

One-Year Life Expectancy Post Thoracic Surgery

1. INTRODUCTION

1.1 OVERVIEW

The introduction of computer applications into the medical enterprise has had an instantaneous effect on doctors' productivity and accuracy in current years. One of those programs is the examiner of fitness consequences. Health consequences are absolutely turning into a lot extra critical within side the shopping and control of healthcare. In maximum nations, most cancers are now one of the main reasons for mortality. Lung cancer is presently the maximum common indication for thoracic surgery.

Massive datasets of most cancers had been gathered and made to be had by clinical experts because of the development of the latest equipment within side the subject of medicine. The maximum tough challenge, however, is exactly

predicting a sickness outcome. As a result, modern-day studies make a specialty of using the system to gain knowledge of strategies to find out and outline fashions, in addition to relationships among them, from huge quantities of facts. The fact is analyzed to extract beneficial data that helps sickness prediction, in addition to enhancing fashions that expect healthcare consequences greater accurately.

1.2 PURPOSE

In order to improve quality initiatives, healthcare administration, and consumer education, it is critical to track health outcomes. Data from patients who have undergone extensive lung resection for primary lung cancer is called thoracic surgery. Attribute ranking and selection are important components for correctly predicting health outcomes when using machine learning algorithms. Researchers have used several techniques, such as early screening, to determine the type of cancer before symptoms appear. The most relevant attributes/characteristics are identified using attribute ranking and selection, and duplicate unwanted attributes/characteristics are removed from the dataset. The goal of our study is to look at patient mortality over the course of a year after

surgery. More precisely, we're looking into the patients' underlying health issues, which could be a powerful predictor of surgical-related mortality.

2. LITERATURE SURVEY

2.1 EXISTING PROBLEM

It has always been a difficult task to accurately predict the life expectancy post-operation. The prediction relies upon numerous fitness elements of which a few have a far more important function in comparison to the opposite elements. A famous approach used within side the beyond became to investigate the CT test photos of the lungs and expect primarily based totally at the everyday check-ups. The thirty-day mortality price is one statistic that has been used to estimate mortality charges within side beyond. This statistic, however, won't be absolutely correct due to the fact many sufferers die or come to be very frail right now after this time period, requiring them to be transferred to any other organization earlier than passing death. As a result, a large number of these deaths go unreported.

2.2 PROPOSED SOLUTION

To improve quality initiatives, healthcare administration, and consumer

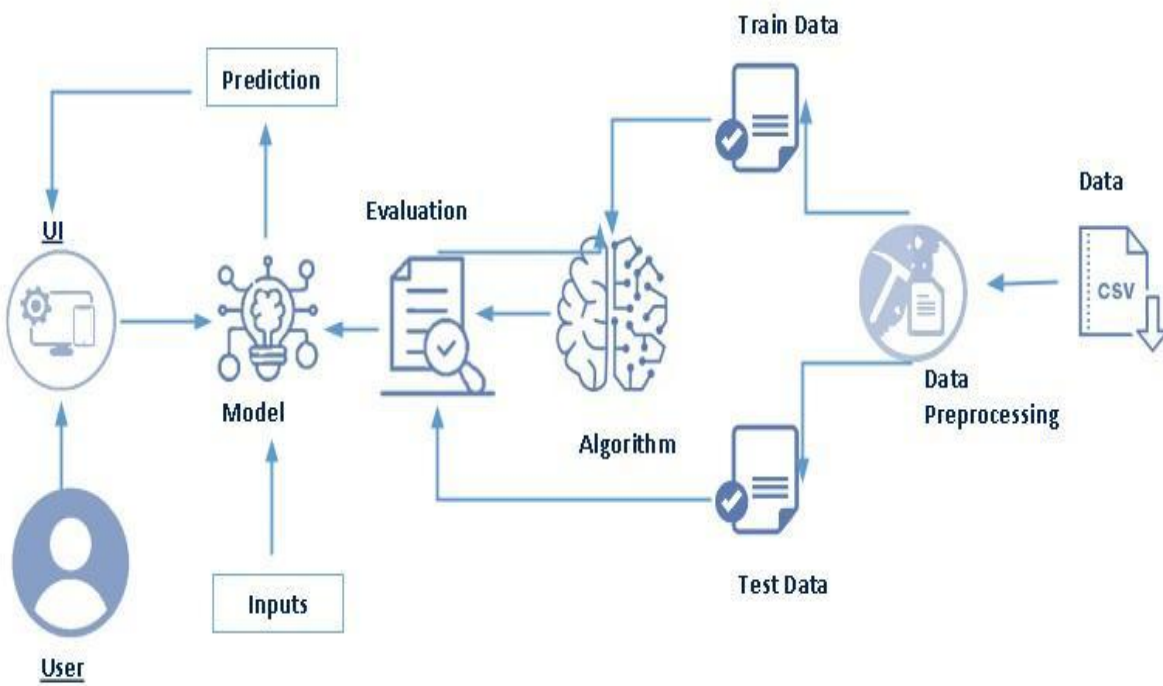
education, it is critical to track health outcomes. Data from patients who have undergone extensive lung resection for primary lung cancer is called thoracic surgery. Attribute ranking and selection are important components for correctly predicting health outcomes when using machine learning algorithms. Researchers have used several techniques, such as early screening, to determine the type of cancer before symptoms appear. The most relevant attributes/characteristics are identified using attribute ranking and selection, and duplicate unwanted attributes/characteristics are removed from the dataset. The goal of our study is to look at patient mortality over the course of a year after surgery. More precisely, we're looking into the patients' underlying health issues, which could be a powerful predictor of surgical-related mortality.

3.THEORETICAL ANALYSIS

Thoracic surgery is considered the consummating operation being performed on carcinoma patients. The survival rate is a key factor for surgeons to determine which patient surgery would be beneficially performed. Patient selection is one of the challenging factors in thoracic surgery decisions, taking into account parameters to determine risk-benefit considerations for the patient both in the short-term (e.g. post-operative complications, including death rate within the first

month) and long-term perspective (e.g. survival for 1-5 years).

3.1 BLOCK DIAGRAM



3.2 Hardware / Software designing

- Hardware Requirement:
- 2GB ram or above

- Keyboard
- CPU
- Dual-core processor
- Software Requirement:
- Anaconda Navigator
- Python Packages
- Open anaconda prompt as administrator

4.EXPERIMENTAL INVESTIGATIONS

The essence of machine learning is model fitting. The outcomes produced by the model will not be accurate enough to be used for actual decision-making if it doesn't fit the data appropriately. Hyper parameters in a correctly fitted model capture the complicated interactions between known factors and the target variable, allowing the model to identify useful insights and generate accurate predictions. The models used in this paper include Decision trees, KNN classifier, Logistic Regression. These models are used both for the classification with all features and classification with the best features

Based on the heatmap the correlation between the attributes can be observed

which help in selecting the attributes having a strong relationship to the output.

The below are the list of the few best attribute/features which increase the performance:

A. Algorithms

Algorithm fitting refers to how well a machine learning model is generalized to data comparable to trained data. A well-fitted model will produce more accurate output / results. The model can be overfitted or overfitted. The overfitting model fits the data too well, and the overfitting model does not fit the data properly.

Every machine learning algorithm/method has a set of basic parameters which can be tweaked to get an improved performance. During this fitting phase, a machine learning model is created by running the algorithm on the data for which one knows the target value, also called "labeled data". The correctness of the end result is then determined by evaluating them to actual, observed values of the target variable.

Then this data is used and tweaked the algorithm's normal settings to minimize the error and improve its accuracy in detecting anomalies and the relation b/w the target and the rest of its features. The procedure is done until the algorithm discovers the best settings for producing valid, practical, and usable insights for

your real-world business challenge.

The following are the algorithms used in this paper:

B. Decision Trees

A decision tree is a type of tree that is a method of supervised learning (that is, an input and the output corresponding to that particular input) in which the data is iteratively classified using specific parameters. It can be defined by two things: the decision area and the leaf. The final decision or termination of the tree is called a leaf. And the node is the decision to separate the functions.

A decision tree is a graphical representation of all solutions that are possible to a decision based on certain conditions. Tree models where the target variable can take a some set of values are known as classification trees and target variable can take continuous values are called regression trees. Decision tree is used as one of the algorithms in this project.

The example of decision tree could be easily explained by the use of the above binary tree. Suppose one wants to know if a person wants to be given a diet, exercise, etc. Decisions here are questions such as 'How old are you?', 'Do you exercise?', 'Have you eaten more no of pizzas'? etc. The end-nodes, which are effects such as 'suitable', or 'unsuitable'. In this case it is a binary split i.e, a

Yes/no type problem. There are mainly two different types of decision trees
Classification trees and Regression trees.

$P(Y/X)$ is approximated as sigmoid function which is applied to linear combination of different features.

Logistic Regression also can be used for multi class classification or for binary classification. In multi- class classification, It has more than one outcomes like some may have the everyday fever or flu, or normal cold or corona virus. Binary classification is when there are possible results like someone is infected with corona virus or isn't infected with corona virus.

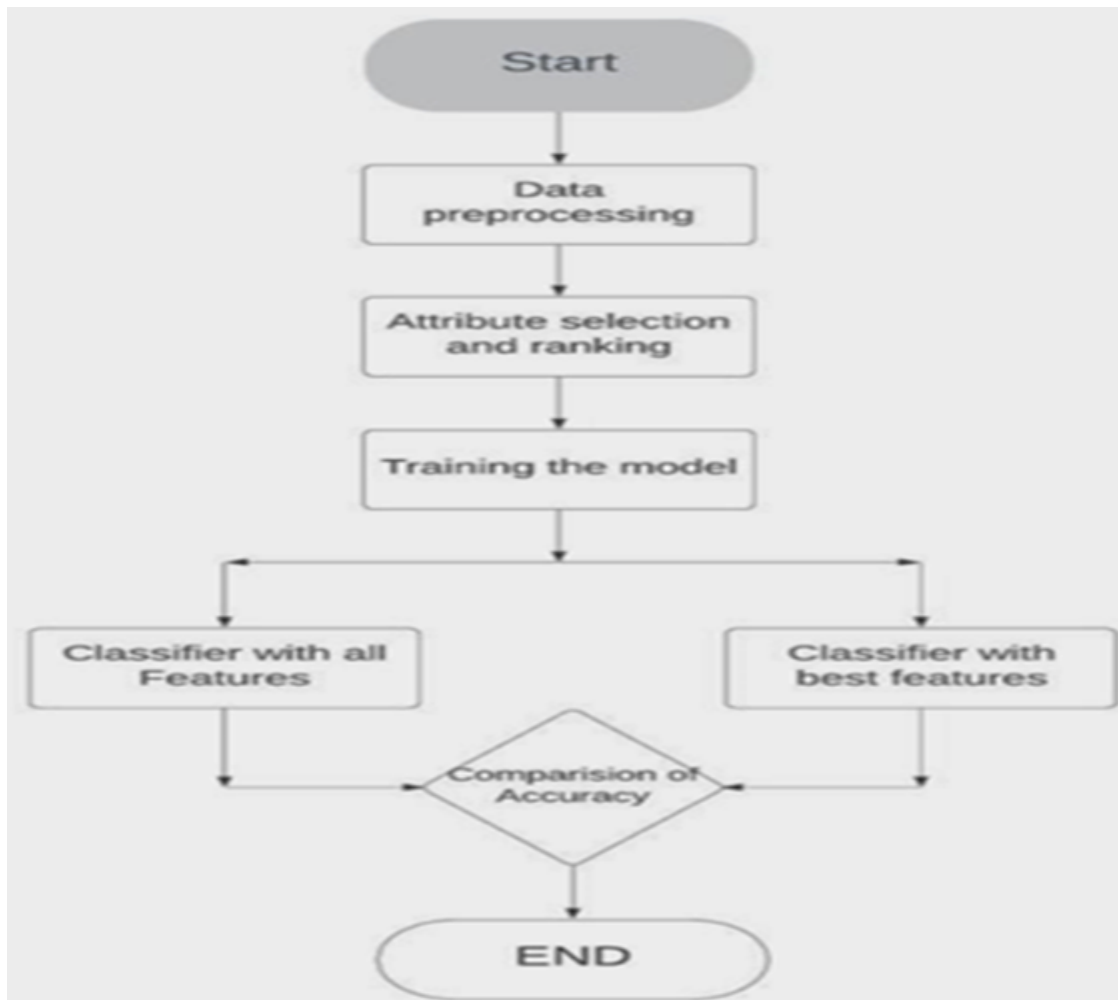
D. K-Nearest Neighbour Classifier

K-Nearest Neighbors Classifier (KNN) is the simplest algorithm used in machine learning for both classification and regression problems. The ANN algorithm takes data and classifies new points based on their similarity to the point (such as Euclidean distance). To do this, the distance is taken and summed with the nearest neighbor. The image below shows the classification of different classes based on distance..

In KNN classifier, 'K' is the no of nearest/closest neighbours. The no of neighbours is the most important aspect of the classifier. 'K' is usually is not

taken as an even number when the number of classes are two. When the value of K is one, then it is called as the nearest neighbour algorithm. It is the simplest case among all others.

5.FLOW CHART



6. RESULT

Input 1:

Thoracic Surgery Survival Status	
FVC	<input type="text" value="2.44"/>
FEV1	<input type="text" value="0.96"/>
Perfomance	<input type="text" value="2"/>
Pain	<input type="text" value="0"/>
Haemoptysis	<input type="text" value="1"/>
Dyspnoea	<input type="text" value="0"/>
Cough	<input type="text" value="1"/>
Weakness	<input type="text" value="1"/>
Tumor Size	<input type="text" value="1"/>
Diabetes Mellitus	<input type="text" value="0"/>

Weakness	<input type="text" value="1"/>
Tumor Size	<input type="text" value="1"/>
Diabetes Mellitus	<input type="text" value="0"/>
MI 6mo	<input type="text" value="0"/>
PAD	<input type="text" value="0"/>
Smoking	<input type="text" value="1"/>
Asthma	<input type="text" value="0"/>
Age	<input type="text" value="73"/>
<input type="button" value="Show Result"/>	

Output 1: Positive indicates that the patient can survive the Surgery.

Survival Status: Positive

Input 2:

Thoracic Surgery Survival Status

FVC

3.19

FEV1

2.5

Performance

1

Pain

0

Haemoptysis

0

Dyspnoea

0

Cough

1

Weakness

0

Tumor Size

1

Diabetes Mellitus

0

MI 6mo

0

PAD

1

Smoking

1

Asthma

0

Age

66

Show Result

Output 2: *Negative indicates that the patient cannot survive the Surgery.*

Survival Status: Negative

7. Advantages And Disadvantages

- Advantages

1. Easily identifies trends and patterns

Machine Learning can review large volumes of data and discover specific trends

and patterns that would not be apparent to humans. For instance, for an e-commerce website like Amazon, it serves to understand the browsing behaviors and purchase histories of its users to help cater to the right products, deals, and reminders relevant to them. It uses the results to reveal relevant advertisements to them.

2. No human intervention needed (automation)

With ML, you don't need to babysit your project every step of the way. Since it means giving machines the ability to learn, it lets them make predictions and also improve the algorithms on their own. A common example of this is anti-virus softwares; they learn to filter new threats as they are recognized. ML is also good at recognizing spam.

3. Continuous Improvement

As ML algorithms gain experience, they keep improving in accuracy and efficiency. This lets them make better decisions. Say you need to make a weather forecast model. As the amount of data you have keeps growing, your algorithms learn to make more accurate predictions faster.

4. Handling multi-dimensional and multi-variety data

Machine Learning algorithms are good at handling data that are multi-dimensional and multi-variety, and they can do this in dynamic or uncertain environments.

5. Wide Applications

You could be an e-tailer or a healthcare provider and make ML work for you.

Where it does apply, it holds the capability to help deliver a much more personal experience to customers while also targeting the right customers.

• **Disadvantages**

1 Data Acquisition

Machine Learning requires massive data sets to train on, and these should be inclusive/unbiased, and of good quality. There can also be times where they must wait for new data to be generated.

2. Time and Resources

ML needs enough time to let the algorithms learn and develop enough to fulfill their purpose with a considerable amount of accuracy and relevancy. It also needs massive resources to function. This can mean additional requirements of computer power for you.

3.Interpretation of Results

Another major challenge is the ability to accurately interpret results generated by the algorithms. You must also carefully choose the algorithms for your purpose.

4.High error-susceptibility

Machine Learning is autonomous but highly susceptible to errors. Suppose you train an algorithm with data sets small and exclusive. You end up with biased predictions coming from a biased training set. This leads to irrelevant advertisements being displayed to customers. In the case of ML, such blunders can set off a chain of errors that can go undetected for long periods of time. And when they do get noticed, it takes quite some time to recognize the source of the issue, and even longer to correct it.

8.APPLICATIONS

Customer Relationship Management (CRM) solutions often require the creation of customer level models that accurately predict buying patterns of customers, based on historical and current data. Or, you may want to determine the probability of a customer purchasing a product based on the price points. Analyzing such historical and current data and generating a model to predict the future outcomes of a product/service is termed as Predictive Modeling.

9.CONCLUSION

By the end of this project:

- You'll be able to understand the problem to classify if it is a regression or a classification kind of problem.
- You will be able to know how to pre-process/clean the data using different data preprocessing techniques.
- You will be able to analyze or get insights into data through visualization.
- Applying different algorithms according to the dataset and based on visualization.
- You will be able to know how to build a web application using the Flask Framework and other Frameworks like Streamlit.

10.FUTURE SCOPE

- This study does have few limitations. The results/outputs which were obtained are particular to a nation or an organization which collected the data set. Results obtained may be time-limited (2007–11).

- The dataset used as part of this project has very less records and may impede the accuracy of few algorithms that are used. In any case, this dataset can serve as a starting point to raise a better understanding of thoracic surgery patients. These tests can be further extended.
- This analysis only makes and uses three data mining methods. Therefore, some more machine learning methods could be used to get more knowledge about the data-set as a future work

11. BIBLIOGRAPHY

- Adam, A., Ivaylo, B., & Peng, J. (2014). Life Expectancy Post Thoracic Surgery. Retrieved from <http://cs229.stanford.edu/proj2014/Adam%20Abdulhamid,%20Ivaylo%20Bahtchevanov,%20Peng%20Jia,Life%20Expectancy%20Post%20Thoracic%20Surgery.pdf>
- American Medical Association. (n.d.). Thoracic Surgery Specialty Description. Retrieved from <https://www.ama-assn.org/specialty/thoracic-surgery-specialty-description>
WWW.GOOGLE.COM

