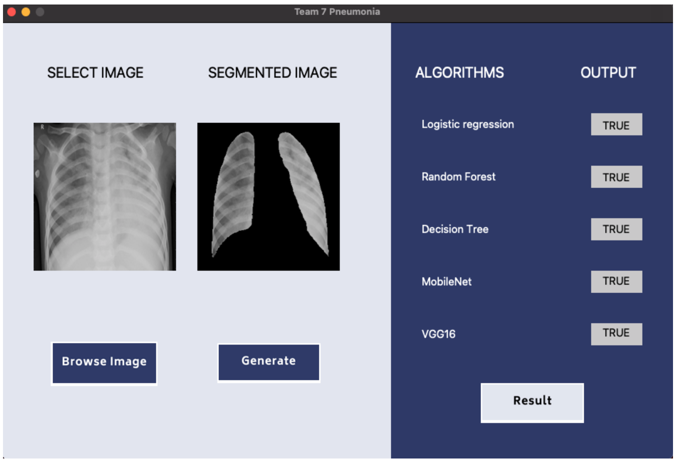


Fig -3: Integrated Application displaying Positive for Pneumonia Disease

6. DELIVERABLES

The final product is a fully integrated desktop application that can classify various lung-based diseases through X-Rays and CT-Scans, these include diseases such as Pneumonia, Lung cancer, Covid-19, and Tuberculosis. This product is designed in such a way that any person with basic knowledge can upload their X-Ray or CT-Scans to get a detailed analysis.

The Application will be used not only as a second opinion but also as a way of initial screening and Cross Verification among radiologists, as detection of diseases from a Chest X-ray and CT-Scan is complex and may go unnoticed.

This is a real-time application that uses machine learning and artificial intelligence (AI) based algorithms to assess the condition of the lungs, return the health of the lungs and the percentage of infection (if affected by the diseases mentioned above). The final application is projected to be implemented once the development and initial testing are done.

This is a project under the collaboration of Dayananda Sagar University, CDSIMER, and Sagar Group of Hospitals.

ACKNOWLEDGEMENT

It is a great pleasure for us to acknowledge the assistance and the support of many individuals who have been responsible for the successful completion of this project work.

First, we take this opportunity to express our sincere gratitude to the School of Engineering & Technology, Dayananda Sagar University for providing us with a great opportunity to pursue our Bachelor's degree in this institution.

We would like to thank Dr. A Srinivas. Dean, School of Engineering & Technology, Dayananda Sagar University for his constant encouragement and expert advice. It is a matter of immense pleasure to express our sincere thanks to Dr. Girisha G S, Department Chairman, Computer Science, and Engineering, Dayananda Sagar University, for providing the right academic guidance that made our task possible.

We would like to thank our guide Dr. Rajesh T M, Associate Professor, Dept. of Computer Science and Engineering, Dayananda Sagar University, for sparing his valuable time to extend help in every step of our project work, which paved the way for smooth progress and the fruitful culmination of the project.

We would like to thank our Project Coordinator Dr. Meenakshi Malhotra and all the staff members of Computer Science and Engineering for their support.

We are also grateful to our family and friends who provided us with every requirement throughout the course. We would like to thank one and all who directly or indirectly helped us in the Project work.

REFERENCES

[1] [1] Qin, Z. Z., Ahmed, S., Sarker, M. S., Paul, K., Adel, A. S. S., Naheyan, T., … Creswell, J. (2021). Tuberculosis detection from chest x-rays for triaging in a high tuberculosis-burden setting: an evaluation of five artificial intelligence algorithms. The Lancet Digital Health, 3(9), e543–e554. doi:10.1016/s2589-7500(21)00116-3

[2] [2] Pasa, F., Golkov, V., Pfeiffer, F. et al. Efficient Deep Network Architectures for Fast Chest X-Ray Tuberculosis Screening and Visualization. Sci Rep 9, 6268 (2019). https://doi.org/10.1038/s41598-019-42557-4

[3] [3] Yoo, Seung Hoon; Geng, Hui; Chiu, Tin Lok; Yu, Siu Ki; Cho, Dae Chul; Heo, Jin; Choi, Min Sung; Choi, Il Hyun; Cung Van, Cong; Nhung, Nguen Viet; Min, Byung Jun; Lee, Ho (2020). Deep Learning-Based Decision-Tree Classifier for COVID-19 Diagnosis From Chest X-ray Imaging. Frontiers in Medicine, 7(), 427–. doi:10.3389/fmed.2020.00427

[4] [4] Chandra T.B., Verma K. (2020) Pneumonia Detection on Chest X-Ray Using Machine Learning Paradigm. In: Chaudhuri B., Nakagawa M., Khanna P., Kumar S. (eds) Proceedings of 3rd International Conference on Computer Vision and Image Processing. Advances in Intelligent Systems and Computing, vol 1022. Springer, Singapore. https://doi.org/10.1007/978-981-32-9088-4_3

[5] [5] van Cleeff, M., Kivihya-Ndugga, L., Meme, H. et al. The role and performance of chest X-ray for the diagnosis of tuberculosis: A cost-effectiveness analysis in Nairobi, Kenya. BMC Infect Dis 5, 111 (2005). https://doi.org/10.1186/1471-2334-5-111

[6] [6] Setio AAA, Traverso A, de Bel T, Berens MSN, van den Bogaard C, Cerello P, Chen H, Dou Q, Fantacci ME, Geurts B, et al. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the LUNA16 challenge. Med Image Anal 2017;42: 1–13.

[7] [7] Bhandary Abhir, et al. Deep-learning framework to detect lung abnormality – a study with chest X-Ray and lung CT scan images. Pattern Recogn Lett January 2020;129:271–8 https://doi.org/10.1016/j.patrec.2019.11.013.

[8] [8] Chouhan V, et al. A novel transfer learning-based approach for pneumonia detection in chest X-ray images. Appl Sci 2020;10(2):559. https://doi.org/10.3390/app10020559.

BIOGRAPHIES

Mr. Mohammed Maaz Ahmed Khan is pursuing a Bachelor of Technology degree in the Department of Computer Science and Engineering from Dayananda Sagar University, located in Bangalore, Karnataka, India. He is currently in his final year of Engineering and will be graduating from Dayananda Sagar University in the year 2022.

Mr. Mohammed Siddiq S is pursuing a Bachelor of Technology degree in the Department of Computer Science and Engineering from Dayananda Sagar University, located in Bangalore, Karnataka, India. He is currently in his final year of Engineering and will be graduating from Dayananda Sagar University in the year 2022.

Miss. Mythri J L is pursuing a Bachelor of Technology degree in the Department of Computer Science and Engineering from Dayananda Sagar University, located in Bangalore, Karnataka, India. She is currently in her final year of Engineering and will be graduating from Dayananda Sagar University in the year 2022.

Mr. Naveen A is pursuing a Bachelor of Technology degree in the Department of Computer Science and Engineering from Dayananda Sagar University, located in Bangalore, Karnataka, India. He is currently in his final year of Engineering and will be graduating from Dayananda Sagar University in the year 2022.

Dr. Rajesh T M is an Associate Professor in the Department of Computer Science and Engineering from Dayananda Sagar University, located in Bangalore, Karnataka, India.