

Deep Learning Algorithm for Classification and Prediction of Lung Diseases using CT Scan Images

Submitted in partial fulfillment of the requirements for the award of

Bachelor of Engineering degree in

Electronics and Telecommunication Engineering

BY

S.SAI PAVAN REDDY(Reg.No.37250021)

N. RANJITH KUMAR (Reg.No.37250017)



**DEPARTMENT OF ELECTRONICS AND TELECOMMUNICATION
ENGINEERING**

SCHOOL OF ELECTRICAL AND ELECTRONICS

SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY)

Accredited with Grade "A" by NAAC

JEPPIAAR NAGAR, RAJIV GANDHI SALAI, CHENNAI - 600 119

APRIL - 2021



SATHYABAMA

**INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)**

Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE

www.sathyabama.ac.in

DEPARTMENT OF ELECTRONICS AND TELECOMMUNICATION

ENGINEERING

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of S SAI PAVAN REDDY (37250021) and Iyer N RANJITH KUMAR(37250017) who carried out the project entitled “DEEP LEARNING ALGORITHM FOR CLASSIFICATION AND PREDICTION OF LUNG DISEASES USING CT SCAN IMAGES” under my supervision from November 2020 to April 2021.

Internal Guide

Dr.T.Bernatin.,M.E., Ph.D.

Head of the Department

Dr. V. VIJAYA BASKAR M.E., Ph.D.

Submitted for Viva voce Examination held on_____

Internal Examiner

External Examiner

DECLARATION

I, S Sai Pavan Reddy (Reg. no. 37250021) and N Ranjith Kumar (Reg. no. 37250017) hereby declare that the Project Report entitled “DEEP LEARNING ALGORITHM FOR CLASSIFICATION AND PREDICTION OF LUNG DISEASES USING CT SCAN IMAGES” done by us under the guidance of Dr.T.Bernatin.,M.E., Ph.D is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering degree in Electronics and Telecommunication Engineering.

DATE:

PLACE:

SIGNATURE OF THE CANDIDATE

- 1.**
- 2.**

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to Board of Management of SATHYABAMA for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. N. M. NANDHITHA M.E., Ph.D.**, Dean, School of Electrical and Electronics and **Dr. V. VIJAYA BASKAR M.E., Ph.D.**, Head of the Department, Dept. of Electronics and Telecommunication Engineering for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **Dr.T.Bernatin.,M.E., Ph.D.** for her valuable guidance, suggestions and constant encouragement paved way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the Department of Electronics and Telecommunication Engineering who were helpful in many ways for the completion of the project.

I wish to express my thanks to the Project Panel members for their valuable feedbacks during the Project Reviews which were useful in many ways for the completion of the project.

ABSTARCT

Covid-19 is one of the dangerous and life-taking diseases in the world. However, early diagnosis and treatment can save a life. Although CT scan imaging is the best imaging technique in the medical field, it is difficult for doctors to interpret and identify the disease from CT scan images. Therefore computer-aided diagnosis can be helpful for doctors to identify them accurately. Many computer-aided techniques using image processing and machine learning has been studied and implemented. Deep learning has been verified as a popular and powerful method in many medical imaging diagnosis areas. In this paper, deep neural networks DNN are designed for lung Diseases classification. Those networks are applied to the CT scan image classification task with some adjustments for the benign and malignant lung nodules. The processing part of the pulmonary nodule is decomposed into the DNN, so that different network layers can be used to obtain the characteristics of the pulmonary nodules with dissimilar sizes. The essential capability of DNN to capture the structure of the image in its feature maps. Further improvements are available in the areas where the abnormalities lie, possibly through the use of deep learning algorithm, and the versatility of an automated discovery system.

TABLE OF CONTENTS

Chapter No	Title	Page No
	ABSTRACT	v
	LIST OF FIGURES	VI
	LIST OF TABLES	X
	LIST OF ABBREVIATIONS	XI
1	1.1 INTRODUCTION	01
	1.2 Project Overview	01
	1.2.1 Lung Disease Detection Application	01
	1.2.2 Research Element	02
	1.3 Project Goals	02

1.4 Project challenges	03
1.4.1 Neural Network Architecture design choices	03
1.4.2 Understanding Deep Learning Concepts	03
1.4.3 Activation Functions	04
1.4.4 Weight Initializations	05
1.4.5 Gradient Descent	06
1.4.6 Variants of Gradient Descent	07
1.5 Overview	08
1.6 OBJECTIVES	08

2

LITERATURE SURVEY

2.1 Deep convolution Pixel-wise Labelling For skin Lesson image segmentation	09
2.2 Early Detection of Melanoma Skin Cancer Using Classifiers	09
2.3 Skin Cancer Detection and Classification	10
2.4 Diagnosis of Diabetes Using classification Mining Technique	10

	2.5 Comparison of Feature Selection Methods	
	For diagnosis of cervical cancer using	
	SVM Classifier	10
	2.6 K-Nearest Neighbour Based Structural	
	Twin SVM	11
	2.7Applying Machine Learning Methods in	
	Diagnosing Hearty disease for Diabetic	
	Patient	11
	2.8 Liver Disease Prediction using SVM	
	And naïve bayes algorithm	12
	2.9 Diagnosis of Lung Cancer Prediction	
	System using Data mining classification	
	Techniques	12
	2.10 Proposed a system that classifies Lung	
	Disease as benign or malignant	13
3	PROPOSED SYSTEM	14

	3.1 salient features	14
	3.2 Proposed Method	16
	3.3 Construction of DNN model	17
4	EXPERIMENTAL METHODS AND RESULT	19
	4.1 Deep Learning Model	19
	4.1.1 Pre-Processing	19
	4.1.2 Splitting the Dataset	19
	4.1.3 Classification	19
	4.1.3(a) K-Nearest Neighbor	20
	4.1.3(b) Decision Tree	20
	4.1.3(c) Random Forest Classifier	20
	4.1.3(d) Logistic Regression	21
	4.1.3(e) Support Vector Machine	21

	4.1.3(f) Cognitive Radio	22
	4.1.3 (g) Deep Neural Network	21
	4.2 Classification	22
	4.3 Strengths	27
5	PERFORMANCE ANALYSIS	28
	CONCLUSION	28
	REFERENCES	29

LIST OF FIGURES

FIGURE NO	TITLE	PAGE No
3.2	Salient Features	10
3.4	Construction of DNN Model	12
4.2.1	Lung CT Scan Images	24
4.2.2	MATLAB Software with Code	24
4.2.3	After Running the Code Output Image	25
4.2.4	Giving CT Scan Image as a Input	25
4.2.5	Detecting the Given CT Scan Image Is healthy	26
4.2.6	Detecting the Given CT Scan Image Is affected	26
4..3.1	Performance analysis for each Algorithm	27

LIST OF TABLES

Table No No	Title	Page
4.3.1	Performance analysis for each algorithm	27

LIST OF ABBREVIATIONS

CT	COMPUTED TOMOGRAPHY
WEKA	WAIKATO ENVIRONMENT FOR KNOWLEDGE ANALYSIS
HU	HOUSE FIELD UNIT
ROI	REGION OF INTEREST
MRI	MAGNETIC RESONANCE IMAGING
SVM	SUPPORT VECTOR MACHINE
DNN	DEEP NEURAL NETWORK
CR	COGNITIVE RADIO
KNN	K-NEAREST NEIGHBOUR

CHAPTER 1

1.1 INTRODUCTION: Lung disease is one of the most prevalent diseases worldwide, causing 1.76 million deaths per year. Covid-19 is a critical disease for one year. This leads to death if not early detected. There is an increased risk of the health due to changing environment, climate and lifestyle. India tops the world in deaths due to lung diseases. Early diagnosis and treatment of lung diseases are critical to prevent complications including death.

Chest X-ray is currently the best available method for diagnosis, playing a crucial role in clinical care for prediction and classification of the lung disease. Using Deep Learning to predict lung diseases from Chest X-rays can be a lifesaving factor for an individual suffering from the disease.

This is possible as the results can be predicted with a high percentage of accuracy immediately. This paper presents an actual way for expert diagnosis of lung diseases using Deep Learning. It focuses on creating a system for the assistance of Radiologists in the recognition of lung diseases.

This will especially advantage rural areas where radiologists aren't easily available. Our system attaches radiology labs with radiologists who can diagnose faster and better with our model.

1.2 Project Overview

1.2.1 Lung Disease Detection Application

A web application has been developed to reveal a proof of concept. The application needs a user to upload a CT Scan. The application then processes the file and displays the images to the user. The user then chooses which scan he or she wants to forecast then the application pre-processes the CT Scan and infers the image to the predictive model. The output of the model is then exhibited to the user. The user has the choice to view the images via a carousel or a gallery mode.

1.2.2 Research Element

This report contains many aspects of research that support deep learning's ability to find lung disease within CT Scans. The data used to train the model was gathered from a deep learning competition called LUNA 2016 which are CT Scans that have been annotated by radiologists. The radiologists determined whether a nodule is lung disease or not and this location has been specified in 3D coordinates.

Deep Learning research has also been conducted to ensure that the correct architecture would be used to train the model. The architecture used is called U-Net. It acts as a deep learning segmentor of images. Using our evaluation methods we could see that the deep learning architecture is able to segment regions of interest relative to its accuracy but found that the model creates many false positives which means that the model required more regularization and training time, although it was trained for 40 hours.

1.3 Project Goals

The completed project should accomplish the following Objectives:

Self Education on Deep Learning is a large part of this project contains a lot of self education, initially the author knew very little about deep learning, as part of this project the author should have a good grip on deep learning concepts.

- **User Research and Evaluation** The project should have a user-centered design aspect. This means that the system should be outlined to help certain users. In this project the user would be medical professionals who work in diagnosing lung disease.
- **System: Upload CT Scans from** the system should be capable of getting CT Scans from Users that will be utilized by the Deep Learning Model.
- **System: Detection of Lung Disease** The system should be able to detect the lung disease within the CT scan images that users have uploaded.
- **System: Display Results** The system should be able to give information that our user can appropriately understand and gain insight from it.

1.4 Project Challenges

This section discusses the challenges that were overcome to complete this project. Each of the challenges are briefly inaugurated and detailed in later sections of the report. This section can be useful for those who focuses to do a deep learning project with large image datasets.

1.4.1 Neural Network Architecture Design Choices

Neural networks are architectures that are designed in of itself and there are many different types of architectures out there that exist to solve different problems. A main challenge of the author is to use an architecture that is able to find malignant tumor patterns in the data. To overcome this, the author has to undertake necessary research to excute the correct model design prior to training. There are many neural network architectures such as multi-layer perceptrons, convolutional networks and sequence models.

1.4.2 Understanding Deep Learning Concepts

Deep learning is an thrilling but new concept for the author. To overcome this challenge the author had performed research and study to deliver the project goals. A useful resource that has been used is Stanford's professor Andrew Ng's deep learning course on Coursera. This course enables the learners to gain a foundation to deep learning.

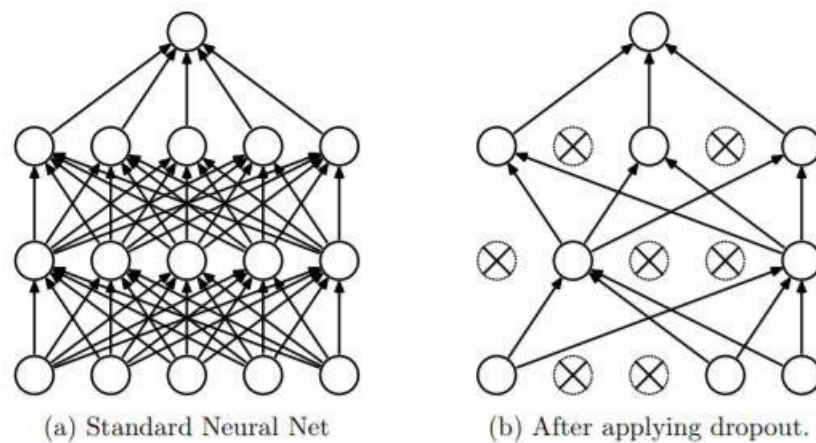
1.4.3 Activation Functions

Activation functions are important part of a neural network. They allow neural networks to create non-linear functions to explain glitches. The 3 utmost used activation functions are Sigmoid, Tan H and Re LU. Activation functions allow or stop neurons from firing into the next layer by comparing it into the activation function thresholds.

Activation functions are used both in the forward and backward propagation. wherein the forward propagation and activation function is used to calculate the loss. The output of a function is related to the real number and in backward

propagation, they are used to update the parameters of the neural network. commonly used activation functions in a neural network.

Neurons in a neural network have a high tendency to adapt co-dependency among each other during training period. This means that certain activations become more heavily weighted and outcomes in a network becoming overfit to the training set. To stop the network from overfitting. Dropout is used.



According to Hinton et al, Dropout is a regularization technique that reduced neuron's ability to be interdependent among one another during training time. This is achieved by shutting off random neurons in the network during period. This forces the network to learn better features that are more robust relative to random subsets of further neurons.

Method	Test Classification error %
L2	1.62
L2 + L1 applied towards the end of training	1.60
L2 + KL-sparsity	1.55
Max-norm	1.35
Dropout + L2	1.25
Dropout + Max-norm	1.05

Fig Comparing Regularization Techniques

Hinton et al also mention that compared to the standard regularization techniques such as L2 with learning decay. Test classification error is critically reduced which means that Dropout is currently state of the art (See

The effects of learned features on the MNIST benchmark can be seen in Fig

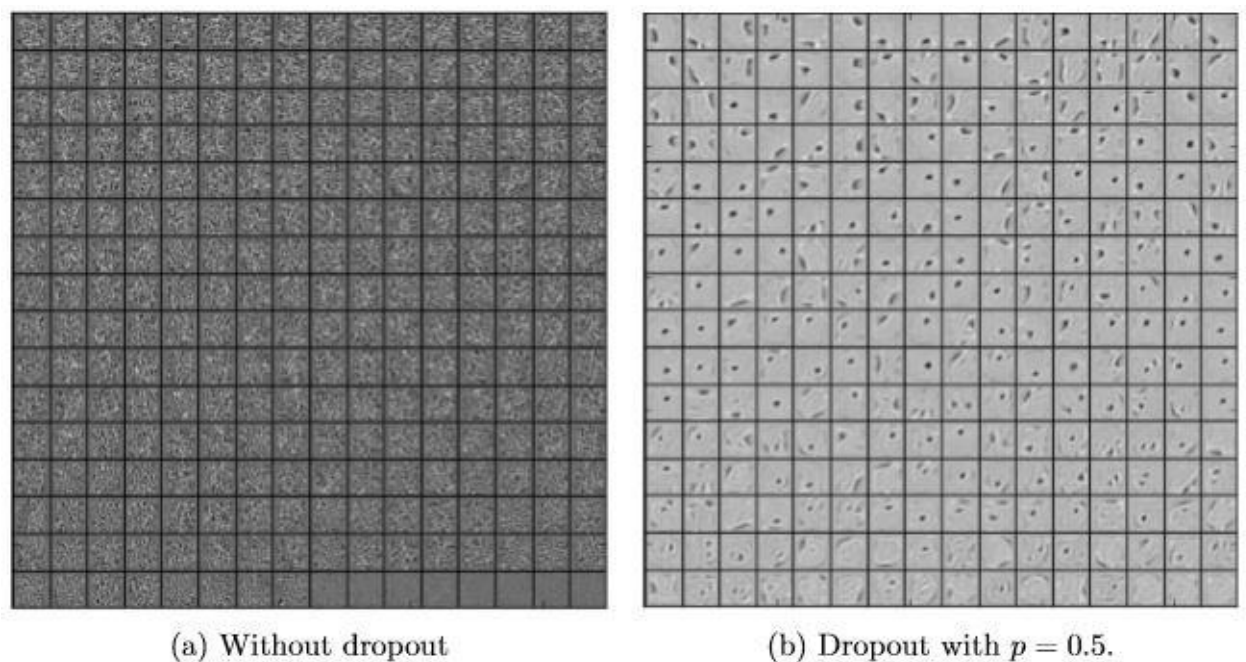


Figure: Dropout Effects on MNIST

1.4.4 Weight Initializations

One of the main problems encountered in deep learning is disappearing/exploding gradients. This is when the activations become lesser or greater at each layer. To summarise briefly, if the weights in a network are too small, then the signals shrink as it passes through each layer until its too small to be useful conversely if the weights in the network start too large then the signals become too large to be useful.

This also has a terrible consequence for the parameter updates on the weights as the network back propagates either giving very small updates or very large into

the network. This has terrible consequence for large neural networks as it means the larger the network is the more likely it will be to be slow to train and inaccurate.

The two main sophisticated ways to initialize the weights is Xavier and He. Both of these initializations put neurons into a Gaussian or uniform distribution and draws weights from the distribution at a certain mean and variance.

After reading Heetal's paper, the author decided to use this method as apparently it performs better than Xavier. (See Figure 3.15)

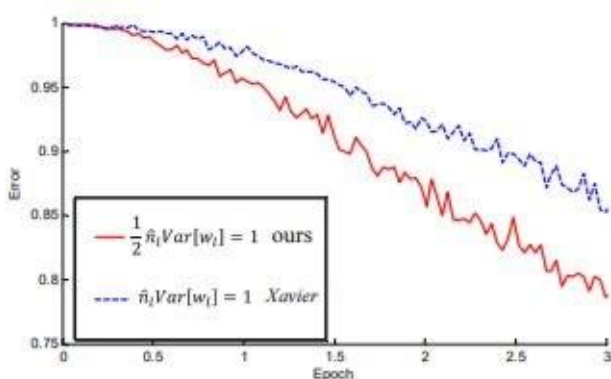


Figure 3.15: Heetal vs Xavier Performance

1.4.5 Gradient Descent

Gradient descent is the learning procedure used in Deep Learning. It is mostly used in the backpropagation step in the training of the neural network. In a neural network, the cost function measures the variation between a predicted value and the actual value. The idea behind gradient descent is that we increase the worth of the weights so that the weights minimize a cost function in the neural network.

Gradient descent takes the loss function and analyses a derivative. This is also identified as calculating the gradient of a loss function. The knowledge behind calculating the gradient is that it computes the top direction. Which the weights must be updated to guarantee the best outcome for minimizing the cost function.

The learning rate parameter then is used to determine the size of the stages we take given the direction. The closer the network gets to convergence i.e. lowest fact in the cost function, the more accurate the neural network becomes.

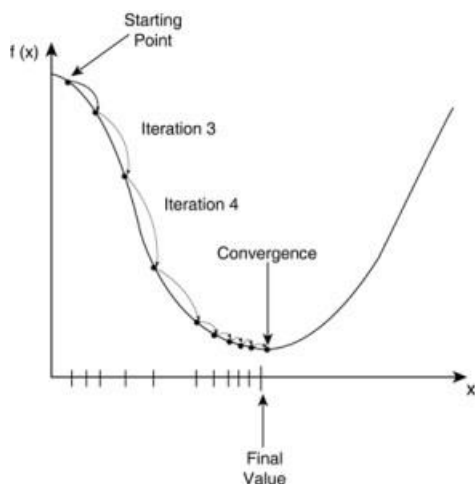


Figure 3.16: Gradient Descent Example

1.4.6 Variants of Gradient Descent

Over time there have been various variations of gradient descent that have improved the algorithm's capability. The main ones are group gradient descent, stochastic gradient descent and mini-batch gradient descent. Sebastian Ruder wrote a great paper on gradient descent optimization algorithms and according to Ruder the trade-off between the algorithms is the accuracy of the parameter update and the time it takes to perform the update..

1.5 Overview

This section discusses the choices made into improving the U-Net model. Deep Learning is a highly iterative procedure and training is extremely slow and can get stuck on plateaus. i.e. the model is not getting better. I've had this experience many times through training the U-Net for hours and getting bad results. To avoid this, different techniques have been applied to the model that helped greatly in training and reaching better convergence. It's also worth mentioning that there are many ways to improve a neural network and that it is constantly evolving as new methods are created. It is completely infeasible to explain each one. The author has decided to only introduce techniques that are effective and currently national of the art.

1.6 OBJECTIVES:

- Different types of cancer detection and classification using machine assistance have opened up a new research area for the initial detection of cancer, which has shown the capacity to reduce manual system impairments.
- This survey presents several sections on state of art methods, analysis and comparisons on benchmark datasets for the brain tumor, chest cancer, lung cancer, liver tumor, leukaemia and skin lesion detection respectively from F-measure, specificity, accuracy, sensitivity, precision points of view

CHAPTER 2

2.LITERATURE SURVEY

2.1 TITLE: Deep Convolution Pixel-wise Labeling for skin lesion Image Segmentation.

AUTHOR: Alehouse, Domenico D. Blois ,Mario Muscio ,Andrea Pennisi ,
Daniele Nardi

YEAR OF PUBLICATION: 2018

conducted an experiment by comparing public data across two network architectures and using the ISIC database to train the network and the PH2 database to demonstrate the used method is not data set related. The results of this experiment show that the proposed method is very successful, and even in the presence of hair and air / oil bubbles performs segmentation. The extra feature of this paper is data annotation which generates many test images, and the implementation of a semi-automatic Interface.

2.2 TITLE: Early Detection of Melanoma Skin Cancer Using Classifiers

AUTHOR: VS.Sabeera, P.Vamsi Krishna

YEAR OF PUBLICATION :2016

proposed a prevention of melanoma and early detection of skin cancer. A test method to assess lesion and to avoid melanoma is suggested. A real- time device has been built to prevent people from skin burns that warns users. User receives the warning message when he/she exposed to sun for longer period and the rays may harm the skin.

2.3 TITLE: Skin Cancer Detection and classification

AUTHOR: Pratik Dubal, Sankirtan Bhatt, Chaitanya Joglekar, Dr. Sonali Patil

YEAR OF PUBLICATION: 2017

Proposed a program which recognizes and detects skin sores based on images of general cameras. The images are portioned; Neural Network is able to predict the injuries with highest accuracy. On a data set of 463 images, the optimized Neural Network achieved a general accuracy of 76.9%.

2.4 TITLE: Diagnosis of Diabetes Using Classification Mining Techniques

AUTHOR: Iyer, J. S R. Sumbaly,

YEAR OF PUBLICATION :2016

Used the Decision Tree and Naïve Bayes algorithms, the goal is to find solutions for diabetes detection by analyzing patterns originating in the data through classification analysis. This becomes an effective method of detecting the disease that will help patients recover well in time. The study concluded that the J48 algorithm provides an accuracy rate of 74.8 percent.

2.5 TITLE: Comparison of Feature selection methods for diagnosis of cervical cancer using Support vector machine (SVM) classifier.

AUTHOR: B. Ashok and P. Aruna,

YEAR OF PUBLICATION: 2016

proposed Comparison of feature selection approaches for cervical cancer diagnosis using a SVM classifier. Image segmentation was performed using thresholding. Feature selection was attained using mutual information, sequential forward search, sequential floating forward search, and random subset feature selection methods. Using the sequential floating forward selection technique, 98.5 per cent accuracy, 98 per cent sensitivity and 97.5 per cent specificity were achieved, higher than the other methods.

2.6 TITLE: K-nearest neighbour based structural twin support vector machine.

AUTHOR: X. Pan, Y. Luo, Y. Xu

YEAR OF PUBLICATION:2015

proposed the development of a structural twin support vector (KNN-STSV). By applying the KNN method intra-class, the samples are given different weights in one class to enhance the structural details. For the other class, the expendable constraints to speed up the coaching cycle are eliminated by the inter-class KNN system. A quick clip algorithm for rate increase is further implemented for large-scale problems. Extensive experimental results on 22 datasets show the effectiveness of their proposed KNN-STSV.

2.7 TITLE: Applying Machine Learning Procedures in Diagnosing Heart Disease for Diabetic Patients.

AUTHOR: Parthiban, G. ,Srivatsa, S.K.

YEAR OF PUBLICATION:2012

have put their efforts into diagnosing heart disease using machine learning methods in diabetes patients. The Naive Bayes and SVM algorithms are implemented with assistance from WEKA. Using collection of data from 500 patients by Chennai Medical Institute. There are 142 patients suffering from the disease and 358 patients without the disease. Algorithm 74 percent of accuracy is achieved by using Naive Bayes. SVM delivers 94.60 maximum accuracy.

2.8 TITLE: Liver Disease Prediction using SVM and Naïve Bayes Algorithms

AUTHOR: Vijayarani ,Dhayanand S.

YEAR OF PUBLICATION :2015

predicted a liver disease based on the various algorithm's accuracy. In 1670.00 ms, Naive Bayes showed 61.28 per cent correctness. SVM achieves 79.66 per cent accuracy in 3210.00 Ms. MATLAB is used for implementation. SVM proves the highest accuracy in liver disease prediction relative to the Naive Bayes. Naive Bayes takes less time in terms of time execution as opposed to the SVM.

2.9 TITLE: Diagnosis of Lung Cancer Prediction System Using Data Mining Classification Techniques

AUTHOR: V.Krishnaiah

YEAR OF PUBLICATION :2013

built a prototype lung cancer prediction method, using classification techniques for data mining. Naïve Bayes is apparently the most powerful model for predicting lung cancer patients followed by IF-THEN regulation, Decision Trees and the Neural Network. Naïve Bayes achieves better outcomes for the diagnosis of lung cancer disease, and has done better than trees.

2.10 TITLE: proposed a system that classifies lung disease as benign or malignant.

AUTHOR: Gonzalez and Ponomaryvo

YEAR OF PUBLICATION :2012

The system uses the priority information and House field Unit (HU) to calculate the Region of Interest (ROI) model along with the edge detection. Shape features like area, eccentricity, circularity, fractal dimension and textural features like mean, variance, energy, entropy, skewness, contrast, and smoothness are extracted to train and analyze the support vector machine to analyze whether the nodule is benign or malignant. The advantage of this model is that it detects cancer as benign or malignant. The performance analysis shows that the proposed model key points.

CHAPTER:3

PROPOSED SYSTEM:

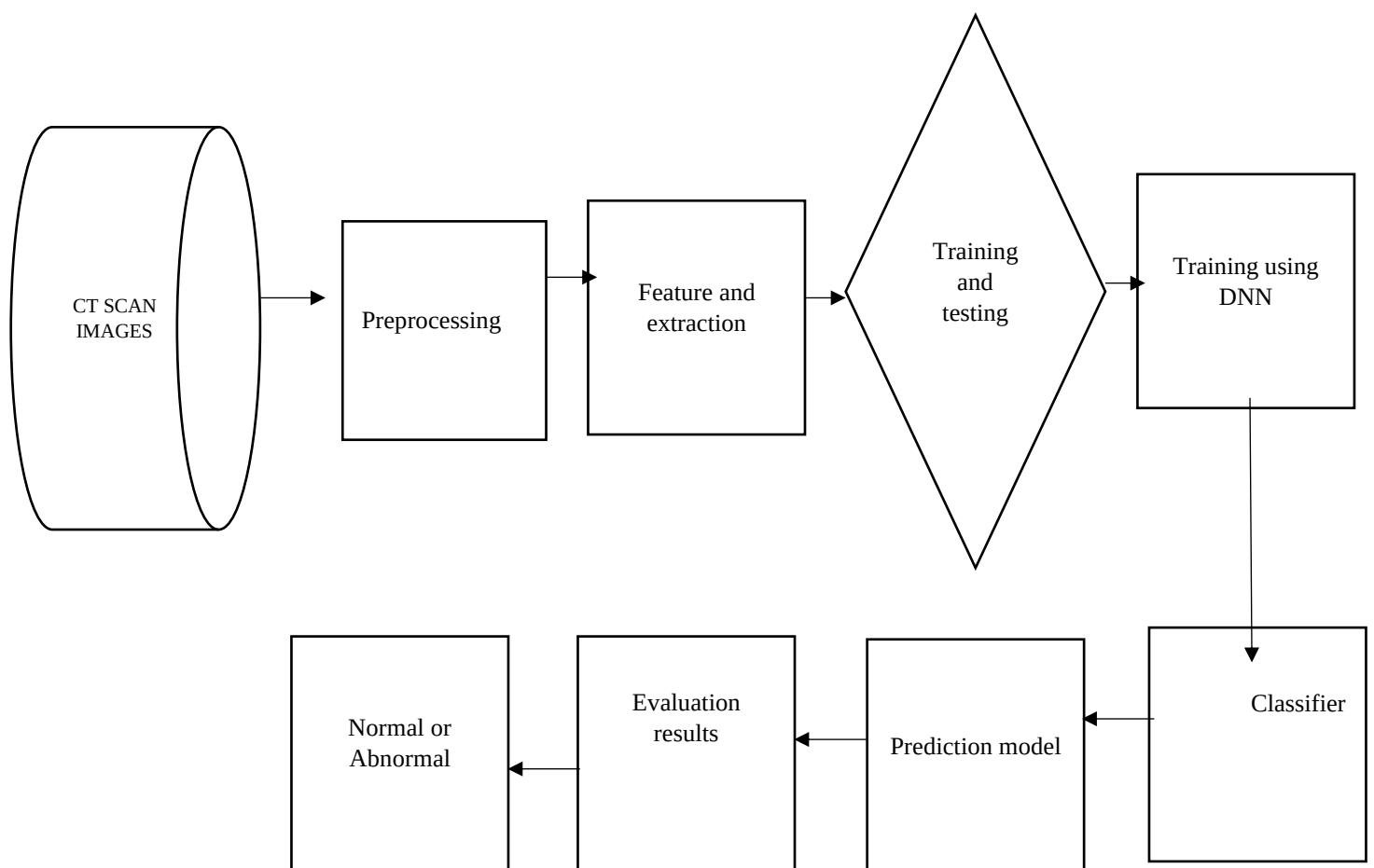
It is of interest to develop a CT scan image processed model for the early detection of lung disease were (1) preprocessing using intensity measure helps to find small particles in an image such as node, speculation and angular margin; (2) high detection and classification accuracy is established; and (3) eliminates noises that create false detection. The first step is pre-processing of the image to find particles using segmentation and feature extraction. The processed image is segmented using a standard segmentation technique. Thus, covid nodules are marked in the image. In addition to features like perimeter and eccentricity, other features like area, diameter and pixel mean intensity have been removed during feature extraction. The classification module follows this where the distinction between malignant and benign tumor's based on CT scan images is established. Extracted features are used as training features and the corresponding trained model is generated for the classification followed by model evaluation for detection and classification with improved accuracy, specificity and sensitivity.

A DNN is an extension in the number of hidden nodes in a simple neural network. The neural network can be used to carry on the more complex input calculation. The nonlinear should be used for each hidden layer, because if the activation function is linear, compared with the single hidden layer neural network, the depth of the hidden layer of the network does not rise the ability to express. The processing part of the pulmonary nodule is decomposed into the DNN, so that different network layers can be handled to obtain the characteristics of the pulmonary nodules with different sizes.

Salient features:

The relevant features of the technique includes (1) image preprocessing by using filtering techniques and it smoothest the image and removes speckle noise (2) It segments the covid-19 nodule from the CT scan image. The segmentation method separates and identifies the touching objects in the images. This feature helps in the proper segmentation of covid-19 nodules when it is touching other

false nodules. (3) Features extraction where features like centroid, diameter, eccentricity, perimeter and mean intensity are extracted from the image. These features are used as training features to develop the classifier, and (4) the classification module classifies the detected nodule as malignant or benign by using the trained sorting method.

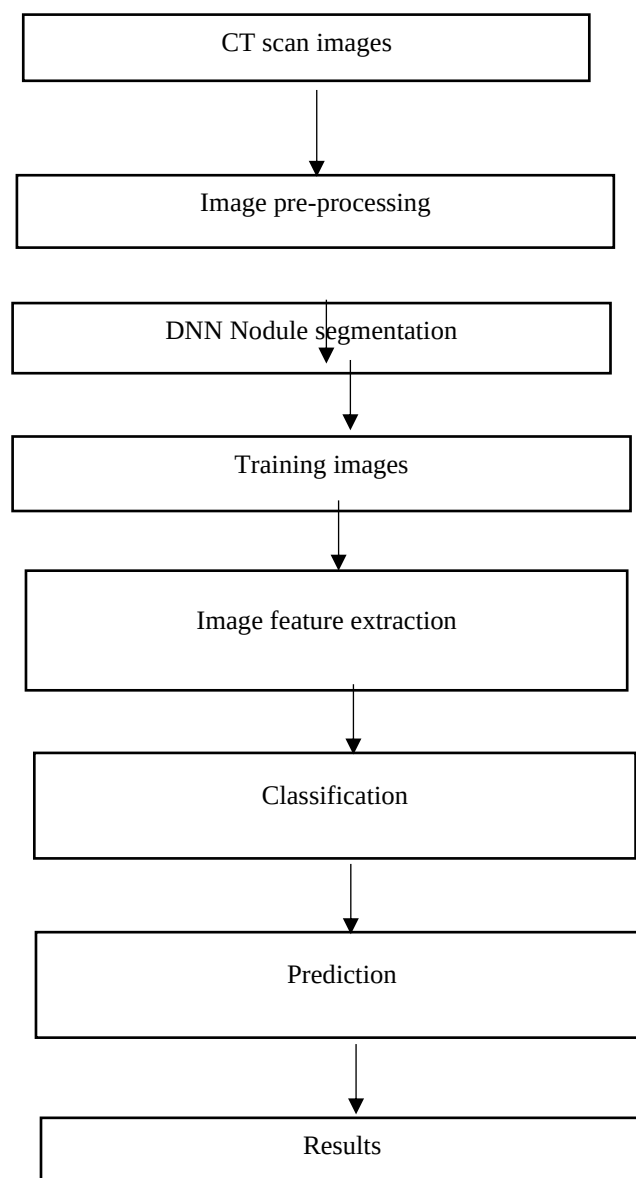


Proposed method:

Image Preprocessing Firstly, in image pre-processing median filter is used on the grayscale image of CT scan images. Some noises are embedded on CT scan Images at the time of image acquisition process which results in wrong detection of nodules. Noise Changes to the current best solution have been made and the new model has been proposed. Instead of Gabor Filter, Median filter and Gaussian filter have been implemented in the pre-processing stage. After preprocessing the processed image is segmented using Edge detection segmentation algorithm. This gives the image with covid nodules marked. In addition to features like perimeter and eccentricity, features like Centroid, Diameter and pixel Mean Intensity have been extracted in the feature extraction stage for the detected covid nodules. The top model ends after the identification of covid-19 nodule, it's feature extraction and calculation of accuracy. Therefore, the additional stage of classification of covid nodule has been performed using Support Vector Machine. Extracted features are used as training features and a trained model is generated. Then, unknown detected nodule is classified using that trained prediction mode May be detected as covid nodules sometimes. Therefore, these noises have to be pulled out for accurate detection of cancer. The median filter removes salt and pepper noise from the CT images. After the median filter, a Gaussian filter is implemented. It smoothest the image and removes speckle noise from the image. Segmentation This process locates objects or boundaries which help in acquiring the region of interest (ROI) in the image. It partitions the image into regions to identify meaningful information. In lung disease detection it segments the covid-19 nodule from the CT scan image. In the proposed model watershed segmentation is implemented. This feature

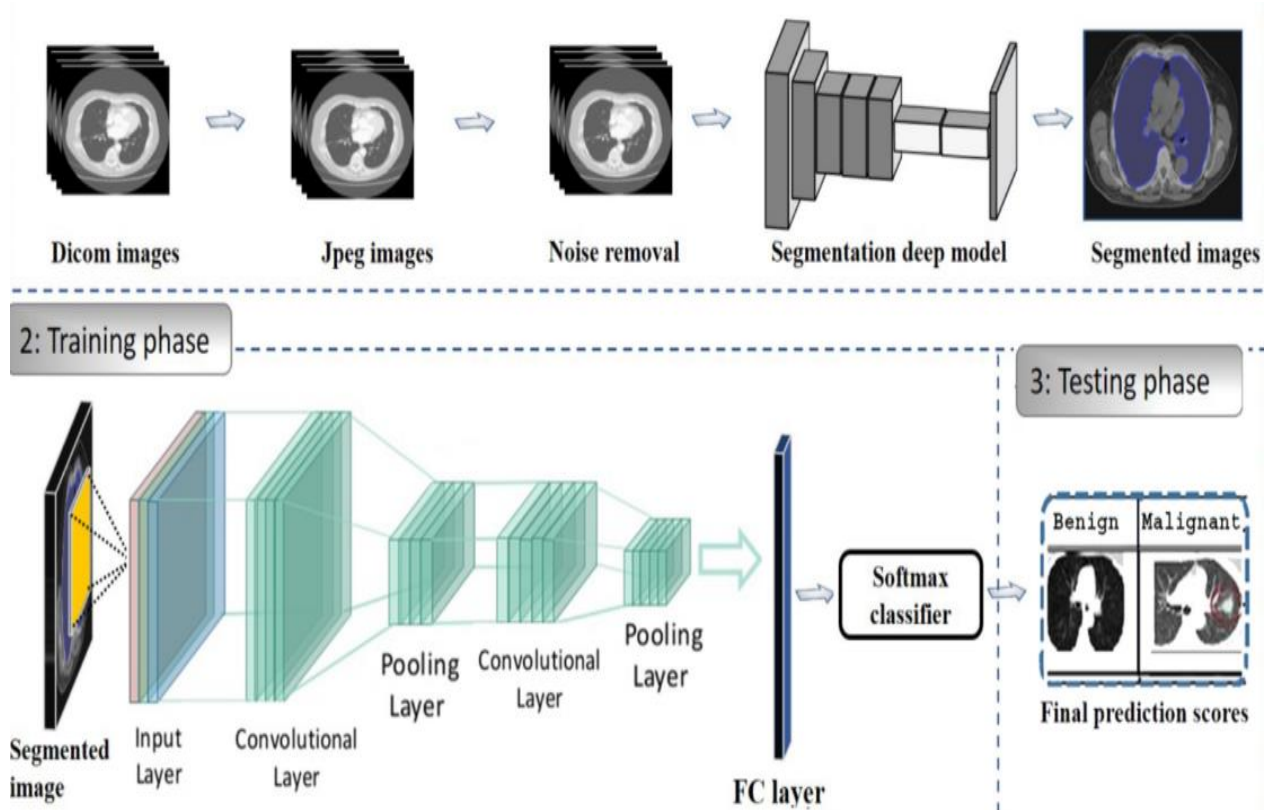
helps in the proper segmentation of covid-19 nodules if it is touching other false nodules. Feature's extraction In this stage, features like perimeter, centroid, eccentricity, diameter eccentricity and Mean intensity. These features, later on, are used as training features to develop a classifier. This stage classifies the detected nodule as malignant or benign from the beginning to the final level. Prediction result shows that classification of algorithms.

Construction of DNN Model:



PREPROCESSING:

- A DNN is an increase in the number of hidden nodes in a neural network. The neural network can be used to carry on the complex input calculation. The nonlinear will be used for each hidden layer because if the activation function is linear, compared with the single hidden layer neural network, the depth of the hidden layer of the network does not improve the ability to express.
- The processing part of the pulmonary nodule is decomposed into the DNN so that different network layers can be used to obtain the characteristics of the pulmonary nodules with dissimilar sizes. There are also local extreme problems and gradient diffusion problems in the DNN.



CHAPTER-4

4. EXPERIMENTAL METHODS AND RESULT:

4.1Deep learning model:

A. PRE-PROCESSING

The obtained data contains missing values which could lead to in- consistency. Data need to be preprocessed to gain better results to increase the algorithm's performance. The outliers must be re- moved, and variable conversion must also be carried out. The data set may contain missing values. The missing values cannot process by the programming hence these values need to convert into numerical values. Pre-processing is aimed at improving the data which suppresses unwanted distortions.

B. SPLITTING THE DATASET

The data used is split into training data set and test data set. For training the 70% data is taken while the remaining 30% data is used for testing. The training set contains a known sample and the model learns from this training dataset. The test dataset (or subset) is used to test our model's prediction on this subset and accuracy is calculated.

C. CLASSIFICATION

The training data is trained by using different machine learning algorithms. The most exciting step in designing any model of machine learning is algorithm selection. We have chosen five different types of classification algorithms.

1. K-Nearest Neighbor
2. Decision Tree
3. Random Forest Classifier
4. Logistic Regression
5. Support Vector Machine
6. Deep Neural Network
7. Cognitive Radio

1. K-Nearest Neighbor(K-NN)

The k-nearest Neighbor algorithm is a simple, supervised machine learning algorithm that can be used to resolve both classification and regression problems. It's easy to execute and understand, but has a major drawback of becoming significantly slower as the size of that data in use grows.

2. Decision Tree

This algorithm works like a tree structure, data is split into smaller and smaller continuously according to certain parameter. In decision analysis a decision tree can be used to represent decisions and Decision-making visually and clearly. A decision tree in data mining represents data but not decisions; instead, the resulting classification tree may be a guide for decision making.

3. Random Forest Classifier

Random Forest is one of the most well-known and powerful machine-learning algorithms. It creates many binary decision trees out of forest. Growing patient moves through each tree before it reaches the node on the root. Plenty of data measurements are taken here, the mean is determined, after all the mean values are averaged to give the real mean value a better approximation.

4. Logistic Regression

Logistic regression is used to describe the relationship among the binary variables (Dichotomous) when the variables depend on each other. And gives result as positive or negative, true or false. It is similar to linear regression, whereas linear regression works on continuous variable. Sigmoid function takes the maximum and minimum of dichotomous values and gives the results.

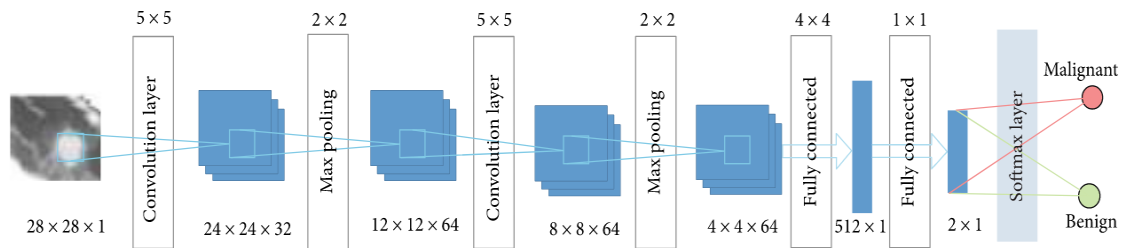
5. Support Vector Machine

Support vector machines are maximum-margin classifiers, which means they find the hyper plane that has the largest perpendicular

6. Deep Neural Network:

A DNN is an rise in the number of hidden nodes in a simple neural network. The neural network can be used to carry on the more complex input calculation. The nonlinear should be used for each hidden layer, because if the activation function is linear, compared with the single hidden layer neural network the depth of the hidden layer of the network does not increase the ability to express. • The processing part of the pulmonary nodule is decomposed into the DNN, so the different network layers can be used to obtain the characteristics of the pulmonary nodules with different sizes. There are also local extremum problems and gradient diffusion problems in the DNN.

6.1 The architecture of the DNN.



6.2 CONSTRUCTION OF THE DNN

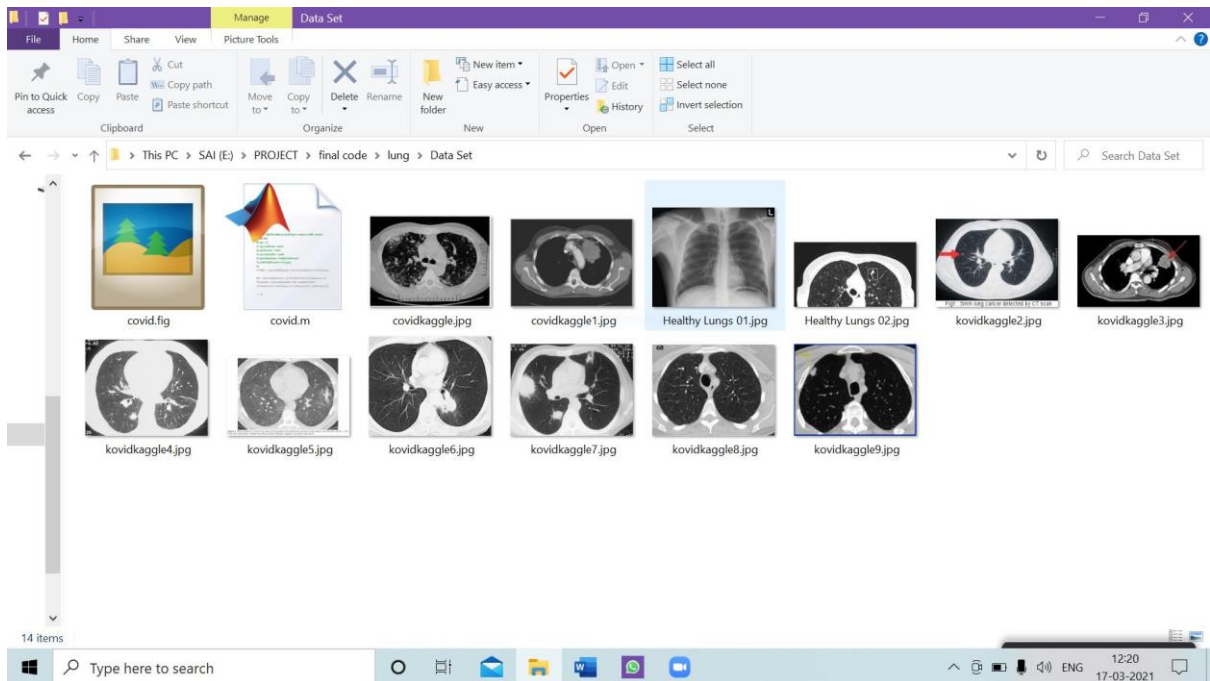
- The DNN consists of a fully connected layer. The input image is a two-dimensional data input 28×28 neural network mapped into 784×1 . The second layer is a fully connected layer of 512×1 .
- The third layer is a fully connected layer of 256×1 . After the third layer, there will be a dropout layer, with a parameter of 0.6, in which the unit will be hidden in 40%. The fourth layer is a fully connected layer of 64×1 , whose activation function is set to ReLU.

7.Cognitive Radio :

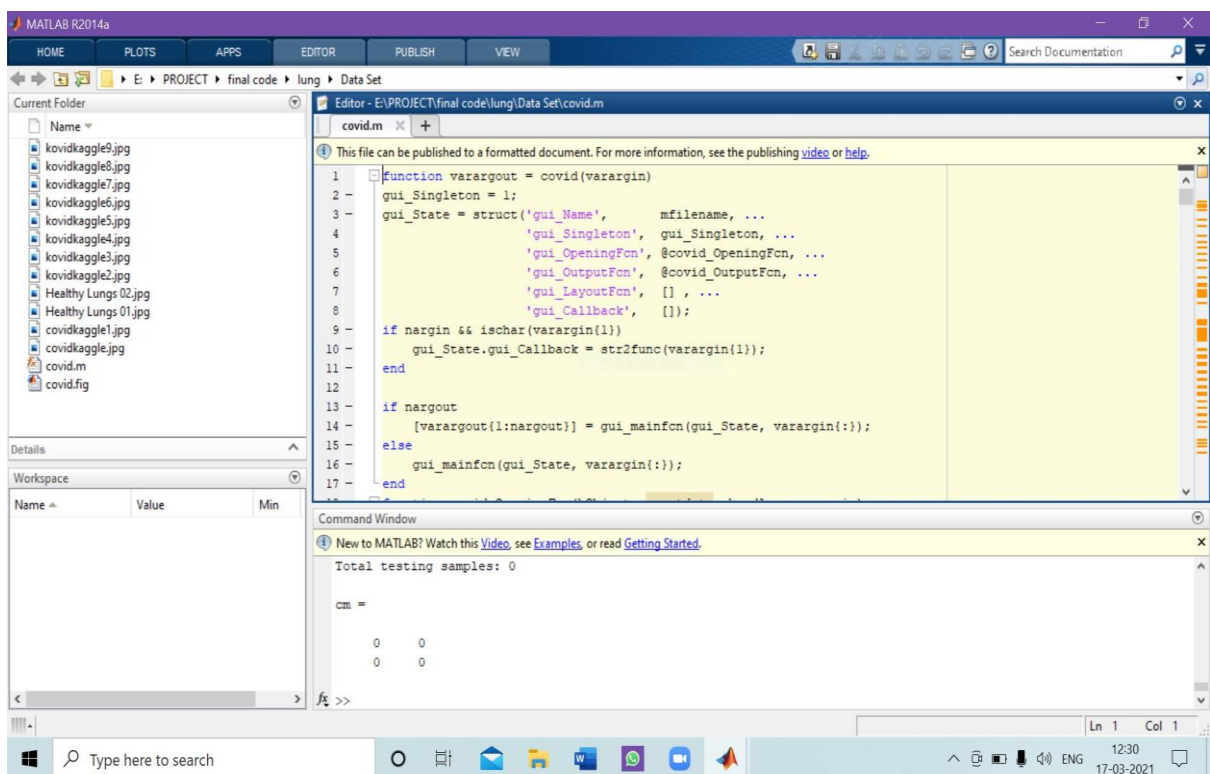
Cognitive radio provides a superior way for utilization of spectrum resource by introducing an opportunistic usage of the frequency bands that are not heavily occupied by a licensed spectrum user or a primary user (PU). In cognitive radio, the detection and estimation of PU channel availability (unoccupied spectrum) are the main challenges that need to be overcome. In order to prevent the interference with licensed spectrum user and improve spectrum resource utilization efficiency. This chapter focuses on developing new ways for detecting and estimating primary user channel availability based on machine-learning techniques.

Classification:

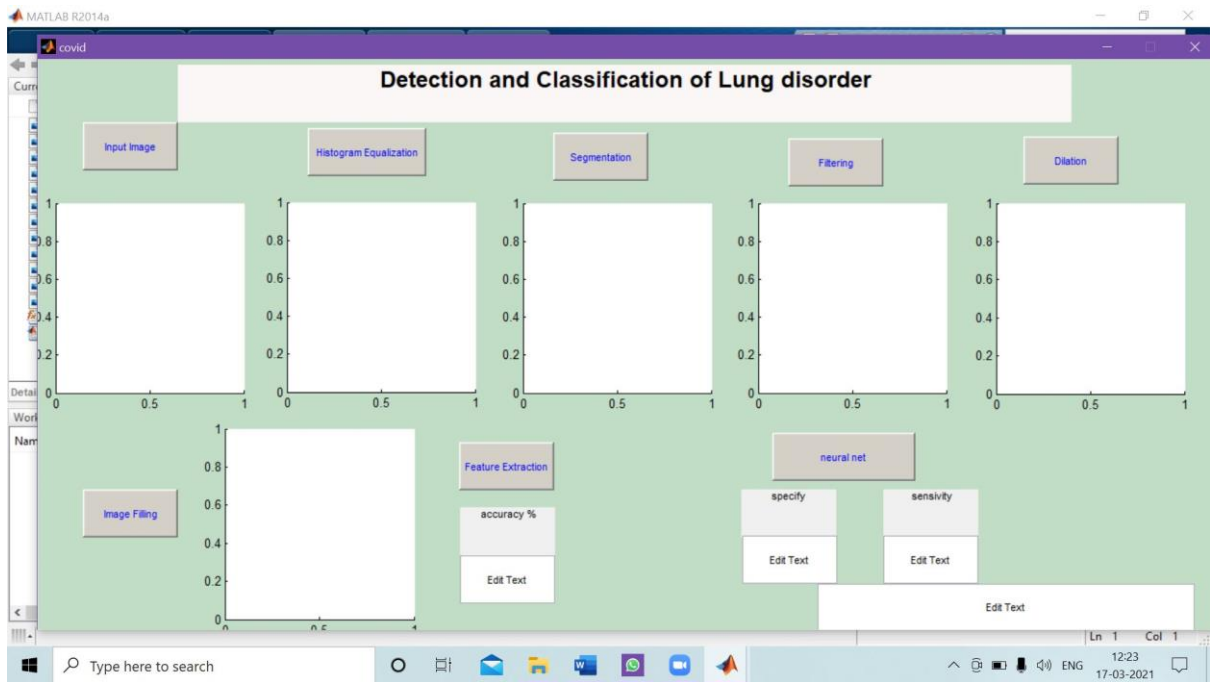
To analyze data and recognize patterns for classification purpose, supervised learning models with related learning algorithms are used also known as support vector machine. It initializes by taking a set of input data and predicts for each given input the specified class among the two choices available, which makes it a non- probabilistic binary linear classifier. For mapping, the given data into a dissimilar space SVM uses kernel function. The simulated steps involve image enhancement, post-processing, feature extraction, image segmentation, feature selection and classification using (support vector machine) SVM kernel. The involvement provided. the development of WOA_SVM to automate the aided diagnosis system for determining whether the lung CT scan image is normal or abnormal. An enhanced technique is developed using a whale optimization algorithm for best feature selection to obtain accurate results and constructing a robust model. The performance of the proposed methodology is evaluated using accuracy, sensitivity and specificity and obtained as 95%, 99% and 92% using racial bias function support vector kernel.



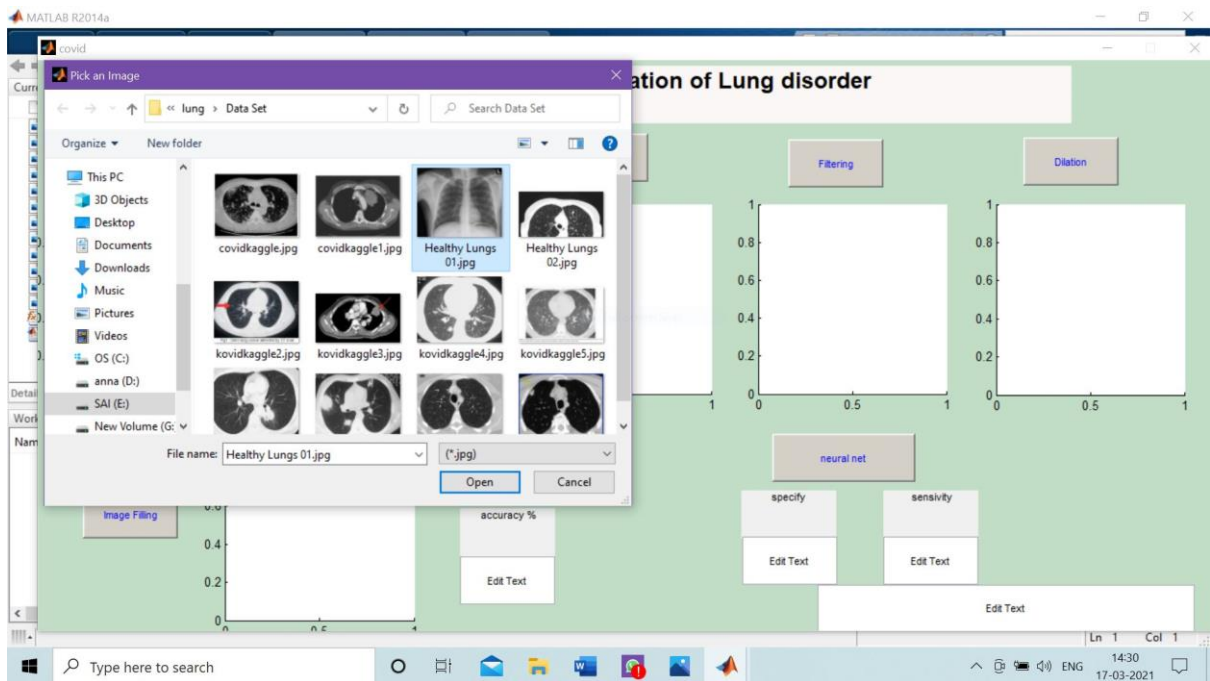
4.2.1 Fig : lung CT scan images



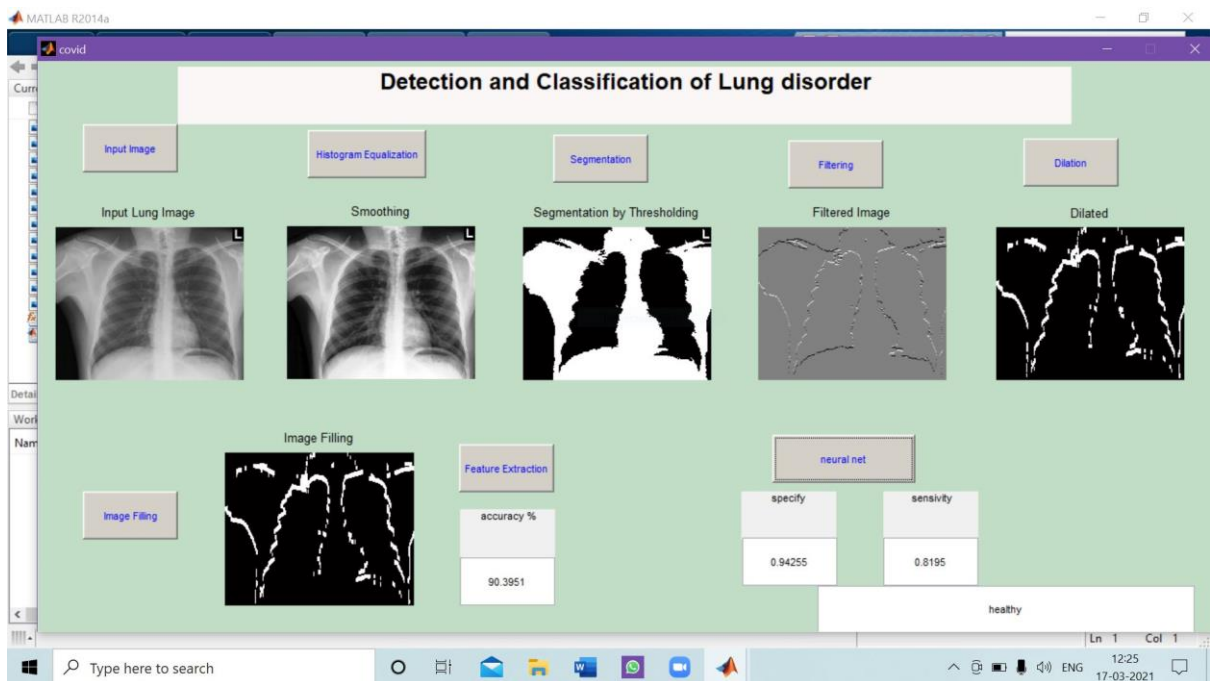
4.2.2 Fig : MATLAB software with code



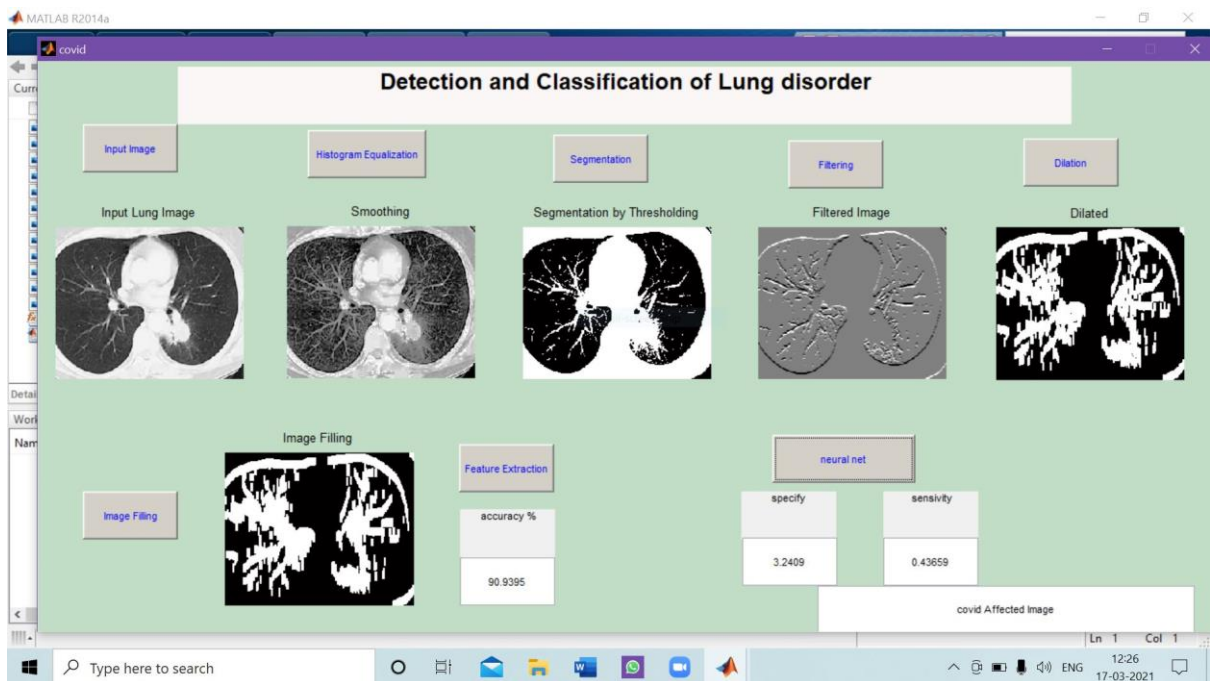
4.2.3 Fig : After running the code out put image .



4.2.4 Fig : giving CT scan image as a input



4.2.5 Fig : detecting that given CT scan image is healthy.



4.2.6 Fig : detecting the given Ct scan image is affected.

Strengths:

Strengths of the method includes:

- (1) The improved accuracy of covid nodule detection;
- (2) classifies the detected lung disease as malignant or benign;
- (3) Removes the noises that create false detection of covid-19;

parameter	LR.	RF	KNN	CR	SVM	DNN
Sensitivity	0.69	0.83	0.87	0.99	1	1
Specificity	0.72	0.85	0.9	.87	1	1
Accuracy (%)	73.2%	89%	92%	88.3%	92.81%	96.5%

4.3.1 Fig: Performance analysis for each algorithm

Performance for each algorithm is analyzed. So the algorithm with the highest precision is selected for the prediction process. In our analysis decision tree and Random forest shows the highest accuracy. We have Chosen Random forest for our Prediction process.

CHAPTER :5

PERFORMANCE ANALYSIS:

There are many ways to diagnose lung diseases, such as Chest Radiograph (x-ray), Computed Tomography (CT), Magnetic Resonance Images (MRI Scan) and Sputum Cytology. First of all, multiple steps are involved in segmentation like; Image pre-processing, histogram analysis, threshold, morphological filling whether the image is normal or abnormal. CNNs were explored to extract representative and discriminative features from X-ray images for classification in different body parts. The essential capability of CNN to capture the structure of the image in its feature charts. Further improvements are available in the areas where the abnormalities lie, probably through the use of a deep learning algorithm, and the versatility of an automated detection system.

CONCLUSION :

In summary, we compared, reproduced, nodule segmentation and classification modules. Results showed that many transfer learning approaches achieved reasonable accuracy in diagnosing chest CT scan images. Future works on additional data collections and validation will further improve the generalizability of the current methods. Lung disease is an extremely complex problem to solve however with early detection a patient has a high increase of survivability. The diagnostics data suggests that the highest frequency of people who get diagnosed have already an advanced form of covid and the second most frequent cohort are have been accidentally diagnosed and have covid that is hardest to determine and is easiest to cure. The research leads the author to believe that deep learning could be a powerful tool in diagnosing very small and very hard to determine nodules to aid the medical decision-making process.

REFERENCES:

- [1] Deep Convolution Pixel-wise Labeling for skin lesion Image Segmentation, AL Youssef; Domenico D. Bloisi ;Mario Muscio ; Andrea Pennisi ; Daniele Nardi ; Antonio Facchiano IEEE International Symposium on Medical Measurements and Applications (MeMeA) Year: 2018 Patchoo, W. 2018 International Workshop on Advanced Image Technology.

- [2] Early Detection of Melanoma Skin Cancer Using Classifiers, VS. Sabeera, P. VamsiKrishna PG Scholar, Dept. of ECE, R.K. College of Engineering, Vijayawada, Andhra Pradesh Assistant Professor, Dept. of ECE, R.K. College of Engineering, Vijayawada, Andhra Pradesh, Year-2016

- [3] Vembandasamy, K., Sasipriya, R. and Deepa, E. (2015) Heart Diseases Detection Using Naive Bayes Algorithm. IJISSET- International Journal of Innovative Science, Engineering & Technology.

- [4] Skin Cancer Detection and classification Pratik Dubal, Sankirtan Bhatt, Chaitanya Joglekar, Dr. Sonali Patil, Department of Information Technology K. J. Somaiya College of Engineering Vidyavihar, Year-2017.

- [5] Iyer, J. S and R. Sumbaly, "Diagnosis of Diabetes Using Classification Mining Techniques", IJDKP, vol.5, no. 1, pp. 01-14, 2015.

- [6] B. Ashok and P. Aruna, "Comparison of Feature selection

methods for diagnosis of cervical cancer using support vector machine(SVM) classifier,” vol. 6, no. 1, pp. 94–99, 2016.

[7] X. Pan, Y. Luo, Y. Xu, “K-nearest neighbour based structural twin support vector machine,” Knowledge Based Systems, vol. 88, pp. 34-44, 2015.

[8] Vijayarani, S. and Dhayanand, S. (2015) Liver Disease Prediction using SVM and Naïve Bayes Algorithms. International Journal of Science, Engineering and Technology Research (IJSETR), 4, 816-820.

[9] V.Krishnaiah “Diagnosis of Lung Cancer Prediction System Using Data Mining Classification Techniques” International Journal of Computer Science and Information Technologies, Vol. 4 (1) 2013, 39 – 45 www.ijcsit.Com ISSN: 0975-9646.

[10] Key Statistics for Lung Diseases, <https://www.Diseases.org/Diseases/non-small-cell-lung-Diseases/about/key-statistics.html>.

[11] A.C. Society, Diseases Facts and Figures, 2015, <http://www.Diseases.org/acs/groups/content/@editorial/document/s/document/acspc-044552.pdf>.

[12] I. R. Valente, P. C. Cortez, E. C. Neto, J. M. Soares, V. H. de Albuquerque, and J. M. Tavares, “Automatic 3D pulmonary nodule detection in CT images: a survey,” Programs in Biomedicine and Computer Methods, vol. 124, no. 1, pp. 91–107, 2016.View at: [Google Scholar](https://scholar.google.com/)

[13]

A. El-Baz, G. M. Beache, G. Gimel'farb et al., "Computer-aided diagnosis systems for lung Diseases: methodologies and challenges review article," International Journal of Biomedical Imaging, vol. 2013, Article ID 942353, 46 pages, 2013.View at: [Google Scholar](#)

[14] H. Chen and W. WuH. Xia, J. Du, M. Yang, and B. Ma, "Classification of pulmonary nodules using neural network ensemble," Advances in Neural Networks, Springer, Guilin, China, 2011.View at: [Publisher Site](#) | [Google Scholar](#)

[15] J. Kuruvilla and K. Gunavathi, "Lung Diseases classification using neural networks for CT images," Computer Methods and Programs in Biomedicine, vol. 113, no. 1, pp. 202–209, 2014.View at: [Publisher Site](#) | [Google Scholar](#)

[16] Roy, T., Sirohi, N., & Patle, A. (2015) "Classification of lung image and nodule detection using fuzzy inference system." International Conference On Computing, Communication & Automation. DOI: 10.1109/CCAA.2015.7148560.

[17] Ignatious, S., & Joseph, R. (2015) "Computer aided lung disease detection system." 2015 Global Conference On Communication Technologies (GCCT), DOI: 10.1109/GCCT.2015.7342723.

- [18] Rendon-Gonzalez, E., & Ponomaryov, V. (2016) "Automatic Lung nodule segmentation and classification in CT images based on SVM." 2016 9Th International Kharkiv Symposium On Physics And Engineering Of Microwaves, Millimeter And Submillimeter Waves (MSMW). DOI: 10.1109/MSMW.2016.7537995.

- [19] Miah, M.B.A., & Yousuf, M.A. (2015) "Detection of lung disease from CT image using image processing and neural network." 2015 International Conference on Electrical Engineering and Information Communication Technology (ICEEICT): 1-6.

- [20] Khobragade, S., Tiwari, A., Patil, C., & Narke, V. (2016) "Automatic detection of major lung diseases using Chest Radiographs and classification by feed-forward artificial neural network." IEEE International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES): 1-5.

- [21] Armato, I., Samuel McLennan, G., McNitt-Gray, F. R., Michael, Charles, Reeves, Anthony P., ... Clarke, Laurenc, (2015) "Data From LIDC-IDRI. TheCovidImaging"
 Archive.<http://doi.org/10.7937/K9/TCIA.2015.LO9QL9>

ranjithkumarchowdary3@gmail.com.docx

ORIGINALITY REPORT

2%	2%	1%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Universiti Malaysia Pahang Student Paper	1%
2	alldokument.com Internet Source	<1%
3	Submitted to Far Eastern University Student Paper	<1%
4	open.unido.org Internet Source	<1%
5	Submitted to University of Teesside Student Paper	<1%
6	Nancy C. Dolan, Vanessa Ramirez-Zohfeld, Alfred W. Rademaker, M. Rosario Ferreira et al. "The Effectiveness of a Physician-Only and Physician-Patient Intervention on Colorectal Cancer Screening Discussions Between Providers and African American and Latino Patients", Journal of General Internal Medicine, 2015 Publication	<1%

journals.plos.org

7	Internet Source	<1%
---	-----------------	-----

35