Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 31s (2025)
https://www.jneonatalsurg.com

OPEN ACCESS

A Rule Based Model to Predict Lung Cancer Risk Level

Achal Sharma[1], Dr. Rahul Mishra[2]

[1]Electronics & Communication Engineering, Dr. APJ Abdul KalamUniversity, Indore, India.

Email ID: achal007sharma@gmail.com

[2]Electronics & Communication Engineering, Dr.APJ Abdul KalamUniversity, Indore, India.

Email ID: rmishra_sati@yahoo.com

Cite this paper as: Achal Sharma, Dr. Rahul Mishra, (2025) A Rule Based Model to Predict Lung Cancer Risk Level. *Journal of Neonatal Surgery*, 14 (31s), 300-312.

ABSTRACT

Lung cancer is one of the most dazedly disease around the world. It is not only affecting the people who are active smokers, but this disease is also affecting the people who are not smoking. Therefore, it is essential to investigate, which features are most responsible for lung cancer. Therefore, the proposed work is providing the two main contributions: (1) providing the detailed understanding about the different essential attributes, which are highly responsible for lung cancer risk. (2) Providing a machine learning model for accurately predicting the lung cancer risk. In this context, first a publically available dataset has been considered for study. Additionally, the feature relevance analysis has been performed using the random forest classifier. Further, by using a fixed threshold (0.01) the less relevant features are eliminated. Further, each selected attribute has been discussed and their details are provided. Finally two classifiers namely decision tree and convolutional Neural network (CNN) has been trained and validated. The validation results provide 100% accuracy for both the machine learning algorithm. Therefore, the prediction rules have been prepared for accurately predict that can predict the lung cancer level accurately.

Keywords: *Decision Tree, Deep learning, Lung Cancer, Machine learning, risk prediction*

1. INTRODUCTION

The lung is a main part of respiration system. In human two lungs are available at both side of the chest. During breathing process, the chest rises and falls because of the lungs swell and exhalant the air. The lungs are mixing the oxygen with blood. The heart sends blood to the lungs, which has rich carbon dioxide. The lungs absorbs oxygen and leaves carbon dioxide. The air reaches to the lungs through the nasal or the oral cavity. Humans never stop breathing because the lungs supply our blood with oxygen, which is vital for human life. The lung diseases are one of the main causes of death. Factors such as smoking, environmental toxins and chronic inflammation cause harmful effects that can lead to permanent damage to the lung. The lungs have the ability to clear themselves using a inbuilt mechanisms. However, for someone who smokes, Environmental factors, genetic, hereditary or a combination are able to affect the lungs and promote various diseases.

Smoking is the leading cause of lung diseases, which is characterized by inflammation and damage to the lungs. The main symptoms are chronic cough, increased mucus production and shortness of breath. The main symptoms of emphysema include coughing, shortness of breath, and limited exercise resistance. Lung cancer is the primary cause of death in both genders. The most common symptom of lung cancer is coughing, which needs special attention because they are smokers and suffer from other disease, which are also causes coughing. In addition, the symptoms caused by lung cancer include expectoration, chest pain, shortness of breath, anorexia, weight loss, fever and hemoptysis. Therefore, timely detection and differentiation of lung cancer risk is essential task.

Nowadays, artificial intelligence (AI) and machine learning (ML) techniques play a critical role in healthcare. Due to the wide applicability of AI/ML in health conditions' risk prediction, a variety of method are available and support the practical deployment of software tools for the early prediction and diagnosis of a disease. This study is concern with lung cancer risk prediction by using ML and historical patient's record.

In this paper, first a detailed understanding of the different essential symptoms of the lung cancer has been discussed. The discussion is performed on the basis of the patient's record of three category i.e. Low(0), Medium (1) and High(2). Each category of the sample is grouped individually and their trend analysis has been performed. Next, two machine learning algorithms namely decision tree and artificial neural network has been used to train with the identified symptoms of the lung cancer and prediction has been performed. Next, in order to predict the lung cancer risk some decision rules has been prepared to manually identify the cancer risk level. Next the performance of the implemented ML techniques have been performed. For the evaluation of the models, the performance metrics accuracy has been considered. The performance analysis revealed that decision tree and artificial neural network both are providing similar performance.

The next sections of the paper are described as: In next section, the dataset obtained has been discussed and preprocessing is performed. Further, the essential features are selected and discussed each feature in detail. Further, a model is provided on the subject under investigation and an analysis of the methodology are given. Finally, conclusions and future directions are reported.

2. EXPERIMENTAL DATASET

The main aim of the proposed work is to investigate the main causes of the lung cancer and also contribute a machine learning model to accurately predict the lung cancer disease risk in three categories low, mid and high. In this context, a dataset from the online public repository namely Kaggle has been identified [1]. The dataset is containing information of patients with lung cancer. The available details of the dataset attributes are given in table 1.

Table 1 Dataset Details

S. No.	Type	Name	Description
1	Numeric	index	Unique index of patient’s record
2	Categorical	Patient Id	Unique id of patient’s record
3	Numeric	Age	patient’s age
4	Categorical	Gender	patient’s gender
5	Categorical	Air Pollution	patient’s exposure to the level of air pollution
6	Categorical	Alcohol use	level of alcohol use
7	Categorical	Dust Allergy	dust allergy level of patient
8	Categorical	OccuPational Hazards	level of occupational hazards
9	Categorical	Genetic Risk	level of genetic risk
10	Categorical	Chronic Lung Disease	Level of chronic lung disease
11	Categorical	Balanced Diet	Level of balanced diet
12	Categorical	Obesity	Level of obesity
13	Categorical	Smoking	Level of smoking
14	Categorical	Passive Smoker	Level of passive smoker
15	Categorical	Chest Pain	Level of chest pain
16	Categorical	Coughing of Blood	Level of coughing of blood
17	Categorical	Fatigue	Level of fatigue
18	Categorical	Weight Loss	Level of weight loss
19	Categorical	Shortness of Breath	Level of shortness of breath
20	Categorical	Wheezing	Level of wheezing
21	Categorical	Swallowing Difficulty	Level of swallowing difficulty
22	Categorical	Clubbing of Finger Nails	Level of clubbing of finger nails
23	Categorical	Frequent Cold	Level of Frequent Cold
24	Categorical	Dry Cough	Level of Dry Cough
25	Categorical	Snoring	Level of Snoring

26	Categorical	Level	Risk of lung cancer category

The dataset is made with the attributes as given in Table 1, which is made with the patient's symptoms. The dataset contains a total of 26 columns and a total of 1000 instances of patient's records. In this dataset, two attributes namely 'index' and 'patient's ID' is able to identify the individual patient's. Therefore, both the attributes are not relevant for learning with the machine learning algorithm.

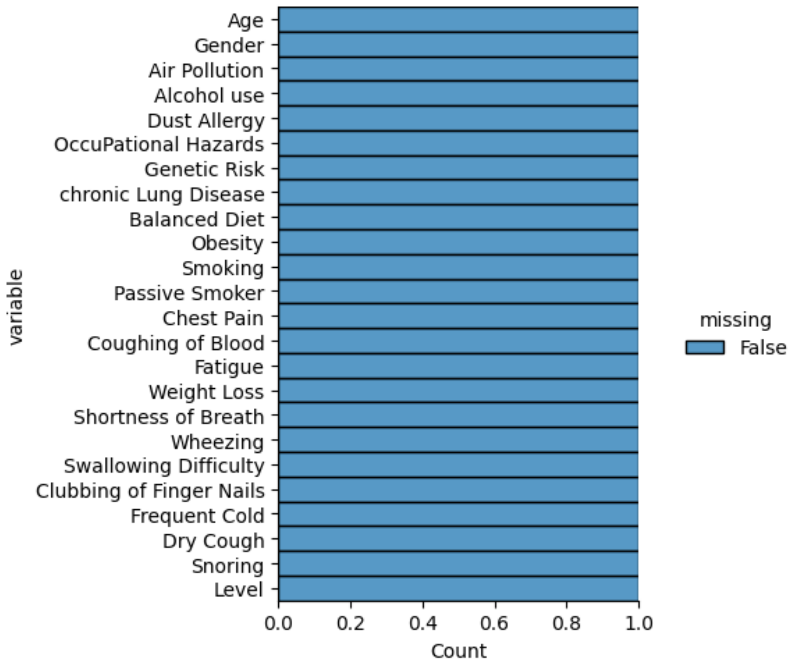

Figure 1 Missing value analysis of the considered dataset

In this context, both the attributes are eliminated from the dataset manually to avoid over fitting and under fitting conditions of machine learning algorithms. The remaining dataset contains a total of 24 attributes, which are used in further investigation. The dataset may also contains the missing values, which can create issues in effective learning process. In this context, the missing value are tested using the data visualization technique. Figure 1 consist of the missing values composition of the dataset. In this diagram, X axis include the % missing values and Y axis shows the attribute names in dataset. According to the analysis, there are no missing value is available in dataset.

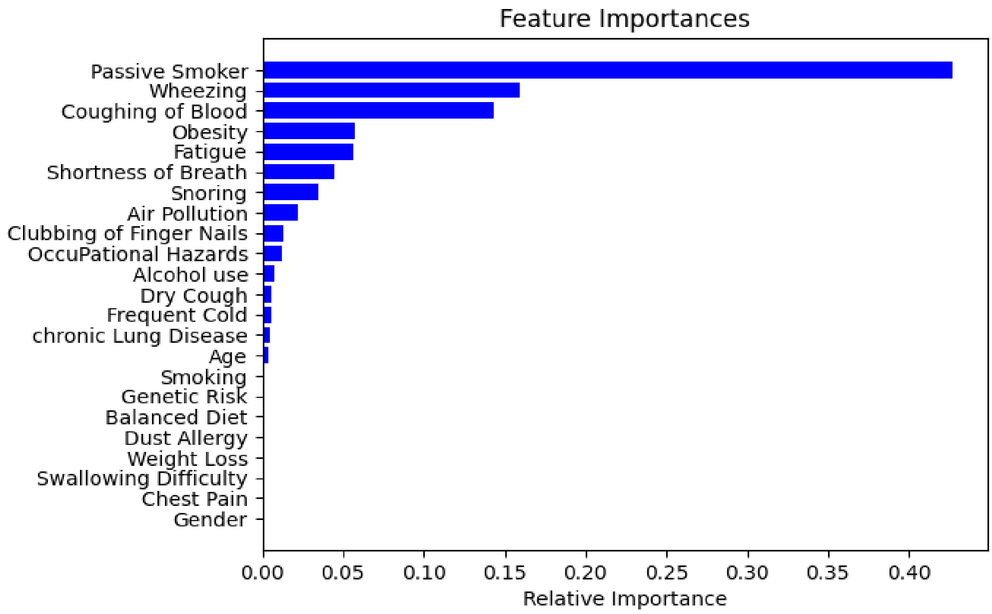

Figure 2 Feature relevance analysis

The dataset has finally 24 attributes, which can be used for performing training to the machine learning algorithms. But, for simplicity and also to minimize the computational resource cost, the most essential features are needed to select. Therefore, feature relevance analysis has been performed using the random forest classifier. The relevance score ranks of the dataset attributes are given in Figure 2. In this diagram, X axis contain the feature relevance score in terms of percentage (%) and Y axis shows the rank wise attributes list. According to the visualized feature relevance score, the highest relevance score is 42% and most of the attributes has very low relevance scores. Therefore, it is decided to eliminate the less relevant features from the dataset. In this context, a threshold of 1% has been considered for eliminating the less potential attributes. Thus, only those attributes are selected, which has the feature relevance score higher than threshold value (1%). The top selected features are 'Air Pollution', 'OccuPational Hazards', 'Obesity', 'Passive Smoker', 'Coughing of Blood', 'Fatigue', 'Shortness of Breath', 'Wheezing', 'Clubbing of Finger Nails', and 'Snoring'. Now, the individual attributes are discussed and described in the next section.

3. DATA EXPLORATION

In this section, the selected top 10 potential attributes have been explored and the detailed analysis has been given. In first step, it is tried to measure the composition of the data attributes. Here, it is found the considered attributes are prepared on the basis of levels which are divided in 9 levels. Figure 3 demonstrate the frequency of attribute's levels. In this context, X axis shows the level score of the attribute values and Y axis shows the count of values. In addition, the different color of line bars are demonstrating the attributes.

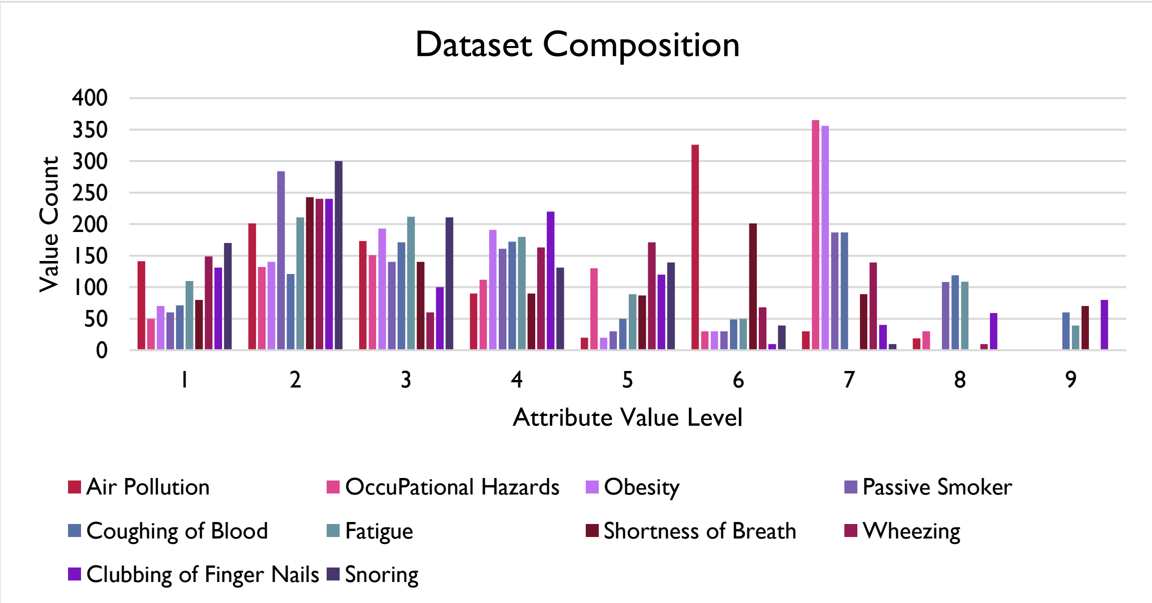

Figure 3 Dataset attribute value's count

According to the prepared diagram, most of the attributes has high count of values based on 2, 3, 4 and 7. Therefore, these values are shows the spikes on the bar graph. In addition, two attributes namely 'occupational hazards' and 'Fatigue' levels are shows the hick on value '7'. It means the patients with these two high level of values has the higher risk of cancer.

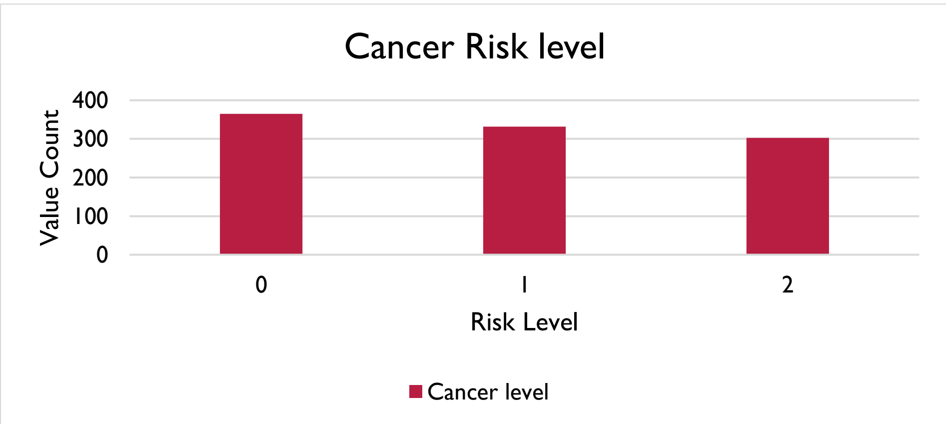

Figure 4 Cancer Risk Levels

Next, the attribute 'Cancer Risk Level' has been considered. This attribute is the target attribute, which need to predict using the machine learning algorithms. The dataset have three risk levels low (0), medium (1) and high (2). The composition of the risk levels are demonstrated visually in figure 4. In this diagram, X axis shows the risk levels and Y axis shows count of levels. According to the given composition of risk levels, we can say the dataset has 36% samples belongs to low cancer risk people, 33.2% samples are belonging to medium risk and 30.3% samples are belonging from high risk patients.

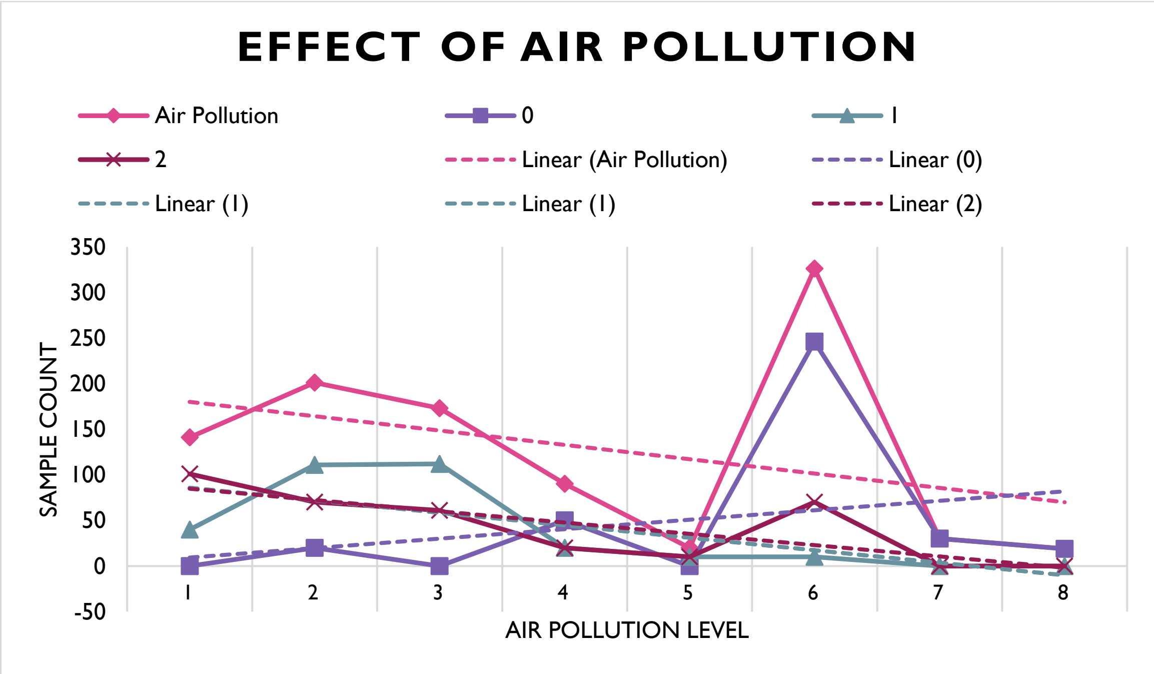

Figure 5 Effect of Air Pollution on Cancer Risk Level

Next, the attribute namely 'Air pollution' has been considered for exploration. In this context, the samples of each risk level low, mid and high are separated. Moreover, the number of samples of each class in different air pollution level has been visualized. The visualization of the data is given in Figure 5. In this diagram, X axis shows the air pollution level and Y axis shows the number of samples. The yellow color line is demonstrating the low risk samples, green line shows medium risk level samples and brown line shows the high risk samples. According to visualized samples, most of the samples are belongs from high pollution level (6) score. In addition, from low air pollution level most of the samples are considered for mid and high risk people. Moreover, the linear trend line patterns of the sample distribution has also been plotted. Based on the sample trends the low risk sample pattern shows increasing trends and other two patterns (mid and high) are showing the decreasing trends.

Further, the next attribute 'Occupational Hazard' attribute has been considered. Figure 6 demonstrate the class wise sample distribution patterns of 'Occupational Hazard'. In this diagram, X axis shows the level of 'Occupational Hazard' and Y axis shows the sample count. According to the sample distribution given in this figure, the low can risk level samples are mostly belongs to 'Occupational Hazard' exposer level (7). On the other hand, the samples belongs to mid and high cancer risk are taken from low 'Occupational Hazard' exposer level (i.e. 2, 3, and 4). In this diagram, the trend line of samples have also been plotted. According to the sample distribution the low cancer risk based samples has increasing Obesity level trends and the samples belonging to mid and high cancer risk has decreasing trend of sample count.

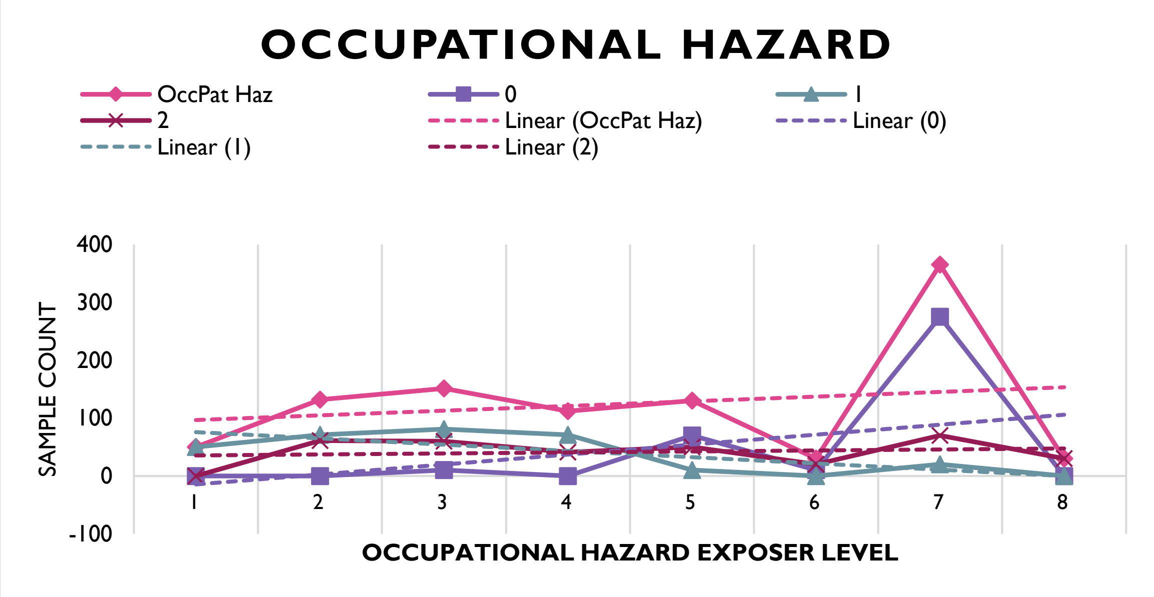

Figure 6 Occupational Hazard based samples

Next, the Obesity level of the patient's samples has been considered. Figure 7 consist of the visual sample patterns of the Obesity level of the patient's. In this diagram, X axis shows the Obesity level and Y axis shows sample count.

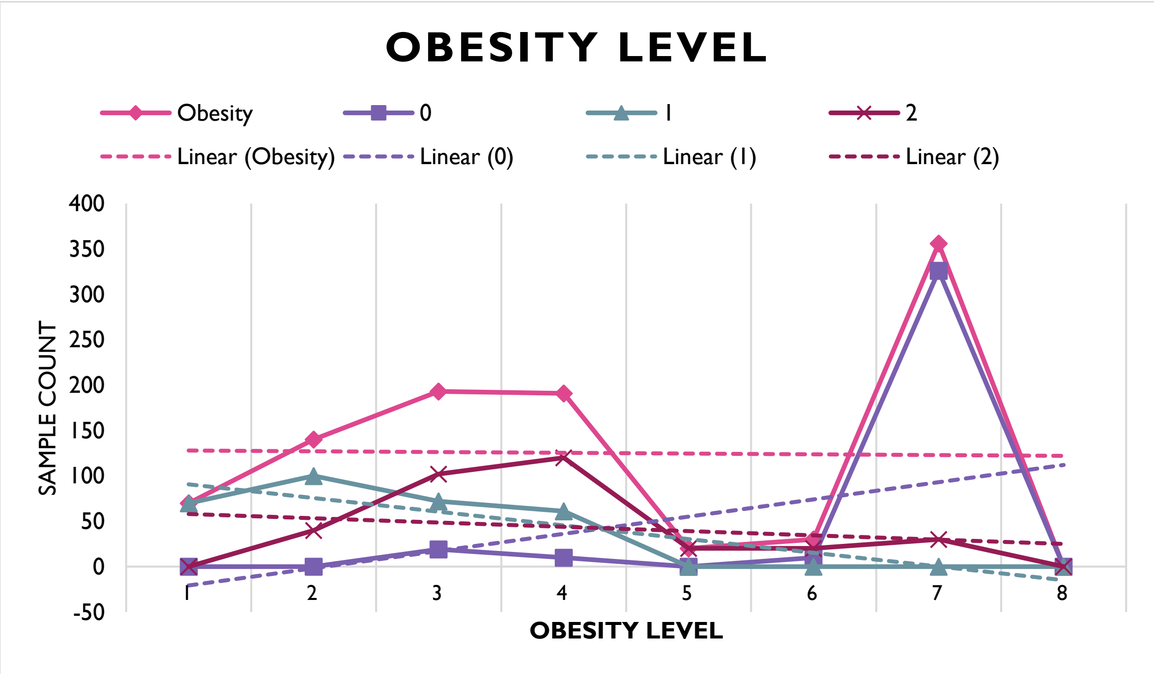

Figure 7 Obesity level of patient's

According to the visualized patterns, most of the low cancer risk level patient samples are has high Obesity level (7). In addition, the samples of mid and high risk level has medium Obesity level (i.e. 3 and 4). In addition, the diagram is also containing the trend lines of sample values. According to the given trends, the trend of low cancer risk level has increasing trend of Obesity level and mid and high can risk level people have decreasing Obesity level trends.

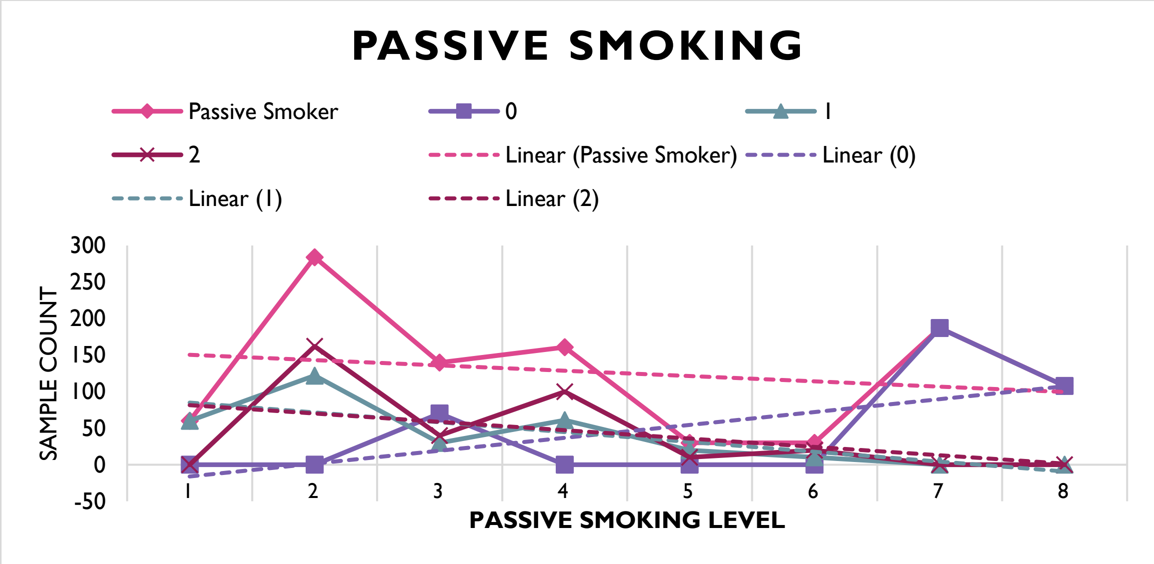

Figure 8 Passive Smoking Level

In Figure 8, the patient's passive smoking exposer level has been discussed. The X-axis of diagram contains passive smoking exposer level and Y axis shows the sample count. The dataset contains the low risk samples which has high passive smoking exposer level (7 and 8) and the samples belongs to mid and high cancer risk has the most of samples which has low passive smoking exposer level (2, 3, and 4). In addition, the trends line of the samples shows the increasing trends of passive smoking exposer level trend and mid and high risk levels shows the decreasing trends of sample count.

Next, the attribute 'Coughing of blood' has been considered. Basically it is a symptom of critical lung disease. The samples are based on cancer risk level has been visualized in Figure 9. The X axis of this diagram contains the Coughing of blood level and the Y axis shows the number of samples belonging to the type of samples. According to given diagram, low can risk level patient's samples contains high level of Coughing of blood (i.e. 7, 8, and 9). On the other hand, the samples belonging to mid and high cancer risk level samples are having low level of Coughing of blood (i.e. 2, 3 and 4). In addition, the trends of Coughing of blood in class wise sample has also been plotted. The trends of the samples are given as dotted line. According to the trends of plot the low cancer risk level samples has the increasing Coughing of blood level samples and mid and high cancer risk level samples have decreasing Coughing level of blood.

Next attribute Fatigue level of patient's has been investigated using visual analysis. Figure 10 contains the cancer risk level based sample distribution of Fatigue level of patient's. In this diagram, X axis shows the Fatigue level and Y axis shows the total number of sample count. According to the given pattern of samples has the similar distribution. But the trends of the sample patterns demonstrate sharp decreasing number of samples for mid and high cancer risk samples.

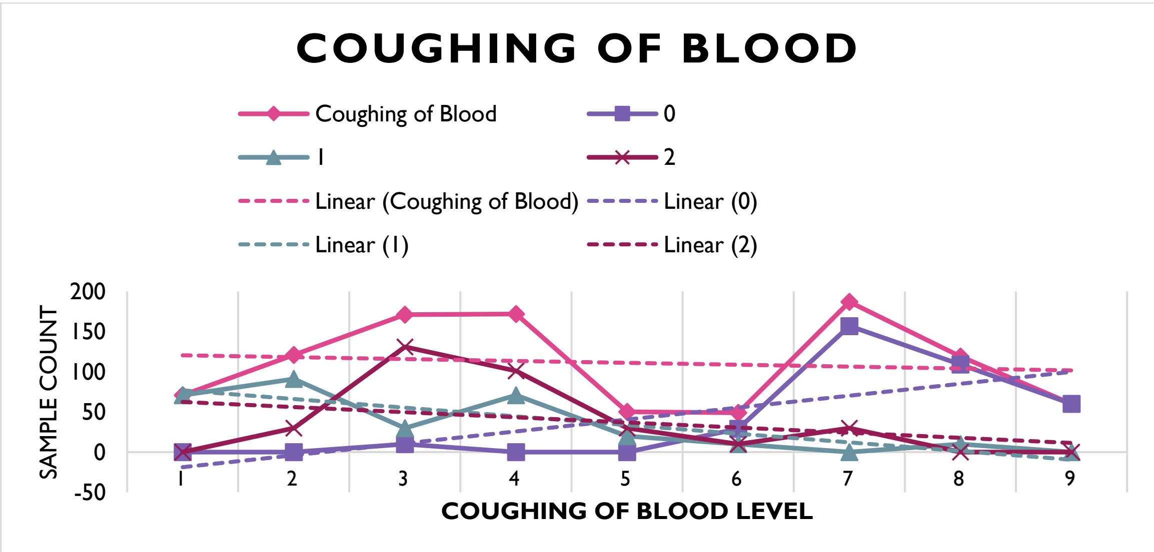

Figure 9 Coughing of blood

Therefore, the trends of low cancer risk samples are linear and slightly increasing. On the other hand, the samples belongs to mid and high cancer risk shows sharp deceasing samples of fatigue. Therefore, we can say the fatigue level of all the type of patient's samples are equally distributed.

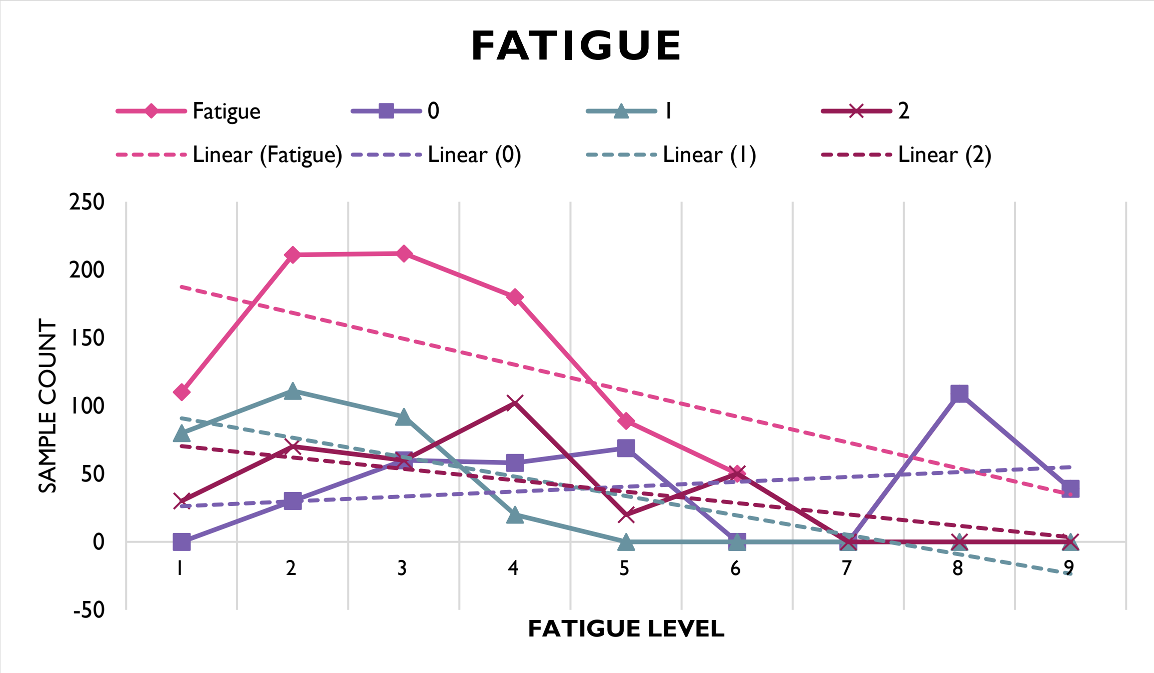

Figure 10 Fatigue level of patient's

Next, the attribute 'short breath level' has been considered. This attribute is patient's symptoms. Figure 11 demonstrate the sample count distribution of shot breath level of patients based on cancer risk level. In this diagram, X axis shows the level of short breath and Y axis shows the number of samples belonging to the short breath level. According to the visual patterns, the low cancer risk level patients are has similar number of samples for short breath levels thus the trend of the low cancer risk level is linear and horizontal. On the other hand, the samples of mid-level of cancer risk is also shows the similar trend as the low risk level but slight deceasing. On the other hand, the samples belongs to high cancer risk level have high number of samples belongs to low short breath level (i.e. 2, 3 and 4). Therefore, the trend of the sample is decreasing.

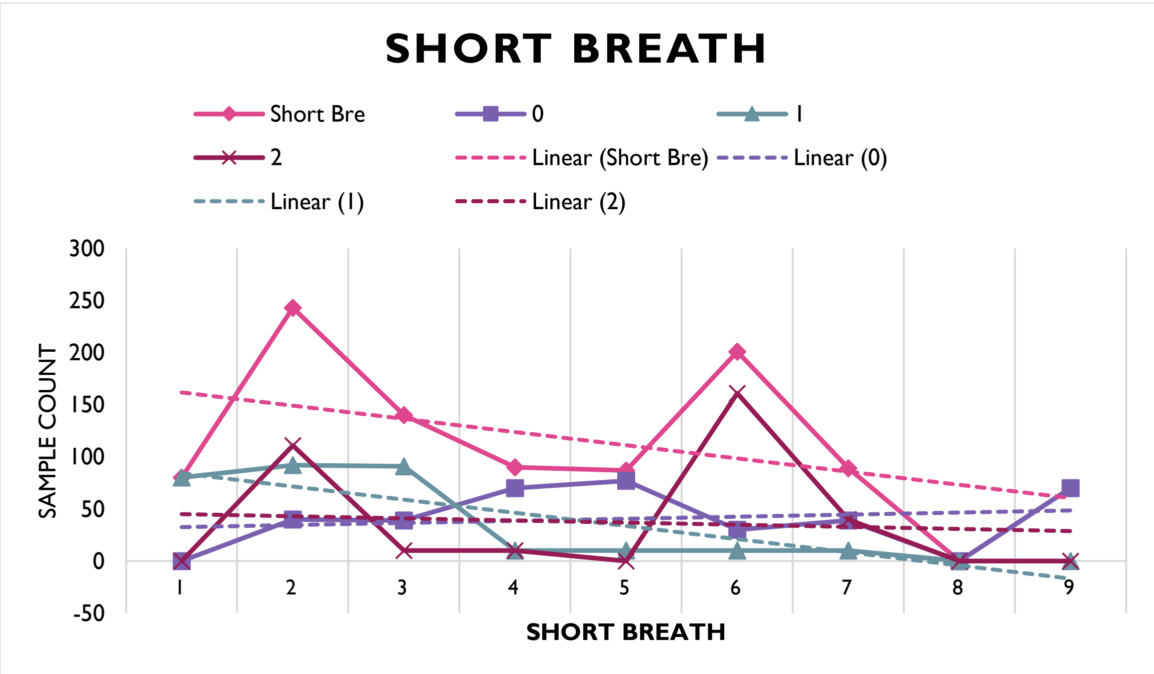

Figure 11 short breath level of patient's

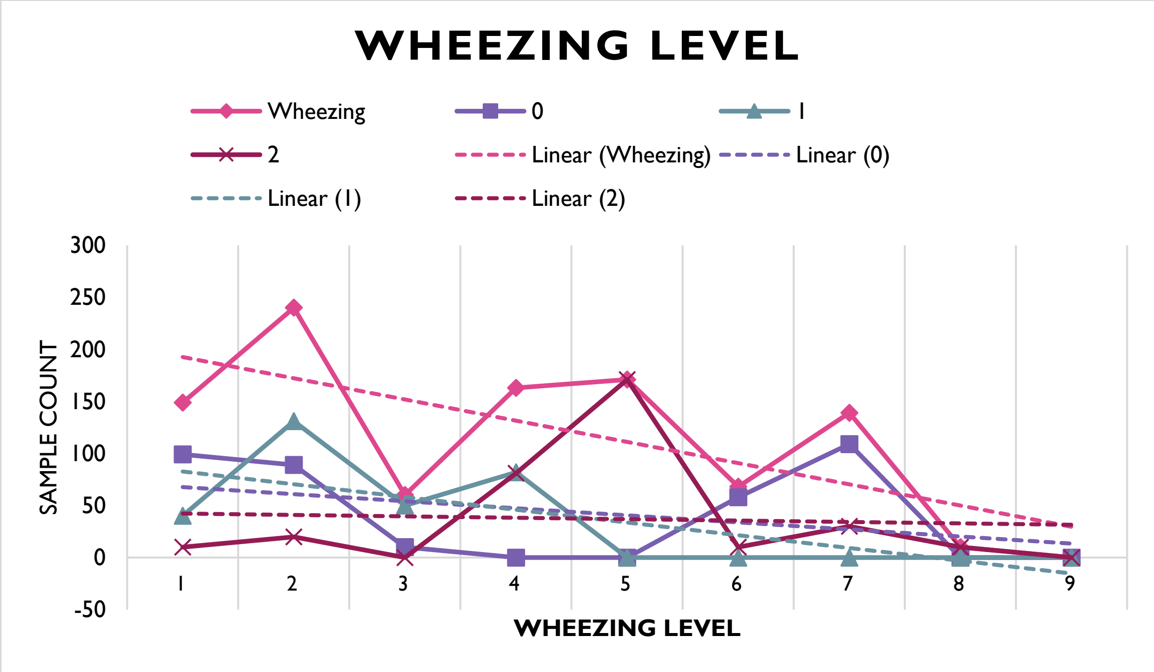

Figure 12 Wheezing level of patients

Next, the wheezing level of cancer risk patient's has been investigated. The cancer risk level based sample distribution has been given in Figure 12. In this diagram, X axis shows the wheezing level of patient's and Y axis shows the number of samples belonging to the wheezing level. According to the visual patterns, wheezing level of patient's belonging to low and high cancer risk level has decreasing level of wheezing. In addition, the medium cancer risk level has medium level of wheezing issues. This pattern is also demonstrated in the trends of the samples. Therefore, mid cancer risk level sample has normal linear trend on the other hand high and low cancer level samples demonstrate the reducing trends in sample count trend.

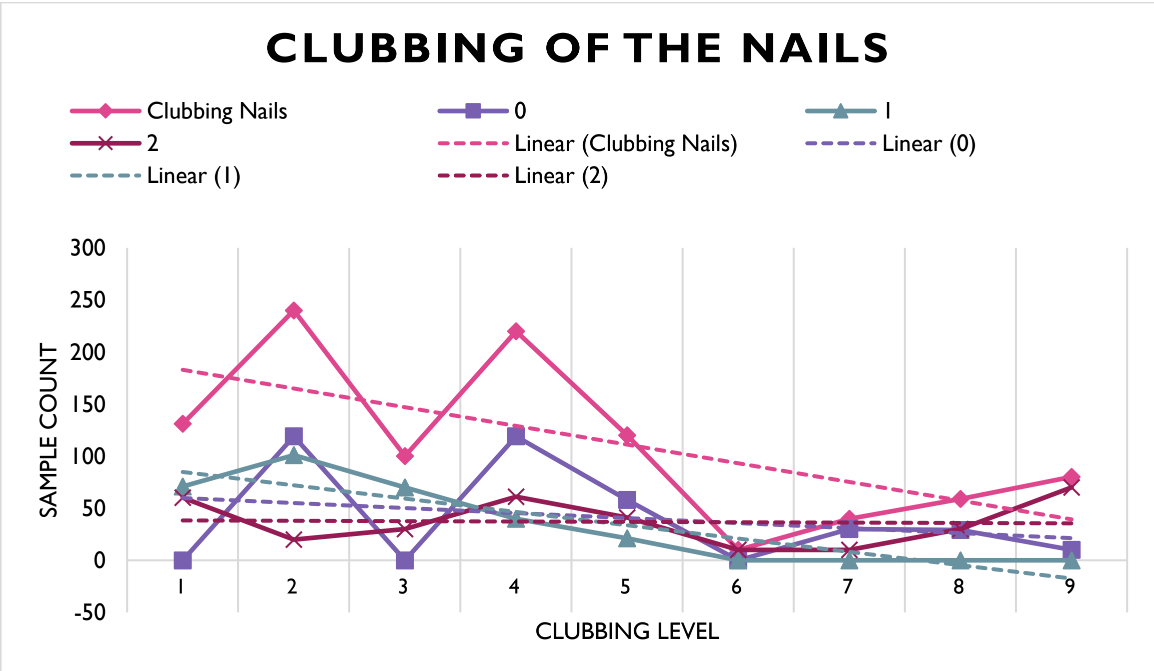

Figure 13 Clubbing of the Nails

Next attribute is Clubbing of the Nails, which is visualized in figure 13. In this diagram, X axis shows the Clubbing level of nails and Y axis shows the number of samples. According to the given sample distribution, the number of samples for high and low cancer risk level patients are reducing and the mid-level samples are equally distributed. Similarly this trend is also observed in the trend lines. According to the given trends of samples the low and high risk sample demonstrate reducing trends of the samples for clubbing of nails but the mid cancer risk level shows linear and horizontal line of trend.

Finally, the last attribute is snoring level of patients. The snoring level for all the group of patients are visualized in figure 14. In this diagram, X axis shows the snoring level and Y axis shows the number of samples. According to the visualized samples most of the samples has less snoring level for all risk level. Therefore, this attribute has less relevance with the respect to cancer risk level. In addition, when we considering the trends of the samples then it is found the sample distribution of all the type of samples are similar and reducing with increasing level of snoring. This section is provided the details about the different attributes of the dataset, which has high relevance with the cancer risk level. Next section is providing the utilization of the discussed attributes to train a machine learning model and predict the possible cancer risk accurately.

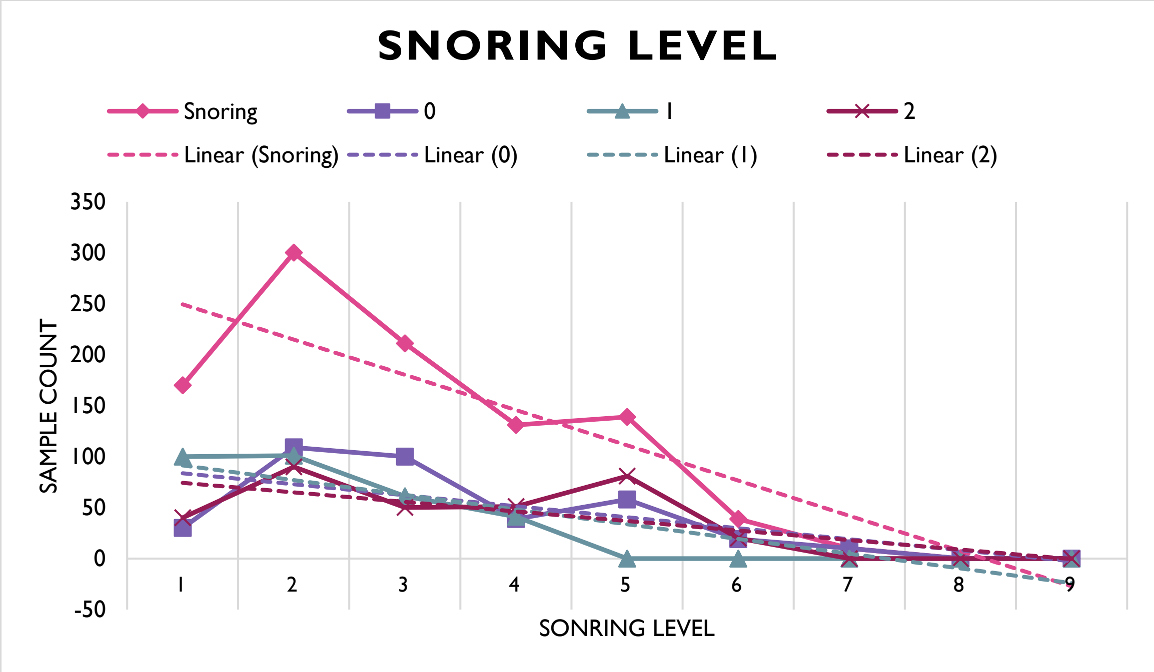

Figure 14 snoring level of patients

4. PROPOSED MODEL

In this section, a machine learning model has been presented for performing training and prediction of possible lung cancer risk by analyzing the features of the patients. The proposed model is demonstrated in figure 15. According to the given diagram of the cancer risk prediction system the first component of the model is historic patient's record. That is a type of database which contains the patient's records. These records are labeled with three type of lung cancer risk level i.e. low, medium and high. The dataset has a total of 26 attributes and 1000 instances. In next step, the dataset is preprocessed. Therefore, two attributes namely index and id have been eliminated. Next, the dataset has a total of 24 attributes. The dataset has high dimension thus a feature selection technique has been implemented. This technique includes a random forest algorithm for measuring the relevance of the attributes.

Based on the measured relevance score of the dataset attributes the attributes are selected, which have higher relevance score than a threshold. In this experiment, the threshold value is 1% considered. Based on this filtering a total of 10 attributes have been remain. These attributes are next used for performing the cancer risk prediction. Therefore, the selected set of dataset is sub-divided into two parts first part is used for performing training of machine learning algorithm. Additionally, second part of data sample is used for performing the validation of the proposed model. Thus, in this presented work, two machine learning models have been considered for training. These algorithms are C4.5 decision tree and convolutional neural network. After training the models are used with the unknown validation samples. The validation samples are classified using the trained machine learning model and the risk prediction has been performed. During this prediction the performance of the proposed model has been measured. The CNN is an opaque model which accept the input attributes and calculate the possible risk level of the given sample. On the other hand the decision tree algorithm is preparing the rules, which can be used for performing the classification.

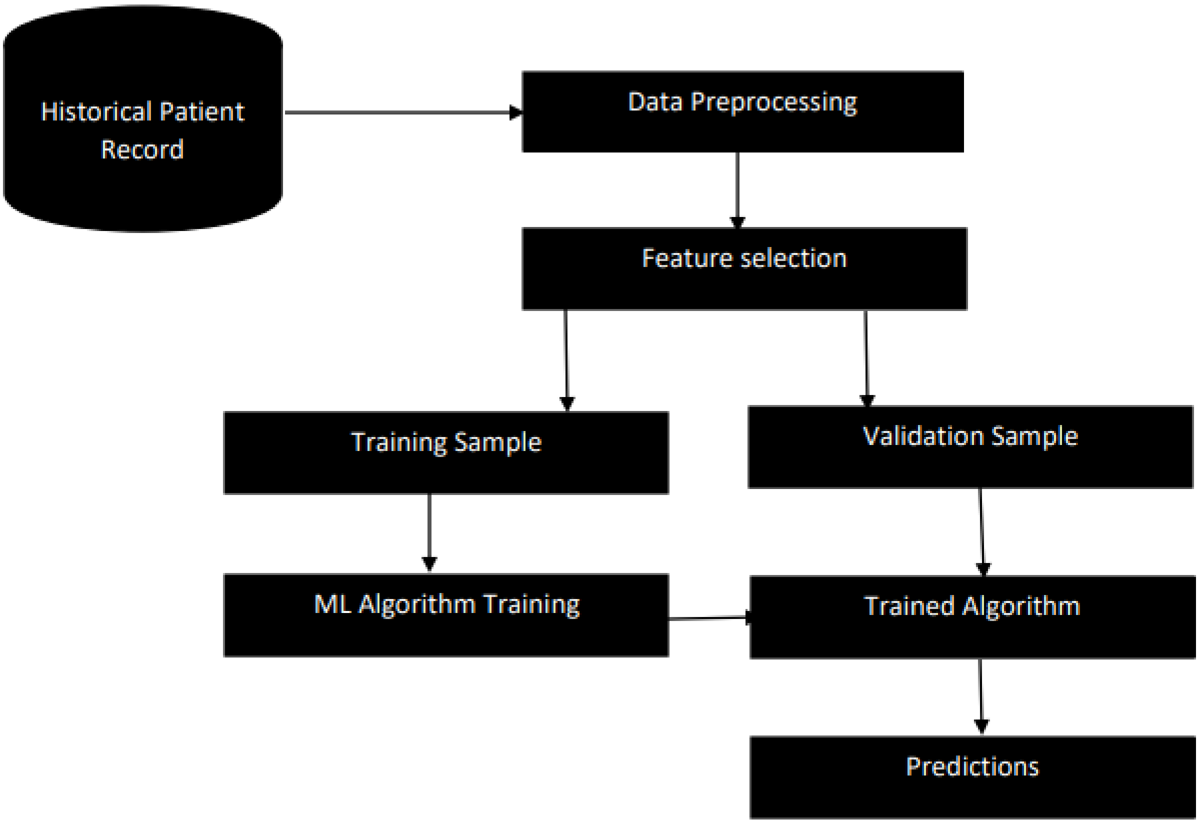

Figure 15 Proposed model for cancer risk prediction

Basically, when the decision tree algorithm takes training than the algorithm utilizes the dataset values and prepare a tree like structure. This tree structure can be transformable to the rules for making the decisions. In this experiment the trained decision tree based extracted rules are given in table 2. These rules can be used to analyze the patient's symptoms manually and predict the cancer risk level with 100% accuracy.

Table 2 shows the extracted rules of the decision

1. If 'Coughing of Blood' <= 5.50 and 'Wheezing' <= 4.50 and 'Snoring' <= 4.50 and 'Obesity' <= 4.50 Then class: 1
2. If 'Coughing of Blood' <= 5.50 and 'Wheezing' <= 4.50 and 'Snoring' <= 4.50 and 'Obesity' > 4.50 and 'OccuPational Hazards' <= 5.50 Then class: 2
3. If 'Coughing of Blood' <= 5.50 and 'Wheezing' <= 4.50 and 'Snoring' <= 4.50 and 'Obesity' > 4.50 and 'OccuPational Hazards' > 5.50 Then class: 0
4. If 'Coughing of Blood' <= 5.50 and 'Wheezing' <= 4.50 and 'Snoring' > 4.50 Then class: 2
5. If 'Coughing of Blood' <= 5.50 and 'Wheezing' > 4.50 Then class: 2
6. If 'Coughing of Blood' > 5.50 and 'Air Pollution' <= 1.50 Then class: 2
7. If 'Coughing of Blood' > 5.50 and 'Air Pollution' > 1.50 and 'Obesity' <= 2.50 Then class: 1
8. If 'Coughing of Blood' > 5.50 and 'Air Pollution' > 1.50 and 'Obesity' > 2.50 and 'Clubbing of Finger Nails' <= 1.50 Then class: 2
9. If 'Coughing of Blood' > 5.50 and 'Air Pollution' > 1.50 and 'Obesity' > 2.50 and 'Clubbing of Finger Nails' > 1.50 Then class: 0

The advantage of these rules is that there are a total of 9 rules by which we can make the prediction about the lung cancer accurately. Finally, the performance of both the machine learning algorithms have been measured in terms of accuracy. Figure 16 shows the accuracy of both the machine learning algorithms. The X axis shows the algorithm names and Y axis shows the accuracy of the machine learning algorithms. Based on the measured results, both the algorithms CNN and decision trees are providing the equal accuracy i.e. 100%.

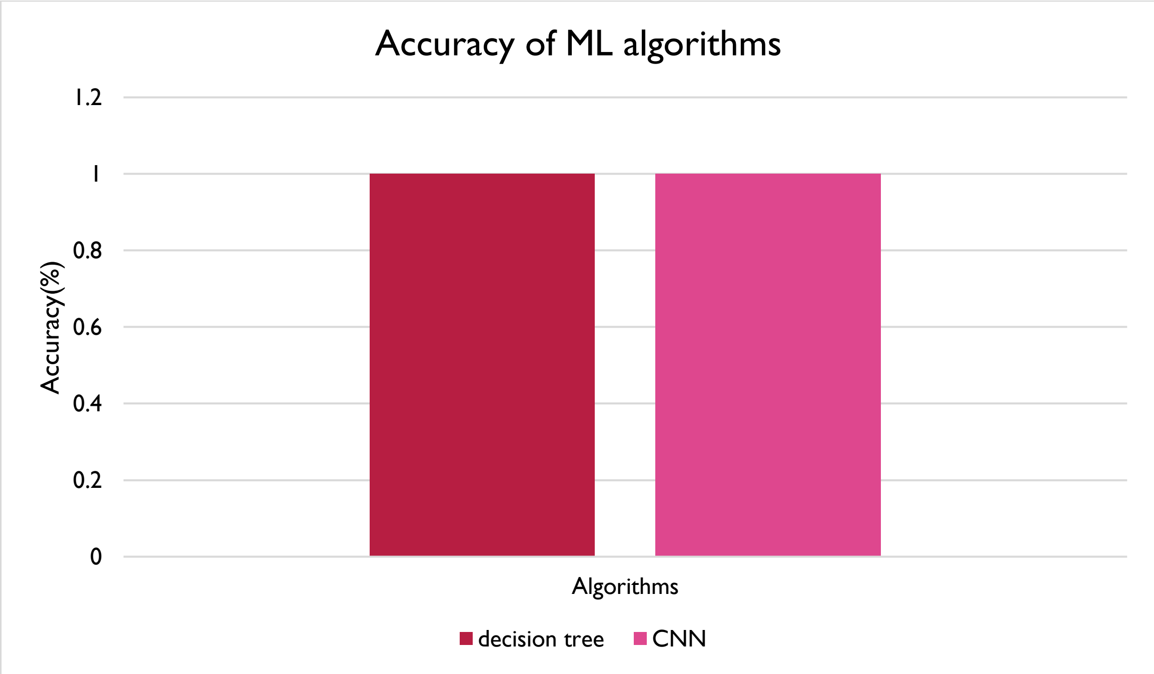

Figure 16 accuracy of the ML algorithms

5. CONCLUSION AND FUTURE WORK

The proposed work is an experimental investigation of lung cancer disease risk factors based cancer risk prediction. In this context, a dataset from the public platform has been identified and used for performing in depth study. Therefore, first the dataset potential features have been identified and based on the selected features two machine learning models have been trained. One model is a deep learning based neural network model for performing classification and second model is based on decision tree classifier. The decision tree classifier is providing a total 9 rules, which also provide accurate classification. Both models deep learning based model and decision tree based lung cancer prediction model is providing 100% accurate classification results.

REFERENCES

[1] Chouhan, S. S., Kaul, A., & Singh, U. P. (2020). A rule-based expert system for lung cancer detection using fuzzy logic. *International Journal of Intelligent Engineering and Systems, 13*(3), 130–140. https://doi.org/10.22266/ijies2020.0630.13

[2] Kumar, D., & Rani, R. (2021). Lung cancer detection using hybrid feature selection and rule-based classification. *Journal of Ambient Intelligence and Humanized Computing, 12*, 7487–7502. https://doi.org/10.1007/s12652-020-02598-4

[3] Patil, S. M., & Sinha, G. R. (2018). Fuzzy expert system for early diagnosis of lung cancer. *International Journal of Applied Engineering Research, 13*(6), 3358–3363.

[4] Achal Sharma, & Dr. Rahul Mishra (2024). Applications Of Deep Learning Models For Image Object Recognition. *Nanotechnology Perceptions* 20(5), 1741-1756

[5] Wang, H., Zheng, B., Yoon, S. W., & Ko, H. H. (2018). A support vector machine-based ensemble algorithm for breast cancer diagnosis. *European Journal of Operational Research, 267*(2), 687–699. https://doi.org/10.1016/j.ejor.2017.12.001
(Related technique applicable to lung cancer)

[6] Islam, M. M., & Rahman, M. M. (2020). Lung cancer prediction using data mining techniques: A comparative study. *International Journal of Computer Applications, 177*(38), 5–9. https://doi.org/10.5120/ijca2020919823

[7] Kavitha, C., & Priya, B. (2020). Diagnosis of lung cancer using rule-based classification and gene expression analysis. *International Journal of Scientific & Technology Research, 9*(1), 2226–2230.

[8] Shankar, K., Perumal, E., & Tiwari, P. (2019). Lung cancer detection using machine learning techniques: A

comparative study. *Multimedia Tools and Applications, 78*(6), 7273–7295. https://doi.org/10.1007/s11042-018-6512-8

[9] Dey, N., Ashour, A. S., & Borra, S. (2019). *Medical Imaging with Deep Learning*. Springer. https://doi.org/10.1007/978-3-030-13969-0
(Contains sections relevant to rule-based AI in medical diagnosis)

[10] Sharma, V., & Kaur, S. (2021). Prediction of lung cancer using hybrid model of artificial neural networks and decision trees. *Materials Today: Proceedings, 49*, 1964–1970. https://doi.org/10.1016/j.matpr.2021.05.645

[11] Yadav, A. R., & Pawar, V. M. (2016). Rule-based classification model for lung cancer prediction using genetic algorithm. *International Journal of Innovative Research in Computer and Communication Engineering, 4*(5), 8355–8360.

[12] Prasad, G., & Raj, A. (2019). A hybrid approach of machine learning and rule-based system for cancer prediction. *International Journal of Recent Technology and Engineering, 7*(6), 50–55.

[13] Jiang, H., Ma, H., & Qiu, Y. (2022). Explainable machine learning model for lung cancer risk prediction. *IEEE Access, 10*, 5801–5812. https://doi.org/10.1109/ACCESS.2022.3142065

[14] Krishnan, M. R., & Nair, S. S. (2020). Predictive analytics for lung cancer detection using hybrid ML models. *Procedia Computer Science, 171*, 2028–2033. https://doi.org/10.1016/j.procs.2020.04.217

[15] Ghosh, R., & Mitra, P. (2021). Fuzzy logic and decision tree based lung cancer classification. *International Journal of Computer Applications, 183*(23), 10–15. https://doi.org/10.5120/ijca2021921453

[16] Zhang, Y., & Yu, C. (2019). Lung cancer diagnosis with explainable rule-based models using Bayesian networks. *Journal of Biomedical Informatics, 94*, 103188. https://doi.org/10.1016/j.jbi.2019.103188

[17] Singh, M., & Gupta, A. (2019). Intelligent rule-based system for lung cancer detection using CT scan images. *International Journal of Advanced Computer Science and Applications, 10*(5), 317–322. https://doi.org/10.14569/IJACSA.2019.0100541

[18] Gupta, D., & Khare, N. (2018). A hybrid intelligent system for lung cancer diagnosis using CT scan images. *International Journal of Computer Applications, 180*(44), 12–17. https://doi.org/10.5120/ijca2018916394

[19] Sharma, P., & Rana, A. (2020). Rule-based prediction of lung cancer using fuzzy inference systems. *Journal of Critical Reviews, 7*(19), 7427–7432. https://doi.org/10.31838/jcr.07.19.834

[20] Rajeswari, S., & Aramudhan, M. (2019). Performance analysis of rule-based classification algorithms for lung cancer prediction. *International Journal of Innovative Technology and Exploring Engineering, 8*(8), 2720–2725.

[21] Bhardwaj, A., & Tiwari, A. (2020). Risk prediction of lung cancer using fuzzy and rule-based expert system. *International Journal of Innovative Research in Computer and Communication Engineering, 8*(5), 2452–2458.

[22] Patel, H., & Prajapati, N. (2021). Deep learning and rule-based approach for lung cancer classification. *Materials Today: Proceedings, 47*, 5026–5030. https://doi.org/10.1016/j.matpr.2021.06.113

[23] Al-Janabi, M., Al-Shourbaji, I., & Al-Janabi, N. (2019). Fuzzy-based expert system for early detection and classification of lung cancer. *International Journal of Engineering and Advanced Technology, 8*(6), 4912–4916.

[24] Murugan, S., & Padmapriya, T. (2021). Efficient detection of lung cancer using hybrid feature selection and fuzzy rule-based system. *Procedia Computer Science, 172*, 358–365. https://doi.org/10.1016/j.procs.2020.05.049

[25] Arora, A., & Dua, S. (2017). Development of an expert system for lung cancer diagnosis. *International Journal of Engineering Research and Applications, 7*(6), 66–72.

[26] Kale, S. A., & Mane, N. B. (2019). Prediction of lung cancer using data mining techniques. *International Research Journal of Engineering and Technology, 6*(3), 3068–3073.