AI BASED HYBRID MODEL FOR THE DETECTION OF LUNG CANCER USING CT IMAGE

Anshu Sharma, Dr.Pankaj Nanglia, Dr. Paramjit Singh , Dr.Vikrant Shokeen, Juhi Priyani

Research Scholar, Maharaja Agrasen University, ahshukaushik10@gmail.com
Associate Professor,Maharaja Agrasen University,Baddi, India,nanglia.pankaj@gmail.com
Assistant Professor,Maharaja Agrasen University,Baddi, India, ppparamjitsingh@gmail.com
Assistant Professor,MSIT,Delhi,India, shokeen18@gmail.com
Research Scholar,Maharaja Agrasen University,Baddi, Indiajuhipy@gmail.com

Abstract: Now a day, one of the major reasons for non-accidental death is the cancer. Lung cancer is uppermost cause of cancer death in all over world. It's possible to reduce the ratio of death if persons are aware for early detection by which appropriate treatment can be done by the doctors within some safe duration. When a group of cells becomes irregular growth uncontrollably and lose balance to make malignant tumours that is invades surrounding tissues, termed as cancer. It can be classified mainly into two parts Non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC). The symptoms, diagnosis and treatments are also discussed here.

This research work focused on the earlier detection of lung cancer in image processing by utilizing some steps such as pre-processing; image segmentation based on thresholding, after got the segmented image that means required region of interest as per this work. Apply feature extraction technique by which we have extracted the features from image data by using SURF all these obtained features are not useful to get the desired output. For data segmentation and pre-processing the k-means using Grasshopper algorithm for optimization and Maximally Stable Extremal Regions (MSER) technique. In this context, we designed a hybrid model for the classification and detection of lung cancer diseases by using classification technique of neural network named as Convolutional Neural Network. The proposed hybrid model compared with the existing model designed by Senthil Kumar et al. (2019) and found the True Positive of the proposed work is 98.7 and Senthil Kumar et al. (2019) is 95.28. The average value of accuracy is 99%.

INTRODUCTION

1.1 Lung Cancer

Lung cancer is a disease that happens because of uncontrolled cell growth in the lung tissues.

This growth might lead to metastasis, known as the adjacent tissue invasion and infiltration ahead of the lungs [1-4]. The prognosis and treatment are dependent on the histological cancer type, the degree of spread (the stage), and the status of the patient's performance. The possible treatments comprise chemotherapy, surgery, and radiotherapy [2-7]. Survival of the patient depends on overall health, onstage, and another factor, but on all only 14% of people diagnosed with lung cancer and survive five years after the diagnosis [7-9].

The symptoms of lung cancer detection include:

- Hemoptysis (coughing up blood)
- Dyspnea (shortness of breath with activity)

1.2 Techniques Used for Lung Cancer

The utilized techniques for detection of lung cancer are explained below;

1.2.1 Grasshopper Optimization Algorithm (GOA)

The various optimization algorithm is available according to the kinds of research work, among them Grasshopper optimization algorithm (GOA) is the one kind of swarm intelligence introduced by Mirjalili. The working of this algorithm is population-based. GOA is interacting with the social behavior of grasshopper swarms. Grasshoppers are insects, and they are considered a pest due to their damage to crop production and agriculture.

1.2.2 Maximally Stable Extremal Regions (MSER)

Several methods or techniques are implemented specifically to blob detection of multiple images. A maximally Stable Extremal Regions (MSER) algorithm is utilized in the detection of the blob in images and is introduced by Matas to see the correspondences among components of images of two images based on different points of view. This method extracts wide ranges of corresponding components of images to contribute to the depth baseline matching by which it leads to the best stereo matching and object-recognition algorithm. A various number of co-variant regions known as MSER are extracted from an image by using the MSER scheme and the

ability of MSER is stable to make the connection between elements for some gray level set of images [9-12].

Extremal regions

In the context of extremal regions included two essential features that are pointed below:

- Transform the image coordinates continue which means the extremal region can keep invariant, without taking care about wrapped or skewed images.
- Like the monotonic transformation of image intensity, this scheme is affected by monotonic effects as per the changes in day-light and moving shadow.
- transformation upon the large domain of intensity.
- MSER offers stability, which means only that kinds of areas are selected support of which regions are almost similar in the range of threshold.
- Detection of multi-scale without considering any smoothing, and having the ability to detect both fine and large structure. And it also noted down that MSER detection in scale-pyramid provides better repetitions along with various amounts of modifications over scale modifications.
- All the extremal areas possibly are enumerated in the worst case of complexity that is evaluated through O (n), where n represents the number of pixels in the image.

1.2.3 Speed-Up Robust Feature (SURF)

SURF is a newly-developed methodology, mainly used for feature detection purposes of an image. To enhance the performance of the feature extraction problem, there is a need to fixed two problems: firstly, there is necessary to focus on the importance of the number of feature pairs after recognition of the object. Because this number of features is already stored in the score of recognition, a set of pairs between the image and each database image can be generated by using the SURF algorithm. Because of the powerful attribute of the SURF algorithm, this is useful for the detection of an object in the image, including invariance of scale, invariance of translation, invariance of lighting, invariance of comparison, and invariance of rotation. This algorithm composes of mainly four parts:

1. Integral Image generation
2. Fast-Hessian detector (point of interest detection)
3. Descriptor orientation assignment
4. Descriptor generation

Figure 0 Flow chart of the SURF algorithm

To locate the corresponding point of images, SURF uses the determinants of Hessian matrices. If the SURF algorithm is used, all the representative points will work with the same weight; this can be accomplished by assigning random weights to representative points. Correct representative points will be produced rarely; the weight of each representative point can be calculated as follows:

$$W_P = \frac{Number\ of\ detected\ images\ with\ respect\ to\ point}{number\ of\ training\ images\ in\ the\ object} \quad (1.4)$$

1.2.4 Convolution Neural Network (CNN)

The Convolutional Neural Network is a mechanism that integrates the extraction and classification of features in itself. The Convolution and sub-sampling layer arrangement constructs the trainable feature detector portion. Convolutional Neural Networks (CNNs) are inspired by our biological system and are essentially a Multilayer Perceptron (MLP) model. CNN has two types of the cell named as a basic cell(s) and complex cell(c). Simple cells within their receptive field respond to different edge-like stimulation patterns to the limit. Complex cells have wider receptive areas and are invariant locally to the exact location of the stimulus. The key benefit of the CNN is that it needs only a few learning parameters, as it consists of the translated version of the same basic function. Convolutional Neural Network is based on connecting the preceding layer previous-layer units define the receptive field width [12-15].

Each layer of a CNN transforms the 3-dimensional input volume to a 3-Dimensional output volume of neuron activations. In the figure above, the red input layer comprises the image, so its width and height would be the dimensions

of the image, and the depth would be 3 (Red, Green, Blue channels) [16-19].

Local area to the next layer. CNN imposes local spatial correlation by implementing a local Handwriting Recognition pattern of communication between adjacent layer neurons. The one-layer units are connected to the preceding layer sub-units. Several

PROPOSED WORK

1.3 OBJECTIVES

1) To study the existing segmentation and features extraction techniques for lung cancer data.
2) To develop a hybrid segmentation algorithm using K-mean with grass hopper optimization and MSER (Maximally Stable Extremal regions) technique.
3) To extract the feature from the segmented RoI of Lung, the SURF descriptor will be used with CNN as a classifier to train the model.
4) To evaluate and compare proposed work based on performance parameters.

The evaluation parameters would be as follows

a) True Positive Rate (TPR)
b) False Positive Rate (FPR)
c) Classification Accuracy

1.4 METHODOLOGY

The flow of the work is that the earlier detection of lung cancer and diagrammatical representation is depicted in figure 2 detailed description is given below. Here we have discussed the proposed methodology that is utilized to detect lung cancer and classify the types of this disease.

Step 1: Design and implement the proposed framework for the simulation of work i.e. early detection of lung cancer.

Step 2: To deploy with this proposed work, collect an image of lung datasets for training and testing.

Step 3: Apply pre-processing for segmentation using k-means with Grasshoperb and MSER.

Step 4: After pre-processing step, the feature extraction technique has been applied to the pre-processed data to extract the feature sets from the loaded dataset. For the feature extraction, the SURF technique is used on the pre-processed data.

Step 4: The uniqueness of the feature is directly proportional to the segmented region of the Lung image. To overcome this problem, CNN is used to segment the Lung images accurately.

Step 5: From the obtained improved images through applied K-means with the grasshopper, we have utilized SURF as a feature extraction technique.

Step 6: CNN is used as a classifier to train the proposed system using the extracted key points. After the feature optimization, the CNN classifier is initialized to train the system based on optimized data using the following steps:

a. Select optimized feature as an input of classifiers for training and testing.

Step7: In the given flow diagram the first part is the training of the system and the second is a classification of particular Lung lesion images such as Malignant and benign types.

The methodology can be understood with the help of the following flow diagram

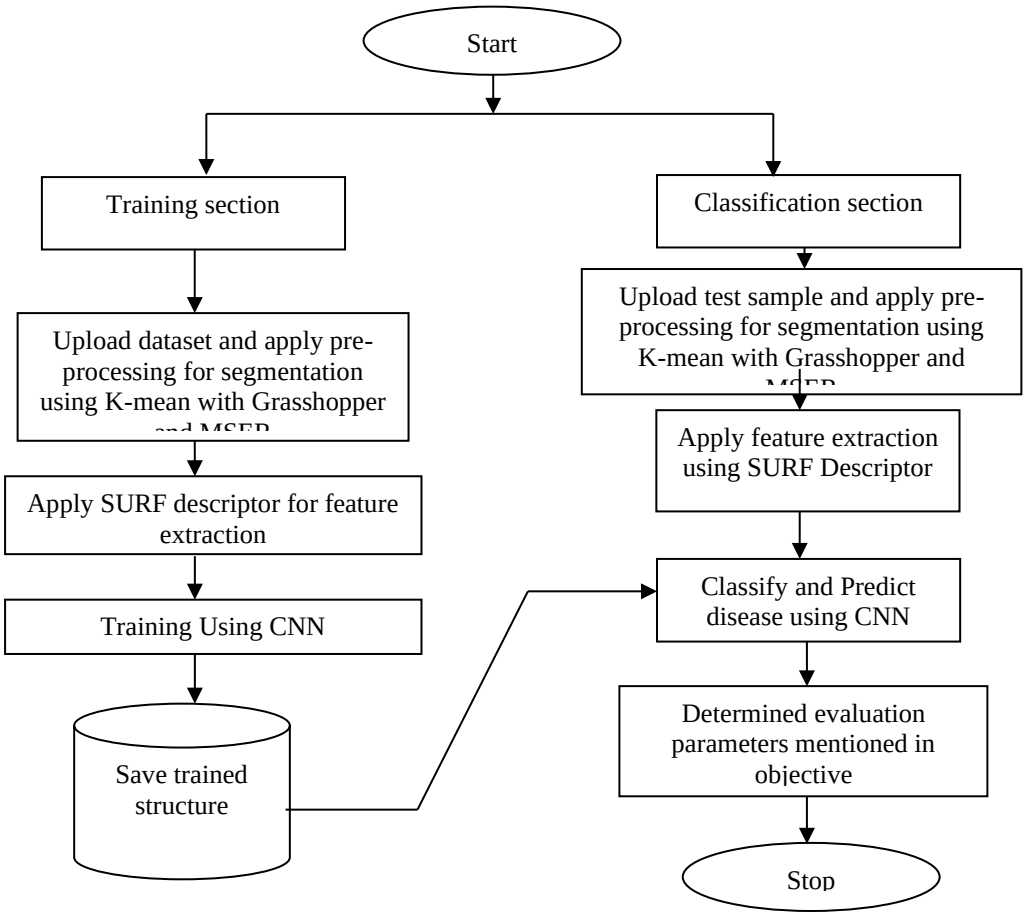

Figure 2. Flowchart of the proposed work

RESULT ANALYSIS

In this section, the research study depicted about the finding and processes or, steps of the task i.e. early detection of lung cancer. If this disease is detected earlier then the ratio of the dead will be reduced. So we proposed and perform this task by using various techniques like segmentation algorithm using K-mean with grass hopper optimization technique that is mainly used to partition an image into a set of pixels by which it's easy to detect a disease from digital images.

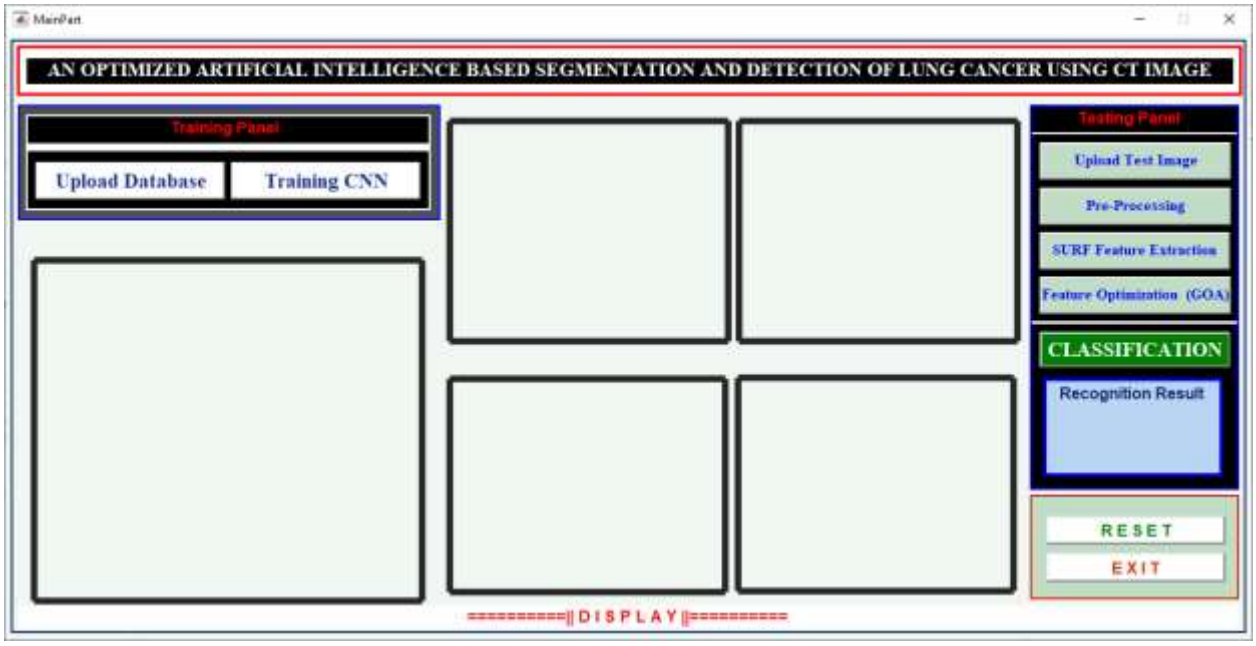

Figure 3. Proposed Model Window

The very first step of this task involves the design and implementation of the model window as shown in figure 3, the outcome of this task completely depends on the classification mechanisms that were utilized here, Artificial Neural Network (ANN) as depicts in the windows upper part. The image datasets of the lung are divided into two panels such as; Training and, testing panel as shown in the model window given. To get the output from this system we have to first train as per our requirements and then match the result obtained from the testing panel.

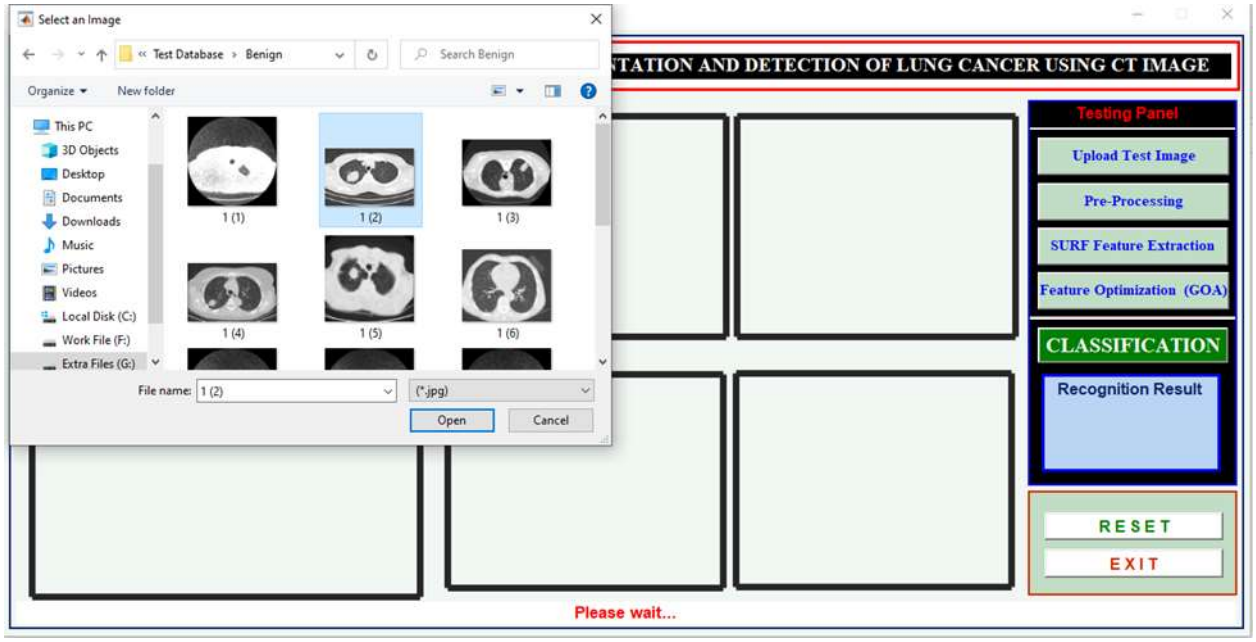

Figure 0. Uploading of Test Lung Image

In figure 4. shows the next steps after the model window that is uploading of test lung images from our collected image datasets. We have various images from which have to upload any one image at one time to perform the further steps of the task.

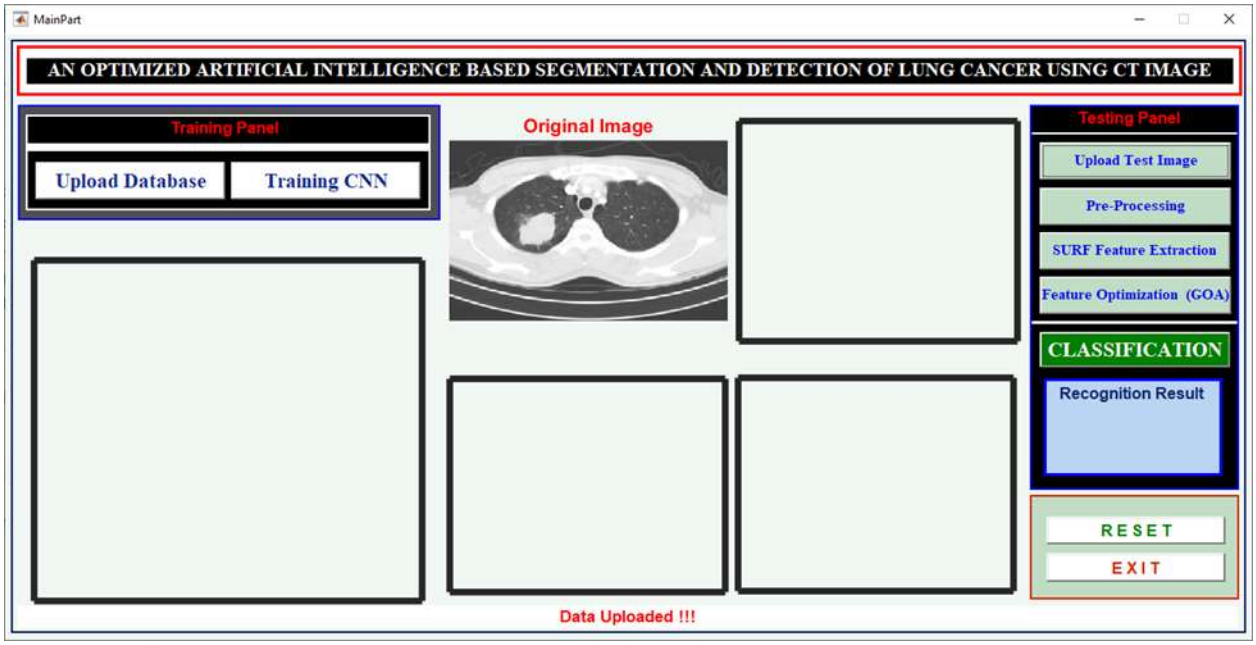

Figure 5. Test Lung Image

As shown in figure 5 the test image is uploaded successfully from our collected image datasets of lung cancer. On the model window the message for the user from command i.e. please wait is changed to data uploaded. That means data is uploaded successfully on the uploaded it has to highlight the by the name due to the reason we have not performed any image processing steps on this uploaded datasets i.e. original image.

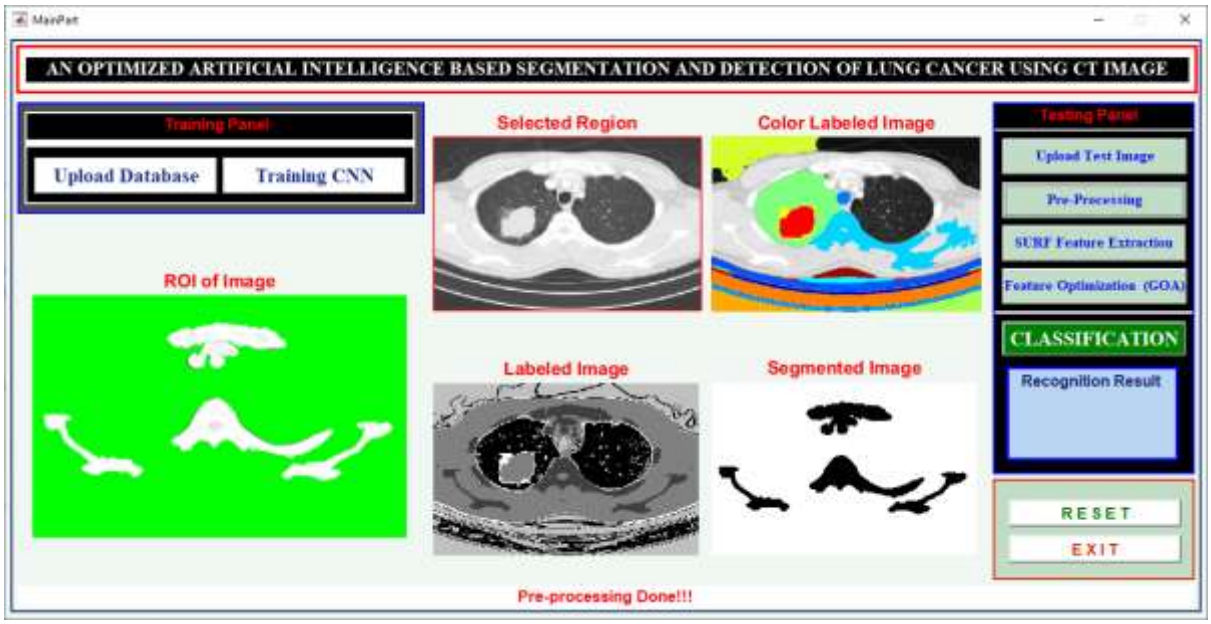

Figure 6. Pre-processing on Uploading of Test Lung Image

After successfully uploaded the test image, the next step is pre-processing as shown in figure 6.

- **Pre-processing:**

The main aim of this step involves an improvement of the image data that includes unnecessary distortions or, enhances some required features of the image for further processing. And also removes the redundancy present in a particular image without affecting the required details that play a significant role in the proposed task. There are four categories given below of image pre-processing methods are exist as per the size of the nearest pixel that is utilized for the computation of brightness of a new pixel.

- Transformations of pixels brightness
- Geometric transformations
- Method of pre-processing that utilizes local neighborhood of the processed pixel and,

- Restoration of an image that needs knowledge about the whole image.
- In a pre-processing step, segmentation is performed using k-means with GOA and MSER. K-means is used to segment lung RoI in terms of foreground and background and for the improvement purpose of the foreground hybridization of GOA and MSER is used. After got the segmented image we have to apply the feature extraction technique named SURF.
- **Feature Extraction**

The extraction of feature is the operation of low-level image processing, in which a feature could be explained as the interesting part of an image. The necessary characteristics of an image are repeatability means the same feature is will be detected in different images on the same scene.

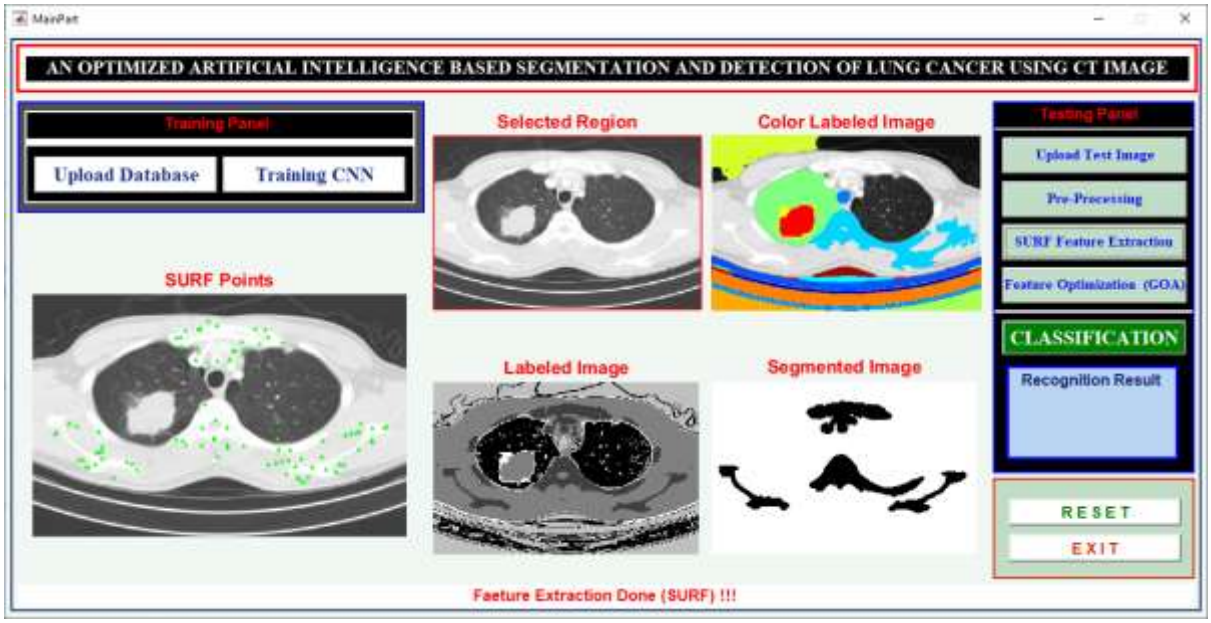

Figure 7 SURF Feature Extraction

The model is partitioned into two parts in left side shows the training panel and on another partition of this window shows the testing panel that involves steps such as; upload test image, pre-processing of an image, segmentation involves the process of partitioning an image into the homogeneous required region of interest concerning different objects within the image. It divides the image into the meaning full region of interest (ROI) that can be performed by using some basic properties of features of images such as intensity, edge, or texture. Various types of

segmentation approaches are available named as; (i) edge-based segmentation, (ii) Region growing segmentation, and (iii) Threshold-based segmentation methods. Here in this proposed work, the threshold-based segmentation method has been utilized in which by using some fixed value termed as the threshold. The colored image is transformed into a black and white image. As shown in figure 7, four boxes have shown in the first box the required region is selected, the next box depicts the colored part of the image, in the third box represents the labelled image in black and white, and at the end in the fourth box save only the segmented region of the image. In which we have extracted only the region of the image that is required according to this proposed work.

After the extracted features of the image, we have to apply the feature selection technique i.e. SURF, the optimized or selected features are shown in figure 8.

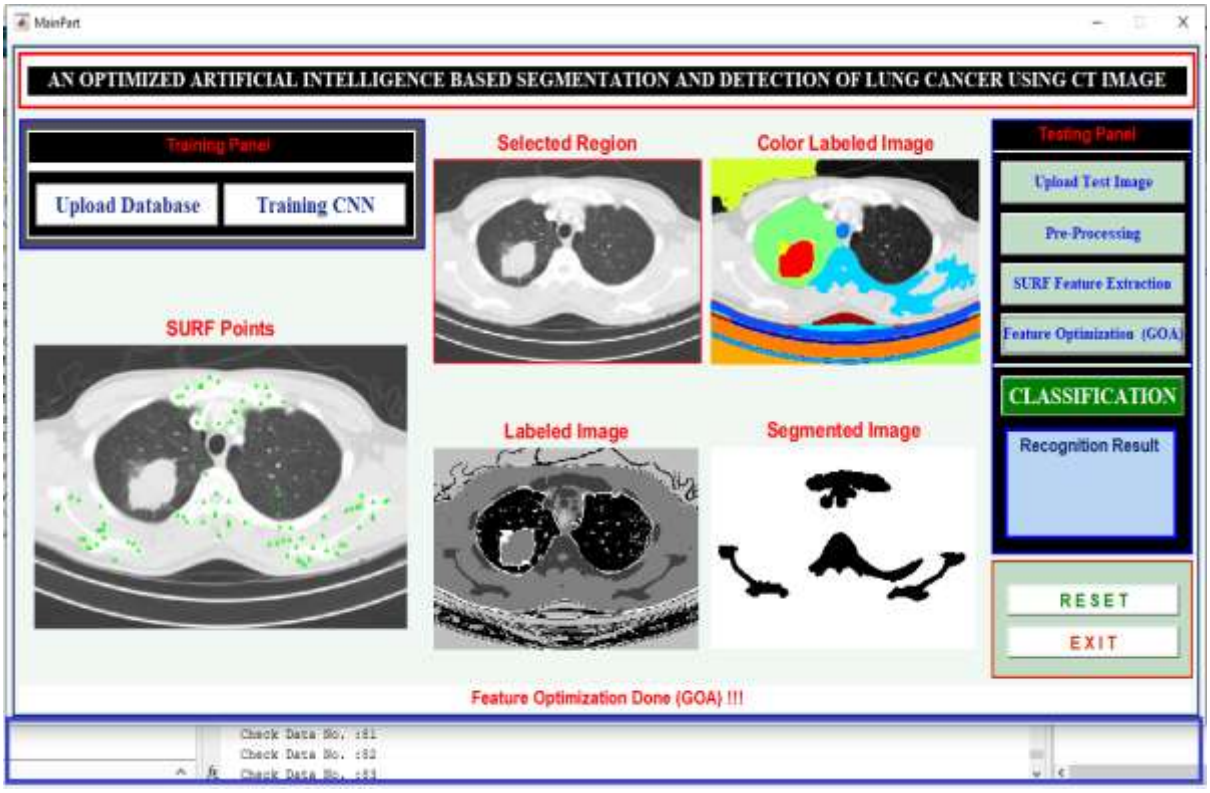

Figure 8 Optimization using GOA

- **Feature Selection**

All features that are extracted do not always produce a better result for the classification of lung cancer disease. So the feature selection is performed to select a required feature subset to minimize the dimensionality of features space, which

enhances the accuracy of the classification and reduces time consumption. Various feature selection techniques are available in this proposed work we have utilized the SURF.

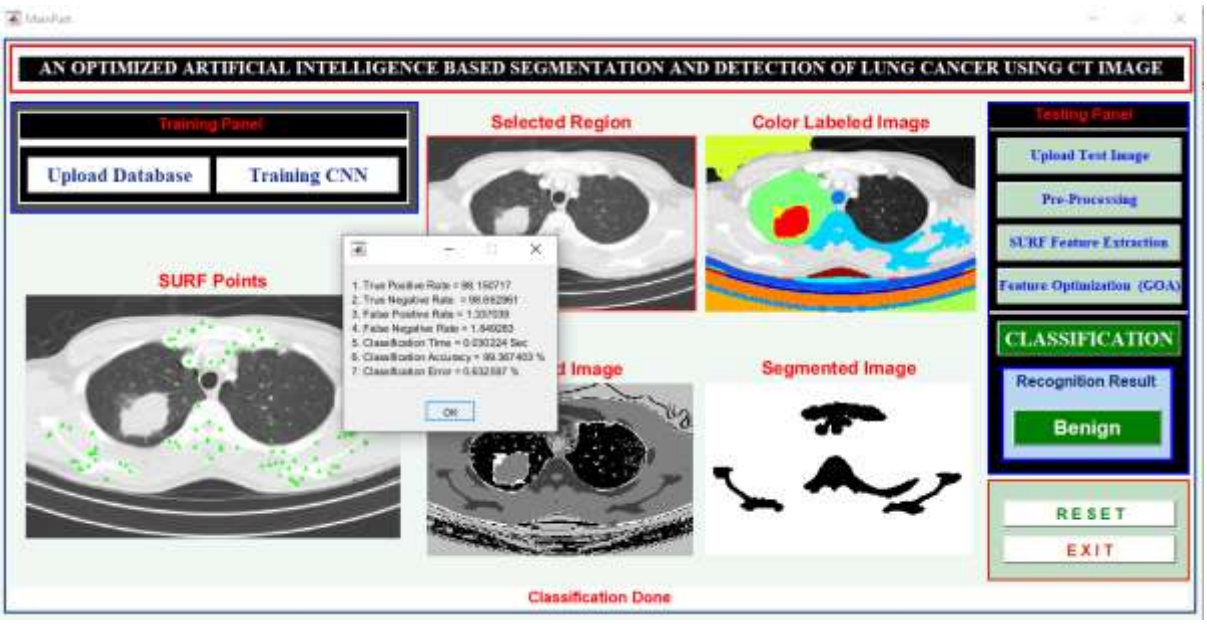

Figure 9 Classification Parameters

Table 1 Classification parameters

No. of samples	True Positive Rate	True Negative Rate	False positive Rate	False Negative Rate	Classification Time (s)	Classification Accuracy (%)	Classification Error
1	99.1114	97.678	1.3098	1.8845	0.2989	99.6842	0.3158
2	98.6789	97.951	1.312	1.8758	0.2974	98.8765	1.1235
3	98.3456	97.936	1.4678	1.845	0.3768	99.7654	0.2346
4	99.0089	98.178	1.345	1.8678	0.2561	98.9889	1.0111
5	99.10	98.556	1.425	1.8345	0.3011	99.898	0.102

6	98.2345	98.795	1.389	1.8623	0.3057	98.657	1.343
7	98.567	98.981	1.452	1.8734	0.3019	99.2816	0.7184

In table 1 shows the value of computed such as Error (%), Time (s), True positive rate, True Negative rate, False positive rate, False negative rate, and Accuracy (%) for 7 numbers of samples. The graphical representation of these parameters is given below one after another.

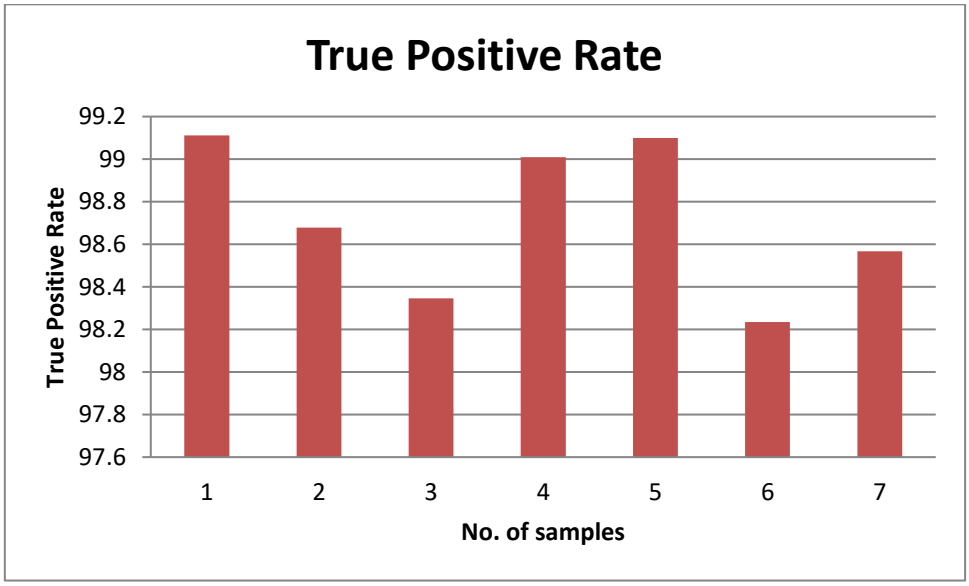

Figure 10 True Positive Rate

In figure 10 represent the True positive rate of the proposed work. In this figure, the x-axis depicts the no. of samples and the y-axis depicts the True positive rate. The true positive rate measures the proportion of actual positives that are correctly identified. In 1st sample, the value of the true positive rate is 99.11 and in sample 4th the value of the true positive rate is 99.008. The average value of the True positive rate is 98.15.

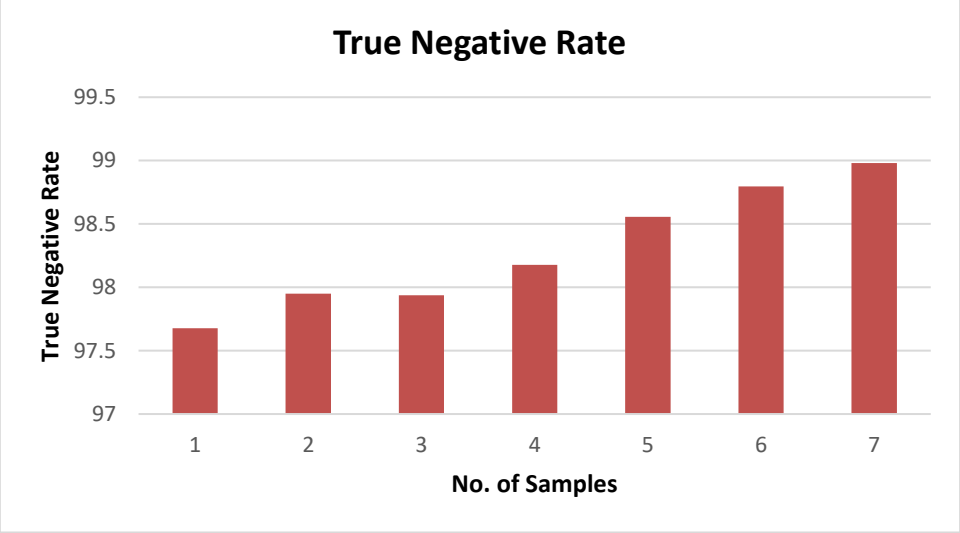

Figure 11 True Negative Rate

In figure 11 represent the True negative rate of the proposed work. In this figure, the x-axis depicts the no. of samples and the y-axis depicts the True negative rate. The true negative rate measures the proportion of actual negatives that are correctly identified. In sample 2nd the value of true negative rate is 97.951and at sample 5th the value of true negative rate is 98.556. The average value of the True negative rate is 98.66.

Figure 12 False Positive Rate

In figure 12 represent the False positive rate of the proposed work. In this figure, the x-axis depicts the no. of samples and the y-axis depicts the False positive rate. False positive rate is the probability of falsely rejecting the null hypothesis for a particular test. The value of the 3rd sample is 1.467 and the value of the 7th sample is 1.452. The average value of the False positive rate is 1.38

.

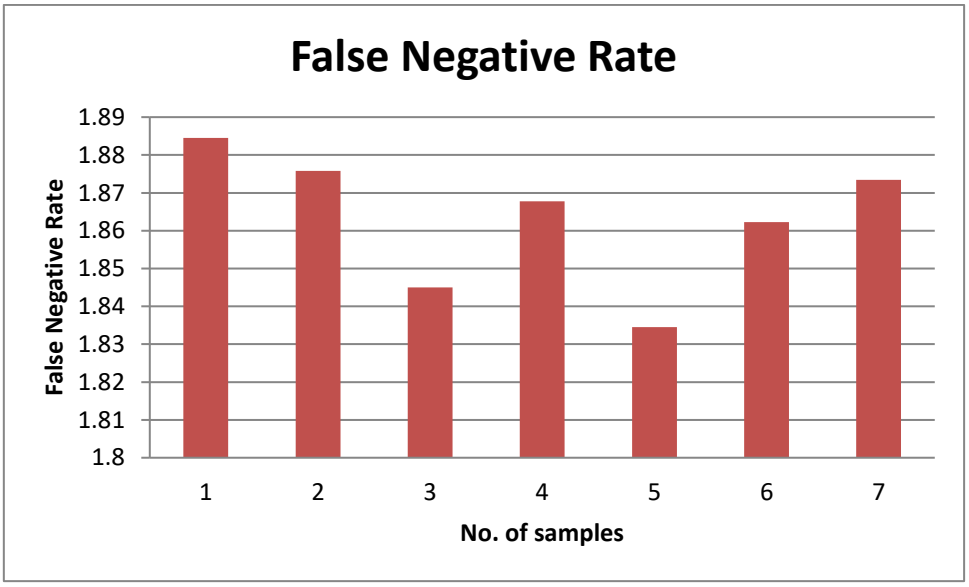

Figure 13 False Negative Rate

In figure 13 represent the False negative rate of the proposed work. In this graph, the x-axis depicts the no. of samples and the y-axis depicts the False negative rate. The false negative rate is the proportion of the individuals with a known positive condition for which the test result is negative. The average value of False negative rate is 1.849.

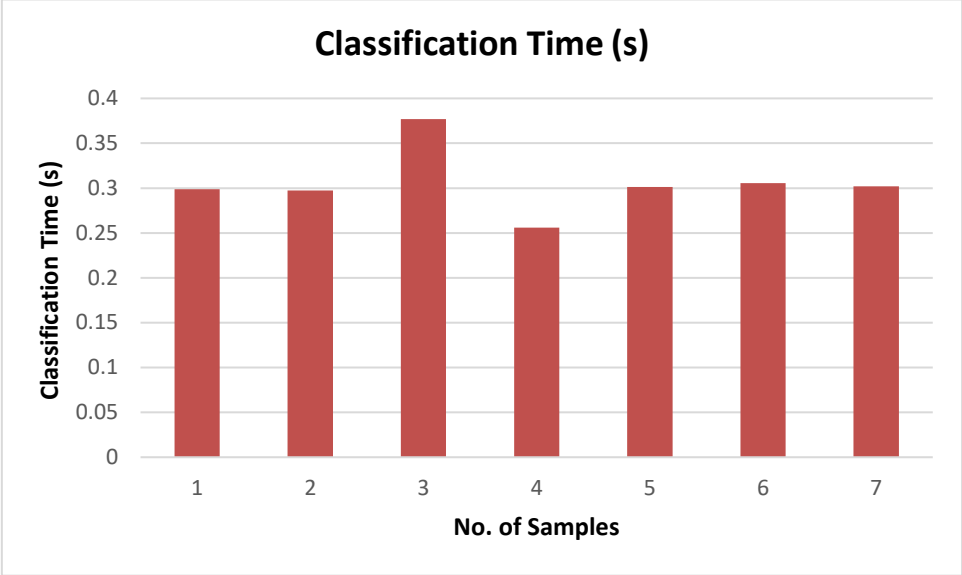

Figure 14 Classification Time (s)

The classification has been defined as the time spent by the system performing that task, including the time spent performing run-time or system services on its behalf. Defined implementation is the system used to evaluate time. Figure 14 shows the classification time against the number of text samples maximum and minimum obtained execution time to perform the task have 0.3768 and 0.2561 in seconds.

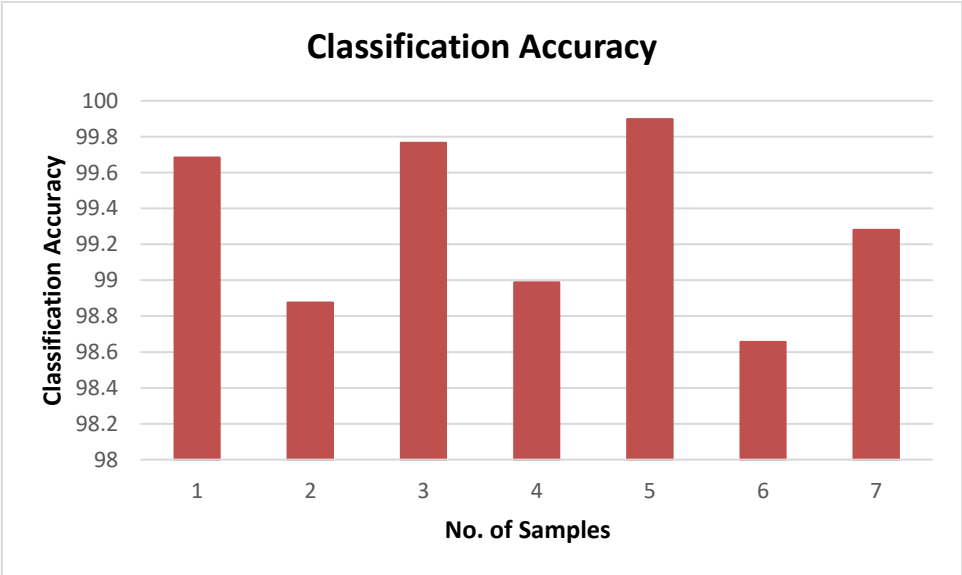

Figure 15 Classification Accuracy

In figure 15 represent the accuracy of the proposed work is the capacity of the machine to measure the precise value. The average value of accuracy is 99.367%.

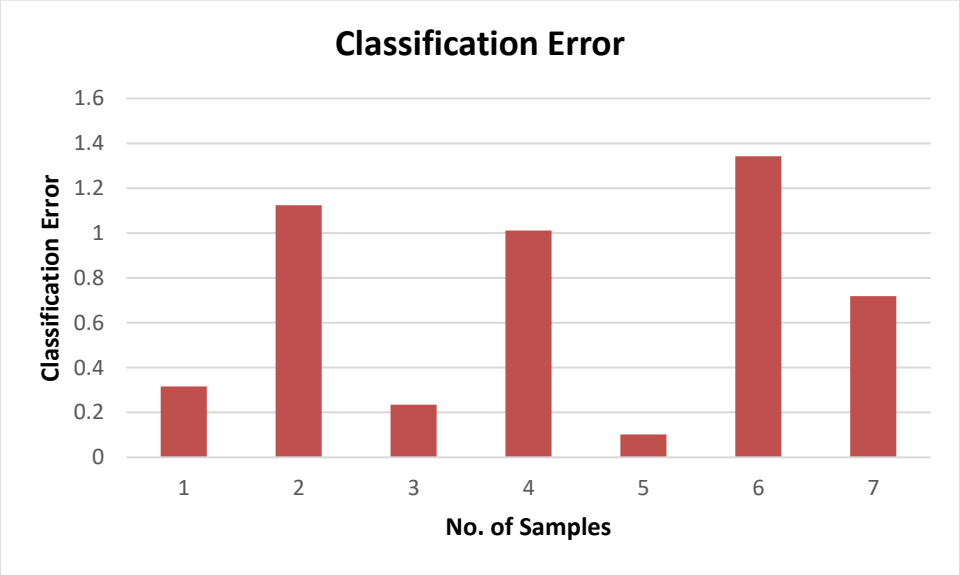

Figure 16 Classification Error

In figure 16 the percentage error has depicted against the number of samples the error doesn't need to be increased according to the increasing number of samples. In the form of a column graph, the error is plotted vertically against several samples horizontally.

Table 2 Comparison with Proposed Work

Parameters	Proposed	Senthil Kumar et al. (2019)
True positive	98.7	95.28

Table 2 shows the comparison of the proposed work. This table shows the value of True Positive of the proposed work is 98.7 and corresponds to Senthil Kumar et al. (2019) is 95.28.

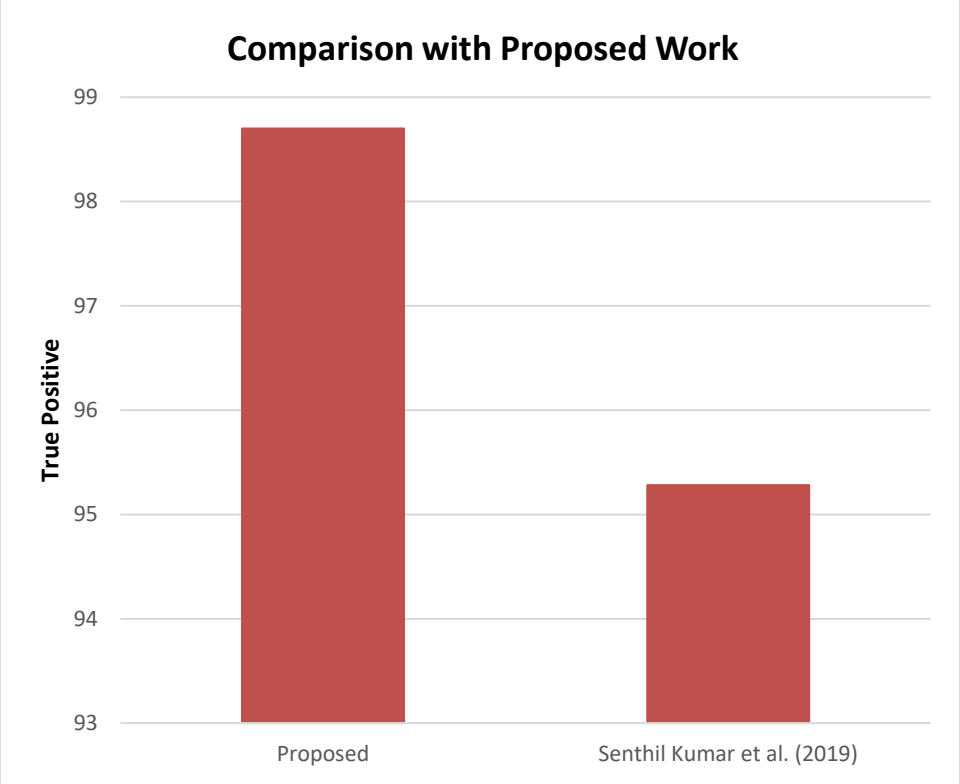

Figure 17 Comparison of Proposed and Senthil Kumar et al. (2019)

In figure 17 represents the comparison of the proposed work and Senthil Kumar et al. (2019) based on a parameter named as true positive. The comparison shows that our proposed work has a high positive rate as compared to Senthil Kumar et al. (2019) which means that our work shows good results.

.

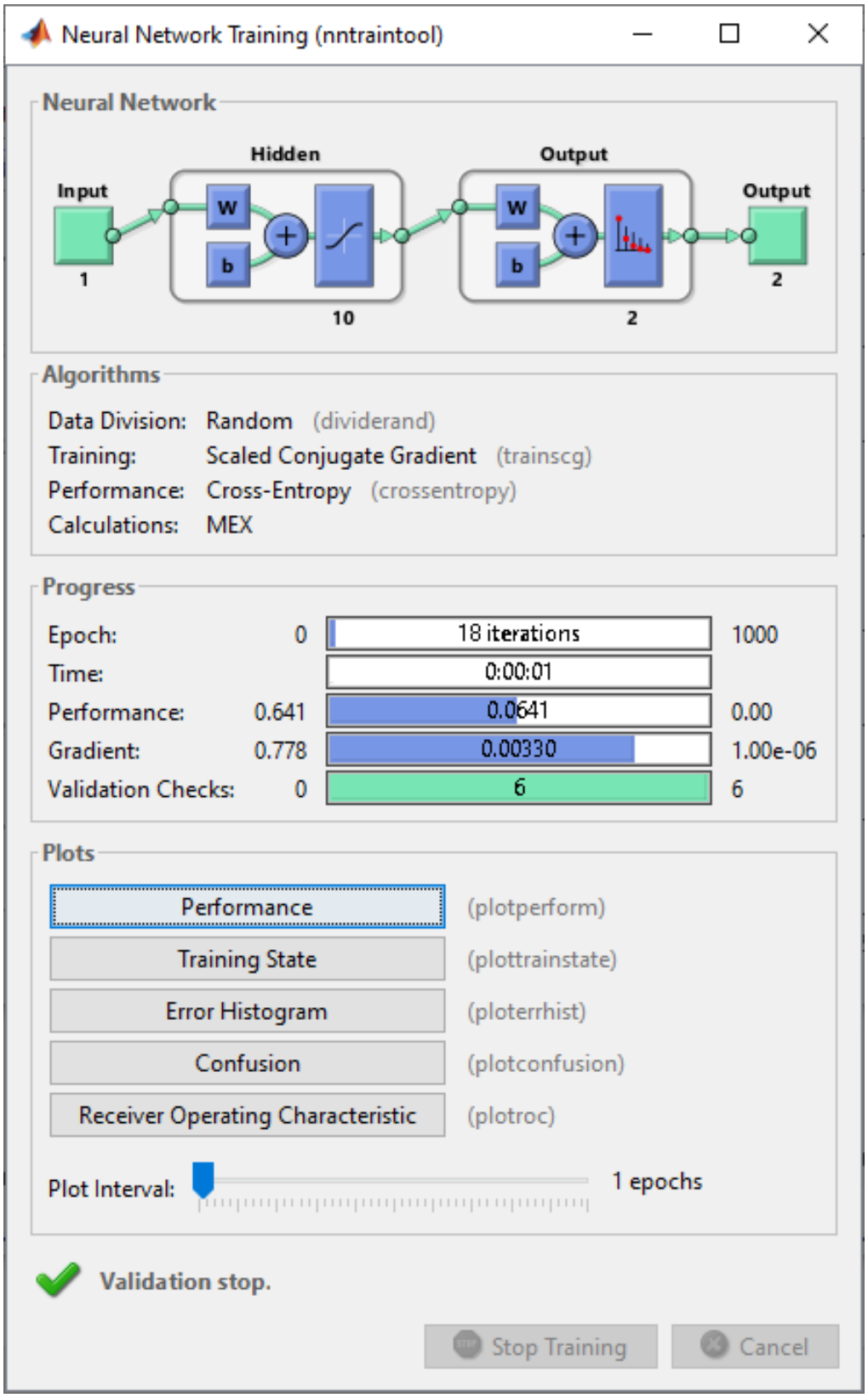

Figure 18 Training using Convolutional Neural Network

Figure 18 shows the working of the Convolutional neural network for detection based on the machine learning approach. The screenshot corresponds to the 18th iteration round out of 1000 simulation rounds with a neuron count of 10. It is also observed that for training Scaled Conjugate Gradient is used while implementing cross entropy-based performance analysis. The progress panel displays the respective progress in an individual iteration round. Convolution Neural network uses five different parameters for the optimization of the job, are Epoch (iterations), Time, Performance, Gradient and Validation checks which are further discussed with the help of figure. In the beginning, the neural network uses a dataset for self-training, and after that testing and validation occur. When the parameter became fully green that it shows that the given parameter completed its task. Here the gradient is completed and hence the dataset is classified. The iteration of the work occurs until it equals the validation checker. Once it crosses the limit of validation checker, neural network dumb that dataset and take another dataset for the scheduling. The process keeps going until the total job gets schedule and it takes 70% of the job for training and 15% for the testing and the rest 15% for the validation.

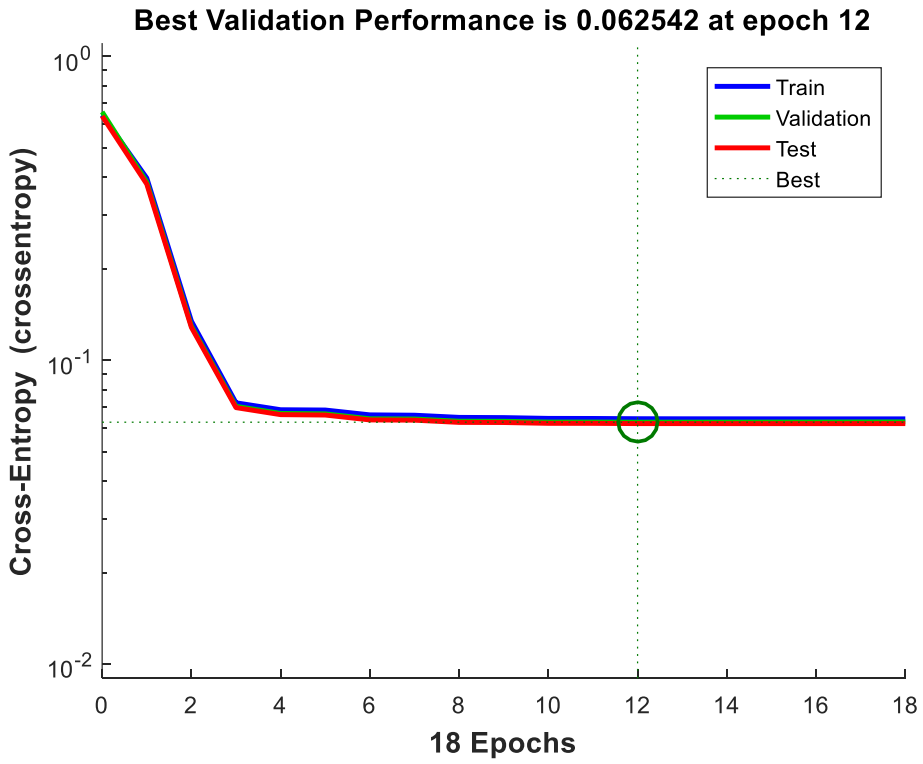

Figure 19 Convolutional Neural Network training performance

Figure 19 disputes the information about the cross-entropy found during the different phases of the scheduling. The blue line indicates the cross-entropy is in the training process. The green line shows the validation value in the validation process, the red line for the testing, and the dotted line for the ideal or the best condition. The Circle shows the best condition among the different iteration in a neural network. Convolutional neural networks divide the job into two groups. Cross entropy shows the variation of the dataset from the category.

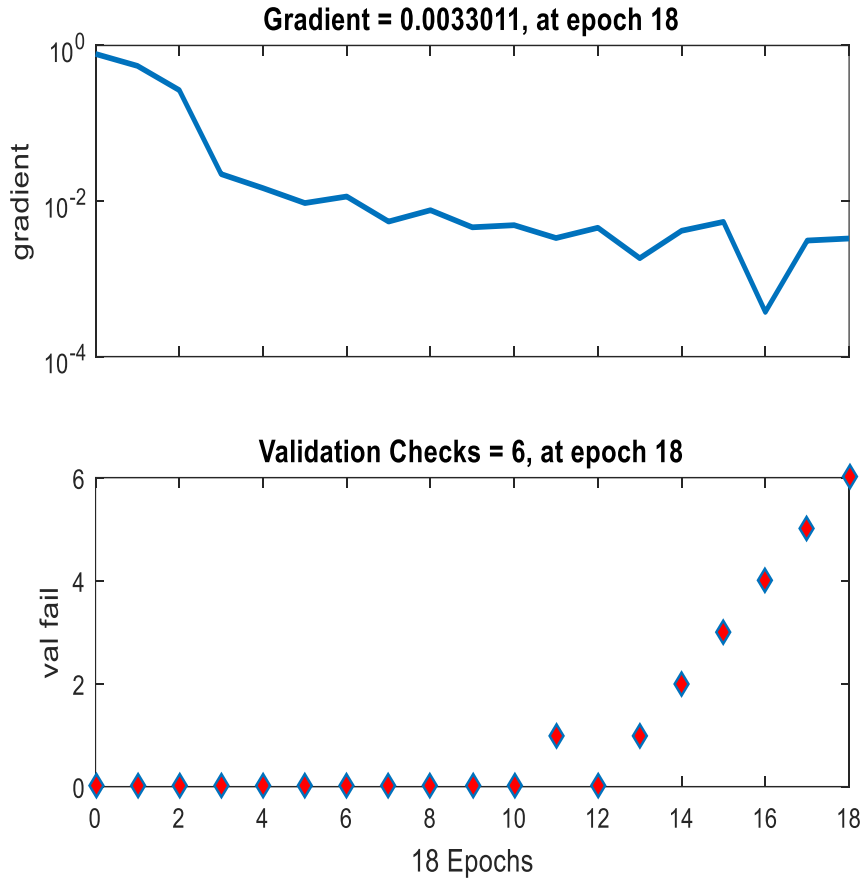

Figure 20 Gradient and Validation of the proposed Convolutional neural network

Figure 20 explains the information about the performance matrix for scheduling the particular dataset of news for the 18th iteration. Gradient shows the actual reading when it competed, the next job is shifted for the optimization. Validation checks the number of iteration for a particular job scheduling. For the first job, it shows zero means it was validated in the first iteration. Similarly, for the second job, it shows o that

is the particular dataset gets a schedule in the first iteration and so on.

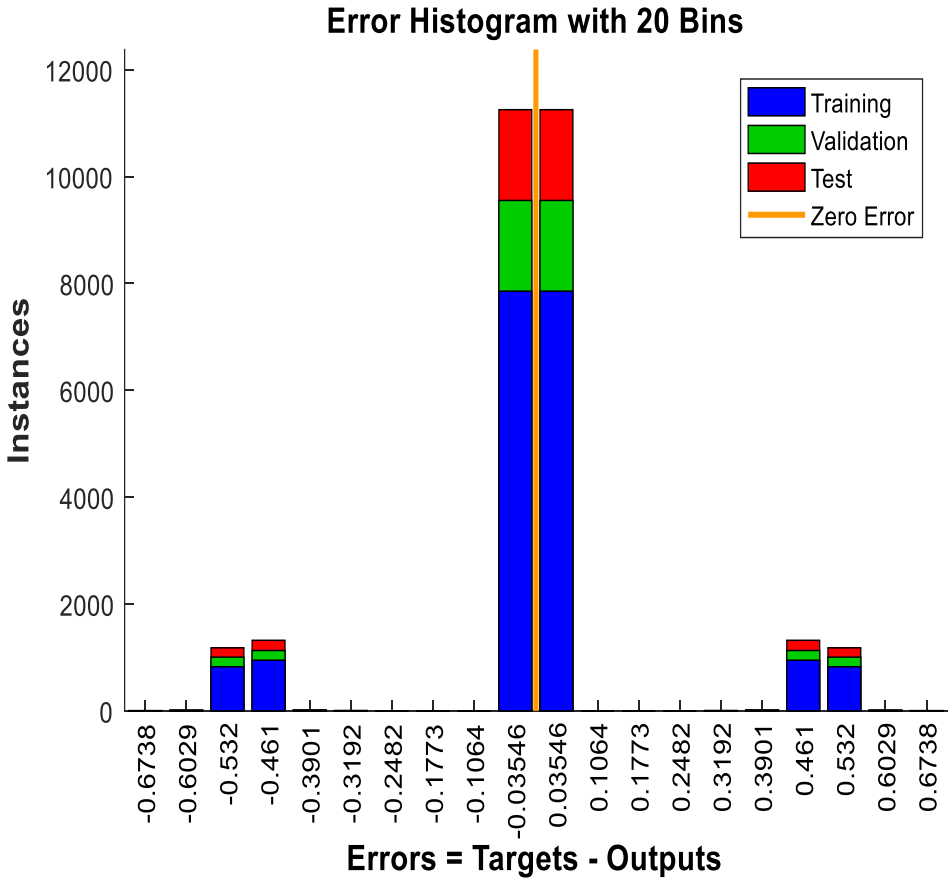

Figure 21 Convolutional Neural network training Error Histogram

Figure 21 shows the Convolutional neural network training error histogram that takes 20 random samples from the total datasets. As it takes the data randomly from the total data then it has the training, testing, and validation data. The blue histogram details the training data, the red histogram details the testing data and green denote the validation data. The center of the histogram shows the ideal value or the zero error value. Here the graph lies near the center that means the proposed model performs well with minimum error. As it shows most of the histogram lies near the center and

outcome despite the higher accuracy of the result.

Figure 22 Confusion matrix of the purposed model

Figure 22 shows the confusion matrix of the proposed Convolutional neural network-based proposed model. Here there are four different matrix formed, first one shows the confusion matrix of the training process.

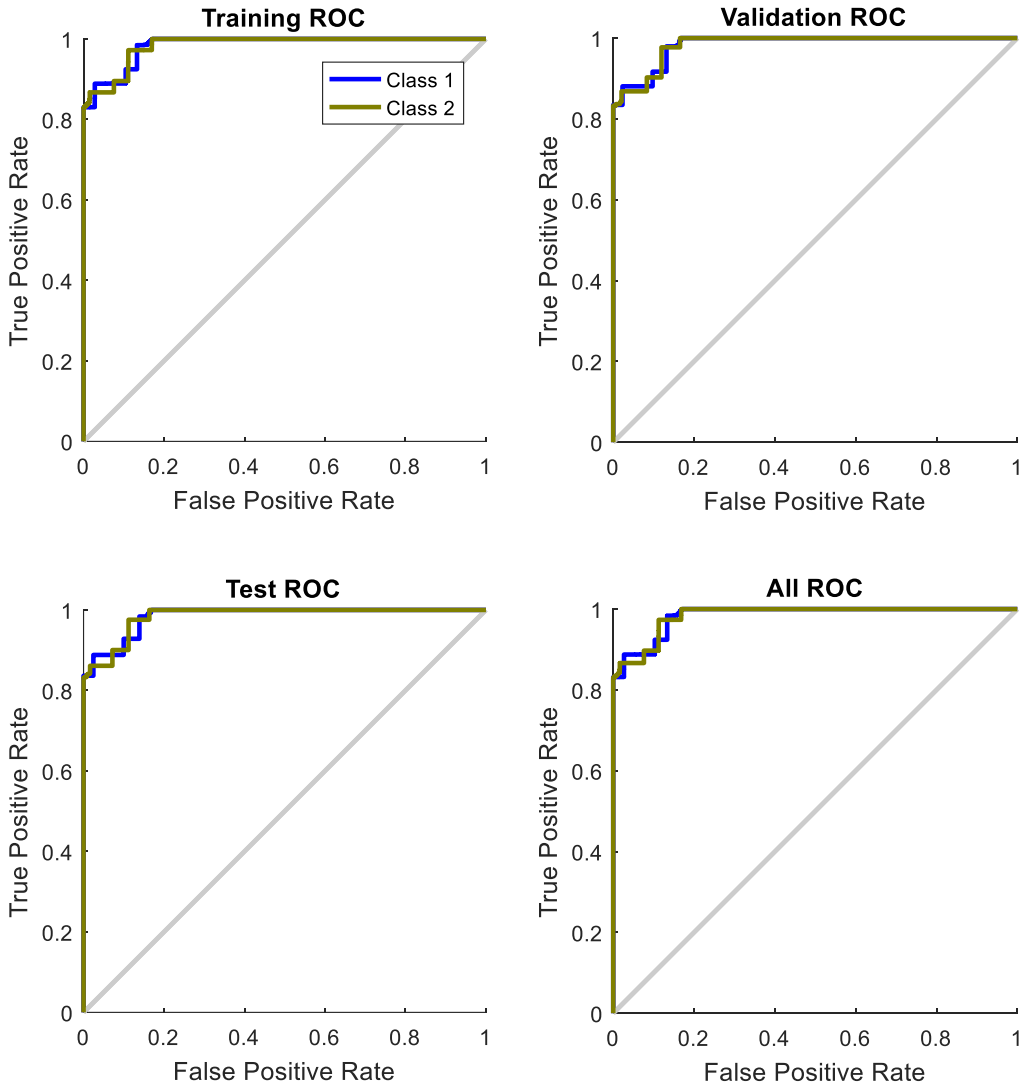

Figure 23 False and True positive

Figure 23 shows the ROC curve for the proposed model. A receiver operating characteristic curve, or ROC curve, is a statistical plot that shows a binary classifier system's diagnostic potential as the threshold for discrimination differs. The ROC curve is developed by plotting the true positive rate in different threshold settings against the false positive rate. The entire ROC graph shown in the figures is the graphical representation of the confusion matrix. Here both the classes lie on the True positive area and overlap due to which it can't be identified easily.

CONCLUSION

The Present research work depicted the Classification of the Lung disease from CT scan images is the key for preventing the life of humans at the earliest stage of Lung disease. The studies of Lung cancer detection signify the studies of visually observable patterns seen on the human Lung image. Strength monitoring and disease detection on the lung is a very critical task for sustainable medical science. It is very complicated to observe lung disease manually. Therefore, image processing is used for the detection of Lung cancer disease using the Lungs CT image of a human. Lung cancer detection using image processing involves several steps like image acquisition, image pre-processing, image segmentation, feature extraction, and classification. For the detection of cancer at its earlier stage, the features must be extracted accurately from CT images. In this research work, early detection of lung cancer classification system uses SURF descriptor along with CNN as classifier have been proposed. The evaluation-based optimization algorithm named Genetic Algorithm is used to optimize the data value based on the evaluation criterion of fitness functions. To extract features from CT images the neural network-based classification mechanism i.e. Convolution Neural Network has been utilized. This classifier is mainly based on three different layers such as input layer, pooling layer, and convolution layer. Previously various works have been proposed for CT images, to evaluate the performance of this proposed work based on few parameters named as Error (%), classification Time, True Positive, True Negative, False Positive, False Negative, and Accuracy (%). The value of True Positive of the proposed work is 98.7 and Senthil Kumar et al. (2019) is 95.28. The average value of accuracy is 99%.

REFERENCES

[1]. Smith, R. A., von Eschenbach, A. C., Wender, R., Levin, B., Byers, T., Rothenberger, D., & Saslow, D. (2001). American Cancer Society guidelines for the early detection of cancer: update of early detection guidelines for prostate, colorectal, and endometrial cancers: Also: update 2001—testing for early lung cancer detection. CA: a cancer journal for clinicians, 51(1), 38-75.

[2]. Al-Tarawneh, M. S. (2012). Lung cancer detection using image processing techniques. Leonardo Electronic Journal of Practices and Technologies, 11(21), 147-58.

[3]. Chaudhary, A., & Singh, S. S. (2012, September). Lung cancer detection on CT images by using image processing. In 2012 International Conference on Computing Sciences (pp. 142-146). IEEE.

[4]. Virmani, A. K., Fong, K. M., Kodagoda, D., McIntire, D., Hung, J., Tonk, V., & Gazdar, A. F. (1998). Allelotyping demonstrates common and distinct patterns of chromosomal loss in human lung cancer types. Genes, Chromosomes and Cancer, 21(4), 308-319.

[5]. Wu, A. H., Henderson, B. E., Thomas, D. C., & Mack, T. M. (1986). Secular trends in histologic types of lung cancer. Journal of the National Cancer Institute, 77(1), 53-56.

[6]. Tian, R. H., Zhang, Y. G., Wu, Z., Liu, X., Yang, J. W., & Ji, H. L. (2016). Effects of metformin on survival outcomes of lung cancer patients with type 2 diabetes mellitus: a meta-analysis. Clinical and Translational Oncology, 18(6), 641-649.

[7]. Turrisi III, A. T., & Glover, D. J. (1990). Thoracic radiotherapy variables: influence on local control in small cell lung cancer limited disease. International Journal of Radiation Oncology* Biology* Physics, 19(6), 1473-1479.

[8]. Graham, M. V., Purdy, J. A., Emami, B., Harms, W., Bosch, W., Lockett, M. A., & Perez, C. A. (1999). Clinical dose–volume histogram analysis for pneumonitis after 3D treatment for non-small cell lung cancer

(NSCLC). International Journal of Radiation Oncology* Biology* Physics, 45(2), 323-329

[9]. Koparal, A. T., & Zeytinoglu, M. (2003). Effects of carvacrol on a human non-small cell lung cancer (NSCLC) cell line, A549. Cytotechnology, 43(1-3), 149-154.

[10]. Alam, M., & Ratner, D. (2001). Cutaneous squamous-cell carcinoma. New England Journal of Medicine, 344(13), 975-983.

[11]. Takei, H., Asamura, H., Maeshima, A., Suzuki, K., Kondo, H., Niki, T., ... & Matsuno, Y. (2002). Large cell neuroendocrine carcinoma of the lung: a clinicopathologic study of eighty-seven cases. The Journal of thoracic and cardiovascular surgery, 124(2), 285-292.

[12]. Hadavi, N., Nordin, M. J., & Shojaeipour, A. (2014, June). Lung cancer diagnosis using CT-scan images based on cellular learning automata. In 2014 International Conference on Computer and Information Sciences (ICCOINS) (pp. 1-5). IEEE.

[13]. Chaudhary, A., & Singh, S. S. (2012, September). Lung cancer detection on CT images by using image processing. In 2012 International Conference on Computing Sciences (pp. 142-146). IEEE.

[14]. Alam, N., Park, B. J., Wilton, A., Seshan, V. E., Bains, M. S., Downey, R. J., ... & Amar, D. (2007). Incidence and risk factors for lung injury after lung cancer resection. The Annals of thoracic surgery, 84(4), 1085-1091.

[15]. Malhotra, J., Malvezzi, M., Negri, E., La Vecchia, C., & Boffetta, P. (2016). Risk factors for lung cancer worldwide. European Respiratory Journal, 48(3), 889-902.

[16]. Gupta, D., Boffetta, P., Gaborieau, V., & Jindal, S. K. (2001). Risk factors of lung cancer in Chandigarh, India. Indian Journal of Medical Research, 113, 142.

[17]. Omenn, G. S., Goodman, G. E., Thornquist, M. D., Balmes, J., Cullen, M. R., Glass, A., ... & Barnhart, S. (1996). Risk factors for lung cancer and for intervention effects in CARET, the Beta-Carotene and Retinol Efficacy Trial. JNCI: Journal of the National Cancer Institute, 88(21), 1550-1559

[18]. Yun, C. H., Boggon, T. J., Li, Y., Woo, M. S., Greulich, H., Meyerson, M., & Eck, M. J. (2007). Structures of lung cancer-derived EGFR mutants and inhibitor complexes: mechanism of activation and insights into differential inhibitor sensitivity. Cancer cell, 11(3), 217-

[19]. Xu, Y., Xing, Y., Chen, Y., Chao, Y., Lin, Z., Fan, E., & Shi, Y. (2006). Structure of the protein phosphatase 2A holoenzyme. Cell, 127(6), 1239-1251.