International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 07 | July 2021 www.irjet.net p-ISSN: 2395-0072

Prediction of Survival of Post Thoracic Surgery

K S SAIKUMAR REDDY[1], BHANU PRAKASH RB[2], MUTHURAJU K[3], ABHISHEK RAJ G[4], SMITHA SP[5]

[1-4]Student, Dept. of Information Science and Engineering, Sai Vidya Institute of Technology, Karnataka, India
[5] Professor, Dept. of Information Science and Engineering, Sai Vidya Institute of Technology, Karnataka, 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- *MachineLearning, Thoracic LungCancer, 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 recognised 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 is the link between smoking and lung cancer 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. According to the United Nations, in the year 2018 2.09 million instances of carcinoma were recorded, with a total of one.76 million people dying as a result of carcinoma. The late identification of cancer is one of the factors for the high death rate due to cancer. In addition, the kind of carcinoma, the stage, and the patient's performance all influence therapy and prognosis. Pectoral surgery, therapy, and radiation therapy are all options if the cancer is discovered. In 1933, the first roaring extirpation

for carcinoma was carried out. Since the 1940s, palliative radiation treatment has been employed. The goal of radical radiation treatment, which was originally used in 1950s, which was to employ higher radiation doses in patients with fairly early stage of cancer who were not unsuitable for surgery. CHART was first thought to be an advance over traditional radical radiation treatment in 1997. Initial attempts at surgical surgery and severe radiation treatment for SCLC in the 1960s were unsuccessful. Roaring treatment regimens were created in the 1970s. Prediction of Survival of Post Thoracic Surgery Dept The discipline of medicine known as cardiothoracic surgery (also known as pectoral surgery) is involved with the surgical operation of organs within the thorax (chest)—generally treating problems of the intestines (heart disease) and lungs (lung disease). For certain COPD and respiratory illness patients, respiratory organ volume reduction surgery, or LVRS, will enhance their quality of life. Components of the respiratory organ that have been damaged by respiratory disease are eliminated, allowing the remaining, relatively intelligent respiratory organ to grow and perform more quickly. The beneficial effects are proportional to the amount of residual volume that is reduced. Typical LVRS entails surgery on the most seriously damaged parts of respiratory sickness, such as the non-bullous respiratory organ (aim is for 20-30 percent). For individuals with end-stage COPD due to an underlying respiratory illness, this is a surgical option that involves a mini-thoracotomy and can enhance respiratory organ elastic recoil as well as diaphragmatic function.

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. Thoracic surgery procedures is done in minimally forward techniques or thoracotomy an open surgical procedure is carried out.

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

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. LITERATURE SURVEY

Authors: Md. Ahasan Uddin Harun, Md. Nure Alam.
Objective: Performance of several data mining techniques are compared to predict the Survival.
Algorithms Used: Naïve Bayes, Simple Logistic Regression, J48, MLP are used.
Results: Naïve Bayes (77.74%), Simple Logistic Regression (84.55%), J48(84.64%), MLP (80.91%) This article describes the performance of different data mining techniques and their upgraded versions, generally giving the best results indicating that a simple logistic regression is increased, or at least that the data mining techniques are competitive. These algorithms are carefully considered and their performance is analysed in various measurements. This study will explain
data mining techniques in medical data analysis as the findings have been explained from the statistical framework. [1]

Authors: Manu Siddhartha, Paramita Maity, Rajendra Nath
Objective: The authors gives the application of based enabled technique i.e., Random Forest which predicts the lung cancer patients survival status who have undergone major lung resections of lung cancer using thoracic surgery.
Algorithms Used: Random Forest Algorithm is used and gave an accuracy of 84.1%. This paper describes, how machine learning pipeline have been developed by handling data problem which is imbalanced and novel approaches that are proposed for explaining that is hard to understand a machine learning model's predicted for survival of patient's. These approaches can be used to have explanations that the humans can understand the rules where each individual can predict the model of complex machine learning which is used to solve the problem that can predict the life expectancy of the lung cancer patients post surgery.[2]

Authors: Abeer S. Desuky and Lamiaa M. El Bakrawy Objective: This paper proposed a methodology to predict the post-operative life expectancy after Thoracic Surgery using attribute and selection ranking technique.
Algorithms Used: Simple Logistic, Multilayer Perceptron and J48 are used. Results: Simple Logistic (84%), Multilayer Perceptron (81%) and J48(84%) In this study, the quality of three attribute ranking and selection methods has been evaluated to improve the prediction for life expectancy of lung cancer patients after thoracic surgery. Five machine learning techniques before and after applying the attribute ranking

and selection methods have been compared with their boosted versions. The results show that boosting is not always the better choice where attribute ranking and selection can perform better in improving prediction accuracy. Other attribute selection and machine learning techniques can be introduced in the future work to gain a better prediction model performance of the dataset. [3]

VII. SYSTEM DESIGN:

A. IMPLEMENTATION:

Python is typically called interactive programming language that is object-oriented and high level. Guido van Rossum founded Python from 1985-1990. Like Perl, Python's ASCII text file was also preferred under a general public licence ante lope (GPL). Python is typically called interactive programming language that is object-oriented and high level. Guido van Rossum founded Python from 1985-1990. Like Perl, Python's ASCII text file was also preferred under a general public licence ante lope (GPL).

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.

Jinja: Jinja guide engine to the python programming language and it is verified under a BSD License. Similarly Django net framework provides the templates for unit area evaluated in the sandbox.

Hypertext mark-up language (Front-End) :
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 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.

CSS (Front-End): Cascading mode sheets (CSS) is a piece of paper used to describe the presentation in marking language of a document. Although the visual sites that are continuously visited are the type and user interfaces of sites written in hypertext and XHTML, the language was often used on any XML documents together with XML SVG, XUL and XUL, which are applicable to speech rendering and other media. JavaScript and CSS are the main technology used by most wedding websites for visually stable web pages, user interfaces of networking applications and user interfaces in the different mobile application fields together with the Hypertext marking language. CSS is used primarily to modify display and content separations along with layout direction, colours and fonts. Separation enhances content accessibility flexibility; management within the presentation specifications; and modifies the hypertext multiple markup language pages; share information by inserting relevant CSS files into a

separate CSS, and scale quality back, repeating the content within the structure.

JavaScript (Front – End): JavaScript is a prototype based scripting language, dynamic, weekly typewritten, general purposed programming language and has major function. And also multi-paradigm language, supporting object-oriented and imperative and programming designs. JavaScript is formalised in ECMA and is mostly used in the form of clients. JavaScript, Conduct a part of the web browsers to provide more dynamic and more powerful user interfaces. It provides programmers with bunch access to the object. Use JavaScript in external website applications. Specific browsers, desktop widgets, instance in a PDF document. For validation purposes like text nox validation, email validation, signal validation JavaScript is used in this app. JavaScript is an intelligent web application verification tool.

B. SOFTWARE IMPLEMENTATION:

Project Modules

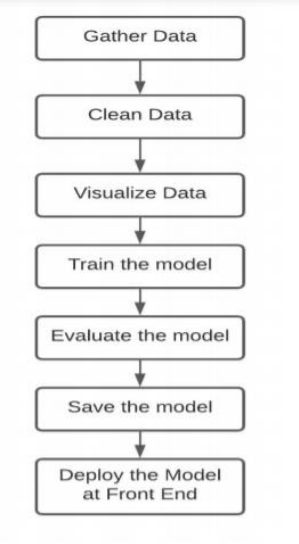

Fig.1.Project Modules

The project is divided into 7 different modules.

- The dataset is gathered and cleaned to eliminate Null Values and redundant values.
- Next the dataset is visualised to find the essential attributes used to coach the model.
- The knowledge set is split into testing data and a few verified algorithms area unit accustomed to train the model.
- The models in the area unit are evaluated by using gross validation techniques to get the important economical models.
- Later the important economical model is saved and deployed at the front- end.
- The front-end website is the newly formed hypertext mark-up language, CSS and JavaScript.
- The front-end and the back-end model is integrated flask framework written in the python language.

Logistic Regression: The logistic regression modelled a chance of failure or loss, loss of victory or loss, live or dead, healthy or sick for a particular class or occurrence. This can extend the model to a number of events such as whether a picture contains a cat dog, a lion. The probability between one and zero is assigned for each object detected in the image, with a total of one. Logistic regression is the statistical model that uses a logistical function in its basic form, although many more complex extensions exist to model a binary dependent variable. In regression analysis, logistic regression calculates a form of binary regression for the parameters of a logistic model. Mathematically, a binary logistic model has two possible values as the pass/fail indicator, in which the values are marked by 0 and 1, dependent on the variable. The log-odds of the logarithm of the odds of the labelled value 1 are linear combinations of one or more independent variables, each of which can be a binary Classes of variables 2 coded by variable indicator or variable continuous. The associated probability of labelled value 1 can certainly differ between 0 and 1 and therefore the labelling of a function which converts log odds to probability in the logistic function. For the log odds scale, the unit of measurements is named a logit by the logistic unit. Similar models with a different sigmoid function can be used, as does the pro-bit model instead of logistic functions The characteristic

defining feature of the logistic model is that the odds of the given result increase on one of the independent multiplicative variables This generalises the odds ratio at a constant rate at which each variable has its own binary dependent variable parameter.

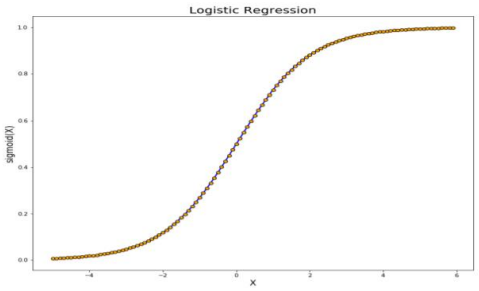

Fig.2.Example graph of the Logistic Regression

Random Forest:

Random forest is a group training system for the classification of regression and other tasks in which a group of decision making bodies is established. Tasks that give random forest output are the classifications of the selected trees. For regression tasks the mean or average forecast of each trees is reported. Random forest decisions for decision making are correct. Random forests generally work, but are less accurate than gradient enhanced trees. Data performance may be affected by its characteristics. The 1995 forestry algorithm created by Tin Kam Ho used the random subespace method, Ho's wording being a way of implementing Eugene Kleinberg's "stochashastic discrimination" classification approach. Random forest is a supervised learning formula. It is trained with the " bagging " Methodology. Final plan of the material methodology is a mixture of learning modules increases the result.

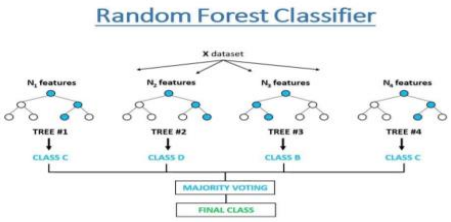

Fig.3.Example graph of the random forest

Support Vector Classifier: support vector machines support SVM as well as support-vector networks are supervised learning models that analyse the classification and regression data with associated learning algorithms; Developed by Vladimir Vapnik with Boser colleagues at AT&T-Bell Laboratories in 1992, in Guyon in 1993, in Vapnik in 1997, SVM is one of the most robust prediction methods based on the Vapnik 1982, 1995, and in Chervonenkis 1974 statistical frameworks or theory on vc. In view of a serial of training examples, each one of its categories is marked with an SVM training algorithm and creates a model which gives new example in one category or another. Although

SVM is used in a probabilistic classification setting by methods like Platt scaling. SVM maps examples of space training to maximise the gap between categories. Examples for space training. New examples are then mapped to the same space to predict which side of the gap is to be part of a category. In addition to linear settings, SVMs can perform a non-linear

classification by using a kernel trick to implicitly map their entries to high-dimensional, future spaces. SVC is controlled machine learning and uses regressive challenges for classification. However, it is somehow used for problems of classification. Between the SVM plot, the item with each feature in the chosen coordinate is in an n-dimensional matrix. Then there is the hyper plane which distinguishes 2 categories.

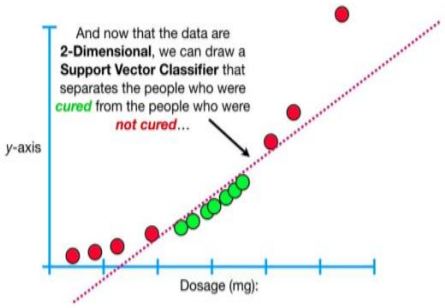

Fig.4.Example graph of the support vector classifier

Linear Discriminant Analysis :
Fisher's generalisation of the linear discrimination analaysis, the method used to identify a linear combination of features that

characterises or separates two or more classes of objects or events, is a linear discrimination analysis (LDA), standardised discriminant analysis (NDA), or discriminant function analysis. The resulting mixture can be used as a linear classifier or, more typically, to reduce dimensionality before further classification. ANOVA analysis and regression analysis are very much linked to LDA, which also seeks to demonstrate one variable as a linear combination of other functions and measurements. However, Anova does employ categorical independent parameters and a constant dependent parameter, while discriminating parameters and categorically dependent parameters are constantly independent. Logistic regression is much closer than LDA, because constant independent variables also explain a categorical variable. In applications that do not necessarily assume that independent variables are generally distributed, this other method is preferable

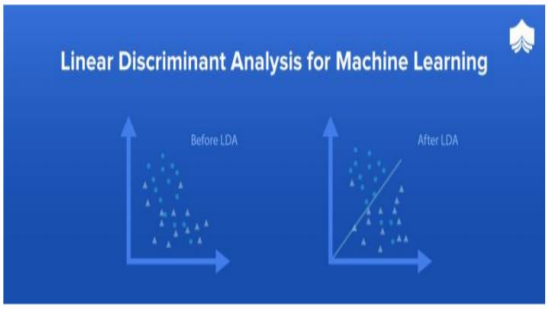

Fig.5.Example graph of the liner discriminant analysis

C. SYSTEM ARCHITECTURE

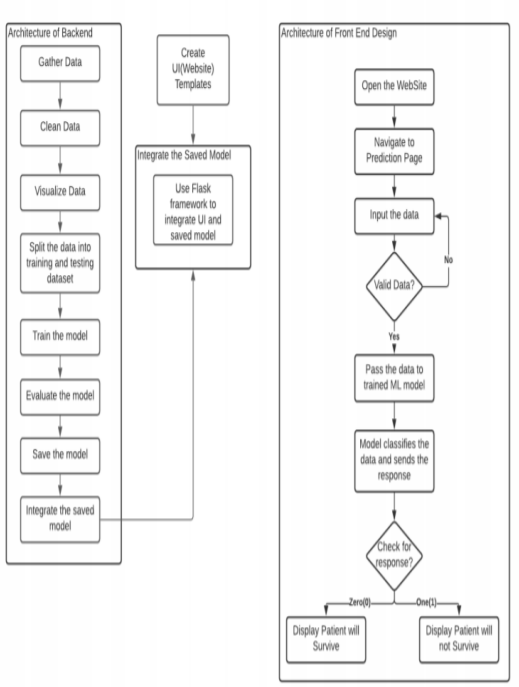

Fig.6.System Architecture

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 the 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.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.

D. USE CASE DIAGRAM

The graphical results of a possible user interact with a system can be a case diagram. Different usage cases and different types of users are shown in the use case scheme.
The various types of diagram are also often accompanied. Either circles or ellipses can be seen in the case of use. The performer and often as stick figures. Simplest user interaction of object representation where the system shows difference between the user and the different use cases where the user and system was involved.

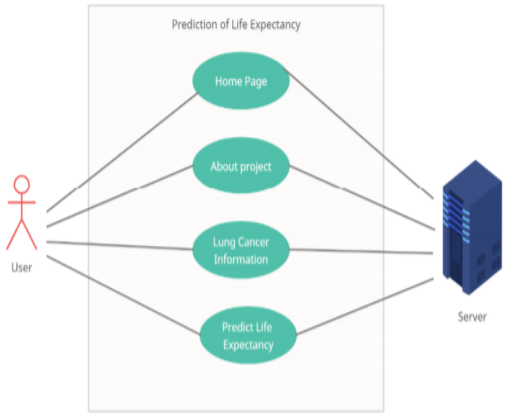

Fig.7.Use Case Diagram

The above figure specifies the Use Case diagram of the system where the diagram mainly contains two actors, one is User and the other is the Server. Both the actors can access all the functionalities provided by the system. We have user at the left and server at the right, In between we have four vowels that is Home page, About project, Lung Cancer Information, Predict Life Expectancy. Here each and every vowels are connected to both User and Server.

E. SEQUENCE DIAGRAM

A sequence chart contains time sequence arranged interactions between objects. It describes the objects involved in the frameworks as well as the sequence of messages between objects needed for the frameworks to function. Sequence diagrams generally are similar in the natural view of the system under development to the use case understanding. Sequence diagrams, event scenarios, are also called event diagrams. A sequence diagram shows different processes or objects living simultaneously in parallel vertical parallel lines and parallel arrows and messages change in their order of occurrence. This provides a graphical definition of simple runtime frames. This shows interaction of object arranged in the timely sequence. Basically it represents the objects which are involved in the scenario and sequence.

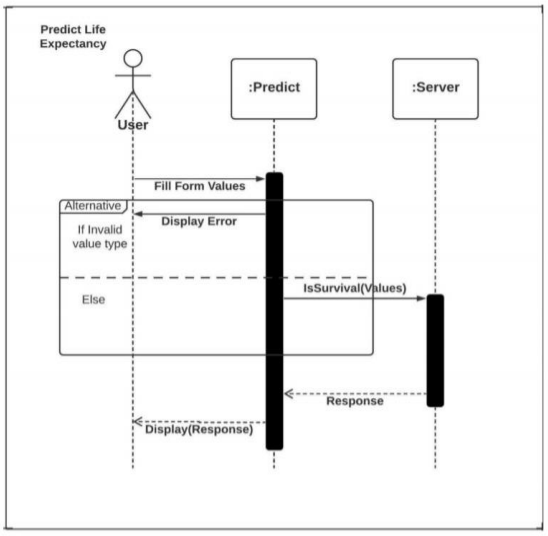

Fig.8.Sequence Diagram

The above figure specifies the sequence diagram of the Life Expectancy Post Thoracic Surgery module. This diagram specifies the timeline of different actors and objects and the sequence of steps undertaken to complete the functionality of the module.

F. ACTIVITY DIAGRAM

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 are 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.

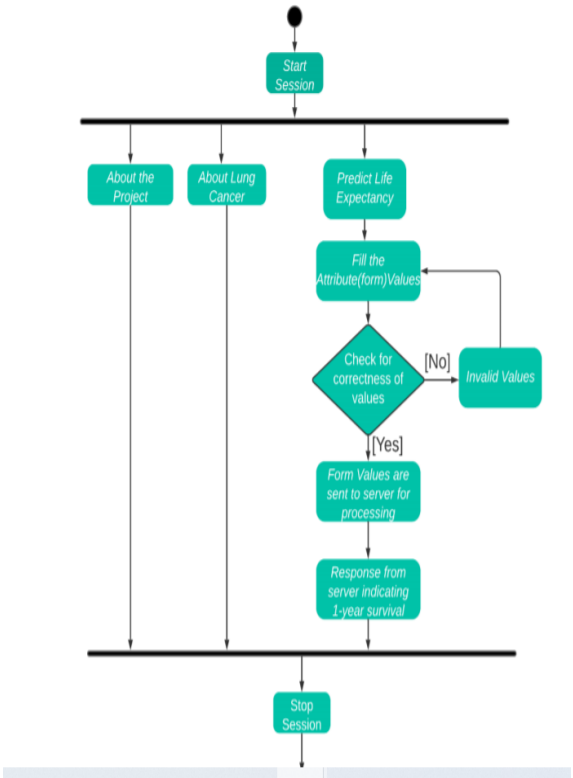

Fig.9.Activity Diagram

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.

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

Table 1: Confusion Matrix

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)

	Measures			
	Accuracy (%)	Precision (%)	Recall (%)	F-Score
Random Forest	81.80	83.09	81.60	0.82
LDA	83.76	82.59	83.76	0.83
SVC	78.43	81.50	78.43	0.80
Logistic Regression	82.60	80.54	82.61	0.81

Table 2: Performance Evaluation of Classifiers

It Specifies the four algorithms' accuracy, precision, recall, and F-Measure. The Random Forest, LDA, and SVC related supply Regressions were found to have accuracy of 82.6%, 83.76 percent, 78.43 percent, and 81.6 percent, respectively. As a result, it's evident that Linear Discriminant Analysis (LDA) provided the greatest accuracy of 83.76 percent with an F-Measure of zero.82 out of the four classifiers utilised.

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.

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.

X. 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).