IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

EARLY DETECTION OF LUNG CANCER USING ADA BOOST ALGORITHM

Dr. K. Merrilaiance and S. Archana

Department of Computer Applications, Sarah Tucker College, Thirunelveli-7.

ABSTRACT

Lung cancer is considered to be the one among the most dreaded disease which will be the main reason for the death of individuals and having greater deterioration of death if it is not identified at primitive stage. Because of the fact that Lung cancer could be identified only after spreading to the parts of lungs to a greater extent and it is very tough to predict the presence of lung cancer at the earlier stage. Moreover, it involves greater error in the diagnosing the presence of Lung cancer by Radiologists and Expert Doctors. Therefor it is compulsory to design an intelligent and automated system for accurately predicting the cancer and stage at which the stage of cancer or enhancing the accuracy of prediction for detecting the cancer at earlier which will be much helpful in deciding the treatment type and depth of the treatment based on the extent of disease. Here this work establishes the idea of preprocessing which used correlation ranking algorithm and Adaboost Algorithm in the classification of the lung cancer and its stages and the predicting the possibility of recurrence.

1. INTRODUCTION

Lung cancer has been one of the deadliest diseases in today's decades. It has become one of the causes of death in both man and woman. There are various reasons for which lung cancer occurs but classification of tumor and predicting it in the right stage is the most important part. This paper focused on the numerous approaches has been derived for lung cancer detection from different literature survey to advance the ability of detection of cancer.

Lung cancer [6] is one of the cancer types that leads to death. The Lung diseases are the disorders which affect the lungs and it is unpredictable medical conditions worldwide especially in India. Lung cancer constitutes 12.8% of all cancer types worldwide, also it constitutes 17.8% of the cancer deaths and it increases by 0.5% every year globally worldwide. The cause of cancer in males represents 38.6%, and in females it represents 5.2%. However, it is suggested that 15% of lung cancer patients live 5 years or more after the diagnosis, together with this early diagnosis and used drugs when they are diagnosed early [1,2]. It makes it possible at computer aided diagnosis of lung cancer by evaluating these parameters.

Lung cancer is a disease which arises due to growth of unwanted tissues in the lung and this growth which spreads beyond the lung are named as metastasis which spreads into other parts of the body. But in some case of cancers the initial growth in lung are named as carcinomas that derived from epithelial cells. If this uncontrolled growth can be detected successfully at early stages, which helps to diagnosis the risk of invasive surgery and increased survival rate. Lung disease which affects the parts of lungs and cause infections such as tuberculosis, influenza, pneumonia and other breathing problems such as asthma, Chronic Obstructive Pulmonary (COPD) disease.

The two major types of lung cancer [2] are Non-Small Cell Lung Cancer (NSCLC) and Small Cell Lung Cancer (SCLC) or oat cell cancer that grows and spreads in various ways which is to be treated differently. If the patients have symptoms of both cancer types, then it is called as mixed small cell/large cell cancer. Non-Small Cell Lung Adenocarcinoma (NSCLA) is more common than SCLC and it commonly grows and spreads slower than SCLC. SCLC is linked with smoking features and grows faster by forming a large tumor that can spread throughout the body. Computer Aided Diagnosis (CAD) system is very helpful for doctors in detection and diagnosing abnormalities earlier and faster [3] and it is a second opinion for doctors before suggesting biopsy test. Several techniques are available for lung cancer diagnosis but those techniques are more expensive, time consuming and having less capability for detecting the lung cancer. Hence, a new prediction method is essential to predict the lung cancer in its early stages.

The major cause of deaths in human beings is Lung Cancer, Since the lung cancer symptoms appear in the advanced stages so it is hard to detect which leads to high mortality rate among other cancer types. Hence the early prediction of lung cancer is mandatory for the diagnosis process and it gives the higher chances for successful treatment. It is the most challenging way to enhance a patient's chance for survival. In this paper a computer aided classification method for lung cancer prediction based an evolutionary system by a combination of architectural evolution with weight learning using neural network.

ii. PROBLEM DEFINITION

There are numerous systems available to analyze lung cancer such as Chest Radiography (x-beam), Computed Tomography (CT), Magnetic Resonance Imaging (MRI output) and Sputum Cytology. Yet, the greater part of these systems is costly and tedious. The majority of these techniques to identifying the lung cancer in its propelled stages, results in low patients' survival. Many researches have been proposed number of techniques for processing of images and automated classification system giving varying results. But still there is need of more classification accuracy, recognition rate and minimum classification error rate. Hence, the essential effective method for diagnosis of lung cancer is required.

2.LITERATURE REVIEW

[1] Divya Chauhan, Varun Jaiswal, " Development of Computational Tool for Lung Cancer Prediction Using Data Mining", 2016

The requirement for computerization of detection of lung cancer disease arises ever since recent-techniques which involve manual-examination of the blood smear as the first step toward diagnosis. This is quite time-consuming, and their accurateness depends upon the ability of operator's. So, prevention of lung cancer is very essential. This paper has surveyed various techniques used by previous authors like ANN (Artificial Neural Network), image processing, LDA (Linear Dependent Analysis), SOM (Self Organizing Map) etc

[2] S. Durga, K. Kasturi, " Lung disease prediction system using data mining techniques", January 2017

Lung cancer is one of the most dangerous diseases in the world. The early detection of lung cancer can cure the disease completely. Data mining plays an effective role by using Naïve Bayes and Artificial Neural Network to massive volume of healthcare of data. The health care industry collects huge amounts of data which unfortunately are not mined to find the hidden data. The Naïve Bayes aims at delivering robust classifications also when dealing with small or incomplete data sets. The aim of the paper is to detect and diagnose the lung diesases as early as possible which will help the doctor to save the patient’s life. This paper describes how lung cancer was predicted and controlled, using data mining techniques. © 2017, Institute of Advanced Scientific Research, Inc. All rights reserved.

[3] Mutiullah Jamil, Mehwish Bari, Adeel Ahmed, " Lung Cancer Detection Using Digital Image Processing Techniques: A Review", April 2019

CV (Computer Vision) play vital role to prevent lung cancer. Since image processing is necessary for computer vision, further in medical image processing there are many technical steps which are necessary to improve the performance of medical diagnostic machines. Without such steps programmer is unable to achieve accuracy given by another author using specific algorithm or technique. In this paper we highlight such steps which are used by many author in pre-processing, segmentation and classification methods of lung cancer area detection. If pre-processing and segmentation process have some ambiguity than ultimately it effects on classification process. We discuss such factors briefly so that new researchers can easily understand the situation to work further in which direction.

[4] Raviprakash S. Shriwas, Akshay D. Dikondawar, " LUNG CANCER DETECTION AND PREDICTION BY USING NEURAL NETWORK", January 2015

The detection of lung cancer in early stage is a challenging problem, due to the structure of the cancer cells, where utmost of the cells are overlapped with each other. It is a computational procedure that sort images into groups according to their similarities. In this Histogram Equalization is used for preprocessing of the images and feature extraction process and neural network classifier to check the condition of a patient in its early stage whether it is normal or abnormal. The performance is based on the correct and incorrect classification of the classifier.

[5] N Numan, S Abuelenin, " PREDICTION OF LUNG CANCER USING ARTIFICIAL NEURAL NETWORK", April 2018

.The proposed model is examined with the conventional methods. Detail error analysis is presented and several data sets are analyzed and compared. In most of the comparisons carried out the proposed model proved to be superior to conventional survival rate prediction methods. However, the generalization of the obtained results cannot be claimed for any problem as the effectiveness of the proposed model has to be examined for more diverse problems before making the conclusion that the proposed model is indeed superior for any problem.

[6] Saravanan K, Sasithra S, " Review on Classification Based on Artificial Neural Networks", December 2014

The back propagation neural network (BPNN) can be used as a highly successful tool for dataset classification with suitable combination of training, learning and transfer functions. When the maximum likelihood method was compared with backpropagation neural network method, the BPNN was more accurate than maximum likelihood method. A high predictive ability with stable and well functioning BPNN is possible. Multilayer feed-forward neural network algorithm is also used for classification. However BPNN proves to be more effective than other classification algorithms.

[7] N. Mohanapriya, B. Kalaavathi , T. senthil Kuamr, " Lung Tumor Classification and Detection from CT Scan Images using Deep Convolutional Neural Networks (DCNN)", Dec. 2019

A classifier based on Deep Convolutional Neural Networks (DCNN), which classifies the lung tumor composed of different fully connected pooling and Convolutional layers. Three architectures were defined for DCNN classifier each one is trained with different patch size. DCNN is applied to the CT image for classification of benign and malignant lung tumor. The proposed architectures were examined on the LIDC database and cross checked with other classifiers result such as Artificial Neural Network Simulation result presents DCNN classifier achieves better performance.

[8] S. Vijayalakshmi, J. Priyadarshini, " Breast cancer classification using RBF and BPN neural networks", January 2017

This paper explores the possible diagnosis of breast cancer using Radial Basis Function (RBF) for the data set. The use of machine learning and data mining techniques has revolutionized the whole process of breast cancer Diagnosis and Prognosis. Breast Cancer Diagnosis distinguishes benign from malignant breast lumps and Breast Cancer Prognosis predicts when Breast Cancer is likely to recur in patients that have had their cancers excised. We analyze the breast Cancer data available from the Wisconsin Breast Cancer WBC, WDBC from UCI machine learning with the aim of developing accurate prediction models for breast cancer using RBF. Overall, the RBF neural network technique has proved better performance than that of the BPN technique.

[9] J.Jamera banu, " Study of Classification Algorithm for Lung Cancer Prediction", February 2016

Lung cancer remains the leading cause of cancerrelated mortality for both men and women and its incidence is increasing worldwide. Lung cancer is the uncontrolled growth of abnormal cells that start off in one or both Lung. The earlier detection of cancer is not easier process but if it is detected, it is curable. We analyzed the lung cancer prediction using classification algorithm such as Naive Bayes, Bayesian network and J48 algorithm. Initially 100 cancer and noncancer patients' data were collected, preprocessed and analyzed using a classification algorithm for predicting lung cancer.The dataset have 100 instances and 25 attributes. The main aim of this paper is to provide the earlier warning to the users and the performance analysis of the classification algorithms

[10] A. Titus, Khanna Harichandran Nehemiah, " Classification of interstitial lung diseases using particle swarm optimized support vector machine", January 2015

The standard deviations of the two Gaussians are estimated using the Expectation-Maximization (EM) algorithm. The feature vectors constructed from the texture features extracted using the GLH and the QWT are applied to the Support Vector Machine (SVM) classifier. The S VM classifier is optimized using particle swarm optimization and is used to classify the different lung tissue patterns. The classifier achieved an overall precision of 90.23%, accuracy of 96.01% andmisclassification rate of 3.99%.

3. METHODOLOGY

This project has following modules

- **Data Pre-processing**
  - **Normalization**
- **Training and Testing samples**
  - **Cross validation**
- **Neural Network Classifier**
  - **ADABOOST classifier**
- **Performance Evaluation**
  - **Performance metrics**

Data Pre-processing

The initial step for lung cancer detection is the preprocessing step to fill the missing data and to eliminate the unnecessary information from the dataset. The missing data are imputed using K nearest neighbor method with three neighbors to make full dataset more reliable.

Training and Testing samples

The input data samples are trained and tested by using neural network. Initially the weights of neural network of the input data are chosen randomly. The neural networks are trained with a sample data for learning and to perform classification process then with testing dataset. The classification result of the tested data is weighed to check the frequency error or the error rate which occur during classification process and the error are resolved by changing the weights in the dataset

Neural Network Classifier

The input sample is classified as cancerous or non-cancerous depend on the extracted features which is accomplished by neural network classifier [19]. Neural network is an interconnected network of neurons that transmits the electrical signals patterns. An Artificial Neural Network (ANN) is a cluster of learning neurons based on biological neural networks (human brain). Generally, a neural network consists about 100 billion neurons and each neuron connected up to 10000 other neurons.

Artificial neural networks are generally presented as systems "neurons" which are all interconnected and which exchange messages between different neurons. The connections have numeric weights that can be adapt based on experience, creating neural nets adaptive to inputs and capable of learning. The advantage of ANNs is that they are often suitable to solve problems that are too complex to be solved by the conventional techniques, or hard to find algorithmic solutions. For the above general model of artificial neural network, the net input can be calculated as follows:

Y = X1 W1 + X2.W2 +⋯……..Xm Wm ----(1)

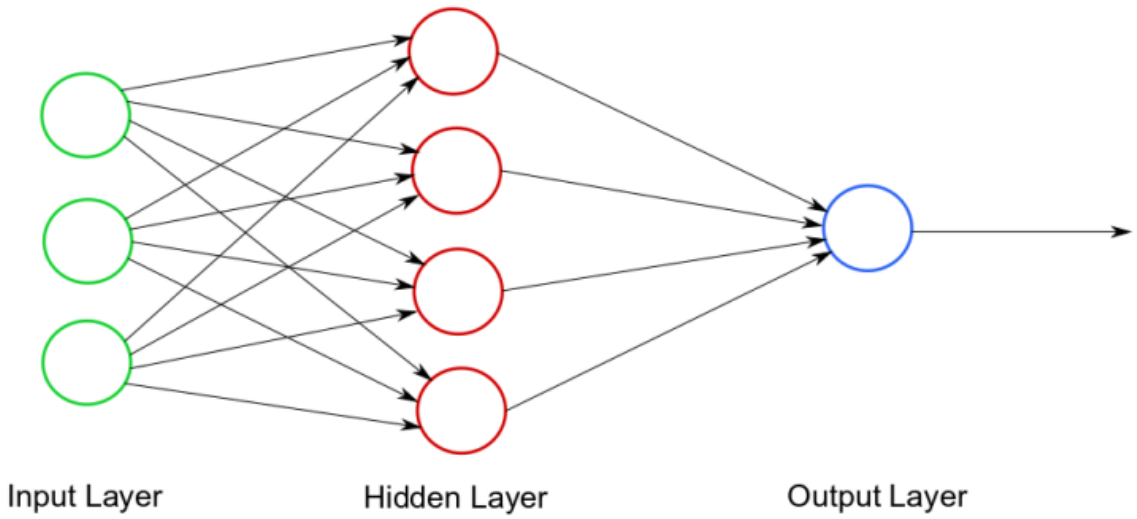
Net input Yin =∑Xi.Wi. Mi. ----(2)

The output can be calculated by applying the activation function over the net input. Y = f(Yin)

Output = function (net input calculated) ----(3)

There is one layer for the input variables and another layer for the output. The layers include:

- Input Layer - Input layer includes input units indicates the unrefined information provided for the network.
- Hidden Layer - This layer includes the hidden units based on the input unit's behavior and the weighted neuron are denoted as which connect these input with the hidden units.
- Output Layer - This layer based on the specificity of the hidden units and the weighted neuron.
- Steps for Neural Network
- Input- Lung Cancer Dataset
- Output- trained Neural Network
- Step 1- Receive an input
- Step 2 -Weight the Input (Each input sent to the network should be weighted)
- Step 3 - Sum all the weighted inputs
- Step 4 - Generate the Output



- The output is analyzed based on a sigmoid function or some suitable function by each neuron. The main advantage of neural networks is their ability to learn from patterns [13]. The two key components of neural network structure are neurons and weighted direct relations, which connect one layer of neurons with another layer of neurons. In the training phase, certain weights of the layers connections are adjusted. ANN models can be trained for these features from sample data and this information can be used to predict or categorize data in a dataset.
- The performance evaluation of proposed Neural Networks are simulated using MATLAB under windows environment. The implementation of this framework is performed on lung cancer dataset obtained from the UCI machine learning repository site [8]. The lung dataset given as input to the neural network and the data is divided into training data and test data. The training set for the neural network consists of 70% of the total dataset and the testing set is 30% of the total data. The proposed method is effectively compared with Back Propagation algorithm, Multi layer perceptron and stochastic gradient descent algorithm in terms of performance metrics obtained from confusion matrix shown in table 1.

Table 1 Confustion Matrix

		Predicted Class	
	Total Population	Prediction Positive	Prediction Negative
Actual Class	Condition Positive	True Positive (TP)	False Negative (FN)
	Condition Negative	False Positive (FP)	True Negative (TN)

Prediction of lung cancer is most challenging problem in the medical field due to structure of cancer cell, where most of the cells are overlapped each other. There are over 100 different types of cancer and one of them is lung cancer. In lung cancer treatment delay results in high mortality rate. Detection of cancer in earlier stage is curable. In this work lung cancer prediction based neural network is implemented. ANN has many advantages such as long training time, high computational cost, and adjustment of weight. The main aim of this system is to provide the earlier warning to the users and it is also cost and time saving benefit to the user. The performance evaluation of proposed method shows effective results and it indicates that neural network can be effectively used for lung cancer diagnosis to help oncologists. The prediction could help doctor to plan for a better medication and provide the patient with early diagnosis.

The potential of our proposed method has been determined by the performance metrics like sensitivity (SN), specificity (SP), and accuracy as shown in Table 1. For any binary classifier, the output can be termed either as positive or negative. Both outputs again can be either true or false, which gives four different possibilities. If the output of the classifier is positive and the actual value is also positive, it is called as true positive (TP), and if the actual value is negative, this output is termed as false positive (FP). If the output of the classifier is negative and the actual value is also negative, it is called as true negative (TN), and if the actual value is positive, this output is termed as false negative (FN). SN is the ability of an algorithm to detect a pixel as a point of interest. It is the ratio of TP and conditional positive values. SP is the ability of an algorithm to detect a pixel as a point of the background pixel. It is the ratio of TN and conditional negative values.

The classification task is to generalize well on unseen/independent data. A classifier is learned on training/learning data and then tested on data that has not been used for learning (unseen test data). There exist many measures to assess the performance of a classifier and a lot of techniques to create training and test data to estimate the generalization ability of a classifier on test (unseen) data.

Performance Evaluation

This is a measurement tool to calculate the performance

$$\text{Accuracy} = \left[\frac{TP+TN}{TP+TN+FP+FN}\right]$$

$$\text{Sensitivity} = \left[\frac{TP}{TP+FN}\right]$$

$$\text{Specificity} = \left[\frac{TN}{TN+FP}\right]$$

Where

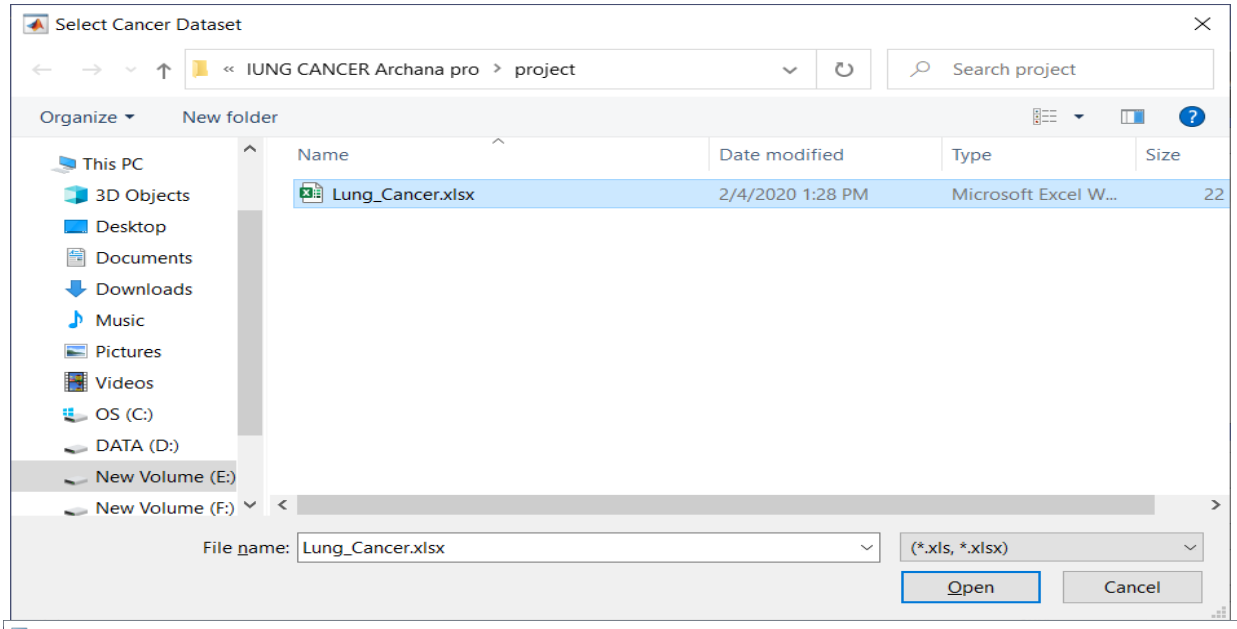
- The *recall* or *true positive rate* (*TP*) is the proportion of positive cases that were correctly identified
- The *false-positive rate* (*FP*) is the proportion of negatives cases that were incorrectly classified as positive
- The *true negative rate* (*TN*) is defined as the proportion of negatives cases that were classified correctly
- The *false-negative rate* (*FN*) is the proportion of positives cases that were incorrectly classified as negative
- The *accuracy* (*AC*) is the proportion of the total number of correct predictions.
- The *Sensitivity or Recall* the proportion of actual positive cases that are correctly identified.
- The *Specificity* the proportion of actual negative cases that are correctly identified.

Table 2 PERFORMANCE METRIC TABLE

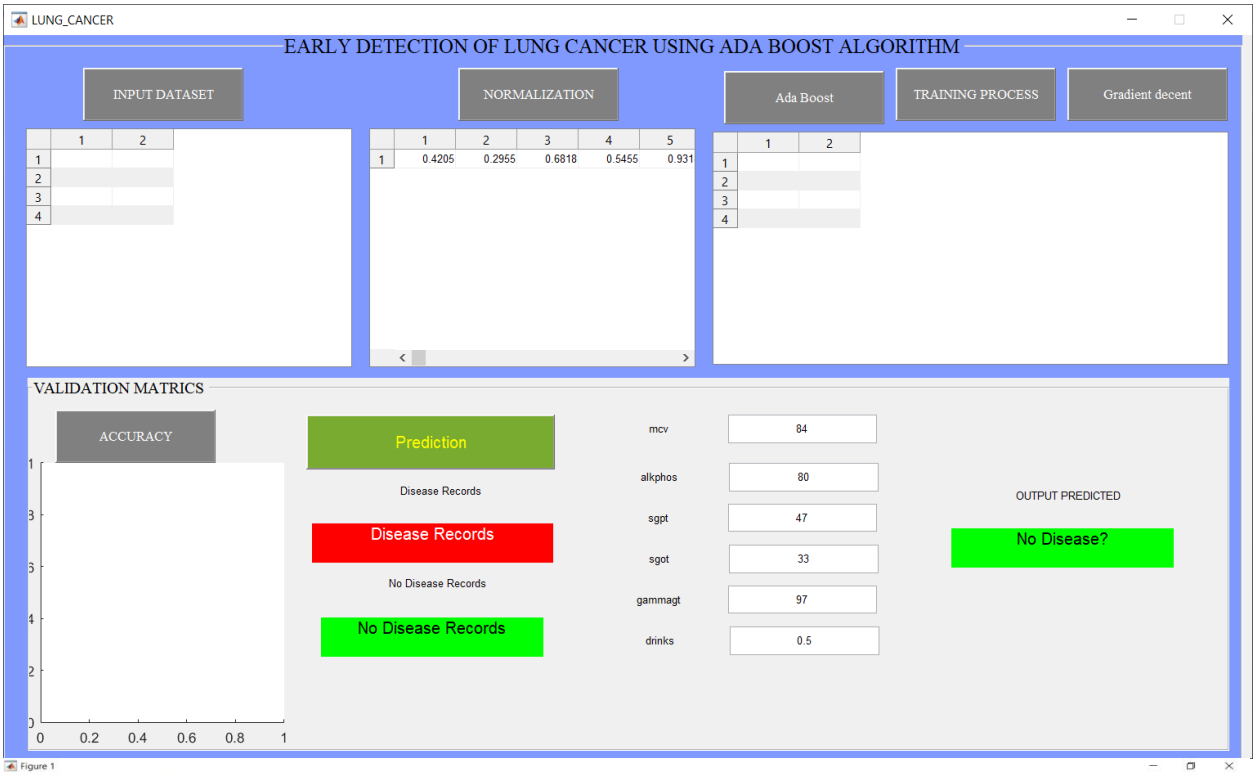
ALGORITHM	ACC	SEN	SPEC
ADABOOST	98.19	99.02	98.05

The above table shows the performance metric of the Adaboost algorithm . this algorithm got the best performance of 98.19% accuracy.

4. RESULTS


Select Cancer Dataset
« IUNG CANCER Archana pro › project
Search project
Organize
New folder
This PC
3D Objects
Desktop
Documents
Downloads
Music
Pictures
Videos
OS (C:)
DATA (D:)
New Volume (E:)
New Volume (F:)
Name
Date modified
Type
Size
Lung_Cancer.xlsx
2/4/2020 1:28 PM
Microsoft Excel W...
22
File name: Lung_Cancer.xlsx
(*.xls, *.xlsx)
Open
Cancel


LUNG_CANCER
EARLY DETECTION OF LUNG CANCER USING ADA BOOST ALGORITHM
INPUT DATASET
NORMALIZATION
Ada Boost
TRAINING PROCESS
Gradient decent
VALIDATION MATRICS
ACCURACY
Prediction
Disease Records
Disease Records
No Disease Records
No Disease Records
mcv 84
alkphos 80
sgpt 47
sgot 33
gammagt 97
drinks 0.5
OUTPUT PREDICTED
No Disease?


LUNG_CANCER
EARLY DETECTION OF LUNG CANCER USING ADA BOOST ALGORITHM
INPUT DATASET
NORMALIZATION
Ada Boost
TRAINING PROCESS
Gradient decent
0.4205 0.2955 0.6818 0.5455 0.931
VALIDATION MATRICS
ACCURACY
Prediction
Disease Records
Disease Records
No Disease Records
No Disease Records
mcv 84
alkphos 80
sgpt 47
sgot 33
gammagt 97
drinks 0.5
OUTPUT PREDICTED
No Disease?

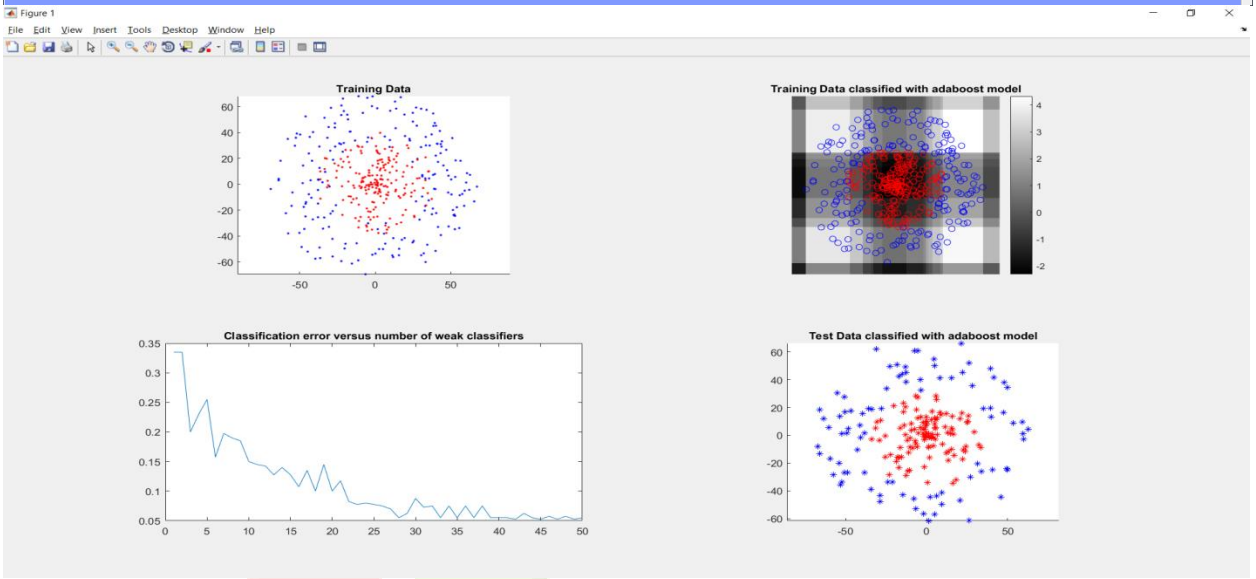


Figure 1
File Edit View Insert Tools Desktop Window Help
Training Data
Training Data classified with adaboost model
Classification error versus number of weak classifiers
Test Data classified with adaboost model

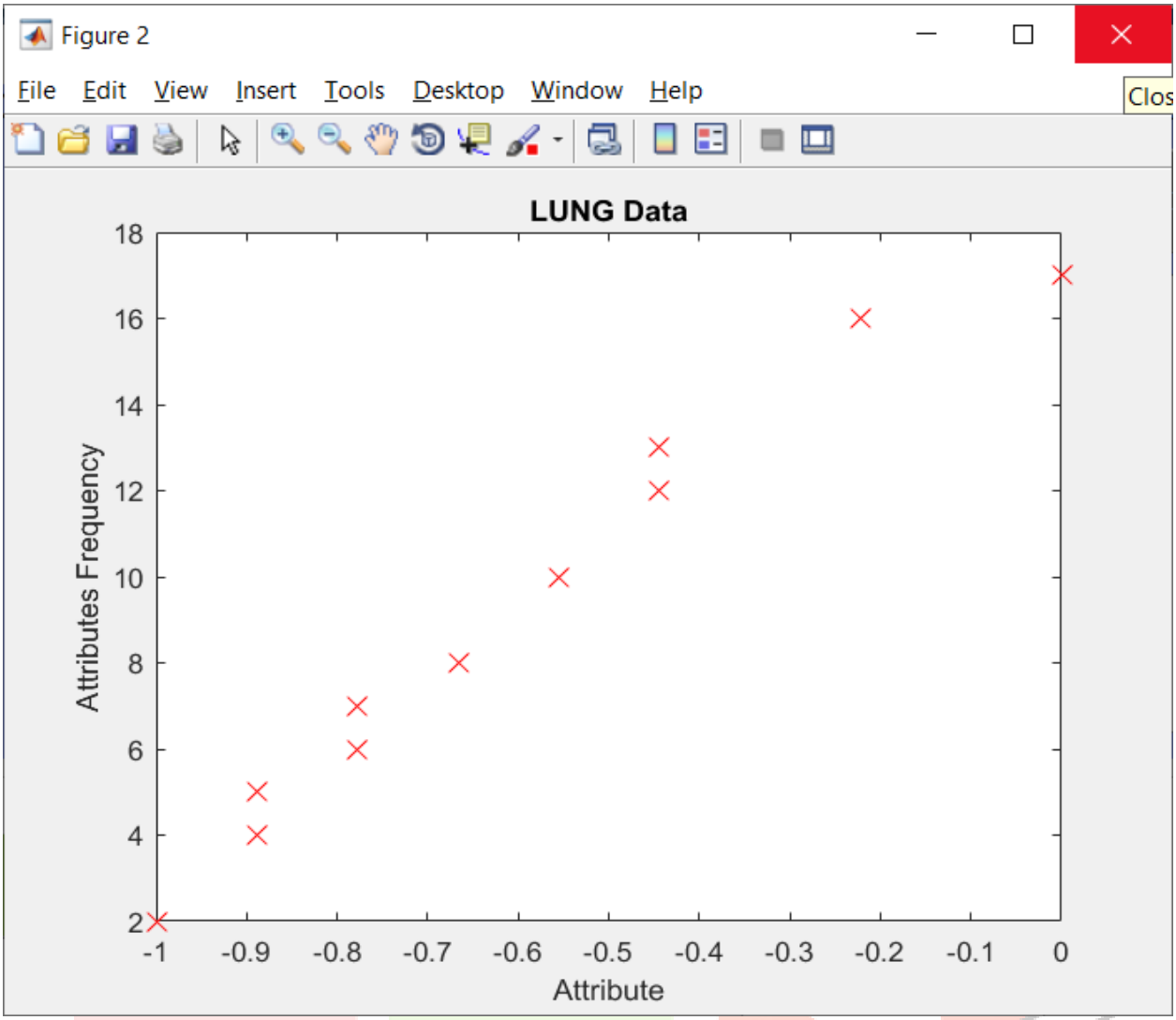
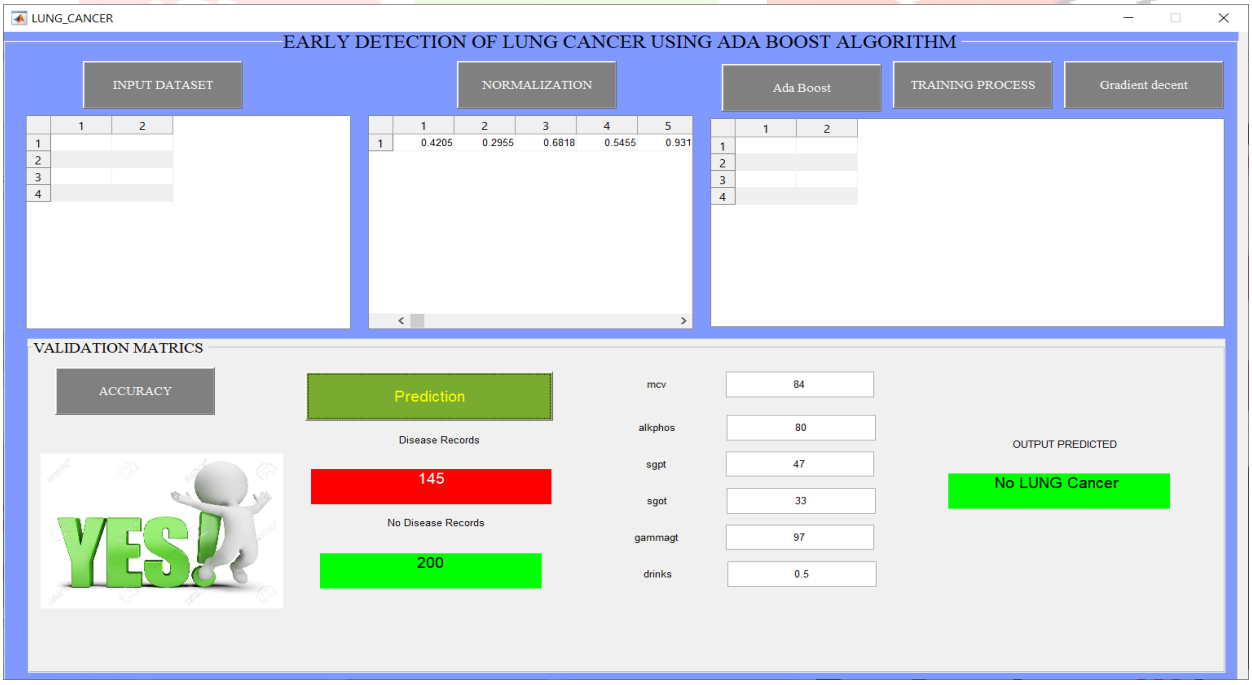
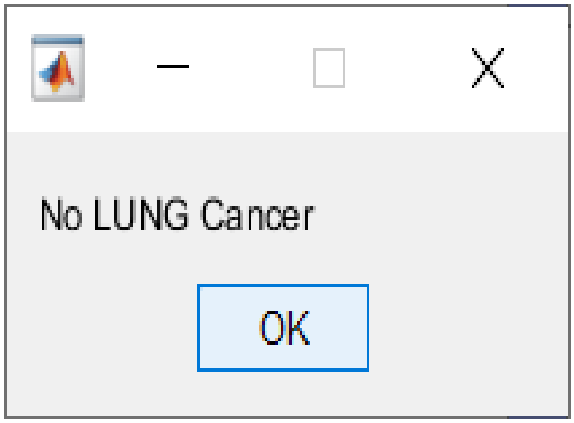


Figure 2
File Edit View Insert Tools Desktop Window Help
LUNG Data
Attributes Frequency
Attribute
18
16
14
12
10
8
6
4
2
-1 -0.9 -0.8 -0.7 -0.6 -0.5 -0.4 -0.3 -0.2 -0.1 0


LUNG_CANCER
EARLY DETECTION OF LUNG CANCER USING ADA BOOST ALGORITHM
INPUT DATASET
NORMALIZATION
Ada Boost
TRAINING PROCESS
Gradient decent
0.4205 0.2955 0.6818 0.5455 0.931
VALIDATION MATRICS
ACCURACY
Prediction
Disease Records
145
No Disease Records
200
YES!
mcv 84
alkphos 80
sgpt 47
sgot 33
gammagt 97
drinks 0.5
OUTPUT PREDICTED
No LUNG Cancer



5. CONCLUSION

An image improvement technique is developing for earlier disease detection and treatment stages; the time factor was taken in account to discover the abnormality issues in target images. Image quality and accuracy is the core factors of this research, image quality assessment as well as enhancement stage where were adopted on low pre-processing this filter shape remarkably reduces the number of calculations. We concluded that a False Positive reduction method using seven characteristic shape features and Support Vector Machines.

6. REFERENCE

[1] Divya Chauhan, Varun Jaiswal, " Development of Computational Tool for Lung Cancer Prediction Using Data Mining", 2016
[2] S. Durga, K. Kasturi, " Lung disease prediction system using data mining techniques", January 2017
[3] Mutiullah Jamil, Mehwish Bari, Adeel Ahmed, " Lung Cancer Detection Using Digital Image Processing Techniques: A Review", April 2019
[4] Raviprakash S. Shriwas, Akshay D. Dikondawar, " LUNG CANCER DETECTION AND PREDICTION BY USING NEURAL NETWORK", January 2015
[5] N Numan, S Abuelenin, " PREDICTION OF LUNG CANCER USING ARTIFICIAL NEURAL NETWORK", April 2018
[6] Saravanan K, Sasithra S, " Review on Classification Based on Artificial Neural Networks", December 2014
[7] N. Mohanapriya, B. Kalaavathi , T. senthil Kuamr, " Lung Tumor Classification and Detection from CT Scan Images using Deep Convolutional Neural Networks (DCNN)", Dec. 2019
[8] S. Vijayalakshmi, J. Priyadarshini, " Breast cancer classification using RBF and BPN neural networks", January 2017
[9] J.Jamera banu, " Study of Classification Algorithm for Lung Cancer Prediction", February 2016
[10] Titus.A, Khanna Harichandran Nehemiah, " Classification of interstitial lung diseases using particle swarm optimized support vector machine", January 2015