Lung Cancer Detection using Supervised Machine Learning Techniques

Rahat Idrees[1], Muhammad Kamran Abid[2], Saleem Raza[3], Muhammad Kashif [4], Muhammad Waqas[5], Mubashir Ali[6] , Laiba Rehman[7],

[1,2]Department of Computer Science, NFC Institute of Engineering and Technology, Multan, Pakistan

[3]Department of Electronic Engineering, Quaid-e-Awam University of Engineering, Science and Technology, Nawabshah, Pakistan

[4]Quaid-e-Awam University of Engineering, Science and Technology, Nawabshah, Pakistan

[5]Department of Computer Science and Technology, Xi'an University of Science & Technology, China

[6]Department of Computer Science, Bahauddin Zakariya University, Multan, Pakistan

[7]Department of Computer Science, Pakistan Institute of Engineering and Technology, Multan, Pakistan

[6]dr.mubashirali@gmail.com

ABSTRACT

In recent times, Lung cancer is the most common cause of mortality in both men and women around the world. Lung cancer is the second most well-known disease after heart disease. Although lung cancer prevention is impossible, early detection of lung cancer can effectively treat lung cancer at an early stage. The possibility of a patient's survival rate increasing if lung cancer is identified early. To detect and diagnose lung cancer in its early stages, a variety of data analysis and machine learning techniques have been applied. In this paper, we applied supervised machine learning algorithms like SVM (Support vector machine), ANN (Artificial neural networks), MLR (Multiple linear regression), and RF (random forest), to detect the early stages of lung tumors. The main purpose of this study is to examine the success of machine learning algorithms in detecting lung cancer at an early stage. When compared to all other supervised machine learning algorithms, the Random forest model produces a high result, with a 99.99% accuracy rate.

Keywords: Machine Learning, Healthcare, Lung Cancer Detection, SVM, ANN, MLR, Random Forest

1. INTRODUCTION

Cancer is a wide term and the most complex disease in people's life. Cancer can cause by the development of uncontrollable cells which can be spread anywhere in the human body. General cancer symptoms are fatigue, weight loss, pain, cough, fever, lumps, etc. Cancer is also called a tumor. To save the

lives of humans it is very important to detect quickly and correctly the initial stage of cancer. There are many kinds of cancer like breast, brain, lung, liver, skin cancer, etc [1]. Lung cancer is a form of cancer which starts that the abnormal cells cluster together. Lung cancer is the main reason for death among both men and women all over the world. Worldwide after heart disease, Lung cancer is the second most frequent disease. Uncontrollable cell growth may be the reason for the death of the cancer patient [2].

Early examination of lung tumors is critical in decreasing the risk of death from lung tumor disease. Lung cancer can spread very quickly to the brain, liver, bones, and adrenal glands, therefore, the initial diagnosis of lung cancer is very necessary. Every year, almost 1.3 million people pass away due to lung cancer around the world. In 2018, approximately 234,030 Americans will be diagnosed of which 121,480 males and 112,350 females. Due to lung cancer, approximately 121,680 deaths will occur of which 83,550 males and 70,500 females [2][3]. Common ways to diagnose lung cancer include Magnetic Resonance Imaging (MRI), CT (Computed Tomography) scan, and Positron Emission Tomography (PET)[4]. Lung carcinoma is another name for lung cancer. Lung cancer has two categories: SCLC(**Small Cell Lung Cancer)** and NSCLC(**Non-Small Cell Lung Cancer**)[5]. **SCLC:** It's a generally rare form of lung cancer. Around 15% of lung cancer cases have a history of small cell lung cancer. Small lung cancer spread very quickly.

NSCLC: It's a quite frequent kind of lung cancer. Around 85% of lung cancer cases have a history of non-small cell lung cancer. This type of lung cancer is more dangerous than SCLC because it spread slowly. NSCLC is further divided into three categories:

NSCLC has four stages which include Stage |, Stage ||, Stage |||, and Stage |V [6].

In **Stage |**, cancer produces 60 t0 80% in a single lung and doesn't develop in lymph nodes or distant organs.

In **Stage ||**, 30 to 50% of cancer grows in lymph nodes, rather than distant organs.

In **stage |||**, the tumor expands 10 t0 15% in nodes and builds in the center of the chest.

In **stage |V**, the Tumor spread throughout the body.

Many studies have been undertaken on the causes of lung cancer, with the conclusion that smoking is the primary cause. Smoking harms the lungs. In the United States, cigarette smoking is a factor of 80 t0 90% of

deaths due to lung cancer. Other tobacco products such as cigars, pipes, or snuff (a powdered form of tobacco) raise the risk of lung cancer and other kinds of cancer such as mouth cancer. The use of tobacco smoking for a long time begins with lung cancer. According to the survey, smoking caused lung cancer in 90% of males and 75 t0 80% of females. It is a reality that 10 to 15% of people who never smoke in their life contain lung cancer. The reason for lung cancer may be passive smoking, air pollution, asbestos, and radon gas [7].

Machine Learning is an artificial intelligence field that applies computer algorithms to conclude various outcomes from data sets. Machine learning provides guidance to execute complex tasks more creatively and efficiently. Many machine learning algorithms are applied to the data set to measure the accuracy of the diagnosis of lung cancer. In the past survey, many machine learning techniques have been tried to detect and predict lung tumors [3][8][9] Support Vector Machine (SVM), Artificial Neural Network (ANN), Naive Bayes (NB), Convolutional Neural network (CNN), Recurrent Neural network, Decision Tree, k-nearest neighbours.

The innovation of this study is too early diagnoses lung cancer from the data set. This research work on four well-known machine learning classifiers which are SVM, ANN, MLR, and RF to detect lung cancer. In the end, we compared the accuracy and find out the best classifier to detect lung cancer. Figure 1 shows the workflow of our study.

Finally, we determined the goals of our proposed research and compared model findings to determine which model works best. Further work divides into sections. The second section contains the literature review, the third section described the methodology of the proposed algorithms, In the fourth section, analyzed the results of the experiment, and the conclusion and future work are in the last section.

2. LITERATURE REVIEW

Machine learning is a method of diagnosis of lung cancer which is the leading cause of death worldwide. Many machine learning applications are already used in the field of Lung cancer detection. In previous work, many researchers have used image processing methods, machine learning approaches, or hybrid techniques to diagnose lung cancer.

In this paper [8] Author used the clinical images dataset for the image processing methodology. The image processing approach is divided into five different steps:

1) collect images

2) image processing and segmentation

3) extraction of features

4) classification of images

5) performance evaluation.

Seven classifiers as KNN(k-nearest neighbours), SVM(support vector machine), DT(Decision tree), MNB(Multinomial naïve Bayes), SGD(Stochastic gradient descent), RF(Random forest), and MLP(Multi-layer perceptron) was used to evaluate the performance based on the following parameters: accuracy, F1 score, precision, recall and then calculated using a confusion matrix. In the end, MLP obtains higher accuracy with the value of 88.55% as compared to other classifiers.

The major goal of this paper[10] is to use image transformation and machine learning methods to identify lung cancer and its phases. Many steps are involved in the image processing approach, First, 200 CT scan reports and blood samples with CT scan images were used as an input for image processing. Secondly, make a Grayscale image out of an RGB image. To convert the RGB image to a Grayscale image first we find out the average of the colour image and then replace the RGB pixel with the average. Formula of average [average = R+G+B/3]. In the third step, a median filter was used to extract the noise from a grayscale picture. In the fourth step, the grayscale picture is converted to a binary image. The grayscale image has(From 0 to 255)pixel range and the binary image has (0,1). When converting a grayscale picture to a binary image, the threshold value 175 is occupied. In the fifth segmentation step, decrease the irrelevant information from the image and convert the duplicate CT scan image into an informative which is easily examined in detail. In the feature extraction step, the segmented output is used as an input for the feature extraction. Feature extraction is a technique for extracting a pattern from a duplicate image using algorithms. These features, such as Area, Perimeter, and Eccentricity, are provided by feature extraction. In the last step, the Support vector machine (SVM) arranges data and manages the pattern. SVM was used for classification and regression problems and categorized positive and negative samples of lung cancer images. In this paper, the SVM classifier is used as compared to other classifiers like Decision trees, Naïve Bayes, and Random forest because SVM is less time-consuming.

In this study [3], To identify the lung tumor access the dataset from Standard Digital

Image Database. The dataset consists of 247 CT images, 154 images labelled with a nodule, and 93 images labelled without a nodule. The dataset was divided into two parts: training and testing. Randomly, for the training set, 70% of the dataset was employed, while for the testing set 30% of the dataset was employed. PCA(Principal component analysis), KNN(k-nearest neighbours), SVM(Support vector machine), NB(Naïve Bayes), DT(Decision tree), and ANN(Artificial Neural Network) are among the machine learning techniques used to detect lung cancer. Then compare all the classifier's accuracy before and after preprocessing. In the end, ANN has higher accuracy of 82,43% than other classifiers after image processing, and the Decision tree has higher accuracy of 93,24% than other algorithms without image processing.

In this paper [11], To save the lives of humans by analyzing the initial stage of lung cancer many machine learning algorithms are used. Lung cancer datasets are sourced from the UCI Machine Learning Repository and Data. World. The K-fold cross-validation method splits datasets into two parts: training and testing. Many classification methods are used to attain accuracy including Logistic Regression, Decision Tree, Naïve Bayes, and Support Vector Machine is used and then compared the accuracy of each classifier with another classifier. SVM has a higher accuracy of 99.2% than all other classifiers.

In this study [7], Author mostly focused on the segmentation and classification method. An Unsharp mask filter is used to sharpen the image. In image processing, an unsharp mask is used to convert the blurred image into original images. For calculating the background and foreground color intensity level of the image many thresholding methods are used but here we use the OTSU'S thresholding. The mean of OTSU'S thresholding to estimate the following parameters (weight, mean, variance) of background and foreground of all levels. To detect the cancer region and actual affected area we used an Adaptive Canny algorithm and this method was estimated over Otsu's global thresholding method. For the segmentation process, the k-nearest neighbour classification method is used. The KNN method has some effective applications like machine learning, pattern recognition, data mining, object recognition, etc. To calculate the classification results from Bayesian Regularization neural network is used. BRNN attains high accuracy with a value of 99.5%.

The Author [5] has proposed a CAD (Computer-aided diagnosis) system using the

SVM to categorise abnormal or normal images using the CT images dataset. The CAD system has four phases to detect lung cancer from CT images. These stages are preprocessing of images, extraction of features, selection of features, and categorization. In this study, a Gray level co-occurrence matrix was applied to extract image information such as colour, shape, texture, etc, and ANT Colony Optimization was employed as a feature selection method for better outcomes. To categorize the normal and abnormal photos, two machine learning classifiers, ANN and SVM, were applied. ANN achieve high accuracy of 98.40% as compared to SVM's accuracy of 93.2%.

Furthermore,[12] Deep Learning Approach is employed to diagnose lung cancer at the initial stages which are very important to decrease the possibility of death. In this study, the TensorFlow library is used to diagnose lung cancer. SPIE-AAPM dataset used consists of 70 CT images. The training set is made up of 40 photographs, while the test set is made up of 30 images. 3D Convolutional Neural Network is utilized to classify the lung nodules which can be benign or cancerous. To calculate the performance of the model 30 CT images are used as a resulting in which 17 patients are malignant and 13 patients are benign. Several metrics are used to assess the performance of the classification method but in this work, a confusion matrix is employed to evaluate the achievement of binary classification. The Confusion Matrix shows 70% accuracy of 3D CNN.

This work [13] has proposed the deep residual network to detect lung cancer using the CT scan images dataset that is in DICOM (digital imaging and communication in medicine) image format. First, perform the preprocessing method on the dataset. In preprocessing, to determine and split lung nodules perform many regions growing and morphological operations and then used the UNet and ResNet models for feature extraction. In segmentation, analyze the distinctive features from the CT scan images. In this work two tree-based classifiers as XGBoost and RandomForest used. In the end, compared the accuracy of the proposed models. UNet+RandoomForest individually get 74% accuracy, ResNet+XSBoost achieved 76% accuracy, and the combination of UNet+RandoomForest and ResNet+XSBoost obtain 84% accuracy.

The Multi-class SVM classifier is used in the work [14] to detect and predict lung cancer. In this work, MATLAB was used for built an algorithm to detect and predict lung cancer. In each phase of classification, image

enhancement and segmentation were done separately. Image enhancement is an approach for improving the quality and information of a data image. Many approaches such as image scaling, spatial transformation, and contrast enhancement have been promoted to improve images. Watershed transformed is used for segmentation to achieve a better resolution in CT images and Gray Level Co-Occurance method is used for feature extraction. For classifying cancer nodules SVM classifier is used that is used for both classification and relapse purposes. Our proposed method achieves higher accuracy.

This research work [15] addressed various segmentation approaches such as Hopfield Neural Network (HNN) and Fuzzy C-Mean (FCM) clustering algorithms methods. To diagnose the early stage of lung cancer used Sputum color images. In this study, a thresholding classifier was used to handle the problem of extreme fluctuation in the grey level and the related contrast between the pictures with 98% accuracy. The results of HNN Segmentation exceed the results of FCM because FCM failed in extracting the nucleus and cytoplasm areas.

[16]Developed a computer-assisted detection approach to distinguish between malignant and non-cancerous tissue in DICOM images.CNN was used for feature extraction. Many machine learning classifiers were used to classify the cancerous and normal images and compared their accuracy with each other to determine which one is best for diagnosing lung and pancreatic tumors.

An Artificial Neural Network [17] was developed to identify the presence and absence of lung cancer. ANN uses the symptoms dataset to perform two layers of training and validation.ANN attains the highest 96.67% accuracy which shows that ANN can detect lung cancer.

A group of Authors [18] proposed a new system to improve the current model’s accuracy. The current model has 88.4% accuracy To increase the accuracy of the existing model, a median filter is employed in the proposed system's preprocessing stage to remove salt and paper noise from CT pictures, which causes challenges in accurate lung cancer identification. After this gaussian filter applies to smooth the images and remove the speckle noise from CT images. In the segmentation step, a watershed technique was utilized to divide the image into areas to examine the meaningful information. In the feature extraction step, extracting features such as area, perimeter, centroid, diameter, eccentricity, and Mean intensity were employed as training features to build the

classifier. The SVM classifier was utilized to identify the nodule as malignant or beginning in CT scan images. The proposed model has higher accuracy of 92% than the current model.

In this paper [19] image processing technique was used to classify the cancerous and non-cancerous nodules using MRI, CT, and Ultrasound images but in the experiment section 6 CT Scan images and 15 MRI images of the lung are used. The image processing method is used to decrease the noise and distortion in the image which is very beneficial for the segmentation step. For both CT and MRI pictures, the Gabor filter was applied to improve the pictures. In the image enhancement step, smoothing of the image, blurring, and elimination of noise is involved. For the next step, picture filtering is highly effective. After image processing, a canny filter is used for edge detection. To decrease the unnecessary detail of images data morphological processing techniques like Erosion and dilation are used. For image segmentation, Superpixel Segmentation was used. Feature extraction is a very important stage and used many algorithms to produce the final output in which we conclude the normality and abnormality of the image. For feature selection and feature classification through GUI (Graphical user interface) in MATLAB are used different algorithms such as Particle swarm optimization (PSO), Genetic Optimization, and SVM(support vector machine) give result as normal or abnormal nodule and attains 89.5% accuracy.

[20] Proposed five steps of image processing technique for the analysis of lung images. In **1st step,** collected data contain the CT scan images of patients who have cancer and non-cancer. In the **2nd step,** the image preprocessing technique is applied which is very important and beneficial for the segmentation of images. In image preprocessing, 2 steps involve the first one is to use the median filter to remove the noise from the images dataset and the second one is to use the contrast adjustment technique to enhance the refined images. In the **3rd step,** an input image is segmented into several parts and morphological operations are utilized to select the appropriate area of interest (ROI) based on the structuring element. After this, masked it to extract the tumor region from the CT scan image. In **the 4th** step, after the segmentation step, to gain the tumor region calculate the features of the tumor-like area, perimeter, and eccentricity and then send them to the classifier. In the **5th step,** an SVM classifier is employed to classify the picture dataset as cancerous or benign.

3. METHODOLOGY

Nowadays, detecting lung cancer is a more challenging task for radiologists but an intelligent computer-aided system helps the radiologist to detect and predict lung cancer very well. Machine learning plays a very important function to detect lung cancer [14]. In literature, many image analyses and machine learning techniques are used to diagnose lung cancer. In this section, we describe our proposed model using different Supervised machine learning methods to diagnose the early stage of lung cancer. Our proposed system is divided into two sections which are as fellow:

- Description of data
- A detailed description of the proposed architecture

3.1 Description of data

The dataset which we used in this study was obtained from the data. world (https://data.world/cancerdatahp/lung-cancer data). The obtained dataset is in CSV format.

There are 1000 instances in this dataset, as well as 25 data columns, 24 of which are predictive and one of which is a class label. 25 column is a class label that contains the high, medium, and low text data. We used the LabelEncoder of the SciKit Learn library in Python to convert the text data of class labels into numbers (0,1,2). In this dataset, many attributes are used as a symptom to classify lung cancer detection. Many variables are employed as a symptom in this dataset to classify lung cancer detection. The name of a property is Occupational Hazards, Dust Allergy, Chronic Lung Disease, Balanced Diet, Genetic Risk, Age, Gender, Air Pollution, Alcohol Use, Occupational Hazards, Genetic Risk, Chronic Lung Disease, Balanced Diet, Obesity is a disease that is affects both men and women. Smoking, chest discomfort, bloody cough Fatigue, passive smoking, weight loss, shortness of breath, finger clubbing Frequent colds, wheezing, difficulty swallowing, dry coughs, and snoring are all symptoms that can indicate lung cancer. In the class label, the '2' value signifies the "Malignant cancer", the '1' value signifies the ''Benign cancer'', and the value '0' signifies the '' No cancer''.

3.2 A detailed description of the proposed architecture:

In this proposed architecture, in figure 2 we are describing the workflow of our proposed work. First, we insert the dataset in the form of CSV. Then, we performed the preprocessing on the dataset to remove the noise. After this, we divided the dataset into two sections: training and testing. Then, using models like ANN, SVM, Random Forest, and Multiple Linear Regression, we

classified our dataset into three categories: malignant tumor, benign tumor, and healthy person with no tumor. If the class label value '2' means Malignant cancer, the class label value '1' indicates Benign cancer, and the class label value '0' signifies No cancer. And results are discussed in Result Section and we compared the results of our models and find out which one is best for diagnosing lung cancer.

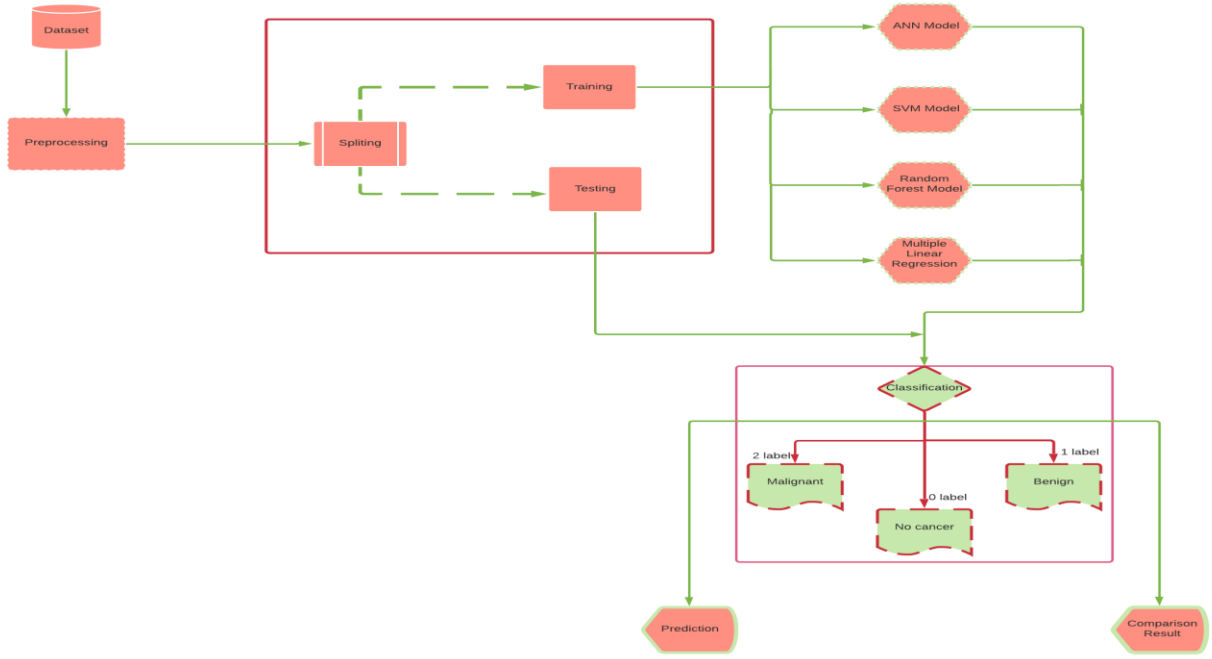


Figure 1. Proposed Methodology of Lung Cancer Detection using Machine Learning Techniques

4. ANALYSIS AND RESULTS

Previous we described our dataset and model working which are used in our research work. In this part, we are visualizing the original data using the Bar graphs. These bar graphs show the level Vs attributes of the dataset. Fig 2a shows the level vs max-age. According to this fig High level has a max-age 64, the Low level has a max-age 62, and the Medium level has a max-age 73.

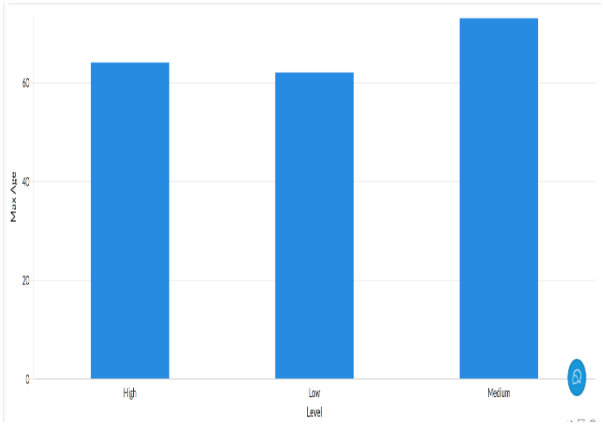


Figure 2a: Level VS max-age

Level(High, Low, Medium) Vs age, air pollution, and alcohol use of data visualization is shown in figure 2b.

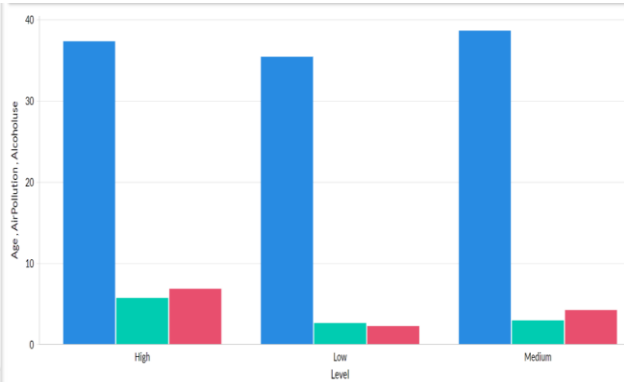


Figure 2 b: Level VS Age, Air Pollution & Alcohol Use

Level(High, Low, Medium) Vs dust allergy, occupational hazards, and genetic risk of data visualisation is shown in figure 2 c.

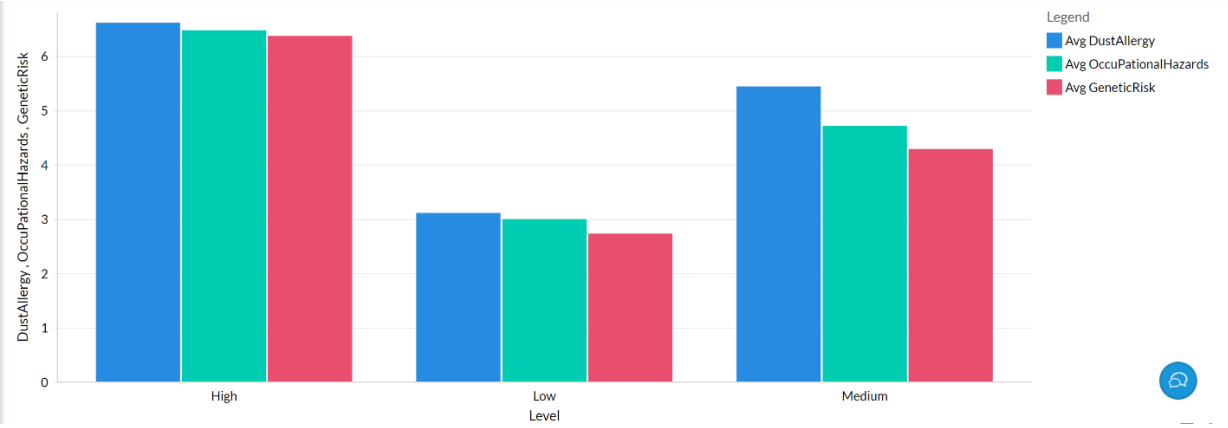


Figure 2 c: Level VS dust allergy, occupational hazards, genetic risk

Level(High, Low, Medium) Vs chronic lung disease, balanced diet, and obesity of data visualization is shown in figure 2 d.

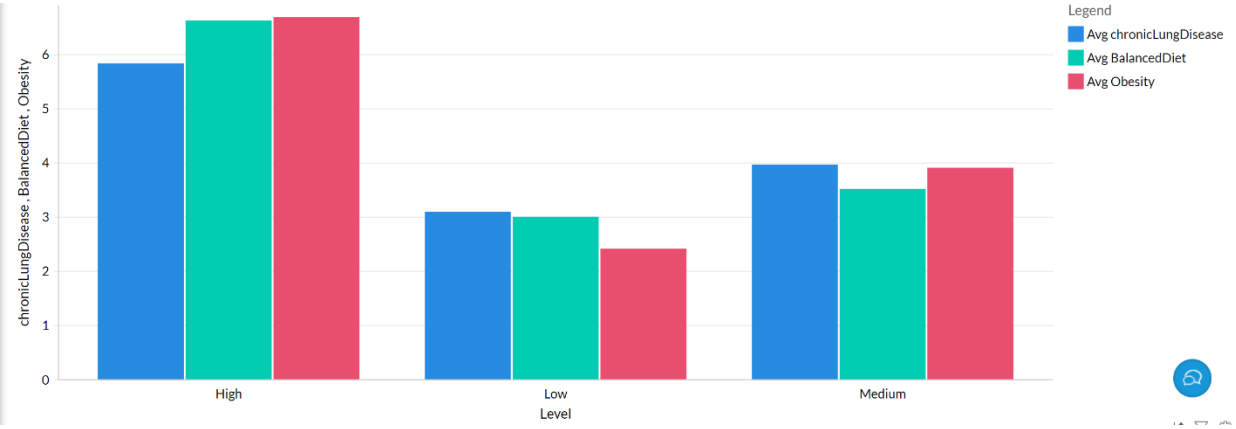


Figure 2 d: Level Vs genetic risk & occupational hazards

Level(High, Low, Medium) Vs smoking, passive smoker and chest pain of data visualisation shown in figure 2 e.

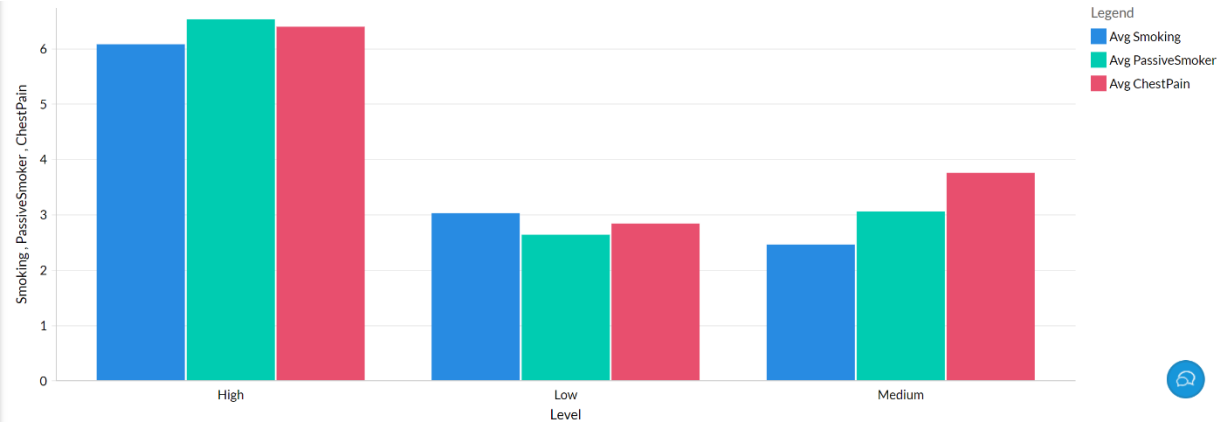


Figure 2 e: Level VS chronic lung disease & balanced diet

Level(High, Low, Medium) Vs coughing up blood, fatigue, and weight loss of data visualization shown in figure 2 f.

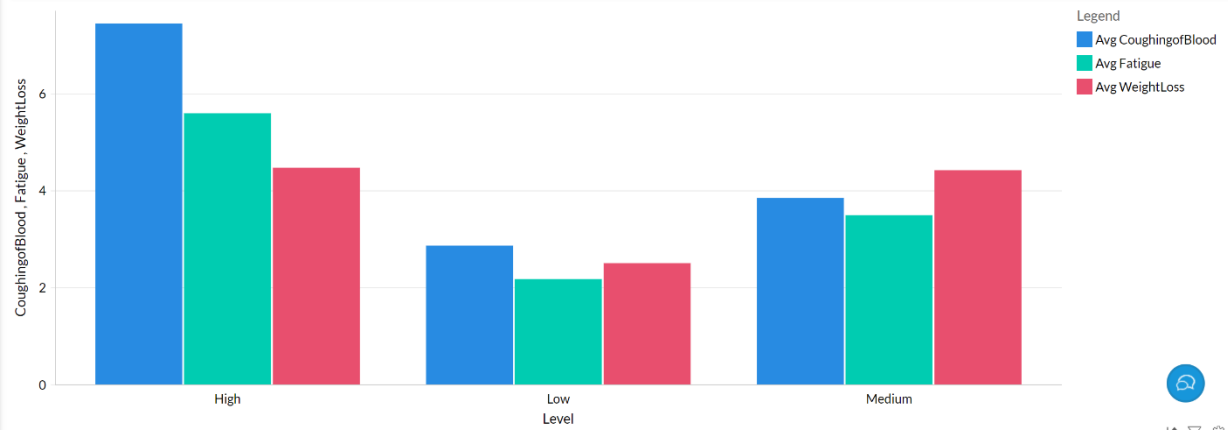


Figure 2 f: Level VS obesity & smoking

Level(High, Low, Medium) Vs shortness of breath, wheezing and swallowing difficulty of data visualization shown in figure 2 g.

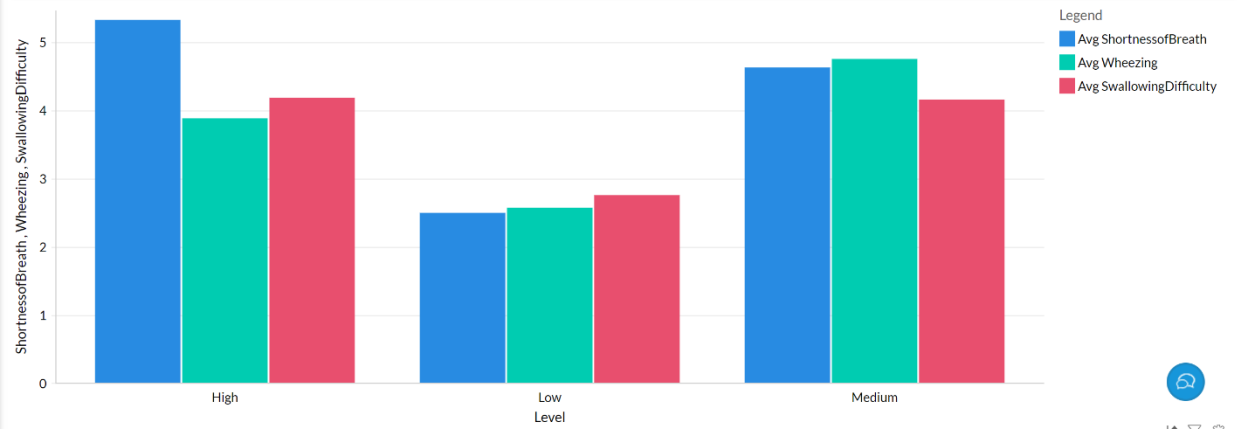


Figure 2 g: Level Vs passive smoker & chest pain

Level(High, Low, Medium) Vs clubbing of fingernails, a frequent cloud, dry cough, and snoring of data visualisation shown in figure 2 h.

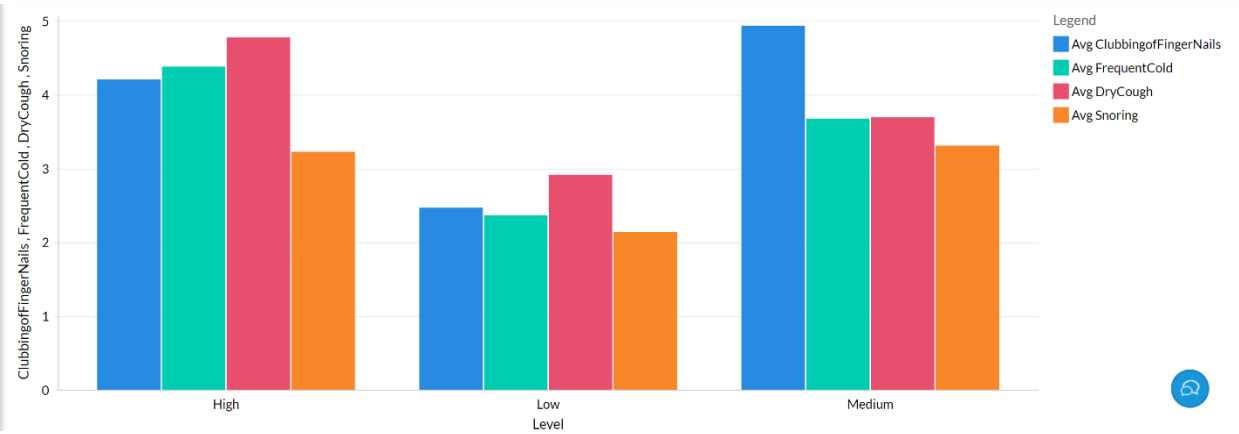


Figure 2 h: Level VS coughing up blood & fatigue

4.1 Performance Evaluation Table:

In this section, we'll create a table that compares the performance of the various machine learning algorithms employed in this study. Random forest beats all other supervised machine learning methods, including multiple linear regression, artificial neural networks, and Support vector machines, as shown in Table 1. For this dataset, the Random forest outperforms other classification algorithms in terms of accuracy.

Machine Learning Algorithms	**Accuracy (%)**
ANN (Artificial neural network)	65.75
MLR (Multiple linear regression)	77.54

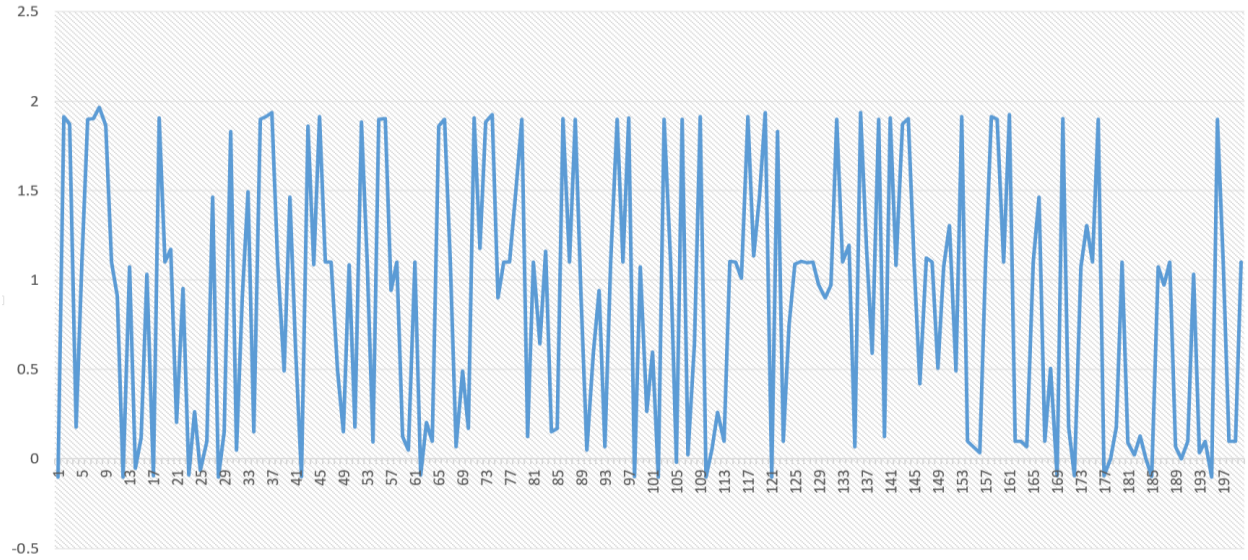
RF (Random forest)	99.99
SVM (Support vector machine)	98.91

Table 1. Performance evaluation table of supervised machine learning algorithms

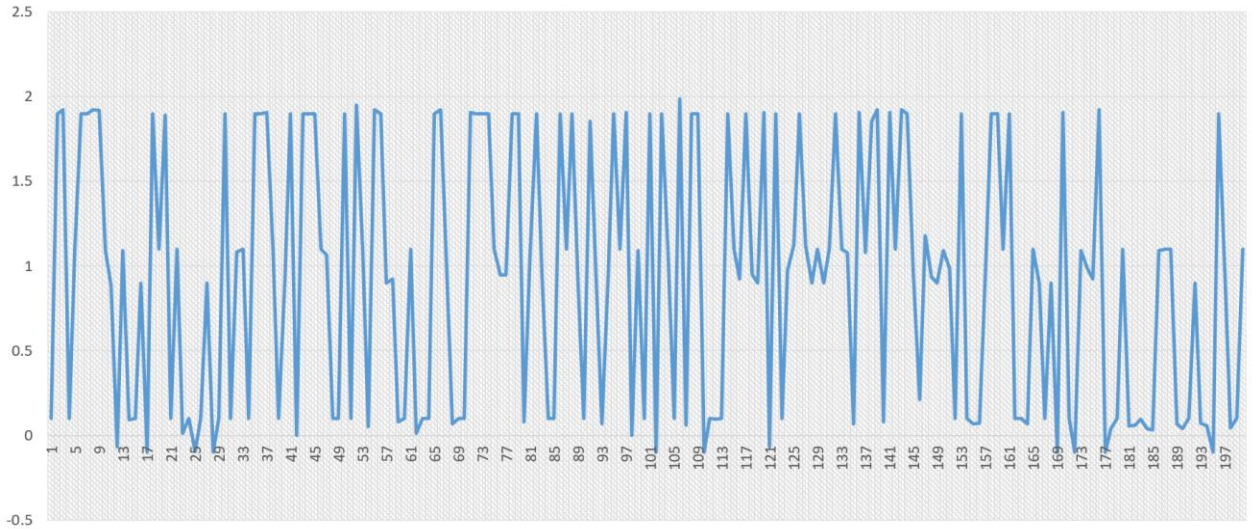
5. RESULT VISUALIZATION

The outcomes of the proposed models such as Artificial neural networks, Multiple linear regression, Random forest, and Support vector machines are visualized in this part. In this dataset, the ANN model performs poorly. This is mainly because ANN is a simple neural network only with limited connected layers. Multiple linear regression model performance is better than ANN model result. But the Random forest model has the highest performance on this dataset of all other machine learning algorithms.

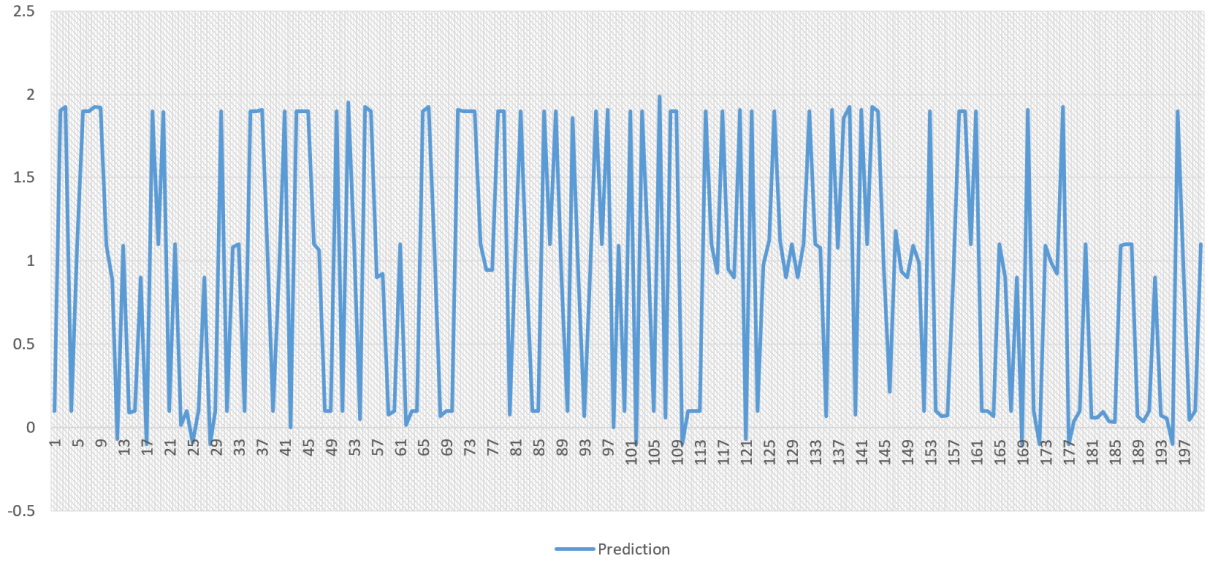
Result visualization is done on the lining graph using the excel sheet. Figure 5 graphs show the prediction results of the models mentioned previously on the Lung Cancer dataset.



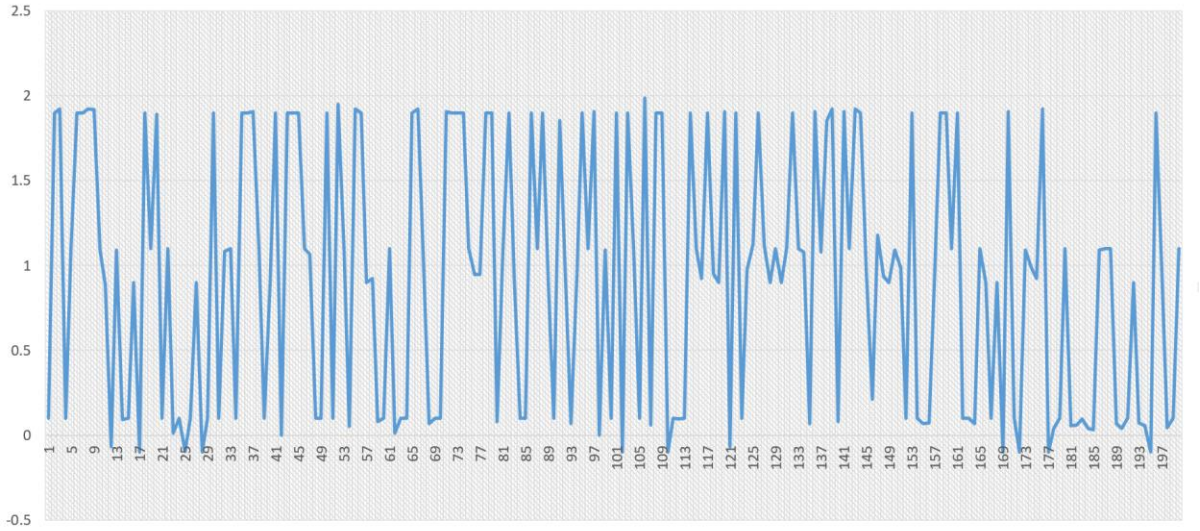
(a) Prediction result of ANN model



(b) Prediction result of Multiple linear regression



(c) Prediction result of Random forest model



(d) Prediction result of Support vector machine model

6. CONCLUSION AND FUTURE WORK

In previous times, Lung cancer detection was a very difficult and time-consuming task because a doctor has done multiple tests on a patient to confirm whether the patient has lung cancer or not. Nowadays, many machine learning classifiers have become more important for diagnosing and detecting lung cancer. The main goal of this work is to find a way to detect lung cancer early. Lung nodule classification is benign, malignant, and has no cancer. The supervised machine learning methods utilized in this study are the Artificial Neural Network, Support Vector Machine, Random Forest, and Multiple Linear Regression. As seen in the performance chart, each machine learning classifier delivers different outcomes on the same lung cancer dataset. When compared to other supervised machine learning algorithms, the Random Forest model gives the best result with an accuracy of 99.99 per cent, according to the proper classification and confusion matrices.

For future work, Additional preprocessing improve the accuracy rate of other supervised machine learning algorithms.

REFERENCES

1. T. Saba, “Recent advancement in cancer detection using machine learning: Systematic survey of decades, comparisons and challenges,” *J. Infect. Public Health*, vol. 13, no. 9, pp. 1274–1289, 2020, DOI: 10.1016/j.jiph.2020.06.033.
2. V. A. Binson and M. Subramoniam, “Advances in Early Lung Cancer

Detection: A Systematic Review," *2018 Int. Conf. Circuits Syst. Digit. Enterp. Technol. ICCSDET 2018*, pp. 1–5, 2018, doi: 10.1109/ICCSDET.2018.8821188.

3. Ö. Günaydin, M. Günay, and Ö. Şengel, "Comparison of lung cancer detection algorithms," *2019 Sci. Meet. Electr. Biomed. Eng. Comput. Sci. EBBT 2019*, 2019, doi: 10.1109/EBBT.2019.8741826.
4. M. ullah, M. Bari, A. Ahmed, and S. Naveed, "Lungs Cancer Detection Using Digital Image Processing Techniques: A Review," *Mehran Univ. Res. J. Eng. Technol.*, vol. 38, no. 2, pp. 351–360, 2019, doi: 10.22581/muet1982.1902.10.
5. R. Kohad and V. Ahire, "Application of Machine Learning Techniques for the Diagnosis of Lung Cancer with ANT Colony Optimization," *Int. J. Comput. Appl.*, vol. 113, no. 18, pp. 34–41, 2015, doi: 10.5120/19928-2069.
6. U. Ahmed, G. Rasool, S. Zafar, and H. F. Maqbool, "Fuzzy rule based diagnostic system to detect the lung cancer," *2018 Int. Conf. Comput. Electron. Electr. Eng. ICE Cube 2018*, pp. 1–6, 2019, doi: 10.1109/ICECUBE.2018.8610976.
7. R. Patra, "Prediction of lung cancer using machine learning classifier," *Commun. Comput. Inf. Sci.*, vol. 1235 CCIS, pp. 132–142, 2020, doi: 10.1007/978-981-15-6648-6_11.
8. G. A. P. Singh and P. K. Gupta, "Performance analysis of various machine learning-based approaches for detection and classification of lung cancer in humans," *Neural Comput. Appl.*, vol. 31, no. 10, pp. 6863–6877, 2019, doi: 10.1007/s00521-018-3518-x.
9. B. Chapaliuk, "' Etection 6 Ystem for / Ung & Ancer," *2018 IEEE First Int. Conf. Syst. Anal. Intell. Comput.*, pp. 1–4, 2018.
10. W. Rahane, H. Dalvi, Y. Magar, A. Kalane, and S. Jondhale, "Lung Cancer Detection Using Image Processing and Machine Learning HealthCare," *Proc. 2018 Int. Conf. Curr. Trends Towar. Converging Technol. ICCTCT 2018*, pp. 1–5, 2018,doi:10.1109/ICCTCT.2018.8551008.
11. P. R. Radhika, R. A. S. Nair, and G. Veena, "A Comparative Study of Lung Cancer Detection using Machine Learning Algorithms,"

Proc. 2019 3rd IEEE Int. Conf. Electr. Comput. Commun. Technol. ICECCT 2019, pp. 1–4, 2019, doi: 10.1109/ICECCT.2019.8869001.

12. E. Cengil and A. Çinar, "A Deep Learning Based Approach to Lung Cancer Identification," *2018 Int. Conf. Artif. Intell. Data Process. IDAP 2018*, 2019, doi: 10.1109/IDAP.2018.8620723.
13. G. S. Rao, G. V. Kumari, and B. P. Rao, *Network for Biomedical Applications*, vol. 2, no. January. Springer Singapore, 2019.
14. J. Alam, S. Alam, and A. Hossan, "Multi-Stage Lung Cancer Detection and Prediction Using Multi-class SVM Classifie," *Int. Conf. Comput. Commun. Chem. Mater. Electron. Eng. IC4ME2 2018*, pp. 1–4, 2018, doi:10.1109/IC4ME2.2018.8465593.
15. F. Taher, N. Werghi, H. Al-Ahmad, and R. Sammouda, "Lung Cancer Detection by Using Artificial Neural Network and Fuzzy Clustering Methods," *Am. J. Biomed. Eng.*, vol. 2, no. 3, pp. 136–142, 2012, doi: 10.5923/j.ajbe.20120203.08.
16. D. Tree, "LUNG AND PANCREATIC TUMOR CHARACTERIZATION IN THE DEEP LEARNING Page No : 123," vol. XVI, no. 123, pp. 123–127, 2020.
17. I. M. Nasser and S. S. Abu-naser, "Lung Cancer Detection Using Artificial Neural Network," vol. 3, no. 3, pp. 17–23, 2019.
18. S. Makaju, P. W. C. Prasad, A. Alsadoon, A. K. Singh, and A. Elchouemi, "Lung Cancer Detection using CT Scan Images," *Procedia Comput. Sci.*, vol. 125, no. 2009, pp. 107–114, 2018, doi: 10.1016/j.procs.2017.12.016.
19. A. Asuntha, A. Brindha, S. Indirani, and A. Srinivasan, "Lung cancer detection using SVM algorithm and optimization techniques," *J. Chem. Pharm. Sci.*, vol. 9, no. 4, pp. 3198–3203, 2016.
20. D. Nagajyothi, R. Addagudi, T. Gunda, S. S. Logitla, and G. Arun, "Detection of lung cancer using SVM classifier," *Int. J. Emerg. Trends Eng. Res.*, vol. 8, no. 5, pp. 2177–2180, 2020,doi:10.30534/ijeter/2020/11385 2020.