International Journal of Engineering Technology and ManagementSciences
Website: ijetms.in Issue: 3 Volume No.6 May – 2022
DOI:10.46647/ijetms.2022.v06i03.007 ISSN: 2581-4621

LUNG CANCER PREDICTION USING DECISION TREE ALGORITHM WITH FLASK FRAMEWORK

T.Aarthi[1],K.Bheemesh[2],N.Raju[3],R.Sruthi[4],A.Divya[5],andT.Swathi[6]

[1]Department of Computer Science and Engineering, ACE Engineering College, Ankushapur, India
[2]Department of Computer Science and Engineering, ACE Engineering College, Ankushapur, India
[3]Department of Computer Science and Engineering, ACE Engineering College, Ankushapur, India
[4]Department of Computer Science and Engineering, ACE Engineering College, Ankushapur, India
[5]Department of Computer Science and Engineering, ACE Engineering College, Ankushapur, India
[6]Department of Computer Science and Engineering, ACE Engineering College, Ankushapur, India

aarthisweety49@gmail.com kbheemesh113@gmail.com
rajunandala2@gmail.com rachurisruthi100@gmail.com
divyaagraram@gmail.com swathiturai12@gmail.com

Abstract: The objective of the project is to make Lung Cancer Prediction using Decision Tree Algorithm. A variety of these techniques, including Artificial Neural Network(ANNs), Support Vector Machines(SVMs) and Decision Trees(DTs) have been widely applied in cancer research for the development of predictive models, resulting ineffective and accurate Decision making. The main idea is to create a novel method using Decision tree technique for cancer prediction.

Keywords: Artificial Neural Network. Support Vector Machine. Decision Tree. Cancer prediction.

1 Introduction

Lung cancer is of disease of abnormal cells multiplying and growing into a tumour. Cancer cells can be carried away from the lungs in blood, or lymph fluid that surrounds lung tissue. Lymph flows through lymphatic vessels, which drain into lymph nodes located in the lungs and in the center of the chest. Lung cancer often spreads toward the center of the chest because the natural flow of lymph out of the lung is towards the center of the chest. Metastasis occurs when a cancer cell leaves the site where it began and moves into a lymph node or to another part of the body through the bloodstreams[1].
Cancer that starts in the lung is called a primary lung cancer. There are various types of lung cancer One is **Benign Tumour** and other one is **malignant** in which benign Tumour is non-Cancerous and the malignant is a cancer Tumour. The most dangerous disease in the world is cancer in which lung cancer is dangerous for women and men. We create a model for classification of symptoms. This model is trained with datasets in 7:3 ratio .We use Decision Tree Algorithm for creating model which gives greater accuracy compared to the machine learning algorithms.

2 Literature Survey

An extensive search was conducted relevant to the use of ML techniques in cancer susceptibility, recurrence and survivability prediction. The majority of these studies use different types of input data: genomic, clinical, histological, imaging, demographic, epidemiological data or combination of these. Papers that focus on the prediction of cancer development by means of conventional statistical methods (e.g., chi-square, Cox regression) were excluded as were papers that use techniques for tumour classification or identification of predictive factors.

Website: ijetms.in Issue: 3 Volume No.6 May – 2022
DOI:10.46647/ijetms.2022.v06i03.007 ISSN: 2581-4621

Maciej Zi¸eba∗, Jakub M. Tomczak, Marek Lubicz, Jerzy ´Swi¸atek, [2], They have proposed a novel boosted SVM method for imbalanced data problem which was further used for rules extraction. They have evaluated the quality of the approach by comparing it with other solutions dedicated for imbalanced data problem. The method is to solve the problem for prediction of the post-operative life expectancy in the lung cancer patients. They have shown that our approach can be successfully applied to the problem by making additional experimental comparison on real-life data.

Tae-WooKimetal .[3],have developed a decision tree on occupational lung cancer. In 1992–2007,153 lung cancer cases were reported by the Occupational Safety and Health Researcher's Institute (OSHRI). The objective parameter was to determine if the situation was accepted as lung cancer linked to age, sex, smoking years, histology, industry size, delay, working time and exposure of independent variables. During the whole journey for indicators for word related cellular breakdown in the lungs the characterization and relapse test (CART) worldview is utilized. Presentation to Known lungs disease specialists was the best pointer of the CART model. As the CART model is not absolute, the functionality of lung cancer must be carefully determined.

In[4]Deep Convolutional Neural Network CNNs is used to identify or label a medical image in some research papers. Diagnosed lung cancer in 2015 with a multiscale two- layer CNN [5] recorded 86.84% accuracy in [6] the CNN architecture, data set characteristics, and transfer learning factors were exploiting and extensively analyzing three significant and previously under studied factors.

N. Picco, R.A. Gatenby, A.R.A. Anderson [7] SVM with Artificial Neural
Networks and Decision-making Trees is identified in this case as the precision predictor(92.85% accuracy). Prostate cancer survival is also examined in context, including artificial neural networks, decision trees and logistical regression. In the segment, data on patients suffering from colon cancer were compared to predict survival and more accurate neural networks were determined.

3 Existing System

Ms. Leena Patil [8], the typical information used by the physicians conclude with a reasonable decision regarding cancer prognosis and included histological, clinical and population-based data. The existing system is created to make the early detection of lung cancer as automated process.

Even though, it is detecting the results accurately. It is not a User-friendly process. User cannot interact directly with the model.

4 Proposed System

The proposed system predicts Lung cancer respectively. We have proposed this cancer prediction system based on data mining techniques. This system is validated by comparing its predicted results with patient’s prior medical information. We have compared algorithms like Logistic Regression, SVM, Random Forest and Decision Tree for predicting the cancer. Among all the algorithms, Decision Tree Algorithm gives high accuracy.

Our main aim of the project is to create an user-interface using Flask Framework for earlier prediction of Lung Cancer. It is User-friendly and cost-efficient.

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 3 Volume No.6 May – 2022
DOI:10.46647/ijetms.2022.v06i03.007 ISSN: 2581-4621

Decision Tree Algorithm

It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome. It is a graphical representation for getting all the possible solutions to a problem /decision based

on given conditions**.** It is called a decision tree because, similar to a tree, it starts with the root node, which expands on further branches and constructs a tree like structure[9].

In a decision tree, for predicting the class of the given dataset, the algorithm starts from the root node of the tree. This algorithm compares the values of root attribute with the record (real dataset) attribute and, based on the comparison, follows the branch and jumps to the next node[9].

For the next node, the algorithm again compares the attribute value with the other sub nodes and move further. It continues the process until it reaches the leaf node of the tree [9].

5 Workflow

Load Dataset: We collected dataset related to lung cancer from the Kaggle whichhasatotalof310recordsin it. We have used panda's library to read the dataset.

Data Visualization: We represent the information and data in a pictorial or graphical form. It helps us to understand trends, patterns in data. We have used sea born and matplotlib algorithms for creating bar graphs for better understanding.

Feature Selection: It selects the most relevant features from the original data by removing the redundant, irrelevant or noisy features. By using bar graphs, we are concluding that all the attributes are relevant features.

Exploratory analysis of Features: It is a part of Feature Selection. It forms hypotheses based on the understanding of the dataset. We used NumPy and pandas for analysis.

Train Test Splitting: We split the modelling dataset into training and testing sets in the ratio of 70:30. We train the model using the training set and the apply the model to the test set.

Model Selection: It chooses one model among many models for a training dataset.

Prediction: It predicts output, whether the person has cancer or not.

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 3 Volume No.6 May – 2022
DOI:10.46647/ijetms.2022.v06i03.007 ISSN: 2581-4621

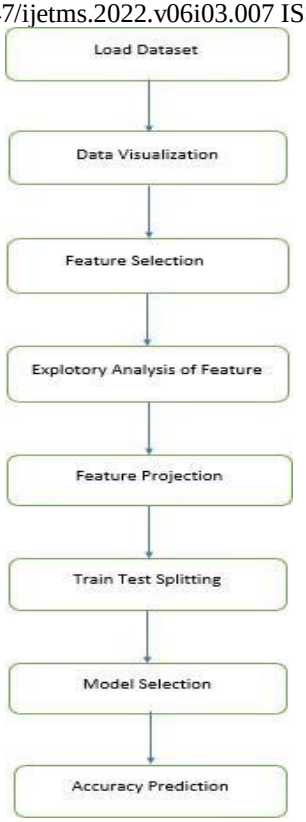

Fig.1. Architecture of Lung Cancer Prediction

Below is the dataset which is used in our project collected from Kaggle. It consists of total 310 records but we have placed only few of them for sample.

	GENDER	AGE	SMOKING	YELLOW_F	ANXIETY	PEER_PRE	CHRONIC	FATIGUE	ALLERGY	WHEEZINC	ALCOHOL	COUGHIN	SHORTNES	SWALLOW	CHEST PAI	LUNG_CANC
2	1	69	0	1	1	0	0	1	0	1	1	1	1	1	1	1
3	1	74	1	0	0	0	1	1	1	0	0	0	1	1	1	1
4	0	59	0	0	0	1	0	1	0	1	0	1	1	0	1	0
5	1	63	1	1	1	0	0	0	0	0	1	0	0	1	1	0
6	0	63	0	1	0	0	0	0	0	1	0	1	1	0	0	0
7	0	75	0	1	0	0	1	1	1	1	0	1	1	0	0	1
8	1	52	1	0	0	0	0	1	0	1	1	1	1	0	1	1
9	0	51	1	1	1	1	0	1	1	0	0	0	1	1	0	1
10	0	68	1	0	1	0	0	1	0	0	0	0	0	0	0	0
11	1	53	1	1	1	1	1	0	1	0	1	0	0	1	1	1
12	0	61	1	1	1	1	1	1	0	1	0	1	1	1	0	1
13	1	72	0	0	0	0	1	1	1	1	1	1	1	0	1	1
14	0	60	1	0	0	0	0	1	0	0	0	0	1	0	0	0
15	1	58	1	0	0	0	0	1	1	1	1	1	1	0	1	1
16	1	69	1	0	0	0	0	0	1	1	1	1	0	0	1	0
17	0	48	0	1	1	1	1	1	1	1	0	1	1	1	0	1
18	1	75	1	0	0	0	1	0	1	1	1	1	1	0	1	1
19	1	57	1	1	1	1	1	0	0	0	1	0	0	1	1	1
20	0	68	1	1	1	1	1	1	0	0	0	1	1	0	0	1
21	0	61	0	0	0	0	1	1	0	0	0	0	1	0	0	0
22	0	44	1	1	1	1	1	1	0	0	0	0	1	1	0	1
23	0	64	0	1	1	1	0	0	1	1	0	1	0	1	0	1
24	0	21	1	0	0	0	1	1	1	0	0	0	1	0	0	0
25	1	60	1	0	0	0	0	1	1	1	1	1	1	0	1	1

Fig.2. Dataset

0 Indicated not having mentioned symptom but affected with lung cancer. 1Indicates having mentioned symptom and affected with lung cancer.

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 3 Volume No.6 May – 2022
DOI:10.46647/ijetms.2022.v06i03.007 ISSN: 2581-4621

Fig. 3. From the bar graph we can observe people with Allergy have more chances of getting cancer rather than the people who does have Allergy.

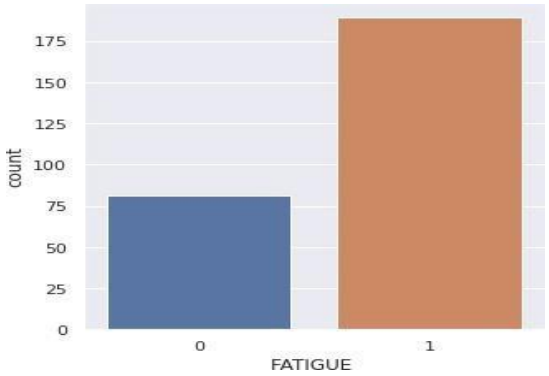

Fig. 4. From the bar graph we can observe people with Fatigue have more chances of getting cancer rather than the people who does have Fatigue.

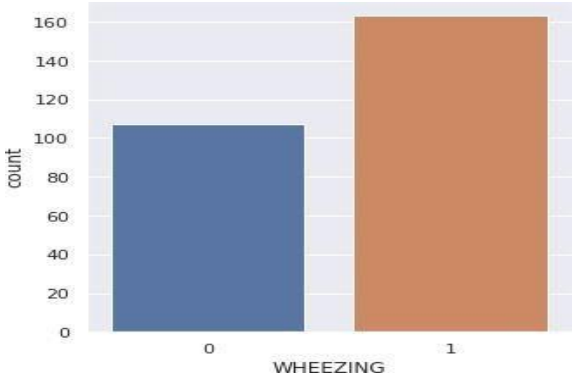

Fig. 5. From the bar graph we can observe people with Wheezing have more chances of getting cancer rather than the people who does have Wheezing.

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 3 Volume No.6 May – 2022
DOI:10.46647/ijetms.2022.v06i03.007 ISSN: 2581-4621

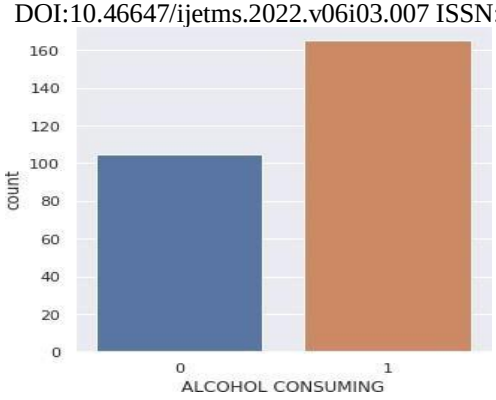

Fig. 6. From the bar graph we can observe people with Alcohol Consuming have more chances of getting cancer rather than the people who does have Alcohol Consuming.

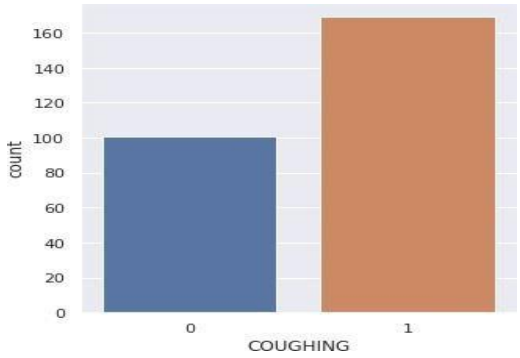

Fig. 7. From the bar graph we can observe people with Coughing have more chances of getting cancer rather than the people who does have Coughing.

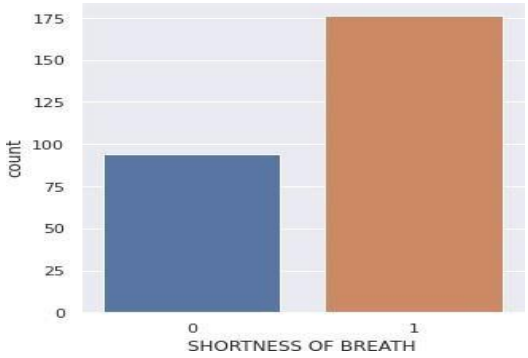

Fig. 8. From the bar graph we can observe people with Shortness of Breath have more chances of getting cancer rather than the people who does have Shortness of Breath.

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 3 Volume No.6 May – 2022
DOI:10.46647/ijetms.2022.v06i03.007 ISSN: 2581-4621

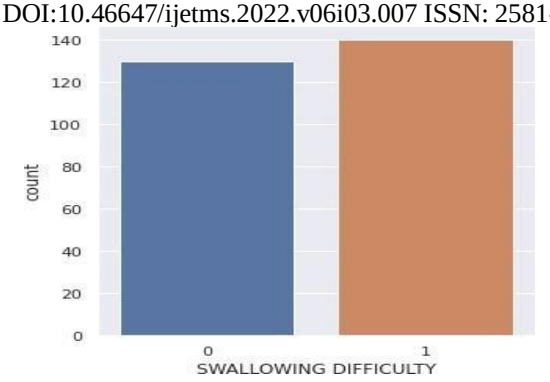

Fig. 9. From the bar graph we can observe people with Swallowing Difficulty have more chances of getting cancer rather than the people who does have Swallowing Difficulty.

Fig. 10. From the bar graph we can observe people with Chest Pain have more chances of getting cancer rather than the people who does have Chest Pain.

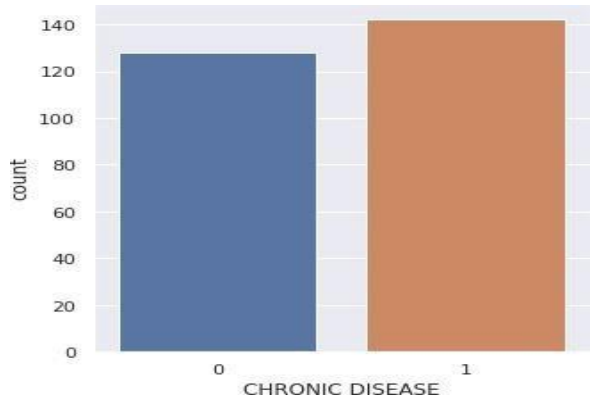

Fig. 11 From the bar graph we can observe people with Chronic Disease have more chances of getting cancer rather than the people who does have Chronic Disease.

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 3 Volume No.6 May – 2022
DOI:10.46647/ijetms.2022.v06i03.007 ISSN: 2581-4621

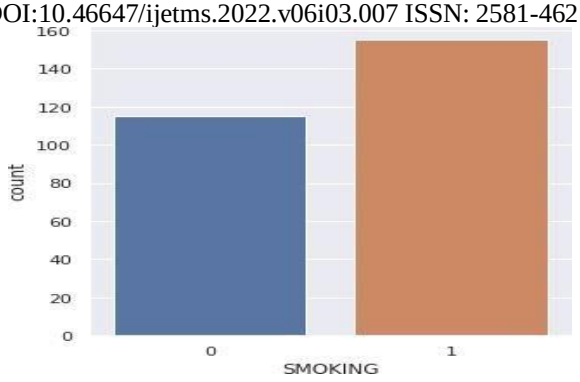

Fig. 12 From the bar graph we can observe people with smoking have more chances of getting cancer rather than the people who does have Smoking.

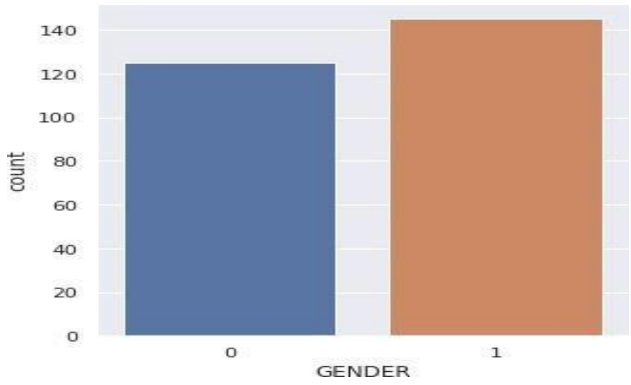

Fig. 13 From the bar graph we can observe males are more prone to get lung cancer than the females.

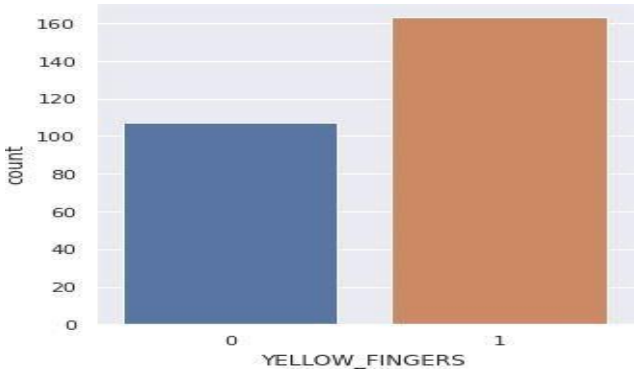

Fig. 14 From the bar graph we can observe people with Yellow Fingers have more chances of getting cancer rather than the people who does have Yellow Fingers.

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 3 Volume No.6 May – 2022
DOI:10.46647/ijetms.2022.v06i03.007 ISSN: 2581-4621

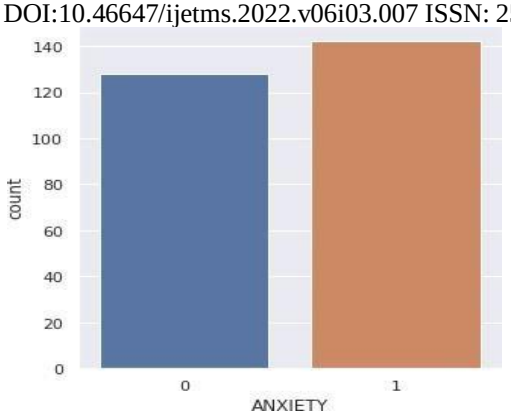

Fig. 15 From the bar graph we can observe people with Anxiety have more chances of getting cancer rather than the people who does have Anxiety.

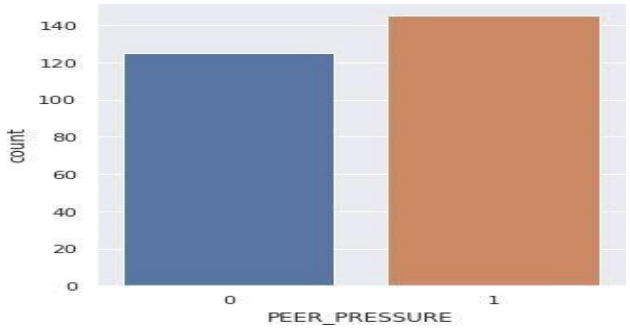

Fig. 16 From the bar graph we can observe people with Peer Pressure have more chances of getting cancer rather than the people who does have Peer Pressure.

By comparing all the bar graphs, we can say that all the considered symptoms play crucial part in determining whether the person has lung cancer or not. Majority, people who does not have these symptoms like Fatigue, Alcohol Consumption, Shortness of Breath are less likely to get lung cancer.

6 Conclusion

This Decision Tree Algorithm aims to be used for predicting cancer, to reduce the time consumption. This technique gives better accuracy (results) when compared to other algorithms. It provides earlier warning with less time consumption.

Acknowledgments

This We would like to thank our guide Mrs. T. Swathi and Mrs. Soppari. Kavitha for their continuous support and guidance. Due to their guidance, we can complete our project successfully. Als
o, we are extremely grateful to Dr. M. V. VIJAYA SARADHI, Head of the Department of Computer Science and Engineering, Ace Engineering College for his support and invaluable time.

Website: ijetms.in Issue: 3 Volume No.6 May – 2022
DOI:10.46647/ijetms.2022.v06i03.007 ISSN: 2581-4621

References

1. https://www.irjet.net/archives/V4/i12/IRJET-V4I12278.pdf
2. Maciej Zi˛eba∗, Jakub M. Tomczak, Marek Lubicz, Jerzy ´Swi˛atek," Boosted SVM for extracting rules from imbalanced data in application to prediction of the post-operative life expectancy in the lung cancer patients", 2014, doi:10.1016/j.asoc.2013.07.016.
3. D.-H. Tae-WooKim, Chung-Yill Park "Decision tree of occupational lung cancer using classification and regression analysis Safety Health Work," Health Work, 1 (2) (2010), pp.140-148
4. Shaikh, F.J., Rao, D.S. Hide details, "Prediction of Cancer Disease using Machine learning Approach"
5. T. Ojala, M. Pietikainen, T. Maenpaa, "Multiresolution gray-scale and rotation in variant texture classification with local binary pattern" IEEE Trans. Pattern Anal. Mach. Intell., 24(7)(2002),pp.971-987
6. V. Krishnaiah, G. Narsimha, C. Subhash Diagnosis of lung cancer prediction system using datamining techniques Int.J. Comp. Sci. Inf.Technol.,4 (1)(2013),pp. 39-45
7. N.Picco, R.A.Gatenby, A.R.A.Anderson, "Stem cell plasticity and niche dynamics in cancer progression" IEEE Trans. Biomed. Eng.,64(3) (2017),pp.528-537
8. https://www.irjet.net/archives/V4/i2/IRJET-V412371.pdf
9. https://www.javatpoint.com/machine-learning-decision-tree-classification-algorithm