SURVIVAL RATE FOLLOWING THORACIC SURGERY

Mr.M.Narendra ,Assistant Professor, Department of Information Technology, CMR Engineering College, Hyderabad, Telangana,
E-Mail-id narendramupparaju06@gmail.com

D.Bhavyasri ,178R1A1210@cmrec.ac.in,
G.Raghuram ,178R1A1213@cmrec.ac.in,
P.Aditya Patel,178R1A1244@cmrec.ac.in

Abstract

Tracking health outcomes is essential for enhancing quality initiatives, healthcare management, and consumer education. Thoracic surgery refers to the collection of information from patients who underwent extensive lung resections for primary lung cancer. When utilising machine learning algorithms to predict health outcomes, attribute ranking and selection are essential elements.
Before symptoms occurred, researchers employed a variety of techniques, such as early-stage examinations, to identify the type of cancer. Utilizing attribute ranking and selection, the most pertinent attributes are found, and the redundant and extraneous attributes are eliminated from the dataset.

Before symptoms occurred, researchers employed a variety of techniques, such as early-stage examinations, to identify the type of cancer. Utilizing attribute ranking and selection, the most pertinent attributes are found, and the redundant and extraneous attributes are eliminated from the dataset.

Many machine learning models like SVM, naïve Bayes, decision tree, random forest, logistic regression have been applied for post thoracic surgery life expectancy prediction based on data set.

Keywords: Thoracic Surgery, Machine learning, Predict, Lung Cancer, Attribute Ranking.

1 Introduction

In recent years, the use of computer applications in the medical field has

directly affected the efficiency and precision of doctors work. Among these applications is the investigation of medical results.

Cancer is currently one of the top causes of death in the majority of countries. The most frequent procedure carried out on individuals with lung cancer is thoracic surgery.

As a result of the development of new medical techniques, enormous datasets on cancer have been gathered and made accessible to medical experts.

Many machine learning algorithms, including KNN, Logistic Regression, Random Forest, etc., are utilised to forecast survival rate following thoracic surgery.

When the lungs quit operating properly, thoracic surgery is performed. Lungs stop exchanging gases in an eloborated manner, which is definitely a death deal. The tiny organs called alveoli in the lungs are essential for the exchange of gases. When alveoli deteriorate or die, the septal cells likewise do so, forming a dead tissue that is typically referred to as a tumour. The tumour that causes lung cancer can be found via CT scans, which are a typical method for finding any form of abnormalities in people. The smoking criteria is one of our attribute in dataset.

2 Related Work

Lung cancer became so clear that we are now working on a project about it as a technological advancement. Despite being aware of how devastating cancer may be, people somehow constantly act carelessly and recklessly. Although there is no known cure for lung cancer, it can be prevented and avoided. After any type of thoracic surgery, postoperative respiratory issues are the leading cause of death.

In order to model cancer risk or patient outcomes, the given prediction models are based on a variety of supervised machine learning approaches, such as logistic regression and random forest.

Their findings showed that simple logistic regression technique has 81% prediction accuracy, which is superior than other machine learning techniques.

The breakthroughs in lung cancer prediction made possible by related and earlier investigations have given rise to these wonderful innovations. With the use of this technical advancement, we can predict a patient's lifespan.

A development in coding called machine learning has made pretty much everything simpler. Python is primarily used for machine learning programming. For implementation, we make use of Jupyter notebook or Jupyter lab. There are three different types of machine learning techniques: supervised, unsupervised, and semi-supervised. Each of these categories uses a separate algorithm. To mention a few: Naive Bayes theorem, Random Forest, KNN, SVM, decision trees, and Logistic Regression. In accordance with our needs, we select algorithms. We compare many algorithms and use the one that provides the maximum accuracy.

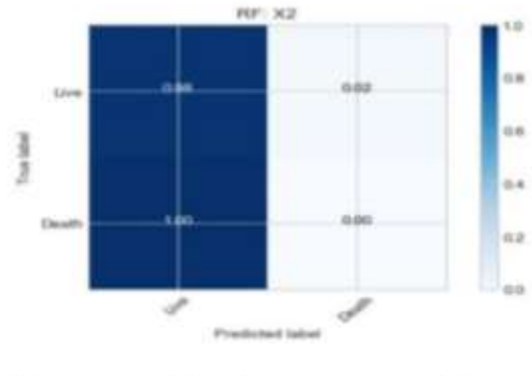
3 Existing System

The ability to precisely anticipate a patient's life expectancy after surgery has never been simple. The outcome is determined by a number of health-related variables, some of which are more important than others in terms of the prediction.

Analyzing lung CT scan images and making predictions based on routine check-ups were once a common technique.

We attempted to forecast life expectancy using existing techniques, but those predictions were inaccurate. Several cases were so unexpected that they gave sufferers false optimism. Yet modern developments have addressed.

Confusion Matrix:



There are numerous dimensions in a single data set for the attributes we use. To avoid or omit outliers, perform feature scaling on all the features. After feature scaling, we interpret graphs using the features we got by comparing each feature to every other feature. Finally, using these, we create a confusion matrix. In our scenario, we opted for 1 and 0 to represent death. This is a typical technique for giving numerical values to attributes in order to fill in the blanks in a data set.

4 Proposed System

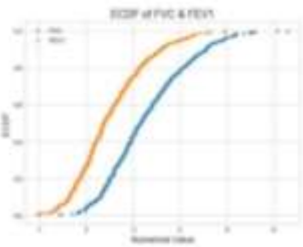
Tracking health outcomes is essential for enhancing quality initiatives, healthcare management, and consumer education .We eliminate some attributes that are unimportant for the prediction since the data is too large and complicated for the model to comprehend. Using attribute ranking and selection, the most important attributes are found, and the redundant and extraneous attributes are eliminated from the dataset.

Our model's objective is to examine patient mortality one year following surgery. More specifically, we're examining the patients' underlying health conditions, which may be an excellent indicator of mortality from surgery. We made significant progress in improving the accuracy of life expectancy projections using the methodologies already in use. Along with removing undesirable qualities, we also took feature mean values into account. As they don't include outliers, mean values are more accurate than singular values. This is what distinguished this approach from the prior one.

5 Research Methodology

On a computer running Microsoft Windows 10 Professional and powered by an Intel(R) CoreTMi5 CPU clocked at 2.30 GHz, we conducted our tests (SP2). We divide the data set into two sections of various sizes. 80% of the subsets are used for the analysis (training), while the remaining 20% are used for further analyses (testing).

Correlation:



These approaches' main goal is to eliminate pointless attributes from the initial list of attributes. In our study, we employ Information Gain (IG) attribute evaluation and attribute ranking techniques. By calculating the Information Gain with regard to the class, Information Gain (IG) Attribute Assessment determines the importance of a given attribute. Entropy, a measure of the system's randomness, forms the basis of IG. The following equation can be used to calculate information gain:

IG(Class,Attribute)=H(Class)−H(Class|Atribute)

where H represents entropy.

Estimating that a relevant attribute should have a different value for cases that belong to distinct classes and the same value for cases that belong to the same class is challenging. In order to solve the problem of predicting life expectancy after thoracic surgery, the objective of this work is to investigate the impact of a number of factors on the precision of machine learning approaches. It is necessary to minimise the amount of features and increase accuracy in order to speed up the computational process of prediction approaches. After implementing the three ranking approaches for life expectancy prediction for post-thoracic surgery, we evaluated the nature of the techniques Simple Logistic Regression, Random Forest, and KNN. In this study, we employed information gain to reduce the number of characteristics. By contrasting the performance of Simple Logistic Regression, Random Forest, and KNN with and without employing attribute ranking methods beforehand, the quality of the proposed approaches is determined.

6 Conclusion

To sum up, we found that simple logistic regression had the highest accuracy when compared to the other two methods, random forest and KNN. Our accuracy rate was 85%. Even after making all these forecasts, the selection process always comes down to one primary factor: the level of personal attention.

We used the KNN, Random Forest, and Logistic Regression algorithms, and we recalculated each process with and without weights. We improved accuracy by straightforward logistic regression by adding weights.

Comparing the accuracy of Random forest versus straightforward logistic regression is the goal of using Random forest. Almost 2% of the precision produced by Random forest. While Random Forest's precision varies during the trial, that of Logistic Regression remains constant.

7 References

[1] KwetisheJoro Danjuma. (2015), "Performance evaluation of machine learning algorithms in post-operative life expectancy in the lung cancer patients", IJCSI International Journal of Computer Science Issues, Vol. 12, No. 2, pp.189-199

[2] Pinar Yildirim. (2015), "Filter based feature selection methods for prediction of risks in hepatitis disease", International Journal of Machine Learning and Computing, Vol. 5, No. 4,pp. 258-263.

[3] KonstantinaKourou , Themis P. Exarchos, Konstantinos P. Exarchos, Michalis V. Karamouzis and Dimitrios I. Fotiadisa. (2015), "Machine learning applications in cancer prognosis and prediction", Computational and Structural Biotechnology Journal, Vol. 13,pp.8-17.

[4] V. Sindhu, S. A. S. Prabha, S. Veni and M. Hemalatha. (2014), "Thoracic surgery analysis using datamining techniques", International Journal of Computer Technology & Applications, Vol. 5 pp.578-586.

[5] Samina Khalid, TehminaKhalil and ShamilaNasreen. (2014), "A survey of feature selection and feature extraction techniques in machine learning", Science and Information Conference(SAI).

[6] Mehdi Naseriparsa, Amir-Masoud Bidgoli and Touraj Varaee. (2013), "A hybrid feature selection method to improve performance of a group of classification algorithms", International Journal of Computer Applications,Vol. 69, No. 17,pp.28-35.

[7] Joseph A. Cruz, David S. Wishart. (2006), "Applications of machine learning in cancer prediction and

prognosis", Cancer Informatics, Vol. 2, pp.59-77 2006.