Prediction of Lung Cancer using Ensemble Classifiers

G Ashwin Shanbhag¹, K Anurag Prabhu², N V Subba Reddy, B Ashwath Rao

Department of Computer Science and Engineering,
Manipal Institute of Technology, Manipal
Academy of Higher Education,
Manipal, Karnataka- 576104

¹ ashwingshan98@gmail.com
² anuragprabhu122@gmail.com

Abstract. Carcinoma detection from CT scan images is extremely necessary for numerous diagnostic and healing applications. Because of the excessive amount of information in CT scan images and blurred boundaries, tumor segmentation and class are extremely laborious. The intention is to categorize carcinoma into benign and malignant categories. In MR pictures, the number of facts is a lot for interpreting and evaluating manually. Over the previous few years, carcinoma detection in CT has grown to be a rising evaluation space in the area of the scientific imaging system. Correct detection of length and site of lung cancer performs a vital position in the designation of carcinoma. In this paper, we introduce a novel carcinoma detection methodology that helps in predicting the carcinoma from the CT scanned images. The methodology has 4 different stages, pre-processing the image data, segmentation, extracting features, and classification stage to categorize the benign and malignant. This work makes use of extraordinary models for detecting carcinoma in a CT test via way of means of constructing an ensemble classifier. Techniques proposed in the paper helped us achieve an accuracy of 85% using Ensemble-Classifier which showcases that model has the capability of predicting the malignant cases correctly. The ensemble classifier consists of 5 machine learning models like SVM, LR, MLP, decision tree, and KNN. The inevitable parameters like accuracy, recall, and precision is calculated to determine the accurate results of the classifier.

Keywords: Cancer Detection, Feature-Extraction, Classification, Machine-Learning Models, Segmentation, Ensemble-Techniques, lung cancer, Ensemble-Classifier, Random Forest.

1. INTRODUCTION
Cancer is a damaging sickness in human life. The fundamental necessary task of the lungs is to require the element to the frame and to put off CO₂ from the body during the very essential activities. Carcinoma occurs due to the out-of-control proliferation of tissues and cells in the lungs. Carcinoma is that the first type of most cancer that reasons dying amongst males and consequently the second type of most cancers amongst females near 1.3 million oldsters die every 12 months in a global way to Carcinoma [1]. The exclusive classes of tumors are benign (non-cancerous) and malignant (cancerous). There is diverse sort of cancer such as Colon Cancer, Carcinoma, Leukemia, etc. [2]. The occurrence of carcinoma has notably multiplied for the reason that early nineteenth century. There is a diverse motive of carcinoma such as smoking, second-hand smoking, exposure to gases like radon, publicity to asbestos, etc. Carcinoma has kinds of NSCLC (Non-Small Cell Lung Cancer) and SCLC (Small Cell Lung Cancer). The dangerous nodules can be detected at partner diplomas in advanced degrees with the aid of using the radiologist's mistreatment automated tomography (CT) and opportunity scanning strategies [3]. They arise within the bronchi near the middle of the chest. Symptoms brought about because of carcinoma are
shortness of breath with activity, fatigue, speech defect, dysphasia, coughing up blood, weight loss, ache withinside the shoulder, chest, arm [4].

The important mission of spotting carcinoma within the early tiers may be very hard considering signs and symptoms seem best within the subsequent superior tiers which reasons the mortality charge of carcinoma to peek amongst all different classes of most cancers. The accurate designation for diverse styles of carcinoma performs an essential function to the doctors to assist them in figuring out and choosing the proper treatment [5]. The picks created with the aid of using the doctors are the most vital elements in designation but recently, the utility of diverse AI class strategies is evidenced in helping doctors to facilitate their techniques. Possible mistakes which can also additionally arise due to unskilled doctors are regularly reduced with the aid of using mistreatment class strategies [6].

Machine getting to know (ML) is a utility of AI which suggests flexibility to robotically study and enhance its performance by experience without programming. ML classifiers are very famous for detecting breast and lung cancers. ML algorithms can be differentiated into three different categories namely Supervised, Unsupervised, and Reinforcement Learning. We used an exclusive model for detecting carcinoma in CT test pictures with the aid of using an ensemble classifier that consists of five exclusive ML supervised algorithms like decision tree, KNN, SVM, RF, MLP, logistic regression, etc. to get more accurate results [7]. ML enables to lessen the wide variety of studying measures in hospitals and clinics. The important goal of this paper is to categorize carcinoma detection into benign and malignant classes. The proposed technique has four tiers: pre-processing of CT test pictures used for noise filtration, segmentation using 'Otsu' thresholding, feature extraction is used to extract a variety of features like area, perimeter, centroid, etc. A predictive model like carcinoma prediction is used to generate alternatives for classifying and making use of implemented arithmetic evaluation on the samples [8].

Ensemble classifiers give the concept of combining decisions from different models to improve the performance and accuracy measures. Ensemble methods usually give more accurate results than a single model. The mechanism for improved performance with ensembles is usually the reduction within the variance part of prediction errors created by the causative models [9].

The planned work for tumor detection within the body uses ML strategies. After extracting the features, ML models are applied for extracting sensitive values and acknowledging tumor cells [10]. Pre-analysis will assist us to identify or narrow down the chance of screening for carcinoma malady. Symptoms and threat elements like smoking, alcohol consumption, obesity, etc had a statistically crucial effect within the pre-analysis degree [11]. The carcinoma diagnostic and prognostic troubles are basically within the scope of the huge range of classification issues [9].

In this paper, we proposed different methods of predicting carcinoma from CT scanned pictures by using the ensemble classifier, and the outcomes of the model with results are analyzed. The rest of the work can be organized as: section 2 offers the brief idea of previously carried works in the field of prediction, section 3 widely discusses the different methods of carcinoma prediction, section 4 offers a certain evaluation of the record to aid the proposed methodology, section 5 concludes the paper, and section 6 has the references used on this paper.

2. LITERATURE SURVEY
A lot of works have already been proposed for the prediction of carcinoma through numerous researchers amongst them, paper [1, 2] projected several approaches for police to detect carcinoma in the early stages. They implemented different ML models like K-NN, Artificial Neural Networks, SVM, Naive Bayes, and
Decision Trees to know the idea of using these approaches, and comparison of results are performed while pre-processing and after it is done. Sumathipala et al., [3] deliberate a version wherever the picture statistics is taken from LIDC-IDRI, as soon as grouping the picture statistics picture filtration has been enforced, filtration is finished supported the affected person United Nations company went through diagnostic check and module degree is good enough to thirty and so snapshots whose module degree is good enough to thirty is split and for predicting they used random forest and logistic regression model.

Paper [4, 5] gives a carcinoma detection machine victimization picture technique and device learning are hired to categorize the presence of carcinoma at some point of a CT- scan images and blood samples. Fenwa et al., [6] proposed a version in which they extracted the properties like brightness, the contrast from picture dataset. The use of extracting texture-primarily based features and on the one’s varieties of ML set of rules are carried out one is ANN any other one is SVM, after which overall performance has been evaluated on each the set of rules to examine which set of rules is giving extra accuracy. In the work [7, 9] goal of the work is to propose a version for early detection and the right designation of the illness which could facilitate the health practitioner in saving the life of the affected person. Maisa Daouda et al., [8] in their work discussed how neural network models can be used in the prediction of cancer and also addressed some issues that are generally encountered while building neural network models for predicting cancer. M.Siddardha Kumar et al., [10] projected pre-managing strategies are likewise applied at some point of this paintings to urge accurate outcomes. In preprocessing approach, the morphological approach has been applied to expel the unwanted statistics from the image. The feature extraction system it's an accustomed restriction the only in all a type dataset via way of means of manipulating a few changed over alternatives. Exceptional Strategies were implemented to find the different ways for extracting geometrical and measurable properties from the image.

Swati Mukherjee et al., [11] the evaluation and study of respiration organ sicknesses has been the most interesting research area of docs from time to present. They addressed this issue by introducing a novel methodology using deep neural mechanisms such as CNN and AI. The paper [12, 13] aims at detection, prediction, and diagnosing of carcinoma has to turn out to be vital as it simplifies resultant clinical board. Wasudeo Rahane et al., [14] Kyamelia Roy et al., [15] aims to erect the development and pills of cancerous situations machine learning strategies are applied because of their accurate outcomes. Various sorts of ML algorithms like, Naive Thomas Bayes, SVM, logistic regression, are implemented in the core region for evaluation and diagnosis of carcinoma. Sanjukta blue blood Jena et al., [16] projected a model consisting of five kinds of feature extraction strategies that had been applied in individual class method to expect at that alternatives extraction approach that system gaining knowledge of method is giving quite a few accuracies. Şaban Oztürk et al., [17] type of histopathologic snapshots and identity of cancerous regions is kind of difficult due to picture heritage first-rate and determination. The difference between conventional tissue and cancerous tissue is extraordinarily tiny in a few cases. So, the alternatives of the tissue patches within the picture have key significance for computerized type.

Dendi Gayathri Reddy et al., [18] projected a version this is cost-effective in predicting the ranges of respiration organ malignant neoplastic disorder through applying the thoughts of cc algorithms. It is a mixture of KNN, decision tree, and NN models beside fabric ensemble technique that helps to boost the accuracy. The obtained results were better than other existing models. Paper [19, 20] talks about the numerous devices gaining knowledge of algorithms that are applied to predict the survivability price of a person, and performance is estimated primarily based on root suggest square error. Using those concepts, we introduce a unique technique to are expecting cancer the use of ensemble strategies which might be mentioned in element within the subsequent section.
3. METHODOLOGY
In this section we discuss the detailed approach for predicting lung Cancer from CT scanned images by extracting the region-based features and an ensemble classifier. The blueprint of the process is represented in Figure 1.

![Figure 1. Process of Predicting cancer from images](image)

3.1 Pre-Processing Layer
Pre-processing is a process that is generally used to increase the quality of the image from all perspectives. Pre-processing also refers to the reduction of noise from the image and also evacuate the unwanted segments using various filters and techniques [16]. As the images in the dataset have some noise in them and some regions are not required. We have initially cropped the margins of CT scan images and then we resized the images by applied OpenCV and Gaussian Blur noise filtration method is used to denoise the input image.

3.2 Segmentation Layer
Segmentation of an Image is the process of dividing an image into multiple segments (image objects). Here to segment the image we use the thresholding technique Otsu which makes binarization of the image and gives us the threshold value that is appropriate to binarization of the image. Then using that threshold value, we segment the image to get the appropriate region. We generated a threshold value based on the intensity of the required region. After this, we remove all the unwanted edges that are present along the border and can become additional noise in the image. Then we generate labels for the obtained region and features are extracted from that region.
3.3 Feature Extraction Layer

The image after segmenting is passed on to an extracting layer where properties are extracted from the labeled regions, we then extract different region-based properties such as area, perimeter, centroid, mean-intensity, solidity, and eccentricity from the labeled regions. The area refers to the number of pixels of a region, perimeter refers to the distance around the boundary of each region, centroid in context is a center of mass of each region which is in the form of a 1x2 vector, mean refers to the average intensity, solidity refers to the ratio of pixels in the region to pixels of the convex hull, and eccentricity refers to the ratio of focal distance over the major axis length where 0 refers to a circle and 1 says that it is straight line using an inbuilt function which computes the perimeter by calculating the distance between each adjoining pair of pixels around the border of the region [17]. These extracted features are used as data for classification.

3.4 Classification Layer

After applying the feature extraction technique, we can then observe the classification method to classify the tumor into benign or malignant by the usage of specific ensemble techniques. After making use of the category method, it is easy to predict whether the tumor is cancerous or not, and this result gives correct prediction [15]. Below stated are different ML models which are used to create a simple ensemble classifier.

SVM [12] is used for prediction, regression, and classification for the given input data. It classifies the input dataset via way of means of introducing a boundary called a hyper-plane that separates the dataset into components. SVM is used to differentiate linear and non-linear areas. A linear separation classifier is hired to split the affected and non-affected areas within the image [14]. In non-linear separation, we’re going to separate the affected element or place via way of means of representing the non-linear type. Logistic Regression [13] (LR) is a famous modeling process utilized in the evaluation of epidemiologic datasets. The LR approach initially calculates the usage of logistic characteristics, learns the coefficients for an LR model, makes accurate predictions. LR is also referred to as a binary classifier that calculates the category of classification which is primarily based on the speculation that works using the sigmoid function.

Decision trees [13] is another technique that is most commonly used for classification. It is a classifier that has a structure of tree where features of a dataset are represented using internal nodes decision rules are represented using branches and outcomes are represented using leaf nodes. It graphically represents all the possible solutions that can be obtained for a problem based on the specified conditions. MLP [18] A multilayer perceptron (MLP) is a feed-forward neural community that generates hard and fast outputs from a hard and fast input. MLP makes use of backpropagation for training the community. To create predictions, backpropagation techniques are incorporated into the neural network. For a given instance the edge weights are corrected by backpropagating the errors. The neural network model has three layers namely input hidden and output layer. The number of features in the dataset is generally considered as the number of neurons in the input layer. There can be n hidden layers and n neurons in each hidden layer based on the requirement and the number of neurons on output layers are generally the number of classes or labels of the given dataset. K-Nearest Neighbor [18] is a statistical technique used for classification. ‘K’ nearest training samples within the featured house acts as associate input. The expected class that the associate item belongs to, relies upon the class of the associates around it.
Using the above discussed five machine learning models, we construct a classifier via the utilization of an ensemble technique called max-voting [7] as shown in Figure 2. The choice ensemble technique may be a common instance of the multi-professional method that enables a combination of different classifiers in a parallel manner. After, each classifier is skilled in all facts and takes part in the decision. At last, the voting approach enables to obtain the unfinished solution. This will help to increase the accuracy by combining the advantages of each classifier. The mechanism for improved performance with ensembles is commonly the reduction in variance part of prediction errors created by the conducive models.

![Ensemble-Classifier](image)

**Figure 2.** Ensemble-Classifier

Random Forest [20] is a technique that generates plenty of decision trees that are allowed to break up arbitrarily from a seed. This finally ends up like a forest of arbitrarily generates the decision trees. Final results of decision trees are ensembled using Random-Forest algorithmic utility that is anticipating to offer extra accuracy in preference to one tree will alone be giving. Individual decision trees are like if-then-else tips that may be generated from the dataset directly.

Then the result of the built ensemble classifier and RF is discussed in the next section. The stepwise algorithm is shown below:

**Input:** Standard CT scan image data  
**Output:** classify the image as Malignant or Benign

**Step 1:** Take the CT scanned image dataset as Input  
**Step 2:** Then denoise the image using noise filtering techniques  
**Step 3:** Generate a threshold value using the Otsu Thresholding technique.  
**Step 4:** Then segment the image based on the obtained threshold value  
**Step 5:** Remove all the unwanted edges that are present along the border.  
**Step 6:** Label the segmented Image  
**Step 7:** Then from the labeled image extract Region-based features like area, perimeter, centroid, solidity, mean, eccentricity is extracted.  
**Step 8:** Use Classification methods for training and predicting the given image as Malignant or Benign  
**Step 9:** Evaluate and analyze the result based on the different parameters.  
**Step 10:** End
To test the algorithm, using some random data may lead to different results each time tested. This may mislead the prediction rate of the model. So, to reduce these shortcomings we have used the standard CT scanned images of lungs [21].

4. Results and Analysis

In this section, we discuss and inspect the results for the built ensemble classifier with Random-Forest based on the different parameters. To analyze the result, we have used the data [21] that consists of CT scanned images of Lungs. It has 561 images belonging to class 1 and 416 images belonging to class 0 where class 0 refers to Benign and 1 refers to Malignant. Images from the dataset are preprocessed and segmented inorder to extract region based features and this data is used as dataset for classification. The dataset is split into an 8:2 ratio for training and testing.

4.1 Confusion matrix

The confusion matrix is used to obtain the result of the classification or misclassification report in the form of a matrix. By using the concept of binary classification, four combinations of data category can be formed which are True-Positive (TP), True-Negative (TN), False-Positive (FP), and False-Negative (FN) as represented in Table 1 which are later used for calculation of different types of performance evaluation metrics. Where the True-Positive (TP) are more concerned samples and False-Negative (FN) are merely rejected/discarded samples.

| Predicted output | Real output | True (1) | False (0) |
|------------------|-------------|----------|-----------|
| Positive (1)     | True-Positive (TP) |          | False-Positive (FP) (11) |
| Negative (0)     | True-Negative (TN) | TN (10)  | False-Negative (FN) (00) |

4.1.1 Accuracy

Accuracy is the proportion of correct predictions versus the total number of predictions made. Accuracy is mainly used for measuring the performance of a classifier.

\[
Accuracy = \frac{Correct\ predictions}{Total\ number\ of\ predictions\ made}
\]

\[
Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)
\]

4.1.2 Recall

The ratio of total correct positive results to the number of total samples that should have been identified as positive is referred to as Recall. It is a number of +ve samples identified in the testing set.

\[
Recall = \frac{TP}{TP+FN} \quad (2)
\]
4.1.3 Precision
It is the proportion of a number of only positive to the total number of predicted positive results.

\[
\text{Precision} = \frac{\text{number of only positive}}{\text{total number of predicted positive result}}
\]

\[
\text{Precision} = \frac{TP}{TP+FP}
\] (3)

4.1.4 F1 Score
It determines the harmonic mean of recall and precision values. This is generally used to get the best recall and precision value at the same time i.e., it tries to finds the balance between precision and recall. The Formula for this is given as shown below.

\[
F1 = 2 \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}
\] (4)

The classification report of the different classifiers is shown in Table 2. It can be observed that the accuracy of the built Ensemble-Classifier is almost nearer to the Random-Forest classifier. Figure 3 showcases the performance graph of Ensemble-Classifier and Random forest classifier when input data size is varied.

![Performance on different size of Input](image)

**Figure 3.** Performance of Ensemble-Classifier and RandomForest classifier on varied input data size

| Model         | Accuracy (%) | Precision (%) | Recall (%) | F1-Score (%) |
|---------------|--------------|---------------|------------|--------------|
| Ensemble-Classifier | 85           | 85            | 89         | 87           |
| Random-Forest  | 88           | 88            | 91         | 89           |
4.2 Time Complexity
The time complexity of the Ensemble-Classifier is $O(n^3)$ for training and $O(np)$ for prediction, where $n$ is the number of training instances and $p$ is the number of features. As we know in KNN all the computation is done during the time of prediction. So, it has a time complexity of $O(1)$ for training and $O(knp)$ for prediction where $k$ is the number of neighbors. As $k$ is constant, time complexity becomes $O(np)$ for prediction. In Logistic Regression the row vector is of size $p$ hence it takes $O(p)$ for its computation and for $n$ instances it takes $O(np)$ for training and for prediction it only requires the computation of row vector so, it takes $O(p)$, where $n$ is the number of training instances and $p$ is the number of features. In the Decision tree, there are different level of tree for the entire training data of size $n$ and complexity at each level is $O(np)$ and for $n$ instances, it becomes $n(np)$. Hence decision tree has the time complexity of $O(n^2p)$ for training and for prediction it has the complexity of $O(p)$. In SVM there is a requirement of inverting a square matrix of size $n$ which takes a complexity of $O(n^3)$ and for evaluating the kernel matrix it takes $O(n^3p)$ hence the time complexity for training can be written as $O(n^2p+n^3)$. When $p<n$, $O(n^2)$ and for prediction, it takes a complexity of $O(p)$. In the case of MLP each layer has a different number of neurons lets assume that in the first layer there is $x_{ij}$ number of neurons hence for training it takes the complexity of $px_{ij}$ where $p$ is the number of features, at the second layer there may be $x_{lj}$ number of neurons hence it takes $x_{ij}x_{lj}$ and so on can be represented as ($px_{ij} + x_{ij}x_{lj} + ......$). Therefore for $n$ instances it takes $n(px_{ij} + x_{ij}x_{lj} + ......)$, so the time complexity of MLP model is $O(n(px_{ij} + x_{ij}x_{lj} + ......))$ and in case of prediction it takes only $O(p(px_{ij} + x_{ij}x_{lj} + ......))$ where $p$ is number of features. Hence the time complexity of Ensemble-Classifier for training is $O(n^3)$ and for prediction, it takes $O(np)$.

4.3 ROC Curve
Receiver Operating Characteristic Curve is a graphical representation of the performance of classification techniques in the slightest degree classification thresholds. It is an evaluation metric for binary classification problems. It is a likelihood curve that plots TPR vs FPR at varied threshold values and primarily separates 'signal' from 'noise'. This curve plots two parameters TPR and FPR. Equation (5,6) gives True Positive Rate and False Positive Rate respectively.

$$TPR = \frac{TP}{TP+FN} \quad (5)$$

$$FPR = \frac{FP}{FP+TN} \quad (6)$$

ROC curve at different classification thresholds plots the TP vs FP rate.Lowering the threshold for classification classifies many things as positive, which leads to the increase of each FP and TP. ROC curve for Ensemble-Classifier and Random-Forest model is shown in figure 4. From Figure 4 it can be observed that the ratio of TPR vs FPR of Ensemble-Classifier is almost nearer to the one obtained by Random Forest.
5. Conclusion
In this work, we propose a novel algorithm for detecting lung cancer in a CT scan by building an ensemble classifier, and then the results are compared with the RF classifier. In Ensemble-Classifier we included five machine learning models like SVM, LR, MLP, Decision-tree, KNN. The proposed methodology gives a detailed knowledge of predicting Lung cancer from a CT scanned image. We extract region-based features from the dataset and then split in the ratio of 8:2 for training and testing. We classify cancer into Malignant or Benign and then generate the classification report that includes accuracy, precision, recall, and F1-score using the confusion matrix. We also plotted the ROC curve for both models. The accuracy of the Ensemble-Classifier is 85% from which we can say that the built ensemble classifier can differentiate between Malignant and Benign cancer. It can be observed that matrices obtained by the Ensemble Classifier are almost nearer to the ones obtained by random forest and from the recall value, it can be observed that the model has identified maximum cases of Malignant tumors correctly.

In the future, Deep-Learning techniques like CNN can be used for the prediction of carcinoma. More range of pictures can be considered from different scanning techniques such as MRI, CT, PET, X-ray, which can evoke additional accuracy, thereby serving to the medical field to supply fast prevention at low value. Continuous information can even be used rather than simply categorical information.

References
[1] Nikita Banerjee, Subhalaxmi Das. “Prediction Lung Cancer– In Machine Learning Perspective” International Conference on Computer Science, Engineering and Applications (ICCSEA) (2020).
[2] Ozge Gunaydin, Melike Gunay, Oznur Sengel. “Comparison of Lung Cancer Detection Algorithms” Scientific Meeting on Electrical-Electronics & Biomedical Engineering, Computer Science (EBBT) IEEE 2019.
[3] Sumathipala, Yohan, et al. "Machine learning to predict lung nodule biopsy method using CT image features: A pilot study." Computerized Medical Imaging and Graphics 71 (2019): 1-8.
[4] Zhang, Junjie, et al. "Pulmonary nodule detection in medical images: a survey." Biomedical Signal Processing and Control 43 (2018): 138-147.
[5] Amrit Sreekumar, Karthika Rajan Nair, Sneha Sudheer, Ganesh Nayar H, and Jyothisha J Nair. "Malignant Lung Nodule Detection using Deep Learning" International Conference on Communication and Signal Processing, July 28 - 30, 2020, India.

[6] Fenwa, Olusayo D., Funmilola A. Ajala, and A. Adigun. "Classification of cancer of the lungs using SVM and ANN." Int. J. Comput. Technol. 15.1 (2016): 6418-6426.

[7] Muhammad Imran Faisal, Saba Bashir, Zain Sikandar Khan, Farhan Hassan Khan. "An Evaluation of Machine Learning Classifiers and Ensembles for Early-Stage Prediction of Lung Cancer" 3rd International Conference on Emerging Trends in Engineering, Sciences, and Technology (ICEEST 2018).

[8] Daoud, Maisa, and Michael Mayo. "A survey of neural network-based cancer prediction models from microarray data." Artificial intelligence in medicine (2019).

[9] Krishnaiah, V., G. Narsimha, and Dr N. Subhash Chandra. "Diagnosis of lung cancer prediction system using data mining classification techniques." International Journal of Computer Science and Information Technologies 4.1 (2013): 39-45.

[10] M. Siddartha Kumar, Prof. Dr. K. Venkata Rao. "Prediction Of Lung Cancer Using Machine Learning Technique: A Survey" International Conference on Computer Communication and Informatics (ICCCI -2021), Jan. 27 – 29, 2021, Coimbatore, India.

[11] Swati Mukherjee, Prof. S. U. Bohra. "Lung Cancer Disease Diagnosis Using Machine Learning Approach" Third International Conference on Intelligent Sustainable Systems [ICISS 2020].

[12] Syed Saba Raoof, M A. Jabbar, Syed Aley Fathima. "Lung Cancer Prediction using Machine Learning: A Comprehensive Approach" Second International Conference on Innovative Mechanisms for Industry Applications (ICIMIA 2020).

[13] Radhika P R, Rakhi.A.S.Nair. "A Comparative Study of Lung Cancer Detection using Machine Learning Algorithms", International Conference on Electrical, Computer and Communication Technologies (ICECCT) IEEE 2019.

[14] Wasudeo Rahane, Himali Dalvi. “Lung Cancer Detection Using Image Processing and Machine Learning HealthCare” IEEE International Conference on Current Trends toward Converging Technologies, Coimbatore, India, IEEE 2018.

[15] Kamelia Roy, Sheli Sinha Chaudhury, Madhurima Burman, Ahana Ganguly, Chandrima Dutta, Sayani Banik, Rayna Banik. “A Comparative Study of Lung Cancer detection using supervised neural network” International Conference on Opto-Electronics and Applied Optics (Optronix 2019).

[16] Sanjukta Rani Jena, Dr. Thomas George, Dr. Narain Ponraj. "Texture Analysis Based Feature Extraction and Classification of Lung Cancer" International Conference on Electrical, Computer and Communication Technologies (ICECCT) IEEE 2019.

[17] Ozturk, Şaban, and Bayram Akdemir. "Application of feature extraction and classification methods for the histopathological image using GLCM, LBP, LBGLCM, GLRLM, and SFTA." Procedia computer science 132 (2018): 40-46.

[18] Dendi Gayathri Reddy, Emmidi Naga Hemanth Kumar, Desireddy Lohith Sai Charan Reddy, Monika P "Integrated Machine Learning Model for Prediction of Lung Cancer Stages from Textual data using Ensemble Method". 1st International Conference on Advances in Information Technology IEEE 2019.

[19] Palani D, K. Venkatalakshmi. "An IoT-based predictive modeling for predicting lung cancer using fuzzy cluster-based segmentation and classification." Journal of medical systems 43.2 (2019): 21.

[20] Lynch, Chip M., et al. "Prediction of lung cancer patient survival via supervised machine learning classification techniques." International journal of medical informatics 108 (2017): 1-8.

[21] alacrity, Shamballa (2020), “The IQ-OTHNCCD lung cancer dataset”, Mendeley Data, V1, DOI: 10.17632/bhmdr45bh2.1