National Conference on Advances in Science, Technology, Engineering, and Management

Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Life Expectancy Post Thoracic Surgery Using Machine Learning

Chandana VS[1], Mr. Roshan Kumar L[2], Ms. Nishmitha R[3], Ms. Nithya A[4], Ms. Nandini V[5]

[1] Assistant Professor, Dept of CSE, KSSEM, Bangalore
Email: chandanavs@kssem.edu.in

[2] Student, Dept of CSE, KSSEM, Bangalore
Email: vnrk24@gmail.com

[3] Student, Dept of CSE, KSSEM, Bangalore
Email: nishmitharaju19@gmail.com

[4] Student, Dept of CSE, KSSEM, Bangalore
Email: nithya24a@gmail.com

[5] Student, Dept of CSE, KSSEM, Bangalore
Email: nandiniv0406@gmail.com

****

Abstract:

The purpose of this project is to determine a life expectancy following thoracic surgery while accounting for the significance of numerous factors that may affect the outcome. The dataset consisted of patient data collected at the time of diagnosis. To better understand the effects of post-surgery, a variety of parameters that influence the outcome have been examined with the aid of random forest and decision tree algorithms. Specific metrics have. Furthermore, there are a number of additional factors that we use for classification, like the patient's smoking status, the existence of asthma, haemoptysis, coughing prior to surgery, and a few more. With improved data feature selection, our classification algorithm forecasts the patient's likelihood of survival with a risk factor of one year.

Keywords — ***Thoracic surgery, Machine learning, Random Forest, Decision trees.***

****

I. INTRODUCTION

Adult cardiac surgery, also known as cardiothoracic surgery, is sometimes confused with thoracic surgery. For people who have set aside time for this procedure, general thoracic surgery should be performed instead. About 80 percent of thoracic surgeries involve some form of cancer treatment. Gathering comprehensive clinical information on Analyzing the aetiology of the disease is a crucial skill that medical professionals must analyze and impart in any hospital environment. However, physicians were unable to locate a thorough examination of the numerous individual medical records since traditional medical records of patients were kept on paper. As said, our feature set consists of continuous data and classification based on the state of the patient's health at the time of surgery.

II. LITERATURE SURVEY

Tomohiro Kawahara, Member, Yoshihiro Miyata, Koichi Akayama, Masazumi Okajima, and Makoto Kaneko, Fellow, IEEE "Design of Noncontact Tumor Imager for Video-Assisted Thoracic Surgery" , DEC 2017

With forceps and an endoscopic camera, a surgeon performs a procedure on a patient through tiny incisions in the chest during video-assisted thoracic surgery (VATS). Lung malignancies are often removed by VATS. Because the surgical wound is tiny, the primary benefit of VATS is that patients recover more rapidly. Prior to surgery, computed tomography (CT) can be used to accurately identify the location of lung cancers. The lung's internal pressure drops during an incision, which causes the lung capacity to decrease. causes the sufferers discomfort, therefore it is challenging to score

several points. When identifying tumors during an operation, surgeons frequently employ the same two-contact probes, which are made of a metal stick with a cotton tip, to minimize causing discomfort to their patients. Surgeons can use forceps to feel the location of the tumor using tactile senses.

Huan-Yu Chen , Hui-Min Wang and Chi-Chun Lee , "Lung Cancer Prediction Using Electronic Claims Records: A Transformer-Based Approach", DECEMBER 2020

For the purpose of recording comprehensive information about a patient's diagnostic/intervention status and results, electronic medical records, or ECRs, are large-scale, longitudinal collections of individual medical service seeking actions. EMRs consist of patient status, biological monitoring data, lab examination results, specimen interpretation, treatments and medication, etc. Since EMRs are a non-intrusive modality that has been collected regularly, EMR datasets are often large-scale in nature, especially satisfying the data-hungry nature of deep learning models. The infrastructure used to host the EMRs is frequently restricted to one hospital or a small number of sites, which isolates the systems within the hospitals and makes cross-site collection challenging. Larger and more uniform data sample collections are made possible by ECRs; yet, claims data that describe therapeutic actions for reimbursement purposes may be skewed toward financial gain.

Truman Cheng , Weibing Li , Calvin Sze Hang Ng, Philip Wai Yan Chiu, and Zheng Li "Visual Servo Control of a Novel Magnetic Actuated Endoscope for Uniportal Video-Assisted Thoracic Surgery" JULY2019

For uniportal video assisted thoracic surgery (VATS), a new magnetically actuated endoscope has been developed. It can be used in limited spaces because it is small and has a workspace that is close to the chest wall. Instrument fencing and port crowding are issues with uniportal VATS, which entails passing an endoscope and several instruments through a single incision. Patient side crowding and staff meddling are other issues with the uniportal technique. Robotic endoscope holders that follow the surgeon's instructions directly and take the place of the human helper can all help reduce these. Currently available control techniques include voice commands, foot pedals, joysticks, head tracking, and body position. A magnetic endoscope is inserted into the patient, and the endoscope is then anchored inside the abdominal wall using an external magnet. This allows the surgeon more freedom to move the camera by substituting a magnetic coupling for the stiff link.

Lal Hussain , Wajidazizi, Abdulrahman A Alshdadi, Malik Sajjad Ahmed Nadeemi, Ishtiaq Rasool Khan Analyzing the Dynamics of Lung Cancer Imaging Data Using Refined Fuzzy Entropy Methods by Extracting Different Features May 2019

Due to inadequate diagnostics at the advanced stage of the disease, lung cancer is the leading cause of cancer-related deaths globally and has a dismal prognosis. A few techniques for extracting features were employed by radiologists to diagnose problems using computer aided diagnosis (CAD) systems that were created in the past. Nonlinear dynamical measurements are the most effective way to study the physiology and behavior of different physiological systems because they capture the intrinsic dynamics that are affected by the deterioration of structural and functional components resulting from many diseases. These dynamical techniques are the most effective means of analyzing hidden information found in cancer images. In this work, we introduced multiscale fuzzy entropy (MFE), multiscale permutation entropy (MPE), multiscale sample entropy (MSE) with mean and KD-tree algorithmic approach, and reweighted composite multiscale fuzzy entropy (RCMFE) with mean, variance, and standard deviation.

Jason L. Causey, Keyu Li , Xianghao Chen, Wei Dong, Karl Walker, Jake A. Qualls Spatial Pyramid Pooling With 3D Convolution Improves Lung Cancer Detection Mar 2022

Lung cancer typically presents at an advanced stage, and the disease's 5-year survival rate is only 18%. Early identification is therefore essential to increasing survival through intervention. Low-dose CT can offer more precise information than radiography and has been shown to result in a 20 percent reduction in death rates. The US Preventive Services Task Force recommends low-dose CT for lung cancer screening. Enhancing the existing clinical practice of CT imaging assessment requires the development of potent computer-aided techniques for early lung cancer screening.

Automated solutions for early lung cancer screening and a lower false positive rate in diagnosis are the goals of computer-aided methods. In the last fifty years, many computer-aided techniques for the analysis of chest images have been created. Finding and analyzing nodules in lung CT scans has been the primary focus of most computational methods to date.

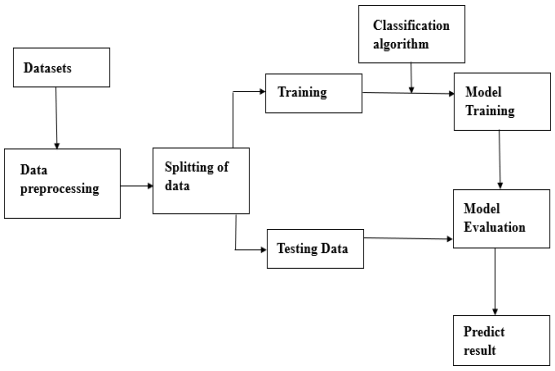
III. PROPOSED SYSTEM

A. Problem Statement

Nowadays, skilled medical personnel identify cancer; these personnel frequently have to go through numerous cases that are identical to one another before determining the appropriate diagnosis or patient's class. Handwritten recognition is a laborious procedure with a difficult to measure subjective element. the creation and application of a reliable and accurate machine learning platform that can be used to forecast the likelihood of lung cancer in its early stages.

Creating a predictive model that uses patient data, surgical parameters, and other pertinent factors to accurately estimate the expected lifespan following thoracic surgery is the problem statement for life expectancy prediction for post-thoracic surgery using machine learning. This will help clinicians make informed decisions and provide individualized care, plans for postoperative care. It also forecasts a lung infected person's life expectancy following thoracic surgery. This will direct the doctors in determining whether or not a patient with lung cancer can be treated with medicine in lieu of thoracic surgery.

The processing methods used are standardization and label encoding. The random decision forest algorithm is used as a Classifier.



The foundation of any web application design is its system architecture. Fig. 1 shows the architecture. The information is gathered from the impacted parties' prior records individuals. The cleaned data is extracted after just the pertinent information is culled from the earlier records. The data was divided into training and test sets using the feature engineering technique. The model is tested using test data. The learning algorithm is used with training data to train the model. The trained data is used to validate the trained model. After that, the model is assessed and predicts the outcome.

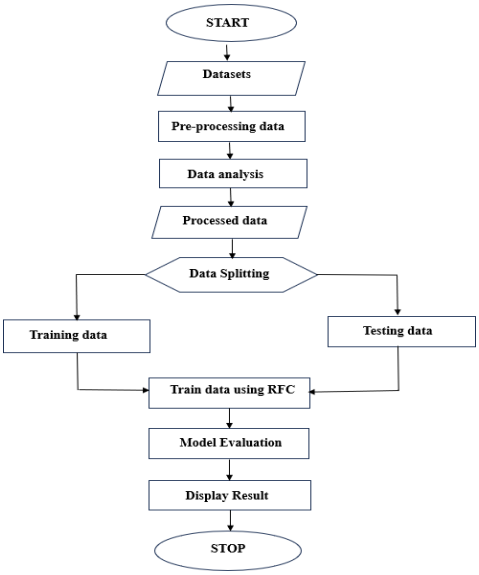


Fig.2 Flow diagram of Random forest technique

It is vital to complete all tasks and attain the boundaries. The system design is based on the flow diagram. For the purpose of scheduling and to process, a flowchart is utilized. The primary benefit of utilizing a flowchart is that it shows every operation that is tangled up in a system inside the linear value chain. The flowchart illustrates the disease prediction.

The sequence of actions that represent one or more inputs and transform them into outputs is depicted in the flowchart. Our project's flowchart is shown in Fig. 2. The lungs dataset is derived from the individuals' historical data.

The feature analysis is used to process the data. The data that has been processed is A data splitting procedure is used to divide an into training and test data. The random forest method is used by the training dataset to train the model. The model is validated using the test dataset. Lastly, the outcome is shown, along with its low, middle, and high scores.

B .Implementation

Fig.3 Level-0 model

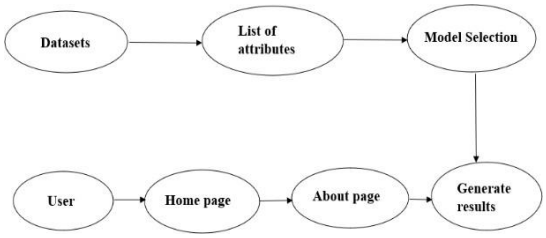


Fig.4 Level-1 model

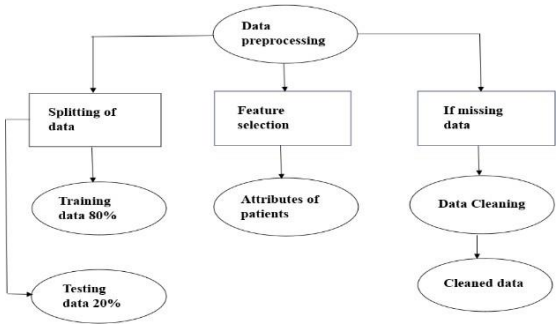


Fig.5 Level-2 model

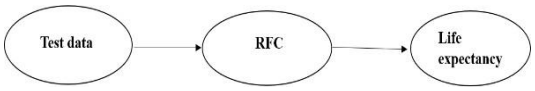


Fig.6 Level-3 model

The model's implementation is shown above. The model's levels are displayed in these diagrams. The model retrieves the data from the data warehouse at level 0. The user has the ability to transmit the model's queries that predict the outcomes as displayed in figure 4.1. Level 2 is created by combining Levels 0 and 1.the division of the cleansed data into two percents, 0.80 and 0.20. The divided data that was utilized for verification and training use.

C. Random Forest Algorithm

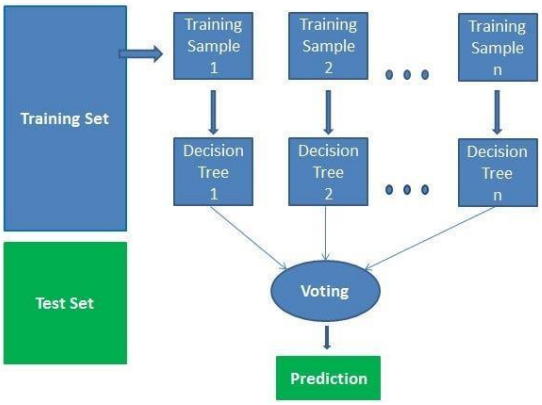


Fig.7 Design of random forest algorithm

The following steps explain the working Random Forest Algorithm :

Step 1: From a given training set or data, choose random samples.
Step 2: For each training set of data, this algorithm will build a decision tree.
Step 3: The decision tree will be averaged to determine the winner.
Step 4: Choose the predicted result that received the most votes to be the final outcome.

Implementation Steps are given below:

- Pre-processing phase for data.
- Aligning the Training set with the Random Forest algorithm.
- Projecting the test's outcome.
- The result's accuracy in testing.
- Presenting the test set result visually.

Benefits of Random Forest:

i. Tasks involving both classification and regression can be completed using Random Forest.
ii. Large datasets with high dimensionality can be handled by it.
iii. It keeps the overfitting problem at bay and improves the model's accuracy.

IV .RESULTS AND DISCUSSION

Thus, the model can vary in its accuracy when predicting the patient's life expectancy over a span of one year . After the patient data is input, with each patient's unique properties specified, the accuracy of the model's prediction is used to estimate each patient's life expectancy. According to the doctor's recommendation, surgery is only advised for patients whose cancer is in the third or fourth stage, at which point their prognosis can only be estimated with the risk factor of one year. Patients in the first and second stages of the disease, on the other hand, receive chemotherapy or radiation therapy and have a higher chance of survival. Once the model has been trained and test data has been utilized for the prediction, the accuracy is estimated using random forest. This prediction is then used to forecast the life expectancy once more.

V. CONCLUSION

Lung cancer is now regarded as one of the most common diseases. Manual recognition is a laborious procedure with a degree of subjectivity that is measured. This study focuses on the development and application of an accurate and efficient machine learning platform that may be used to forecast the likelihood of lung cancer in its early stages. Future improvements to this system could yield a more accurate model using the Random Forest method in conjunction with KNN and Naive Bayes. Rather than using repositories, datasets directly from hospitals and other agencies can be used to automate the diagnosis of lung cancer. It is possible to use algorithms like CNN, ANN, and AI. Whether a patient has non-small cell lung cancer (NSCLC) or small cell lung cancer (SCLC), the survival rate after surgery is generally poor. Therefore, before recommending surgery, a thorough evaluation and analysis based on past patient historic data and the patient's medical state are required. Additional aspects to be taken into account are the surgeon's experience and age, among others. Operative life span must be studied independently.

This can be achieved by creating a mobile application that allows for patient interaction and the development of an intelligent, even self-automated, recommendation system that helps patients lead happier, less disruptive lives and adopt healthier lifestyles for the years they may still have left. Following the application of multiple statistical tests and Random Forest classification—one on the complete set of features, the other on the ten most significant features—the latter method yielded a higher mean AUC score of 0.66 than the initial one of 0.55. Therefore, it is advantageous to estimate the key variables for patient risk classification following thoracic surgery

ACKNOWLEDGMENT

We are grateful to Mrs. Chandana VS, Assistant Professor, for serving as our project guide and for her capable leadership in making this project work a success.

REFERENCES

[1] Adam, A., Ivaylo, B., & Peng, J. (2014). "Life Expectancy Post Thoracic Surgery".

[2] American Medical Association. (n.d.). "Thoracic Surgery Specialty Description".

[3] Desuky A. El Bakrawy L. (2016). Improved prediction of post-operative life expectancy after Thoracic Surgery. Advances in Systems Science and Applications, 16(2), 70–80.

[4] Kokulu, M., Kahramanli, H., & Allahverdi, N. (2015)."Applications of Rule Based Classification Techniques for Thoracic Surgery".

[5] Nachev A. Reapy T. (2015). "Predictive Models for Post-Operative Life Expectancy after Thoracic Surgery". Mathematical and Software Engineering, 1(1), 1–5.

[6] Sarna L. Cooley M. Brown J. Chernecky C. Kotlerman J. (2008). "Symptom severity one to four months post-thoracotomy for lung cancer".

[7] Sindhu V. Sathya Prabha S. A. Veni S. Hemalatha M. (2014). "Thoracic Surgery Analysis Using Data Mining Techniques".

[8] Zieba M. Tomczak J. Lubicz M. Swiatek J. (2014). "Boosted SVM for extracting rules from imbalanced data in application to prediction of the post-operative life expectancy in the lung cancer patients". Applied Soft Computing, 14, 99–108. 10.1016/j.asoc.2013.07.016

[9] Chence Zhang, Yi Shen, Yucheng Wei, Bingsen Zhang and Qian Kong, "Research on the Design of a General Thoracic Surgery Clinical Database System for Evaluating Operation Risk", 2010 3rd International Conference on Biomedical Engineering and Informatics BMEI, 2010.

[10] UCI Machine Learning Repository, [online] Available:

[11] Leo Breiman, "Random Forests". [9] A. Mark, "Hall Correlation-based Feature Selection for Machine Learning".

[12] "UCI ML Repository Thoraric Surgery Data", [online] Available:

[13] Md. A U Harun and Md. N Alam, "Predicting Outcome of Thoracic Surgery by Data Mining Techniques", IJARCSSE, vol. 5, no. 1, January 2015 Sindhu V.SAS Prabha, S Veni and M Hemalatha, "Thoracic Surgery Analysis Using Data Mining Techniques", IJCTA, vol. 5, no. 2, 2017

