Journal of
Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.777644

SEGMENTATION AND VALIDATION OF LUNG CANCER IMAGES BY USING ARTIFICIAL INTELLIGENCE

Anshu Sharma[1], Dr.Pankaj Nanglia[2], Dr.Vikrant Shokeen[3],Dr.Paramjit[4], Dr.Devinder Rathee[5]

[1]Research Scholar,Maharaja Agrasen University,Baddi
[2]Associate Professor, Maharaja Agrasen University,Baddi
[3]Assistant Professor,MSIT Delhi
[4]Assistant Professor, Maharaja Agrasen University,Baddi
[5]Associate Professor, Adama Science and Technology University,Ethiopia

Corresponding Author: Dr.Pankaj Nanglia, Maharaja Agrasen University, Baddi

Abstract-Segmentation and validation of lung cancer images using artificial intelligence (AI) is a crucial area of medical imaging that can aid in early diagnosis, treatment planning, and monitoring of lung cancer patients. This research work designed a Computer Aided design model that enabled for the detection of Lung cancer with minimum execution time and least mean square error supported by SURF, GA, and FFBP-NN. The hybrid model also tested SIFT and PCA for feature vector in the preliminary phases. The PSNR after adding the white Gaussian noise to the feature vectors stood 44.16, 46.94 and 48.68 namely for PCA, SIFT, and SURF. The present research work used 75% for the training and rest for the classification. The feature vectors obtained by SURF are modified using GA. GA results into a reduced set of SURF feature vector by 22%. The optimized feature set is then passed to SVM and trained by FFBP-NN. The classification parameters are sensitivity and classification accuracy. The proposed algorithm obtains a classification accuracy of 97.89% whereas the obtained average sensitivity is 95.8%. It is then compared with previous implemented nanglia et.al (2020) algorithm and stands a growth of 1-2.5% for both the parameters. The future possibilities may include variation in the fitness function in the present work. The future research workers may also try their hand at varying neuron count for training.

Introduction

Lung cancer remains one of the most common cancers worldwide. It is a significant public health concern, with millions of new cases diagnosed each year. Lung cancer is broadly categorized into two main types:

- Non-Small Cell Lung Cancer (NSCLC): This is the most common type, comprising around 85% of all lung cancer cases. It includes subtypes like adenocarcinoma, squamous cell carcinoma, and large cell carcinoma.
- Small Cell Lung Cancer (SCLC): This type is less common but tends to be more aggressive.

Smoking continues to be the leading cause of lung cancer, although non-smokers can also develop the disease. Other risk factors include exposure to second hand smoke, occupational exposures (e.g., asbestos, radon), and genetic factors [1-4].

CAD can be very useful for early detection of lung cancer. CAD divides the detection process into two phases namely Training and Classification. Figure 1.1 describe the working of CAD. Figure 1.1 shows the two processes namely training and classification. The classification process is often termed as detection process as well [5-7]. It starts with the training mechanism. The training involves feature extraction, followed by optimization.

The optimized feature vector is passed to the learning model of the classification algorithm. As shown in Figure 3.1, after training, the trained set is stored in a database. There are several feature extraction and optimization algorithms. Some of them are listed in the literature [8-9] utilized geometric feature descriptors namely SIFT and SURF is utilized to extract feature vectors. These extracted feature vectors are then utilized in a ratio by the proposed algorithm.

Later, the classification is done using k-Nearest Neighbour (K-NN). The designed model has used data-driven template-matching approach over the CT images [10-13] utilized SURF and presented a neighbourhood selection algorithm for the optimization of key points. The extracted features are utilized for the automated registration process. It took scaled data in the first step.

The extracted feature set is optimized using GA and then a hyperplane selection algorithm for Support Vector Machine (SVM) [14-17] is applied. The training and classification both the activities are performed by SVM utilized both (MRI) and CT images for the analysis SIFT and SURF is used for feature vector extraction. The evaluated results show that SURF depicts a number of key-points as compared to SIFT and it gets an edge over SIFT in this comparison & has been Utilized GA for the betterment of feature vectors [18-22].

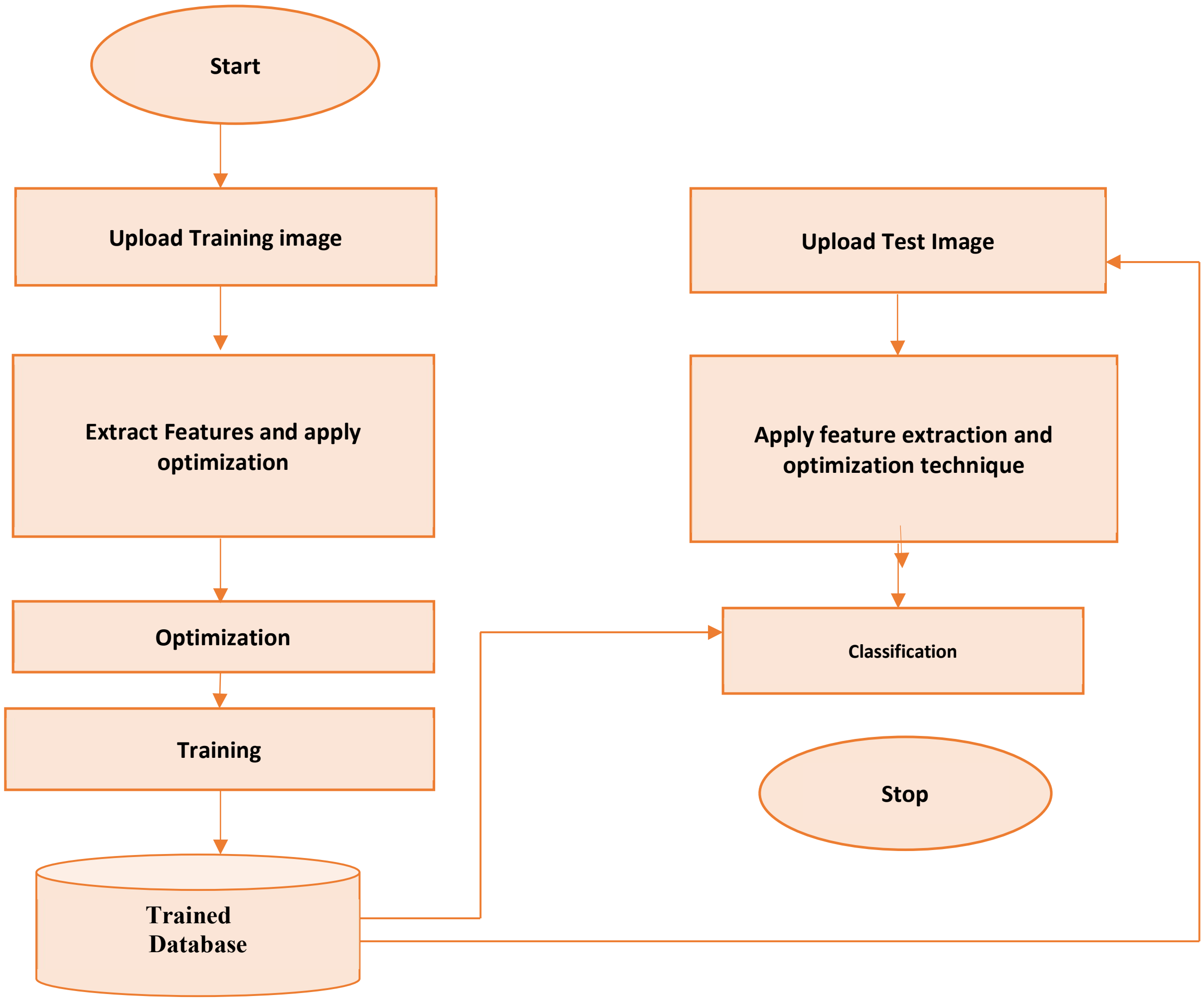


Figure 1.1: Working of Computer Aided Design Model (CAD)

The optimized vector is passed to the Radial Basis Function (RBF) of SVM for classification. The performance evaluation is done on the basis of accuracy weight, True Positive (TP), and False Positive (FP) ratio, also utilized GA for optimization of feature vectors. Further, the work is carried out for template matching, proposed lung cancer classification mechanism using K-NN and cuckoo search. The utilized fitness function aims to reduce the complexity of classification proposed the classification mechanism using Artificial Neural Network. The proposed work is compared with the nanglia.et.al algorithm [23-24].

Proposed Solution

The proposed solution uses SURF for feature vector extraction. As per the literature, optimization is necessary for feature vectors and therefore, GA is applied for the same. The proposed solution trains the optimized feature set using the FFBP Neural Network. The trained set is then passed for the classification.

Feature Extraction and Training:

SURF results in key point feature vector. A total of 800 affected lung cancer images are binarized and then passed as input to it.

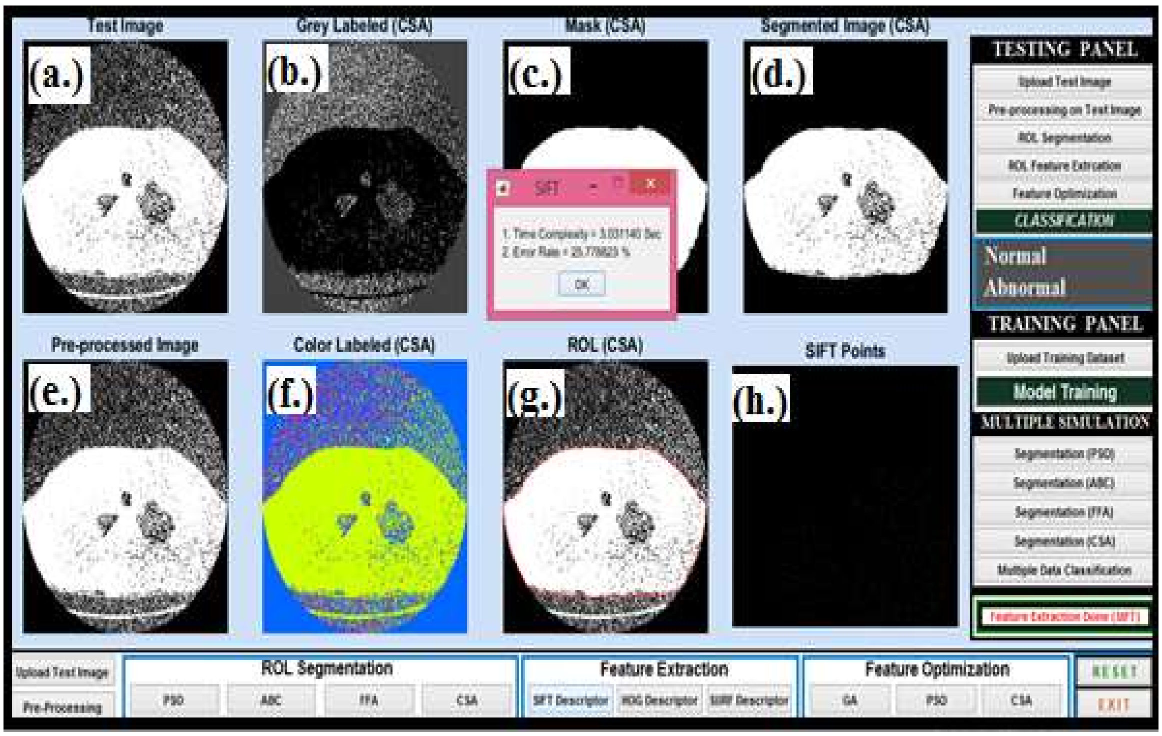


Figure 1.2: Demonstration of different stages of Binarization of the CT image

Table 1.1: Feature vector of SURF, SIFT and PCA

Image	PCA->Principle Eigen Vector)	SIFT-> Key Points	SURF -> Key Points
Image 1(255 * 394)	256 *1	252*252	221*64
Image 2 (194*256)	256 *1	253*252	166*54
Image 3 (182*278)	256 *1	252*252	169*32
Image 4 (196*200)	256 *1	253*252	192*52

Here table 1.1. Illustrates the feature vectors extracted by SIFT, PCA, and SURF.

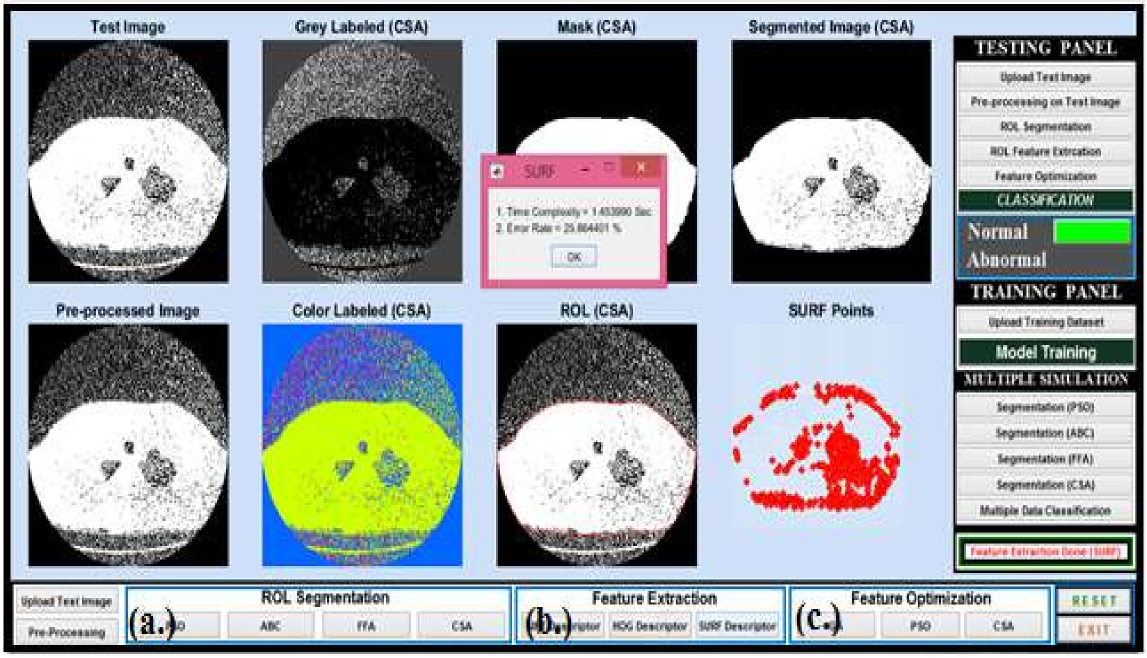


Figure 1.3: Implementation of Feature Extraction algorithms in Testing & Training Panel

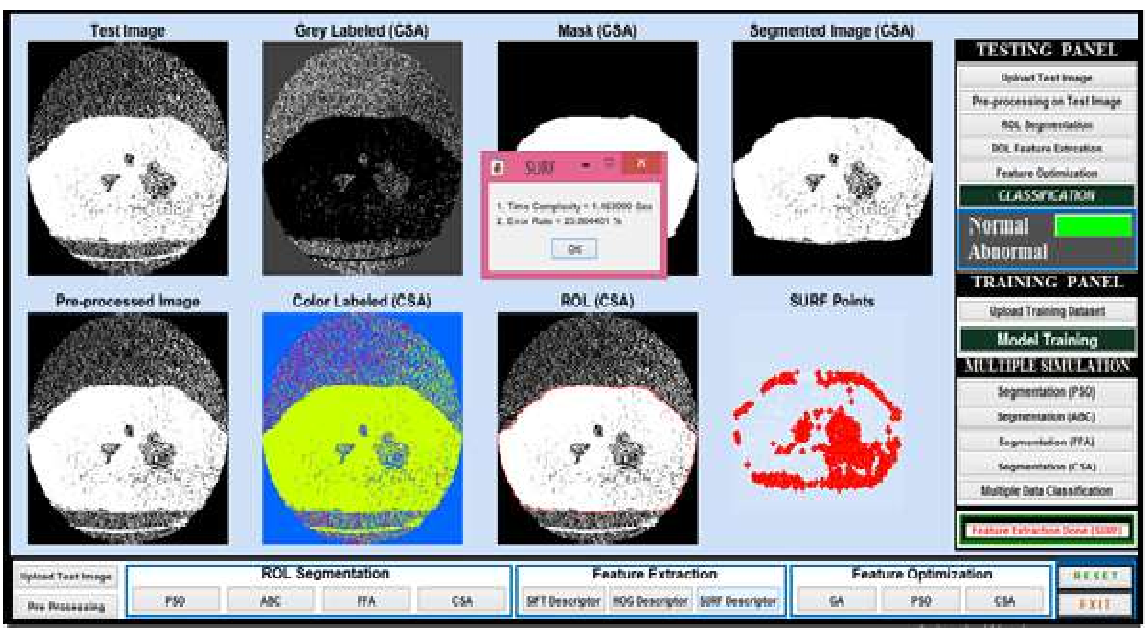


Figure 1.4: Feature vectors extracted by SURF, SIFT and HOG algorithms

SURF key points maps to most of the ROI part extracted in Figure 1.4 whereas, for the same input image, the key points of other algorithms do not provide close competition to SURF. The proposed work has tested the key points by adding some noise to all the key features of each algorithm. How much any feature is pronto noise, can be evaluated by Peak Signal-to-Noise Ratio (PSNR). Each feature set is passed with the white Gaussian noise of intensity 0.1 followed by the PSNR evaluation.

$$\mathrm{PSNR} = 10 * \mathrm{Log10}\left(\frac{256^2}{\mathrm{MSE}}\right)$$

Table 1.2: Peak Signal- to -Noise Ratio (PSNR) with white Gaussian Noise

PCA---- Average PSNR	SIFT----- Average PSNR	SURF---- Average PSNR
44.16	46.94	48.68

The PSNR of SURF is recorded to be the highest among all the three feature extraction technique. Hence SURF is finally selected as the key feature holder for further processing. The extracted feature set is passed to GA for the optimization and the following fitness function is utilized for the processing [24-28].

Table 1.3: Used Genetic Fitness Function

GA Fitness	if $f(r) > f(t)$
0 otherwise	

Where f (r) is the current feature vector and f (t) is the threshold of selection. The SURF features are reduced by 22-25% after implementing GA. The optimized feature set is passed to FFBP NN for training. FFBP NN is a multiclass classifier with Feed forwarding in order to satisfy the stopping constraints and it back propagates in order to validate the Mean Square Error (MSE). Neural Network mainly encompasses the three-layer mechanism. A brief description of layers of Neural Network is as follows.

1. Input Layer: Takes the feature vector as input. No processing is done at this layer other than forwarding the data to the intermediate layer with ‘n’ number of neurons.
2. Hidden Layer: The hidden layer utilizes weight function to convert data into a format, which is understandable by the architecture of machine learning. The weight function may be of the following types:

Linear: $w = ax + b$

Quadratic: $w = ax^2 + bx + c$

Polynomial: $w = (ax + b)^k$

Where w is the converted weight of the input set a, b and c are the arbitrary constants where k is the polynomial power constraint. The simulation is performed by using MATLAB's Neural Network Toolbox. The toolbox helps in the selection of the type of weight function depending upon the type of data utilized [29-30].

1. Output Layer: It can be further divided into two sub-categories namely the output layer for the training data and output layer for the test data. The training architecture of utilized Feed Forward Back Propagation Neural Network is as follow.

Algorithm Train_NN (N1, Opt_set1, N2, Opt_set2)

Where N1= Total number of cancerous images

Opt1= Optimized Feature vector for N1

N1= Total number of cancerous images

Opt1= Optimized Feature vector for N1

1. gcount=1; Trainset= [ ]; //group count
2. For each fvec in Opt1.Row
3. Train_set(gcount,:)=Opt1_set1.Row.Value // Placing feature vector
4. Group(gcount)=1;
5. Gcount=gcount+1;
6. End For
7. Foreach fvec1 in Opt_set2.Row
8. Train_set(count,)=Opt_set.Row.
9. Group(count)=2
10. End For
11. Net=newff (Train_set, Group,20) // Initializing Neural Network(NN) Machine Learning with the train set, its target set named as a group and 20 Neurons. The neurons will help the hidden layer to initiate the weight machines [31].

12. Net. Training parameters. Add parameters (total iteration) =50; // Initialized neuron framework with 50 iterations at max. If the provided set of iterations is 50 even, then it is not necessary that it will have to complete 50 iterations. It depends upon the validation parameter value that how much iterations are required to process.

13. Trained Machine (Net); // Performing Training. The train set is initialized as empty and group set is accordingly set. The train set will contain both the values of the cancerous and the non-cancerous feature set. This data is passed to Neural training with varying weight function and the following training architecture is obtained [89].

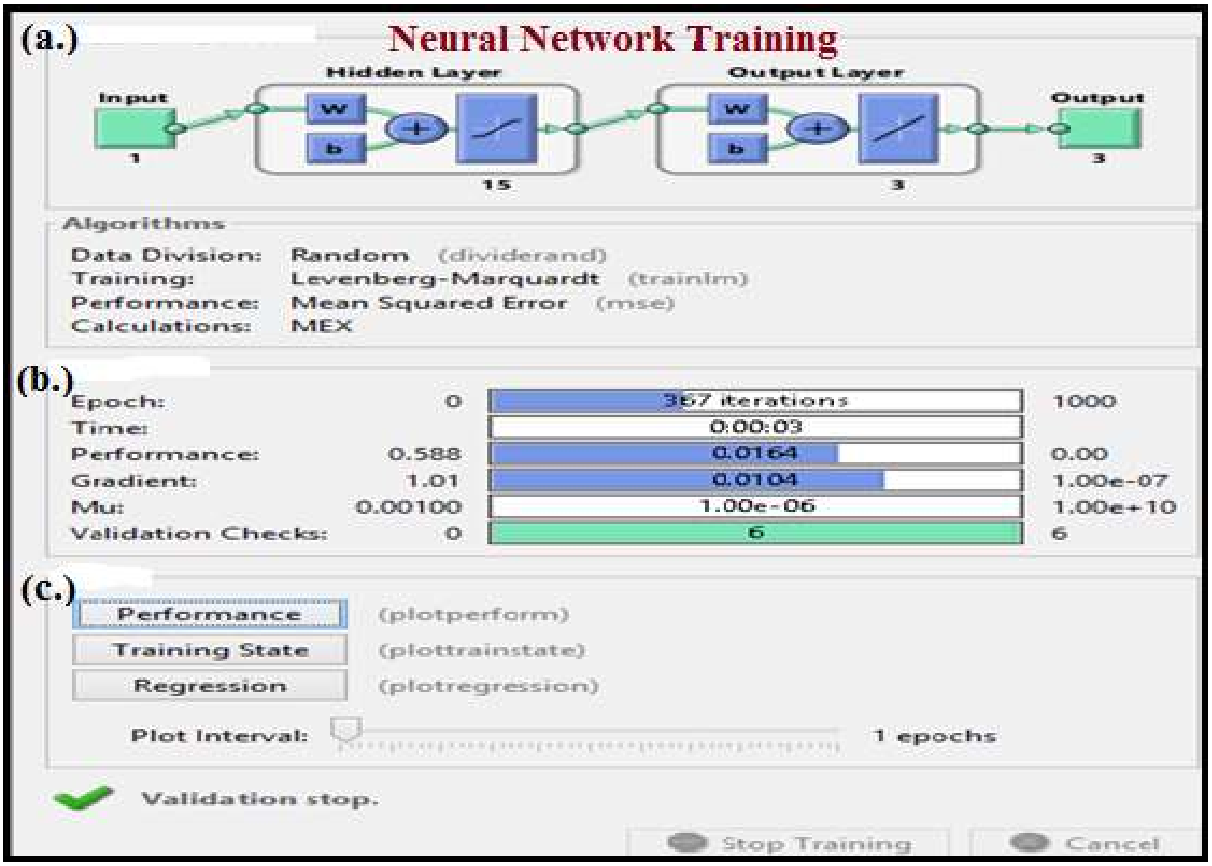


Figure 1.5: Feed Forward Back Propagation Neural Network (FFBPNN) Training

Figure 1.5 demonstrates the complete working of the Neural Network training mechanism. The parametric architecture for Neural Network is as follows:

i. Total Iterations Max=600;
ii. Validation Parameters: 4
iii. Passed Number of Neurons:15
iv. Attained Iteration: 350-400
v. Numbers of Times Training done: 100

The utilization of NN gives an edge of checking the back-propagation mechanism. Figure 1.6 demonstrates the results of back propagation [31].

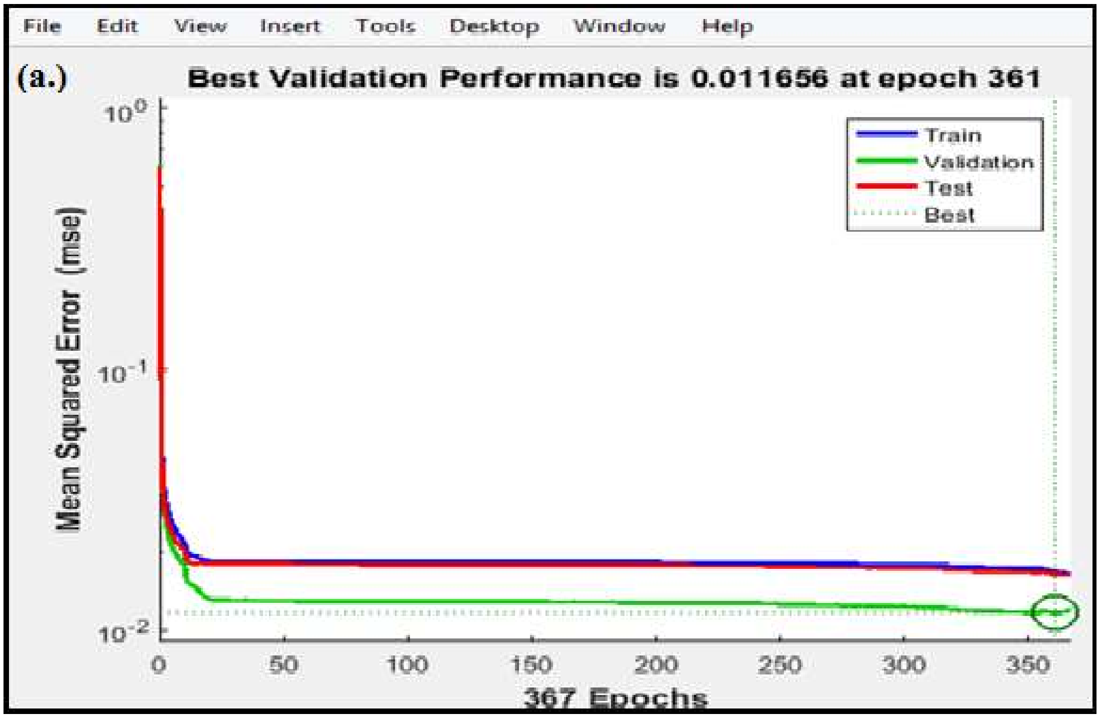


Figure 1.6: Back Propagation (BP)

Figure 1.6 represents the back propagation of NN. Out of 8 iterations, the structure of 2nd iterations are selected. This is because of the cross-reference of train data and validation data generated by the tool itself [92-94].

Experimental Results

The train set contains around 75% of the total data and rest is used as a test set. Sensitivity and accuracy are calculated and is compared with (Kuruvilla and Gunavathi,). Sensitivity is the measure of true classified images out of true-labelled images.

$$\text{Sensitivity} = \text{True Positive} / \text{True Positive} + \text{False Negative}$$

Accuracy is the measure of performance in terms of true classification. In other words, how efficiently a classifier judges the affected samples, can be measured by accuracy.

$$\text{Accuracy} = (\text{True Positive} + \text{True Negative})/(\text{True Positive} + \text{True Negative} + \text{False Positive} + \text{False Negative})$$

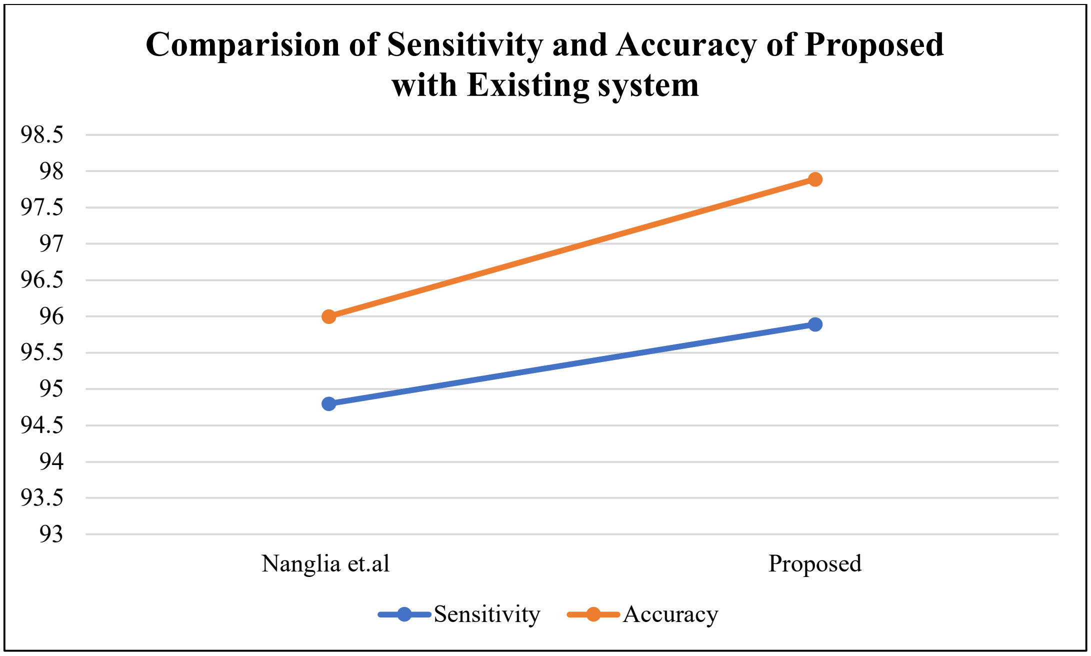
True Negative= Left not affected samples out of total samples

False Negative = Left affected samples out of total samples

True Positive = True classified samples out of total samples

False Positive= False classified samples for true samples

The average classification accuracy obtained by the proposed algorithm is 97.89%. The average sensitivity of the proposed algorithm is 95.8%. Compared to (nanglia et.al, 2021), the percentage growth is about 1-2%.



Conclusion

Segmentation and validation of lung cancer images using artificial intelligence (AI) is a crucial area of medical imaging that can aid in early diagnosis, treatment planning, and monitoring of lung cancer patients. In this research article, lung cancer prediction models have been proposed to assist clinicians in managing incidental or screen detected indeterminate pulmonary nodules. If, the malignancy in the lung nodules is recognized at the beginning phase, then survival chances may increase. Machines images that have generally been used to diagnose cancer within the body are X-Ray, CT, MRI, and PET. Radiologists have proposed different lung cancer prediction models like, feature fusion mechanism, reference-model, 3D-CNN, optical flow methods, and cloud-based CNN CAD system. This study also enables us to measure the challenges, benefits, and limitations of existing techniques. In the medical industry, Computer Aided Detection (CAD) aims to optimize the classification process.

This research work designed a Computer Aided design model that enabled for the detection of Lung cancer with minimum execution time and least mean square error supported by SURF, GA, and FFBP-NN. The hybrid model also tested SIFT and PCA for feature vector in the preliminary phases. The PSNR after adding the white Gaussian noise to the feature vectors stood 44.16, 46.94 and 48.68 namely for PCA, SIFT, and SURF. The present research work used 75% for the training and rest for the classification. The feature vectors obtained by SURF are modified using GA. GA results into a reduced set of SURF feature vector by 22%. The optimized feature set is then passed to SVM and trained by FFBP-NN. The classification parameters are sensitivity and classification accuracy. The proposed algorithm obtains a classification accuracy of 97.89% whereas the obtained average sensitivity is 95.8%. It is then compared with previous implemented nanglia et.al (2020) algorithm and stands a growth of 1-2.5% for both the parameters. The future possibilities may include variation in the fitness function in the present work. The future research workers may also try their hand at varying neuron count for training.

REFERENCES

1. Shubham Dodia, Annappa B., Padukudru A. Mahesh,Recent advancements in deep learning based lung cancer detection: A systematic review, Engineering Applications of Artificial Intelligence, Volume 116,2022,105490,ISSN 0952-1976,https://doi.org/10.1016/j.engappai.2022.105490.(https://www.sciencedirect.com/science/article/pii/S0952197622004808).
2. P. Nanglia, S. Kumar, A. N. Mahajan, P. Singh and D. Rathee, "A hybrid algorithm for lung cancer classification using SVM and Neural Networks," ICT Express Science Direct, vol. 3, no. 2, pp. 1-7, 2021.
3. P. Nanglia, A. N. Mahajan, D. S. Rathee and S. Kumar, "Lung cancer classification using feed forward back propagation neural network for CT images," International Journal of Medical Engineering and Informatics, vol. 12, no. 5, pp. 447-456, 2020
4. P. Nanglia, A. N. Mahajan, P. Singh and D. Rathee, "Detection & Classification of Lung Cancer at an Early Stage by Applying Feature Extraction Optimization and Neural Network on Hybrid Structure," International Journal of Innovative Technology and Exploring Engineering (IJITEE) Scopus, vol. 9, no. 2, pp. 3737-3745, 2019.
5. P. Singh, P. Nanglia and A. N. Mahajan, "Improved Lung Cancer Segmentation Using K-Means and Cuckoo Search," International Journal of Innovative Technology and Exploring Engineering (IJITEE) Scopus, vol. 9, no. 2, pp. 3746-3758, 2019.
6. P. Nanglia, S. Kumar, D. Rathee and P. Singh, "Comparative Investigation of Different Feature Extraction Techniques for Lung Cancer Detection System," in International Conference on Advanced Informatics for Computing Research (ICAICR) Springer, Himachal Pradesh, 2018.
7. Shallu, P. Nanglia, S. Kumar and A. K. Luhach, "Detection and Analysis of Lung Cancer Using Radiomic Approach," in International Conference on Computational Strategies for Next Generation Technologies Springer, Punjab, 2017.
8. In Palani, D., &Venkatalakshmi, K. (2019). An IoT Based Predictive Modelling for Predicting Lung Cancer Using Fuzzy Cluster Based Segmentation and Classification. Journal of Medical Systems, 43(2), 21.
9. Lakshmanaprabu, S. K., Mohanty, S. N., Shankar, K., Arunkumar, N., & Ramirez, G. (2019). Optimal deep learning model for classification of lung cancer on CT images. Future Generation Computer Systems, 92, 374-382.
10. D. Palani and K. Venkatalakshmi, "An IoT based predictive modelling for predicting lung cancer using fuzzy cluster based segmentation and classification," J. Med. Syst., vol. 43, no. 2, p. 21, 2018.

11. Lakshmanaprabu, S. N. Mohanty, Shankar, Arunkumar, and G. Ramirez, "Optimal deep learning model for classification of lung cancer on CT images," Future Gener. Comput. Syst., vol. 92, pp. 374–382, 2019.
12. Shallu, Nanglia. P., Kumar. S., Shukla. P.: Detection and analysis of lung cancer using radiomic approach. In: International Conference on Computational Strategies for Next Generation Technologies. in the press. (2017).
13. Radhakrishnan, M., Kuttiannan, T.: Comparative analysis of feature extraction methods for the classification of prostate cancer from the Trus Medical images. International Journal of computer science, vol. 9, pp. 171-179 (2012).An Information Resource on Lung Cancer Testing for Pathologists and Oncologists, https://Www.Verywell.Com/Non-Small-Cell-Lung-Cancer-2249281

14. Makaju, Suren., Prasad, P.W.C., Alsadoon., Singh, A. K.: Lung Cancer Detection using CT Scan Images, 6th International Conference on Smart Computing and Communications Kurukshetra. pp. 107-114. (2017)
15. Prem, Chander, M., Venkateshwara, Rao, M., Rajinikanth, T.-V.: Detection of Lung Cancer Using Digital Image Processing Techniques: A Comparative Study. International Journal of Medical Imaging. vol.5, pp. 58-62 (2017). doi: 10.11648/j.ijmi.20170505.12
16. Ma, L., Wang, Debby, D., et al.: An eigen-binding site based method for the analysis of anti-EGFR drug resistance in lung cancer treatment. IEEE/ACM transactions on computational biology and bioinformatics. vol. 14, pp. 1187-1194 (2017). doi-10.1109/TCBB.2016.2568184
17. Hussein, A, El., Tynga, I.-M., Harith, M.-A.: Comparative study between the photodynamic ability of gold and silver nanoparticles in mediating cell death in breast and lung cancer cell lines. Journal of Photochemistry and Photobiology Biology. vol. 153, pp. 67–75 (2015). doi.org/10.1016/j.jphotobiol.2015.08.028
18. Song. Y., Cai. W., Zhou. Y., Feng. D.D.: Feature-based image patch approximation for lung tissue classification. IEEE Trans. Med. Imag. vol. 30, pp. 797-808. (2013)
19. Wang, J., Watada, J.: Panoramic image mosaic based on SURF algorithm using Open CV. IEEE 9th International Symposium on Intelligent Signal Processing (WISP) Proceedings. pp. 1-6 (2015). doi: 10.1109/WISP.2015.7139183
20. Saini, A.-K., Bhadauria, H.-S., Singh, A.: A Survey of Noise Removal Methodologies for Lung Cancer Diagnosis. Second International Conference on Computational Intelligence & Communication Technology (CICT) Ghaziabad. pp. 673-678 (2016). doi: 10.1109/CICT.2016.139
21. Teramoto, A., Fujita, H.: Fast lung nodule detection in chest ct images using cylindrical nodule-enhancement filter. Int J Comput Assist Radiol Surg., vol. 8, no. 2, pp. 193–205. (2013)
22. Kulkarni, A., Panditrao, A.: Classification of lung cancer stages on CT scan images using image processing. IEEE International Conference on Advanced Communications Control and Computing Technologies. pp. 1384-1388 (2014). doi: 10.1109/ICACCCT.2014.7019327

23. Han, G., Liu, X., Han, F., et al.: Lung Diseases for Computer-Aided Detection and Diagnosis Research and Medical Education. IEEE Transactions on Biomedical Engineering. vol. 62, pp. 648-656 (2015). doi: 10.1109/TBME.2014.2363131.
24. Shivakumar, B.-L., S, Santhosh.: Automated Forensic Method for Copy-Move Forgery Detection based on Harris Interest Points and SIFT Descriptors. International Journal of Computer Applications, vol. 27, pp. 9-17 (2011).DOI: 10.5120/3283-4472.
25. Chumerin, N., Van, Hulle, M.-M.: Comparison of Two Feature Extraction Methods Based on Maximization of Mutual Information. In: Proceedings of the 16th IEEE Signal Processing Society Workshop on Machine Learning for Signal Processing. pp. 343-348 (2006). doi: 10.1109/MLSP.2006.275572
26. P, Lambin., Velazquez, E, Rios., et al.: Radiomics: extracting more information from medical images using advanced feature analysis. European journal of cancer, vol. 48, pp. 104-118 (2012). doi:10.1016/j.ejca.2011.11.036
27. Cateni, S., Vannucci, Marco., et al.: Variable Selection and Feature Extraction through Artificial Intelligence Techniques. In: Multivariate Analysis in Management, Engineering and Science. vol. 6, pp. 103-118 (2013). DOI: 10.5772/53862.
28. Kang, H., A, Efros, Alexei., Hebert, Martial., et al.: Image Matching in Large Scale Indoor Environment. IEEE International Conference on Computer Vision and Pattern Recognition Workshop. pp. 33-40 (2009). doi: 10.1109/CVPRW.2009.5204357
29. Huijuan, Z., Qiong, H.: Fast image matching basedon improved SURF algorithm. IEEE International conference Electronics Communications and Control (ICECC). pp. 1460-1463 (2011). doi:10.1109/ICECC.2011.6066546
30. Jin, X., Zhang, Y., &Jin, Q.: Pulmonary Nodule Detection Based on CT Images Using Convolution Neural Network. 9Th International Symposium On Computational Intelligence And Design (ISCID). (2016) doi: 10.1109/ISCID.2016.1053
31. Yin, Y., Sedlaczek, O., Muller, B., Warth, A., Gonzalez-Vallinas, M., Grabe, N.,Drasdo, D.: Tumor cell load and heterogeneity estimation from diffusion-weighted MRI calibrated with histological data: an example from lung cancer. IEEE Transactions on Medical Imaging (2017).