International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 1, Issue 3, pp 45-49, December-2017

Determination of Post Thoracic Surgery using Supervised Machine Learning Algorithms

Sujatha Godavarthi

Associate Professor, Department of Computer Science and Engineering, Malla Reddy College of Engineering for Women, Hyderabad -500100, Telangana, India

Abstract - Lung Cancer is a disease characterized by the uncontrolled cell growth in tissues of the lungs. It is one of the dangerous and life taking disease in the world. One of the major causes of the death in human beings is Lung Cancer. The early detection of lung cancer and also the proper medication is important for the diagnosis process and it gives the higher chances for successful treatment. Therefore, the objective of this project is to develop a system that predicts the life expectancy post thoracic surgery for the lung cancer infected patient. Once the cancer is detected, the Thoracic Surgery is one of the treatment options for the diagnosis of Lung Cancer. The project involves the analysis of the patient's dataset who underwent Thoracic Surgery and an attempt is made to model a classifier that will predict the survival of the patient post the surgery. The dataset will be trained using four Supervised Machine Learning Algorithms Namely Linear Discriminant Analysis (LDA), Support Vector Machine (SVM), Random Forest and Logistic Regression. The classifier classifies the input attribute values and will predict whether the patient will survive for a minimum span of one-year post the surgery. This project may be considered as a promising tool to support the medical specialist to make a more precise diagnosis and prognosis concerning the lung nodules.

Keywords- Machine Learning, Thoracic Lung Cancer, Thoracic Surgery, Trained Datasets, Flask Framework, Post Method.

I. INTRODUCTION

Lung Cancer was an uncommon before the invention of smoking cigarette, and it wasn't even recognized as a disease until 1761. In 1810, distinct characteristics of carcinoma were defined in greater detail. Malignant respiratory organ tumours accounted for just one percent of all malignancies seen at autopsy in 1878, but by the first decennium, they had increased to 10–15 percent. In 1912, there were only 374 case reports in the medical literature globally, but a survey of autopsy revealed that the prevalence of carcinoma had quadrupled from zero.3% in 1852 to 5.66 percent in 1952 In Germany, Fritz Lickint a physician proved what the link between smoking and lung cancer is in 1929, resulting in a strong anti-smoking campaign. The Study of British Doctors' which was published in the 1950s, was the first large-scale epidemiological study to show a link between smoking and lung cancer. In 1964, the US General Surgeon recommended that smokers give up their habit. The first mention of noble gas was made in the Ore Mountains by miners in Schneeberg, Saxony. Since 1470, silver was well-mined there, and mines are made of metallic elements, with their associated noble gas. Within a decade, miners had a disproportionate amount of respiratory organ malady, which was later diagnosed as carcinoma. Inspite this, mining which was continued throughout 1950s, owing to desire the Soviet Union's for metallic elements. In the 1960s, noble gas was identified as a possible cause of cancer.

Thoracic surgery is an operation that is done on organs which is located in the chest, that includes heart, lungs and esophagus. The lung and esophageal cancer are treated by, specialized thoracic surgeons while treat the heart is treated by specialized cardiac surgeons. Thoracic surgery is also called as chest surgery, which is used to diagnose or repair lungs that are affected by cancer, trauma.

II. OVERVIEW

Lung Cancer is a malicious lung tumor is characterized without controlling cell growth in lung tissues. It is one of the most dangerous and life threading disease in the world. There are some treatments available for Lung Cancer Disease like Thoracic Surgery, Radio Therapy etc. Therefore, its important to decide on the proper medication or treatment which has to be given to the Lung Cancer infected patient for the greater chances of survival of the patient.

Therefore, the project focuses on predicting whether the patient will survive or not, at least for a minimum span of one year post thoracic surgery.

© 2017 IRJIET All Rights Reserved

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 1, Issue 3, pp 45-49, December-2017

III. PROBLEM STATEMENT

The goal of this study is to construct a prediction system that uses supervised Machine Learning Algorithms to estimate the survival of carcinoma infected patients after pectoral surgery.

IV. EXISTING SYSTEM

There are just a few systems that attempt to predict survival after pectoral surgery. Existing solutions just provide mil models for matter statements, but no front-end interface for end users. There isn't a lot of information on programming at the United Nations. Also, the types of gift systems that are now available have an accuracy range of 70% to 85%

V. PROPOSED SYSTEM

The system aims at forecasting the survival of lung cancer who are infected patient by post thoracic surgery. The dataset has been included with 17 attributes which are specified in Table-1. Among all those attributes, Risk 1YEAR is the target class mentioning zero if the patient survives for at least one-year post thoracic surgery and one for those who died before completing one-year post surgery. The visualization of dataset is been done using the Matplotlib and Seaborn libraries of Python. Furthermore the essential attributes are found based on the Information Gain (IG) attribute evaluation which is been used to find the importance of attribute by using the Information Gain with respect to the target class. IG (Class, Attributes) = E (Class) – E (Class | Attributes). where, E stands for Entropy After the IG Attribute Evaluation on all the 16 independent attributes in the dataset, the essential attributes would be decided which will be used to train the model.

Firstly to train dataset the following algorithms are used:

1. Random Forest: It’s an ensemble method of training for regression, classification that performs by constructing a multitude of decision trees at the time of training and gives the output for the class that is mode of the classes of the single trees.
2. Linear Discriminant Analysis: It’s a Technique for dimensional reduction that attempts to reduce the number of data sets while retaining as much information as possible.
3. Support Vector Classifier (SVC): It tries to identify a hyperplane in an N-dimensional space that clearly classifies the points of the data.
4. Logistic Regression: It’s an algorithm which is used to forecast the prospect of a target variable.

To train the dataset, the principle of 10-fold Cross Validation is used. Here at the first ten equal sized datasets is created from the original and real data. Further each data set is partitioned into mainly two parts, that is 90% for training and remaining 10% for testing. After this, a classifier is produced from 90% labelled data and applied to 10% test data for set 1. The above procedure is repeated for remaining sets that is from 2 to 10.

VI. SYSTEM DESIGN

Flash Black: Flask is a small python-language web framework. Flask can be classified as a small framework and requires no specific libraries or tools. Flask supports extensions that are enforced in flasks alone by adding application options. Extensions will exist for relational object mappers, transfers, licenced and many other similar tools connected to authentications. Unit area extensions update more than Core Flask. MongoDB uses the Flask, which provides extra data history management. Flask Framework is used to use Pinterest LinkedIn and the flask itself for community-based web content.

Werkzeug: Werkzeug is a utility library of the python programming language in other words it is a toolkit for the Net server entry interface (WSGI) applications are verified under a BSD License. Werkzeug notices the software package objects are for request response and the utility functions and It is easy to build a custom software package frame work and also supports a pairs.

For establishing sites and net applications, HTML is a normal nomenclature. Language hypertext markup describes the site markup structure. Language labels that include items such as "heading," "paragraph," "table" and so on can be marked in hypertext. Browser does not show the hypertext tags used to render the page contents. The traditional cornerstone technologies for the worldwide network are the cascades of vogue sheets(CSS) and JavaScript. Browser receives hypertext markup language

© 2017 IRJIET All Rights Reserved

International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 1, Issue 3, pp 45-49, December-2017

documents which are stored regularly and rendered to the Web sites by the web server. Semantically express the structure of an internet page originally surrounded for the outline document with the hypertext labelling language.

VII. SYSTEM ARCHITECTURE

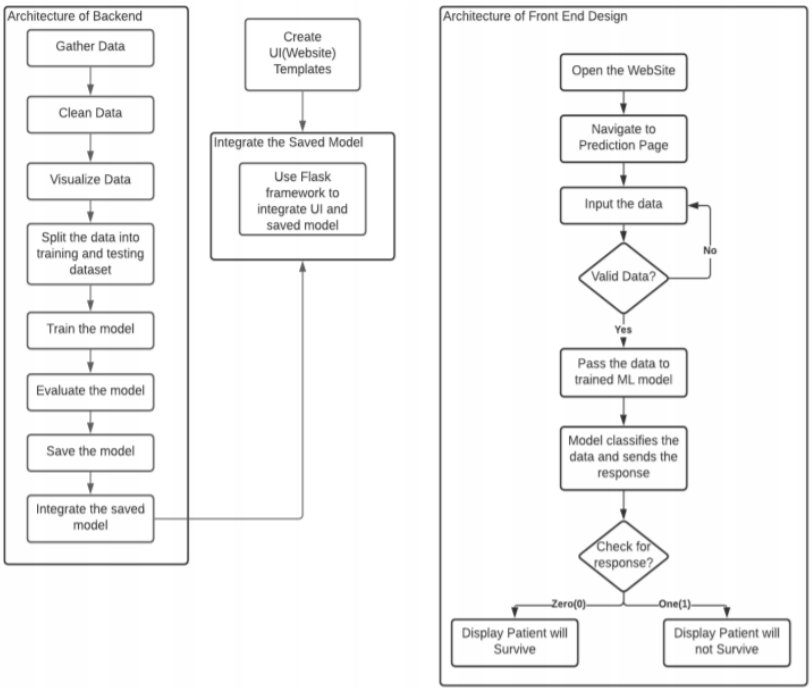

Fig. 1: System Architecture

In the above diagram specifies System Architecture of the project where this diagram contains two actors, one is front end and the other is backend. First we should gather the data from the data sets and we should clean the null values from the data set then we should visualize the data. After visualizing the data we should split the data into two parts for testing and for training the dataset after they should train the model again should evaluate the model again should save model the saved model should be integrate by using the flask framework to integrate UI.

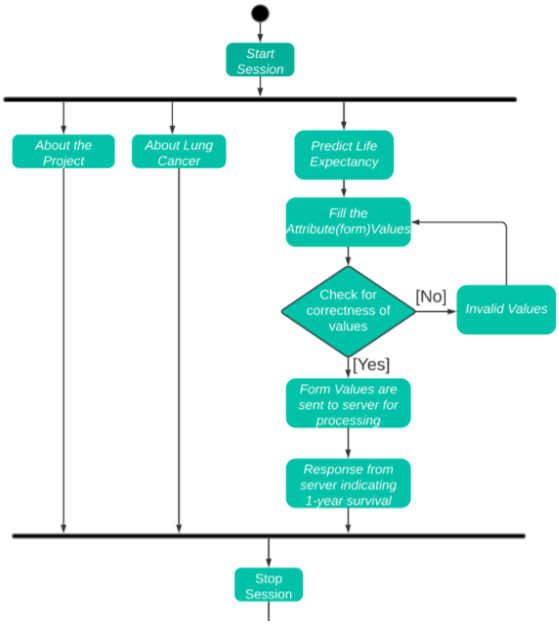

Fig. 2: Activity Diagram

© 2017 IRJIET All Rights Reserved

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 1, Issue 3, pp 45-49, December-2017

In the backend at the end of the program if the code is correct we will get one website link after open we come to the frontend in the we will be having the introduction and also it navigates to the prediction page in the we should insert the data if the data is valid it pass the data to the trained model if not again we should enter the data .If the data is valid the model classifies the data and sends the response from that response (zero , one)zero means patient will survive , one means patient will not survive All these controls using the Local server without using the Internet. The control speed is very fast as everything is happening locally.

A diagram of activity can be built in a number of forms connected to arrows. Ellipses are actions. Shapes are the main types. Decisions are made by diamonds. Split start or end of the concurrent activities is shown in bars. The initial workflow node is the black circle. A circled black circle is the final node. Arrows start towards the end and it represent the order in which activities will take place. Activity diagrams can be viewed as a structured flow chart with a prescription flow chart. Typical techniques of flowchart are designed to express competition. However, only simple cases solve the division and joining symbols in activity diagrams when casually combined with decisions or loops, the meaning of the model will not be clear. It is a graphical representation of workflow of stepwise activities and actions which tells the exact flow of an activity. The above figure specifies the activity diagram of the system where each functionality is considered as an activity indicated in ellipse.

VIII. RESULTS

The Cross Validation of 10-fold method that is used to train and examine the dataset using just four techniques before. Because accuracy alone isn't enough to determine which model is the best, the Fmeasure is calculated in addition to the accuracy to assess the model. The parameters determined from the Confusion Matrix graphic are accuracy and F-measure area unit.

Table 1: Confusion Matrix

CONFUSION MATRIX		Predicted Value	
		P	N
Actual Value	P	TP	FN
	N	FP	TN

Accuracy: It's the percent of observations that the model correctly predicts.

Accuracy = (TP + TN) / (TP + TN + FP + FN) where,
TP – True Positive
FP – False Positive
TN – True Negative
FN – False Negative
F-Measure: It's also known as F-Score, which is a measure of precision and recall.
F-measure = (2 * Precision * Recall) / (Precision + Recall)
where,
Precision = TP / (TP + FP) Recall = TP / (TP + FN)

IX. CONCLUSION

Lung cancer is one of the challenging problems in medical field due to structure of cancer cells. Therefore, the proper medication has to be given to the patient for increasing the survival chances of the patient. Once the cancer is detected, the Thoracic Surgery is one of the best treatment options for the diagnosis of Lung Cancer. The project involves the analysis of the patient's dataset who underwent Thoracic Surgery and an attempt is made to model a classifier that will predict the survival of the patient post the surgery. The dataset will be trained using four Supervised the Algorithms of Machine Learning that are Linear Discriminant Analysis (LDA), Support Vector Machine (SVM), Random Forest and Logistic Regression. Among the four algorithms used, its observed that the LDA algorithm which gives the highest accuracy of 83.76% compared to the other algorithms. data in the future to analyse the system.

© 2017 IRJIET All Rights Reserved

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 1, Issue 3, pp 45-49, December-2017

X. FUTURE WORKS

Overall, the system acts as an aid in the medical field which will assist the doctors to whether or not consider the thoracic surgery as the treatment option. The system can be enhanced for the following functionalities to make the system more efficient, effective and applicable to meet the requirements of real world.

1. In the prediction of survival part, higher accuracies of the classifier can be achieved by considering large amount of dataset.
2. The system can be further enhanced to suggest on the type of treatment to be given for the lung cancer infected patient keeping in mind about the most applicable treatment depending on the input values which are given.

REFERENCES

[1] Data Mining Techniques for Predicting Outcome of Thoracic Surgery (2015).
[2] The Prediction Of Post-operative Life Expectancy In Lung Cancer Patients (2020).
[3] Data Mining Techniques for Prediction Of Post-surgical Survival Of Lung Cancer Patients After Thoracic Surgery Using Data Mining Techniques for (2017).
[4] Analyzing a Lung Cancer Patient Dataset with the Focus on Predicting Survival Rate After One Year Of Thoracic Surgery(2017).
[5] Improved Prediction of Post-operative Life Expectancy after Thoracic Surgery(2016).
[6] Evaluation of Machine Learning algorithms performance in post-operative life expectancy in the lung cancer patients (2015).
[7] Life Expectancy Post Thoracic Surgery (2015).
[8] The Prediction of Post-operative Life Expectancy in Lung Cancer Patients (2020).

******

© 2017 IRJIET All Rights Reserved