Lung Cancer Prediction System

Kshirsagar Shraddha Popat
Department Of Computer Technology
Sanjivani K.B.P Polytechnic,
Kopargaon,India

Prof. S. A. Patil (Lecturer)
Department Of Computer Technology
Sanjivani K.B.P Polytechnic.
Kopargaon, India
.

Saloni Santosh Parkhe
Department of computer Technology
Sanjivani K.B.P Polytechnic.
Kopergaon , India
.

Parth Dagadu Vahadane.
Department Of Computer Technology.
Sanjivani K.B.P Polytechnic,
Kopargaon , India

Vaibhav Sampat Wagh
Department Of Computer Technology
Sanjivani K.B.P Polytechnic
. Kopargaon, India,

Abstract

Cancer is a rising deadly disease which causes the death of 10% among all the diseases in Bangladesh. There are over 100 Types of cancer. Predicting Cancer plays a vital role for progressing data mining applications. Naive Bayes, k-nearest neighbor and j48 algorithm are used in this paper for predicting cancer disease. Naive Bayes is easy to build and really useful for very big dataset. K-nearest neighbor is uses dataset and create a dataset by separated into different classes and also predicting classification of new points. J48 Classifier are based on the decision Tree from training datasets, using the fact that each of them and data sets can be used for decision-making it into smallest subset. Weka tool is used for the purpose of measuring the accuracy of the cancer disease dataset including 09 types of cancer. 10-fold cross-validation is used for predicting cancer disease. In Naive Bayes the accuracy is 98.2%, k-nearest neighbor accuracy is 98.8% and j48 accuracy is 98.5%

Keywords- Data Mining, Lung Cancer Prediction, Classification, Naive Bayes, Bayesian Network, J48.

I. INTRODUCTION

Lung cancer is the one of the leading cause of cancer deaths in both women and men. Manifestation of Lung cancer in the body of the patient reveals through early symptoms in most of the cases. [1]. Treatment and prognosis depend on the histological type of cancer, the stage (degree of spread), and the patient's performance status. Possible treatments include surgery, chemotherapy, and radiotherapy Survival depends on stage, overall health, and other factors, but overall only 14% of people diagnosed with lung cancer survive five years after the diagnosis.
Symptoms that may suggest lung cancer include:

- dyspnea (shortness of breath with activity),
- hemoptysis (coughing up blood),
- chronic coughing or change in regular coughing pattern,
- wheezing,
- chest pain or pain in the abdomen,
- cachexia (weight loss, fatigue, and loss of appetite),
- dysphonia (hoarse voice),
- clubbing of the fingernails(uncommon),
- dysphasia(difficulty swallowing),
- Pain in shoulder ,chest , arm,
- Bronchitis or pneumonia,
- Decline in Health and unexplained weight loss.

Mortality and morbidity due to tobacco use is very high. Usually lung cancer develops within the wall or epithelium of the bronchial tree. But it can start anywhere in the lungs and affect any part of the respiratory system. Lung cancer mostly affects people between the ages of 55 and 65 and often takes many years to develop

II. LITERATURE SURVEY

The approach that is being followed here for the prediction technique is based on systematic study of the statistical factors, symptoms and risk factors associated with Lung cancer. Non-clinical symptoms and risk factors are some of the generic indicators of the cancer diseases. Initially the parameters for the pre-diagnosis are collected by interacting with the pathological, clinical and medical oncologists (Domain experts).

Lung cancer symptoms:-
1. Smoking
2. Yellow fingers
3. Anxiety
4. Peer pressure
5. Chornic disease
6. Fatigue
7. Allergy
8. Wheezing

9. Alcohol Consuming
10. Coughing
11. Shortness of breath
12. Swallowing difficulty
13. Chest pain

A.
III. Proposed System

In Proposed Systemhis system, users to get instant guidance on their Cancer disease through an intelligent system online. The Cancer Disease Prediction web application is an end user support and online consultation project. The application is fed with various details and the Cancer disease associated with those details. It allows user to share their Cancer related issues. It then processes user specific details to check for various cancer disease that could be associated with the inputs received from user. Here we use Naïve Bayes algorithm to predict the most accurate cancer disease that could be associated with user's details.

I.V ALGORITHM DESIGN

N figure 1 first the process is get started with the start. Symptoms are indicated as the training part. If the symptoms Are true then it showed the result. The test part is also working As the main part of predicting cancer. It the test part indicate True then the biopsy portion is performed otherwise the Process is stopped. The test part and the biopsy both show Negative then the process is stopped
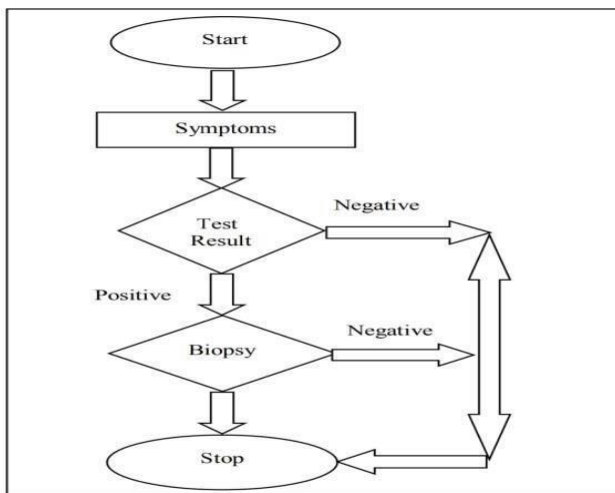

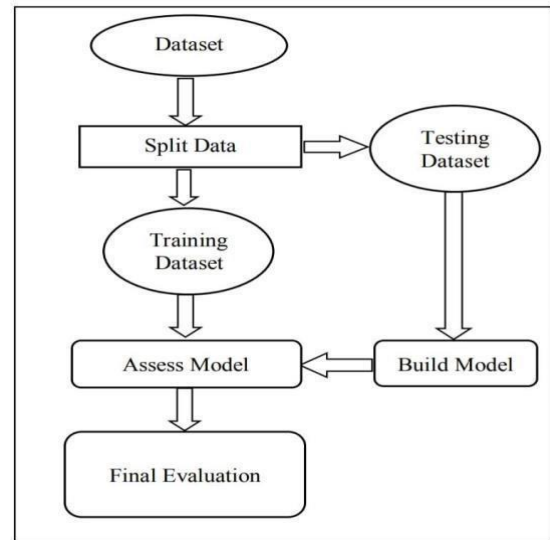Fig 1: Dataflow Diagram of predicting cancer

Data Mining Process

Fig 2: Data mining Diagram

The data mining process works on a large number of datasets. It also predicts where the dataset will be. It is used for Prediction, clustering, classification. There are two main Section training and testing.

☐ Training is the part where the dataset is trained by Input with expected output.

☐ Testing is a part where the evaluation of the model

In Fig 3 the Black one represents K-Nearest Neighbor (KNN), the orange one represents J48 algorithm and the blue One represents Naïve Bayes Algorithms accuracy level.
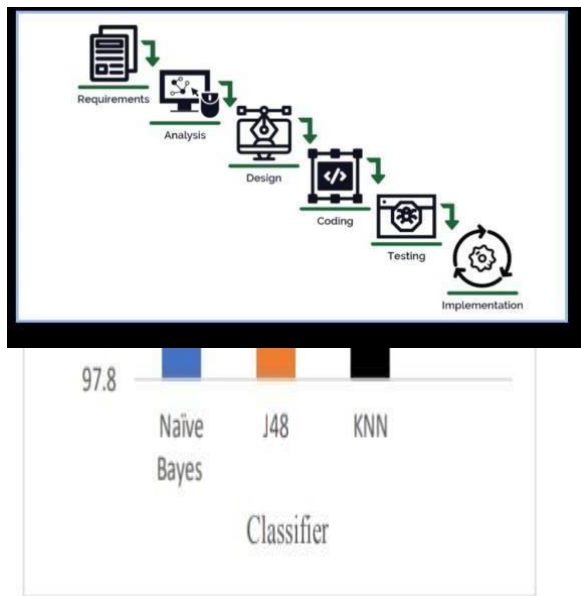
Fig 3: Graph of Accuracy

The waterfall model is a classical model used in system development life cycle to create a system with a linear and sequential approach. It is termed as waterfall because the model develops systematically from one phase to another in downward fashion. The waterfall approach does not define the process to go back to the previous phase to handle changes in requirement. The waterfall approach is the earliest approach that was used for software development

Advantages

1. It can easily get the cancer disease result on single click
2. It will tell that probably person can have lung cancer so patient can check from doctor.

Software Requirement

1. Laptop or PC .
2. Language :- Python
3. Intel core i3 processor
4. 500 GB HDD
5. 4 GB Ram

Result

In this project we got 88% accuracy and its predict the cancer by using neavy bayes algorithm

Conclusion

We processed the dataset to differentiate the Affected patient and its level of the growth of The cancer by the machine learning system. Here it presented an approach to find best

Accuracy of the cancer result to assist the Radiologist and for the future enhancement. Further loads ought to be directed at Improving the classifying accuracy levels of

Future Scope

The cancer detection system using the machine learning techniques is much efficient and gives the betterment results to radiologist and assist them. This enhance with the additional features for upgrading in the future. On this processing system to support the radiologist to detect the affected patients as accurate as the result.

Reference .

https://ieeexplore.ieee.org/document/89506 50
 [1] Ayyadurai.P, Kiruthiga.P, Valarmathi.S,

Amritha.S , Respiratory Cancerous Cells Detection Using TRISS Model and
Association Rule Mining, International
Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 2, Issue 3, March 2013.

[2] Priyanga.A, Prakasam.S, *Effectiveness of*

*Data Mining - based Cancer Prediction System (DMBCPS),*InternationalJournal of Computer Applications Volume 83 – No 10, December 2013.

[3] ShwetaKharya, *Using Data Mining Techniques for Diagnosis And Prognosis of Cancer Disease* International Journal of

Computer Science, Engineering and Information Technology (IJCSEIT), Vol.2, No.2, April 2012.

[4]HarleenKaur and SiriKrishanWasan,

Empirical Study on Applications of Data Mining Techniques in Healthcare, Journal of Computer Science Vol.2 (2), 2006.

[5]Sang Min Park, Min Kyung Lim, Soon Ae Shin & Young Ho Yun 2006, *Impact of prediagnosis smoking, Alcohol, Obesity and Insulin resistance on survival in Male cancer Patients:* National Health Insurance corporation study, Journal of clinical Oncology, Vol.24, Number 31 November 2006.

[6]YongqianQiang, YouminGuo, Xue Li,

Qiuping Wang, Hao Chen, &DuwuCuic, *The Diagnostic Rules of Peripheral Lung cancer Preliminary study based on Data Mining Technique*, Journal of Nanjing Medical University,Vol. 21(3):190-195.

[7]   Krishnaiah,V.,   Narsimha,G., SubhashChandra,N., *Diagnosis of Lung Cancer Prediction System Using Data Mining Classification Techniques*, et al,/(IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 4 (1) , 2013.

[8]PrashantNaresh, *Early Detection of Lung Cancer Using Neural Network Techniques*, Journal of   Engineering Research   and Applications Vol. 4, Issue 8, August 2014.

[9]Thangaraju P, Karthikeyan T, Barkavi G,*Mining Lung Cancer Data for Smokers and Non-Smokers by Using Data Mining Techniques*,International   Journal of Advanced   Research in Computer and Communication Engineering Vol. 3, Issue 7, July 2014.

[10]Ravi Kumar G.,Ramachandra.A, Nagasaki,   *An Efficient Prediction of Breast Cancer Data Using Data Mining Techniques*, International Journal of Innovations in Engineering and Technology (IJIET) Vol. 2 Issue 4 August 2013.

[11]   Tapas RanjanBaitharu, Subhendu Kumar Pani A, *Comparative Study of Data Mining Classification Techniques using Lung Cancer Data,* International Journal of Computer Trends and Technology (IJCTT) – volume 22 Number 2–April 2015.

[12]Mary KirubaRani.V,SafishMary.M,

*Predicting Progression of Primary Stage Cancer to Secondary Stage Using Decision Tree Algorithm*International Journal of

Advanced Information Science and Technology

(IJAIST) Vol.26, No26, June 2014.

[13]Ada, RajneetKaur, *Early Detection and*

Prediction of Lung Cancer Survival using Neural Network Classifier International Journal of Application or Innovation in Engineering & Management (IJAIEM)Volume 2, Issue 6, June 2013.

[14]Sowmiya.T, Gopi.M, Thomas Robinson,

Optimization of Lung Cancer Using Modern

*Data Mining Techniques,*International Journal of Engineering Research Volume No.3, Issue No.5.

[15]Vijaya.G,   Suhasini.A,   Priya.R,

Automatic Detection of Lung Cancer In CT Images IJRET: International Journal of Research in Engineering and Technology Volume 03, Issue:07|May-2014.