

AN EXTENDED FRAMEWORK OF LUNG CANCER CLASSIFICATION USING HYBRID ARCHITECTURE OF SURF AND SVM

Paramjit Singh¹ Dr. Pankaj Nanglia² Vikrant Shokeen³ Dr. Aparna N Mahajan⁴

¹Research Scholar, Maharaja Agrasen University, Himachal Pradesh, India

²Assistant Professor, Maharaja Agrasen University, Himachal Pradesh, India

³Assistant Professor, MSIT, Delhi, India

⁴Professor, Maharaja Agrasen University, Himachal Pradesh, India

Email: ¹ppparamjitsingh@gmail.com, ²deputyregistrar@mau.edu.in,

³vikrantshokeen@msit.in ⁴directormait2014@gmail.com

Abstract: The present research work focussed on the Lung Cancer disease classification by the potential usage of hybrid model in which segmentation, feature extraction, optimization and classification techniques has been performed on the dataset of CT scan images of 1000 images. A set of 1000 images are to be utilized in which 75% data is used for the test purpose and rest 25% is used for classification. The present research article measured the performance of hybrid model by applying the post segmentation techniques Particle Swarm Optimization (PSO), Artificial Bee Colony (ABC), FFA (Fire Fly Algorithm) Cuckoo Search (CS), and best features extraction technique Speed Up Robust Feature (Surf) in the terms of minimum execution time and minimum error rate with classifier Support Vector Machine (SVM) is also used as cross validator for the evaluation of the performance of hybrid model in the terms of parameter accuracy, error rate, precision, recall and execution time. The overall accuracy of hybrid model has 98.90%, recall value 91.46%, F-measure 94.73% and minimum execution time .00031 secs has been achieved for the hybrid model.

Keywords: Lung Cancer Detection, Feature Extraction, SIFT, SURF, PCA, ABC, PSO, FFA, CSA and SVM.

I Introduction

Lung cancer is one of the most precarious and critical disease, which is often in the modern time. More than 10,000 deaths per year occurs only in India and rate of deaths is increasing by 10 to 12 % every year [1-2]. The dangerous diseases have momentum and provide the scope for researchers in the medical industry to diagnosis diseases at an early stage. Only the early stage treatment is the

remedy to save the patient and a lot of lives have been saved. Now the researchers and doctors have only the one alternate to automate the complete detection process for the diagnosis of the lung cancer. A lot of attempts have been done by the scholars, doctors and medical experts in this field to carry out the research at next level.

Lung cancer classification requires two phases: In the first stage is called the training phase and the second stage is called the classification stage. The figure 1 shows the architecture diagram for lung cancer classification. The figure depicts that in the training stage an efficient segmented valid key feature is required for the system and to achieve these valid key features various optimization techniques with feature extractions techniques have been implemented. Similarly, same techniques have been implemented on the test images for the best classification and best accuracy. In this context extraction algorithm is applied followed by optimization techniques. Then efficient optimized feature set is passed into the training algorithm, which makes a training supervised architecture. The supervised training architecture is used for the disease classification mechanism [3-4]

There are many studies have been conducted and here some studies discussed related to the present work in this part. Here dataset of the images, advantages, disadvantages and the proposed method accuracy, execution time and error rate have been discussed so that this research work can be carried out on next level. A study conducted by Nanglia et.al (2020) entitled "A hybrid algorithm for lung cancer classification using SVM and Neural

Networks". In this study a dataset of 500 images have been considered. 75% data is used for the purpose of training and rest 25% of the data have been considered for the for the classification. In this research work SURF technique has been for feature extraction, GA algorithms implemented for optimization and SVM classifier with Feed Forward Back Propagation Neural Network (FFBPNN) has been applied on the dataset. This designed structured model is named as hybrid model in which classification accuracy 98.08% have been achieved [1].

Similarly, a study has been conducted by Paramjit et.al (2019) entitled "Improved Lung Cancer Segmentation using K-means and Cuckoo Search". In this study the dataset of CT scan images has been considered and implemented Artificial Bee Colony(ABC) and Cuckoo search applied on this dataset. Here SVM has been used as a cross-validator to find the Accuracy, Precision, Recall and F-measure. The study revealed that improved accuracy by 5%, precision by 6%, recall 3% and F-measure by 4% have been achieved for ABC and accuracy by 10%, precision by 11%, recall 12% and F-measure by 11% have been achieved for Cuckoo search [3].

Another study has been conducted by Nanglia et.al (2019) entitled "Detection & Classification of Lung Cancer at an Early Stage by Applying Feature Extraction Optimization and Neural Network on Hybrid Structure". In this article dataset of 400 CT scan images have been considered for the analysis of best feature extraction techniques and best optimization technique. Further a cross validator Support Vectors Machine(SVM) has been utilized for the better classification. The study reported that hybrid model calculates the execution time of 1.94 sec. which is least and minimum Error rate 29.25 have been noticed in the case of SURF and GA algorithm. The overall classification accuracy for the hybrid model calculated i.e. 99.670% by using this unique hybrid model [2].

Another study which has been carried out by Nanglia et.al (2018) entitled "Comparative Investigation of different feature extraction techniques for lung cancer detection system". The study expressed the comparative analysis between the different feature extraction techniques SIFT, SURF and Principle Component Analysis (PCA). The study found that average execution time.448sec. and average error rate 25.704 have been calculated

for the SURF techniques. Thus the study conclude that SURF technique is the best techniques in the terms of least time execution and least mean square error [4].

II Material and Method:

In this section implementation process of the hybrid model has been discussed. Moreover, the dataset specification has been discussed in that part. The implementation steps to achieve the unique hybrid model have for the detection of lung cancer at an early stage has given below:

- a) Segmentation and Optimization of the images
- b) Comparative analysis of Segmentation and Optimization of the images
- c) Selection of the valid key feature by feature extraction techniques
- d) Comparative analysis has been done between the features extraction techniques and implemented best feature extraction technique on hybrid model.
- e) Training and Classification have been done by using a classifier Support Vector Machine.
- f) Find out results in the terms of parameters accuracy, sensitivity, f-measure, precision, execution time and recall.

2.1 Dataset:

In this study the proposed hybrid model is using the ELCAP lung image dataset which is originally designed and developed by Cornell University. The first release of the dataset was made in 2019, which contains 100 CT Scan Images (documented) with 1 mm slice thickness [5-8].

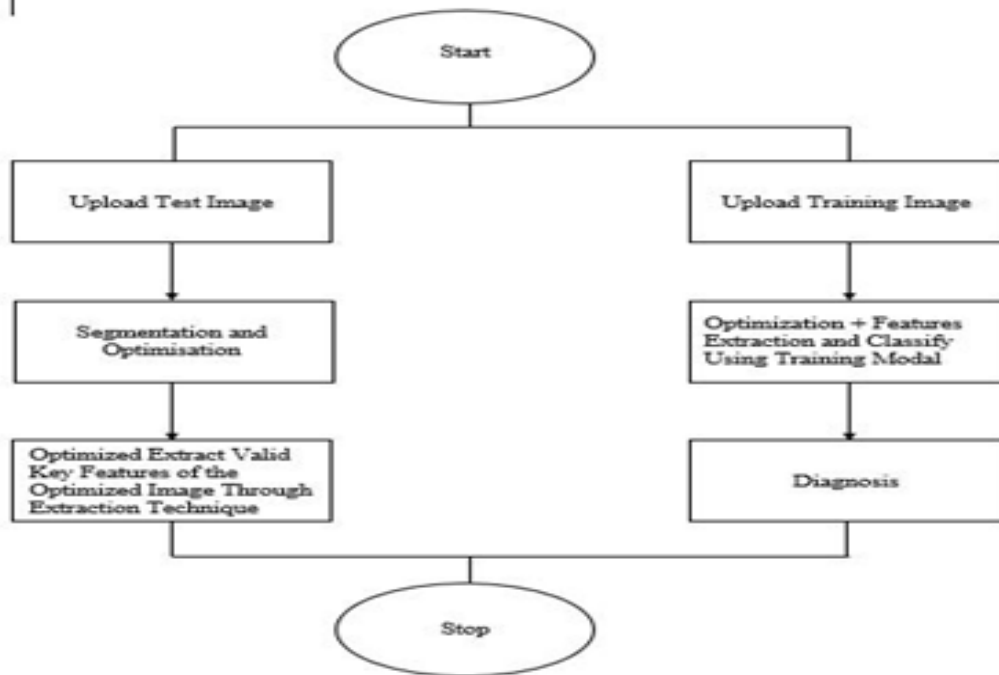
2.2 Method:

The following steps have been performed for the classification of cancerous and non-cancerous images. A graphical use interface has been prepared in which following steps have been implemented step by step: [9-15].

Research methodology have been used for the Classification Process (Train Samples [Cancerous, Non-Cancerous], Test Set)

- Segmentation and Optimization of the images to extract the ROI.
- Extract valid key features for cancerous and non-cancerous images
- Store to database
- End for

- Train Optimized valid key features
- Upload Test Sample
- Extract Feature Vector and Optimize
- Use trained database for classification
- Classify and diagnose
- End



III Experimental Results and Conclusion

3.1 Segmentation and Optimization of the images

In this section various segmentation techniques have been performed on the images to find out the Region of Interest (ROI). Region of interest played a vital role in the early detection and diagnosis of the lung cancer. It provides the region of interest where need to be perform all the operations and it reduce the complexity of the hybrid model and enhanced the execution speed of the operation. This helps the doctors and scholars to save many lives of the patients at the detection on an early stage. In this context Artificial Bee Colony(ABC), PSO (Particle swarm optimization), FFA and CSA (Cuckoo Search) techniques has been performed. The figure 2 shows the testing and training panel of GUI which has been created in the Matlab.16. As clearly shown in the figure there are two panels testing and Model training panel. Segmentation has been done in both the phases simultaneously.

3.2 Comparative analysis of Segmentation and Optimization of the images

A comparative analysis has been done among these techniques in the terms of parameter Accuracy, Sensitivity, F-measure and Precision. Then average mean of all these values have been calculated and analysed. In first case average mean of accuracy for the PSO is 97.14, average mean of sensitivity has been calculated i.e..96, average mean of F-measure has been calculated .97 and average mean of

Precision value .94 has been calculated. Similarly, average mean of all these values have been calculated and analysed in the case Artificial Bee Colony(ABC) and clearly shown in the figure 3. In this case average mean of accuracy for the ABC is 96.12, average mean of sensitivity has been calculated i.e..95, average mean of F-measure has been calculated .96 and average mean of Precision value .94 has been calculated. Similarly, average mean of all these values have been calculated and analysed in the case (FFA) and clearly shown in the figure 5.4. In this case average mean of accuracy for the FFA is 95.07, average mean of sensitivity has been calculated i.e..94, average mean of F-measure has been calculated .93 and average mean of Precision value .93 has been calculated. Similarly, average mean of all these values have been calculated and analysed in the case of Cuckoo Search(CSA) and clearly shown in the figure 5.5. In this case average mean of accuracy for the CSA is 98.14, average mean of sensitivity has been calculated i.e..97, average mean of F-measure has been calculated .98 and average mean of Precision value .98 has been calculated. So, here it has been concluded that on the basis of above mentioned results Cuckoo search is the best optimization techniques and in this hybrid model Cuckoo search(CSA) technique has been applied for the better results.

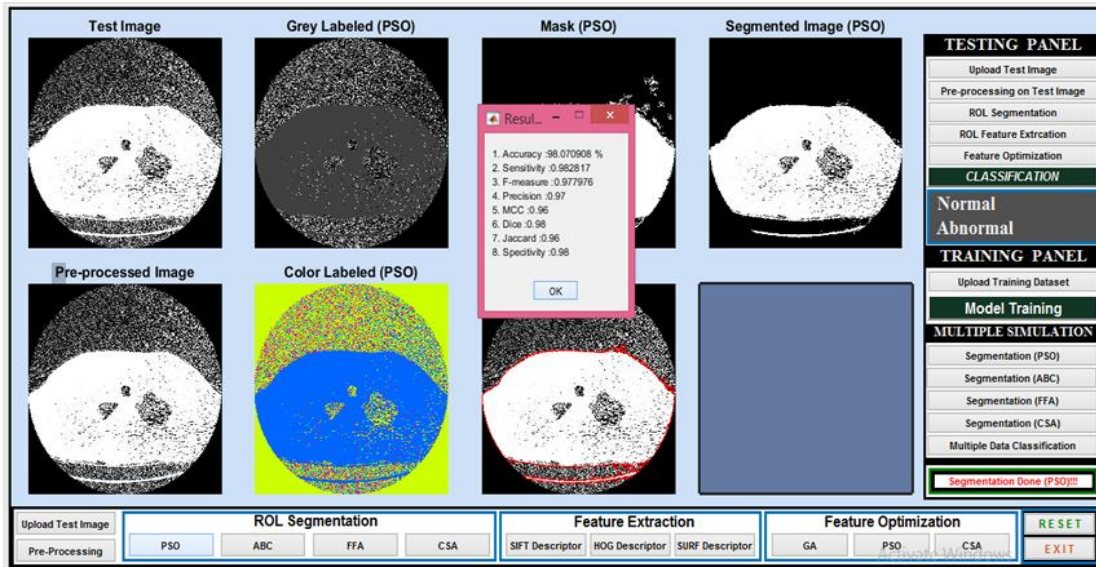


Fig.2. Implementation of PSO technique

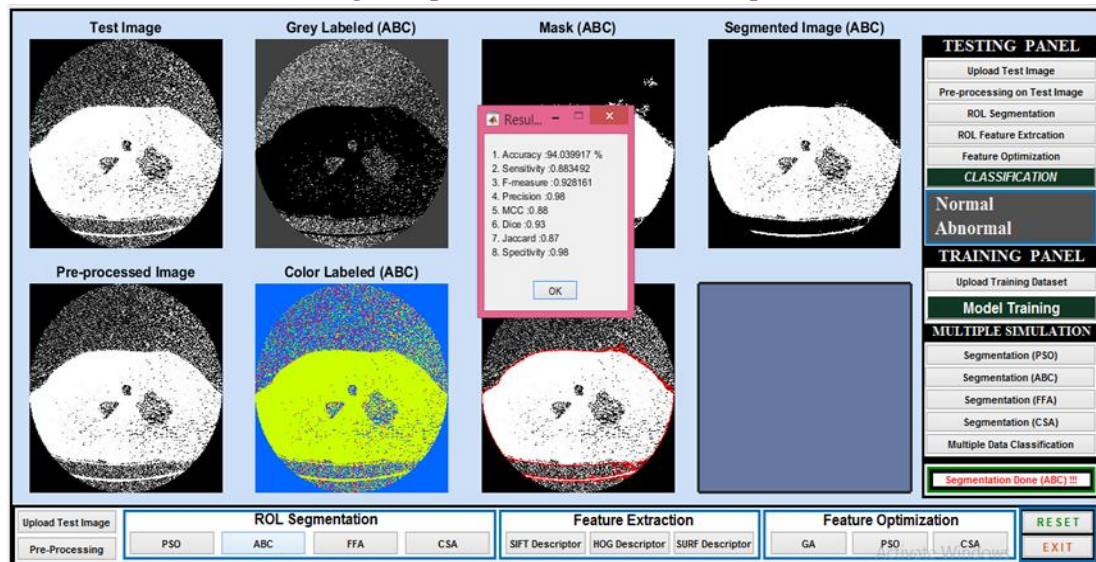


Fig.3. Implementation of ABC technique

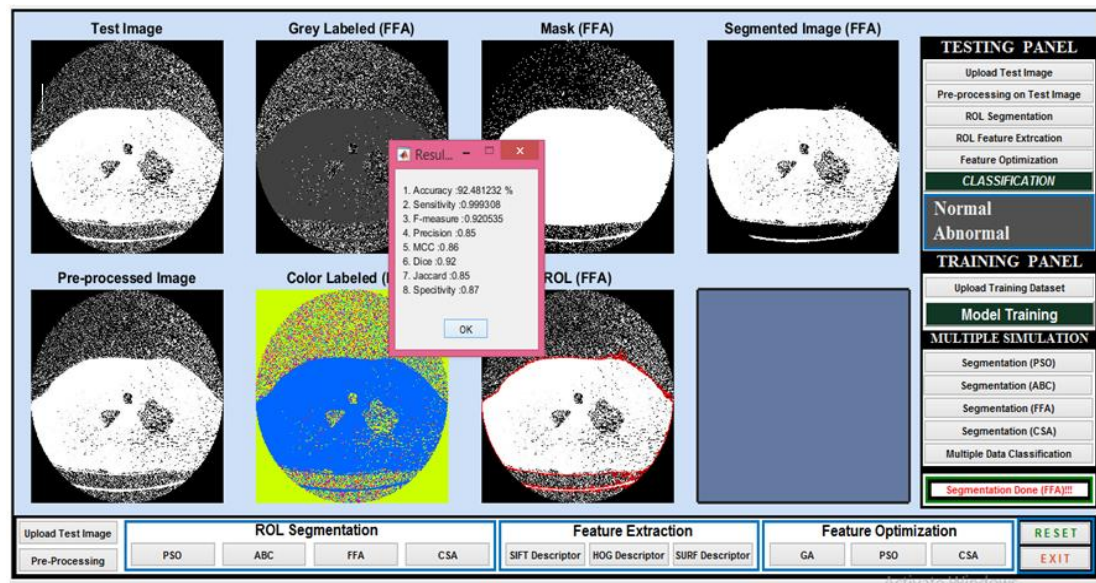


Fig.4. Implementation of FFA technique

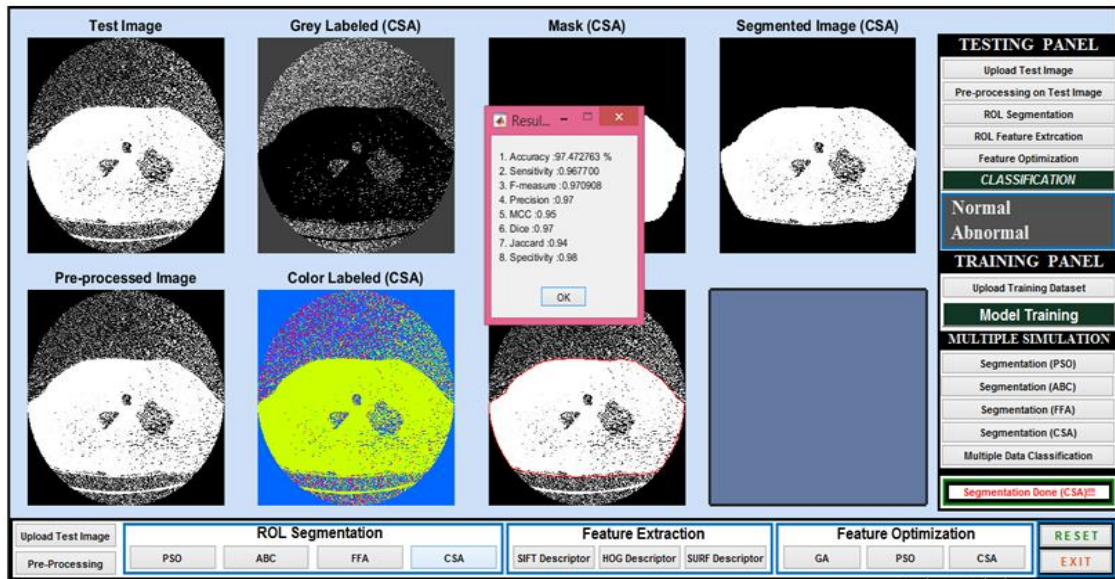
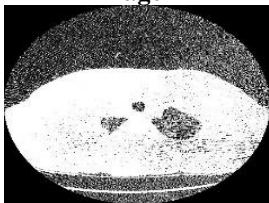
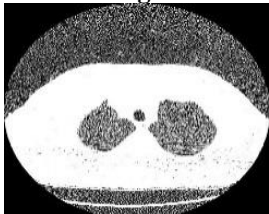
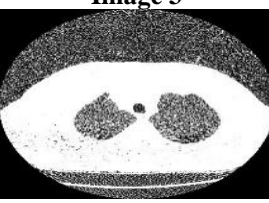
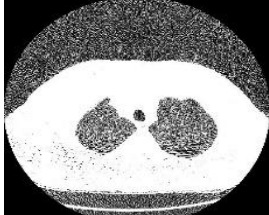
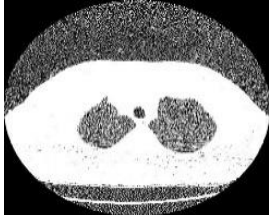
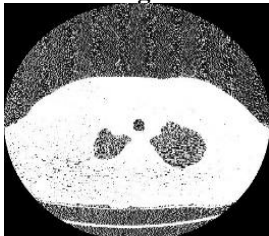
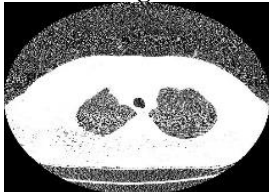
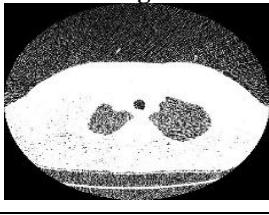
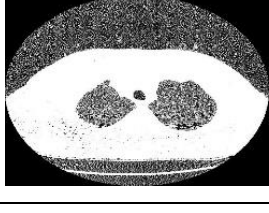


Fig.5. Implementation of CSA technique

Table 1
Comparative analysis among PSO, ABC, FFA and CSA

	Parameters	PSO	ABC	FFA	CSA
Image 1 	Accuracy	98.07	94.04	92.48	98.12
	Sensitivity	0.98	0.88	1.00	0.99
	F- Measure	0.98	0.93	0.92	0.97
	Precision	0.97	0.98	0.85	0.97
Image 2 	Accuracy	97.33	96.81	93.82	98.50
	Sensitivity	0.94	0.93	1.00	0.97
	F- Measure	0.97	0.97	0.94	0.98
	Precision	0.97	0.97	0.94	0.98
Image 3 	Accuracy	94.53	93.83	89.56	94.82
	Sensitivity	0.97	0.95	1.00	0.98
	F- Measure	0.94	0.93	0.89	0.94
	Precision	0.91	0.91	0.81	0.91
Image 4 	Accuracy	98.24	95.99	92.70	98.62
	Sensitivity	0.99	0.94	1.00	1.00
	F- Measure	0.98	0.96	0.93	0.97
	Precision	0.97	0.97	0.86	0.97
Image 5	Accuracy	98.41	80.46	99.87	99.66

	Sensitivity	0.96	0.55	1.00	0.95
	F- Measure	0.98	0.71	0.99	0.97
	Precision	0.91	0.93	0.98	0.99
Image 6	Accuracy	91.91	85.55	93.37	98.37
	Sensitivity	0.83	0.68	1.00	0.97
	F- Measure	0.90	0.81	0.93	0.98
	Precision	0.91	0.82	0.89	0.94
Image 7	Accuracy	99.09	96.24	94.27	98.70
	Sensitivity	0.98	0.92	0.99	0.97
	F- Measure	0.99	0.96	0.94	0.99
	Precision	0.92	0.93	0.92	0.91
	Accuracy	98.48	96.70	96.01	98.49
Image 8	Sensitivity	0.97	0.93	0.92	0.97
	F- Measure	0.98	0.96	0.96	0.98
	Precision	0.93	0.98	0.93	0.96
	Accuracy	98.18	97.27	94.32	98.70
	Sensitivity	0.96	0.94	0.95	0.97
Image 9	F- Measure	0.98	0.97	0.94	0.99
	Precision	0.96	0.97	0.92	0.93
	Accuracy	97.14	93.80	94.24	97.17
	Sensitivity	0.94	0.87	0.89	0.94
	F- Measure	0.97	0.93	0.92	0.97
Image 10	Precision	0.93	0.95	0.92	0.96

3.1 Selection of the valid key feature by feature extraction techniques

In this section different feature extraction techniques have been discussed and calculate their value in the terms of parameters time complexity and error rate. Key-features of an image specially in the deadly disease of lung cancer played a very important role. Only the key-features revealed that where the operation need to be performed and reduce the

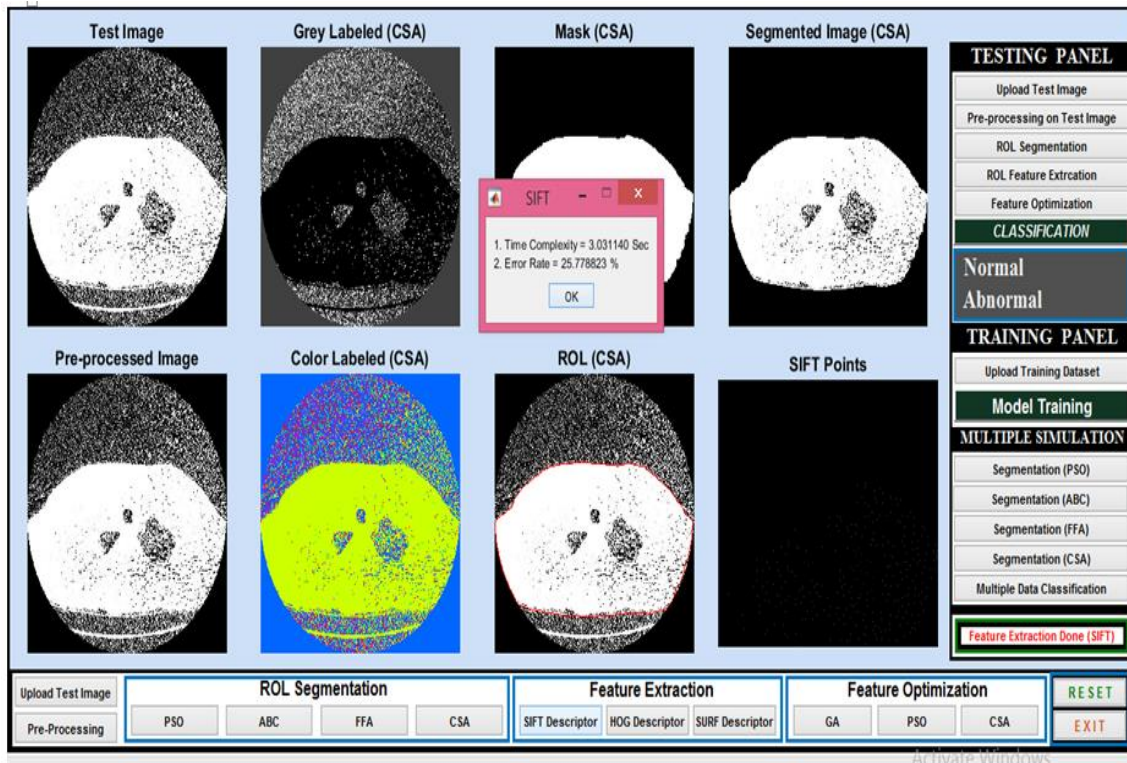
complexity of a network. In this section Scale Invariant Feature Transformation(SIFT), HOG and Speed Up Robust Feature(SURF) techniques have been discussed.

3.4 Comparative analysis has been done between the features extraction techniques and implemented best feature extraction technique on hybrid model.

Average mean of time 4.5 sec. and mean error rate 26.01 has been calculated in case of SIFT. Similarly,

in case of HOG average mean of time 2.5 sec. and mean error rate 25.01 has been calculated. Similarly, in case of SURF average mean of time 1.02 sec. and mean error rate 23.01 has been calculated. In this

feature extraction technique for the extraction of valid key features of an image and it is clearly shown by the table 2. Fig.6,7 and 8 shows the different feature extraction techniques.



section it has been revealed that SURF is the best

Fig.6. Implementation of SIFT feature extraction technique

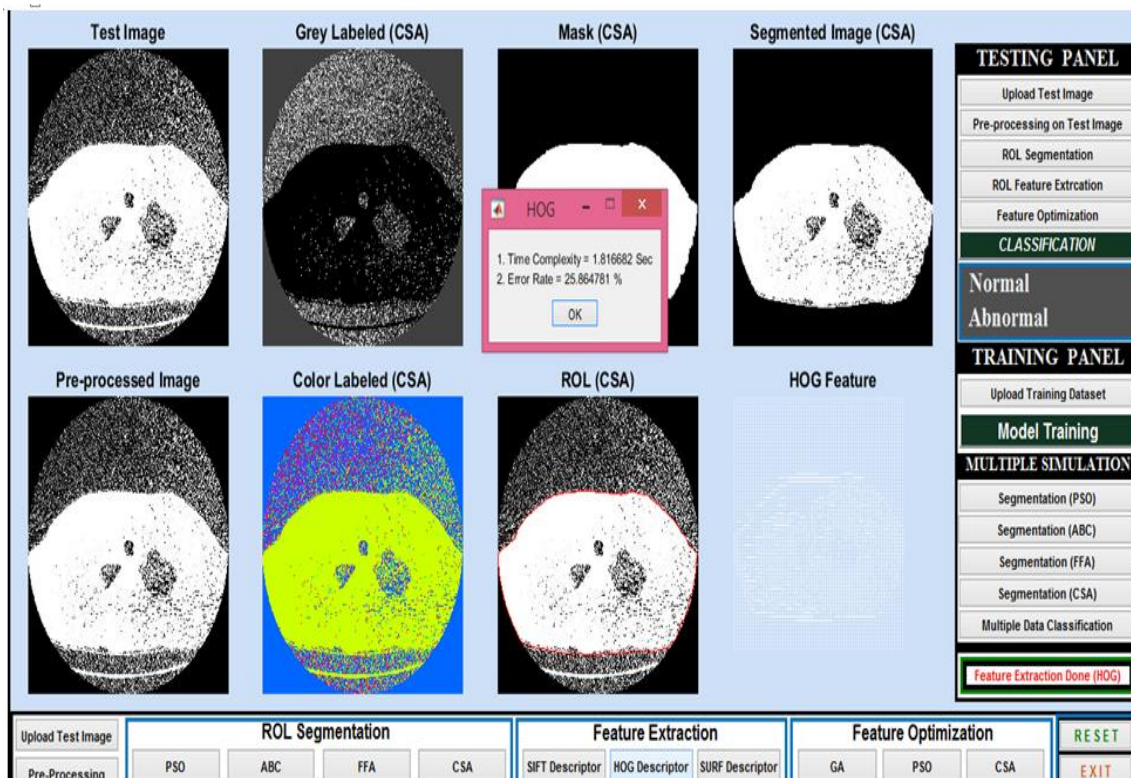


Fig.7. Implementation of HOG feature extraction technique

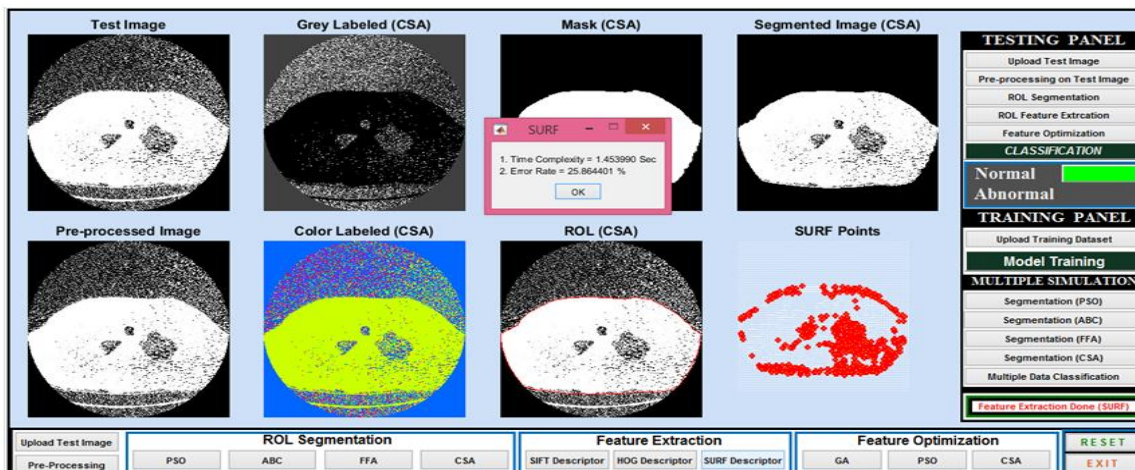
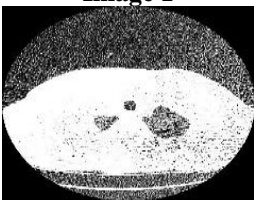
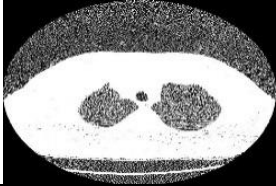
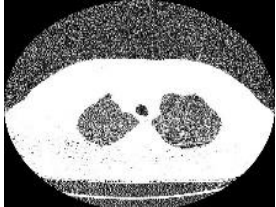
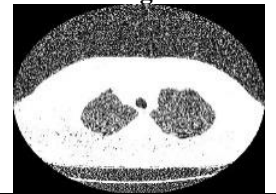
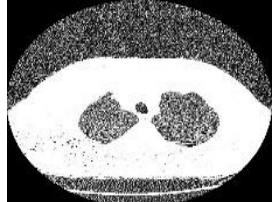
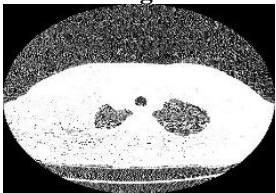
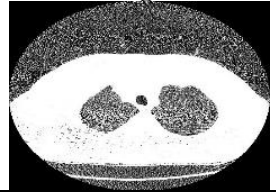
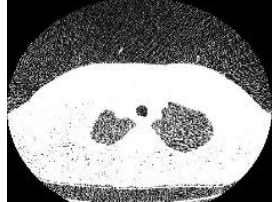
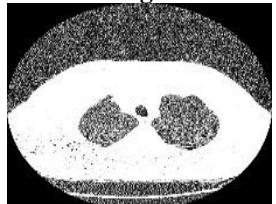
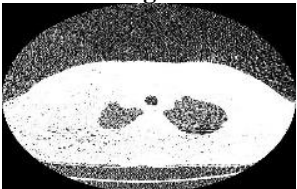


Fig.8. Implementation of SURF feature extraction technique

Table 2
Comparative analysis of Different Feature Extraction techniques

	Parameters	SIFT	HOG	SURF
Image 1 	Time Complexity	3.03	1.82	1.45
	Error Rate	25.78	25.86	25.86
Image 2	Time Complexity	0.92	0.52	0.16

	Error Rate	26.57	26.60	26.60
Image 3 	Time Complexity	10.89	0.49	0.17
	Error Rate	26.41	26.44	26.44
Image 4 	Time Complexity	5.90	0.70	0.13
	Error Rate	26.41	26.44	26.43
Image 5 	Time Complexity	4.77	0.80	0.15
	Error Rate	25.38	26.38	26.12
Image 6 	Time Complexity	4.90	0.47	0.14
	Error Rate	26.45	26.50	26.50
Image 7 	Time Complexity	3.92	0.47	0.13
	Error Rate	26.41	26.44	26.44
Image 8 	Time Complexity	5.90	0.49	0.14
	Error Rate	26.67	26.71	26.71
Image 9 	Time Complexity	1.90	0.48	0.13
	Error Rate	26.41	26.44	26.43

	Time Complexity	5.90	0.48	0.13
	Error Rate	26.33	26.39	26.39

3.5 Training and Classification have been done by using a classifier Support Vector Machine

In this section, Support Vector Machine (SVM) a classifier has been applied on the optimized valid key feature of an image after the after implementing the best feature extraction technique SURF and best optimization technique Cuckoo search on the hybrid model [16-20]. Here SVM classifier is using two kernel functions linear and polynomial as shown in the figure 9. In this context the SVM prefer the feature sets that are closed to the kernel and presented in the figure 9. With the help of SVM selection property here polynomial kernel function has been utilized to get the accurate results for the bulky size of the data set. The polynomial kernel function reduces the data size and complexity of the network. It also enhances the valid feature count and increase the speed of computation in pre-processing of valid key points of the image. Thus in this way only selected kernel support vectors are to be passed through neural network in the hybrid model. As

earlier mentioned that hybrid structure has both training and classification module. The selected kernel feature set value need the training and Levenberg-Marquardt model applied on the feature set [20-24].

An efficient training of the dataset provides the better results and elucidate the results in a better way. In the next phase, the selected feature sets are then passed to the Neural Network that extracts the weight values for the past feature set. Here, Mean Square Error (MSE) acts as a cross validator in propagating back for neural network and the most efficient learning Epochiterations values have been stored in the trained structure [24-26].As per Nanglia et. al (2020) they calculate the accuracy i.e.98.08%.for 500 data samples. On the basis of above analysis, table 3 presents the overall accuracy of hybrid model has 98.90%, recall value 91.46%, F-measure 94.73% and minimum execution time .00031 secs. Has been achieved for the hybrid model for the dataset of 1000 samples.

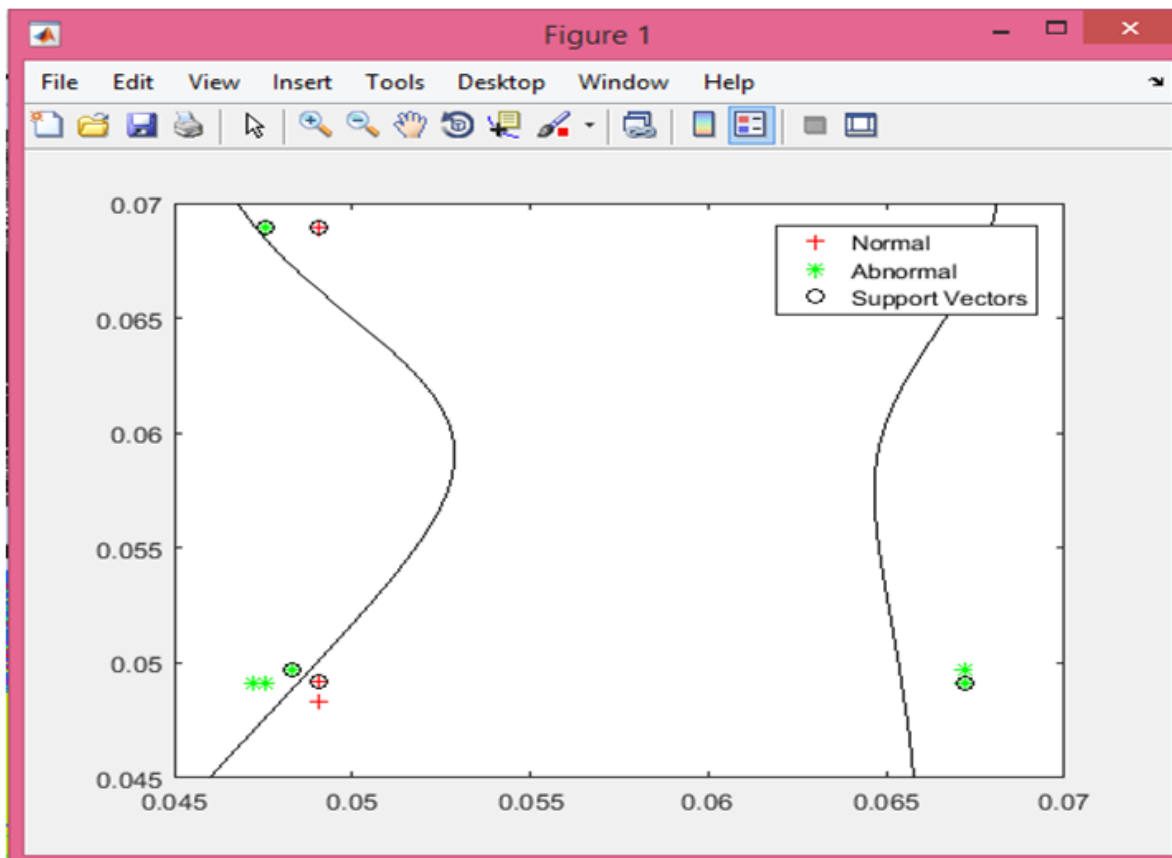


Fig.9. Implementation of SVM classifier

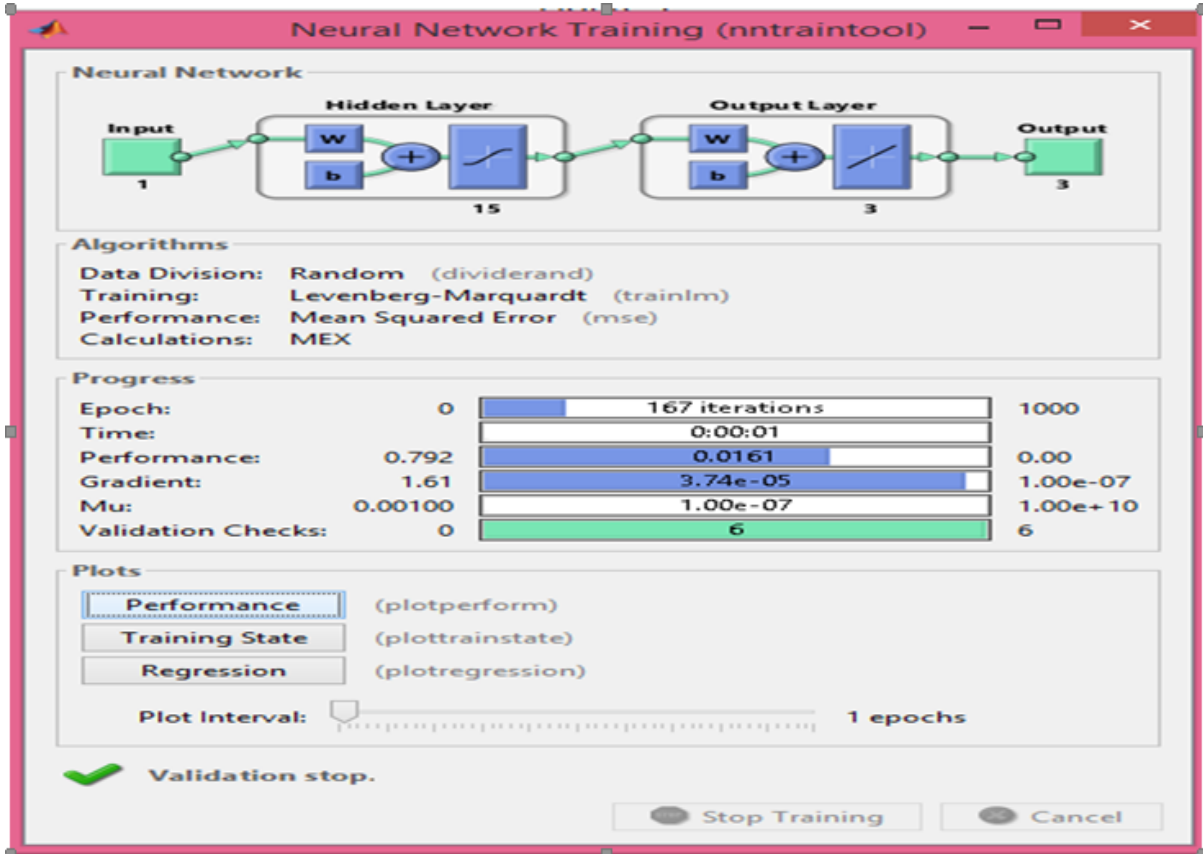


Fig.10. Performance of Neural Network

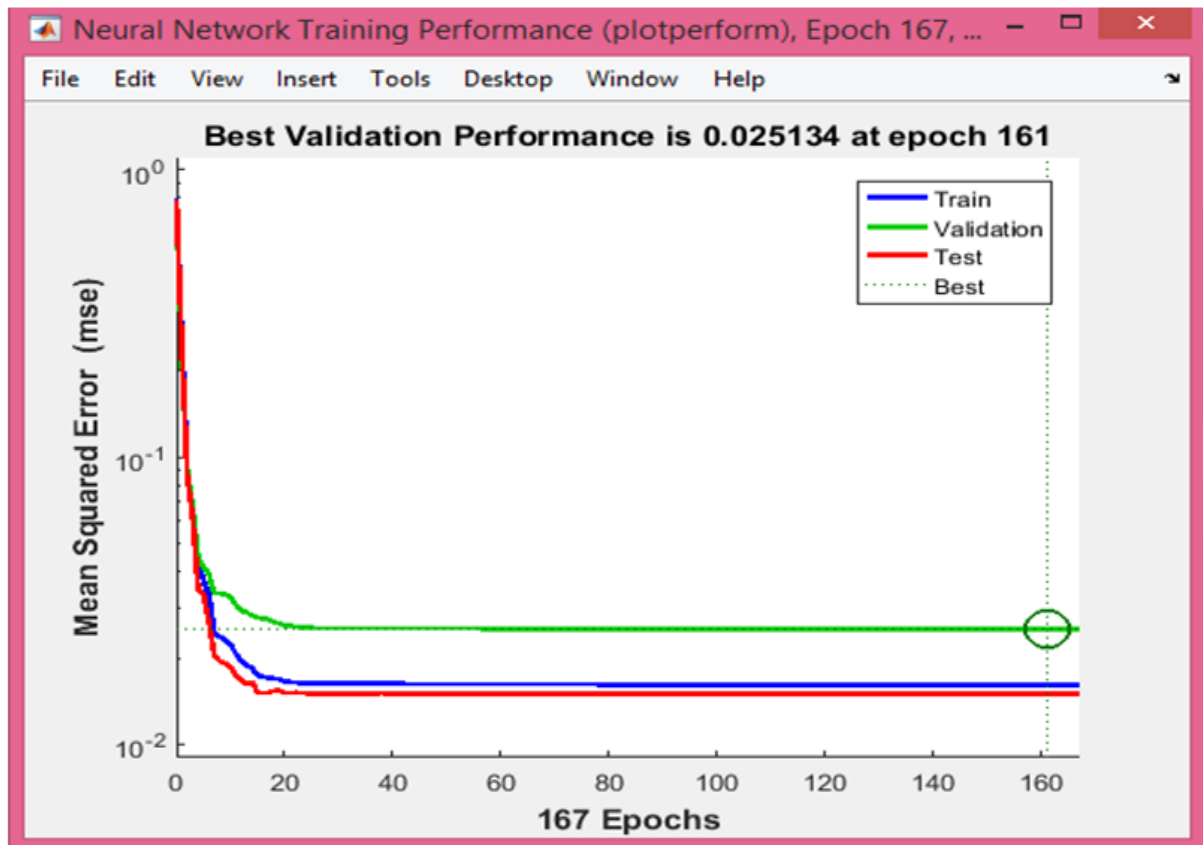


Fig.11. Neural Network Training Performance in the training and validation phase

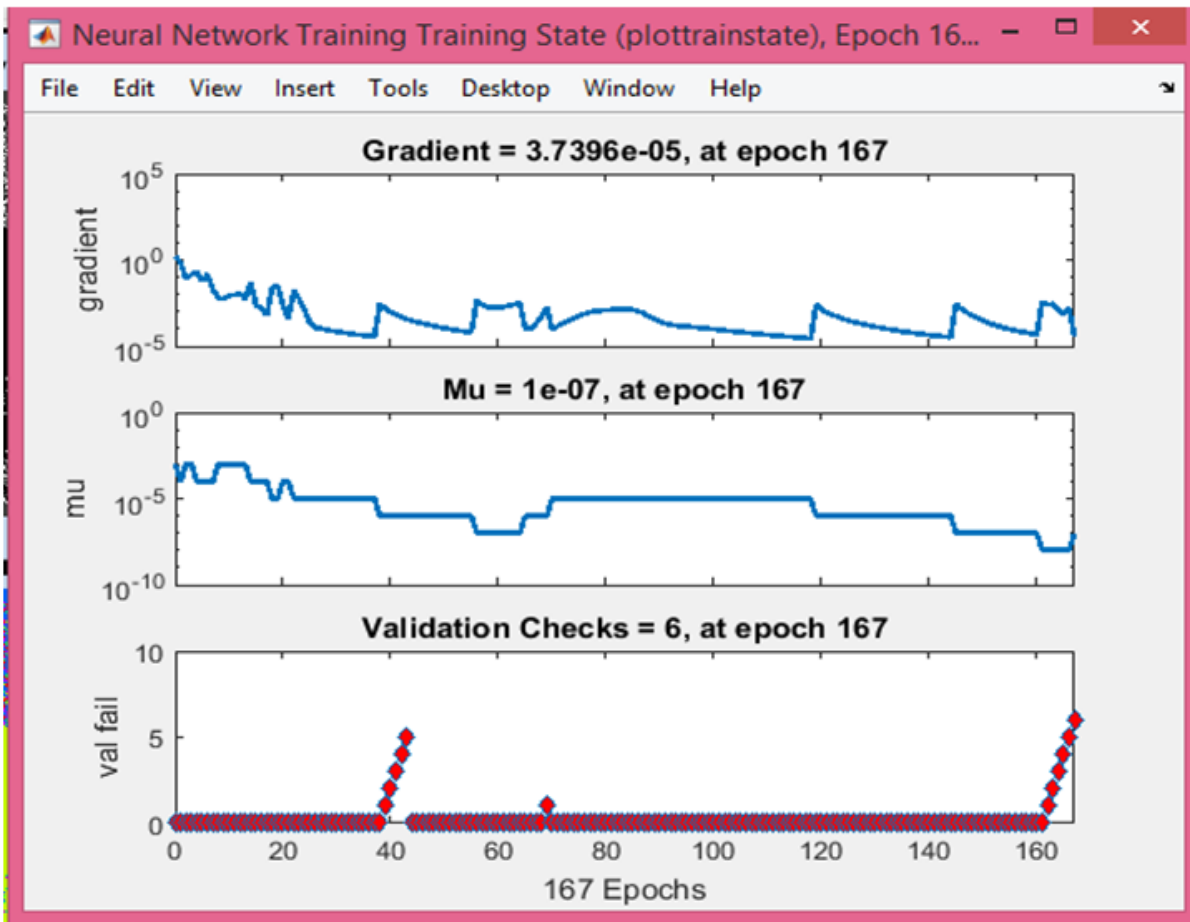


Fig.12. Graphical Representation of Neural Network Training state and cross validation check

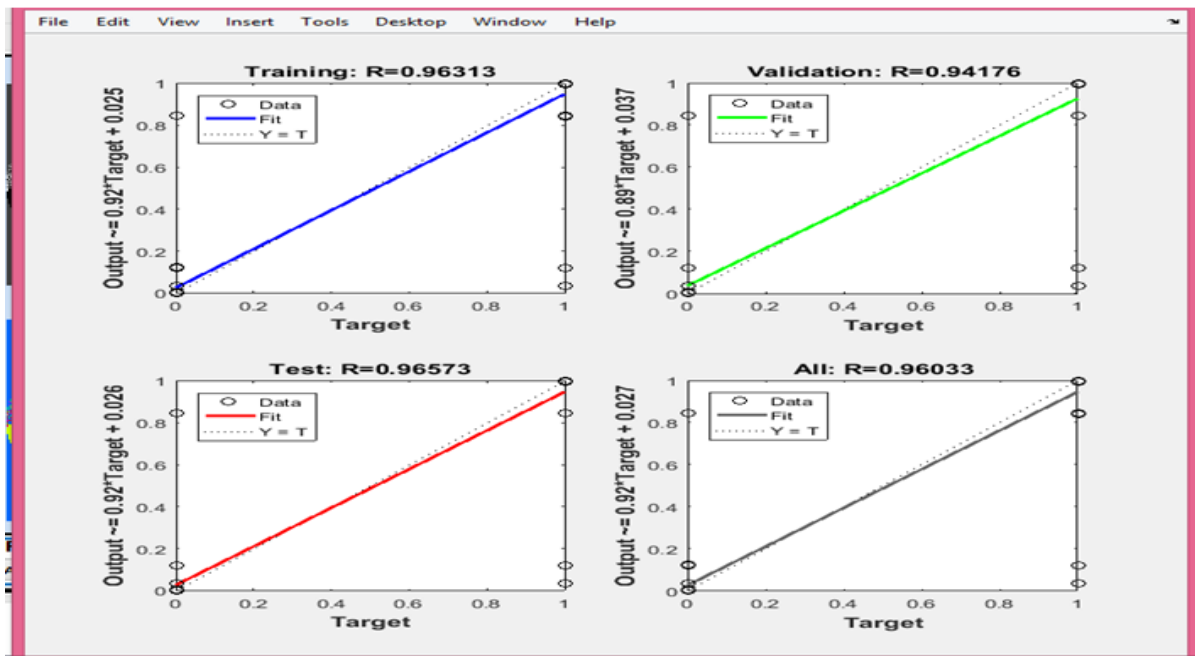


Fig.13. Graphical representation of Regression value for Training and Validation state.

Table 3 Performance of Hybrid model in the terms of different parameters

IV CONCLUSION

The article describes the detailed analysis of segmentation, optimization, feature extraction and classification process through neural network for the

early detection of lung cancer. The paper also expressed the comparative analysis between different feature extraction techniques and different optimization techniques. Here the study concludes

that on the basis of minimum execution time i.e. 1.02 sec. and minimum mean square error i.e. 23.01, SURF is the best technique. Similarly, Cuckoo search is the best optimization technique on the basis of parameters maximum Accuracy i.e. 98.14%, sensitivity i.e. .97, F-measure i.e. 0.98 and Precision i.e. .98. After implementing the most efficient feature extraction technique and most efficient optimization technique then SVM classifier

with polynomial kernels selection property through neural network overall accuracy 98.90%, recall value 91.46%, F-measure 94.73% and minimum execution time .00031 secs. have been achieved for the hybrid model. Thus present research article prepared a hybrid model for the early detection of Lung Cancer and provide good accuracy in each case.

Parameter	Image 1	Image 2	Image 3	Image 4	Image 5	Image 6	Image 7	Image 8	Image 9	Image 10
Accuracy	99.74	99.79	98.91	99.61	99.71	99.81	99.40	99.61	99.65	99.38
Error	0.26	0.21	1.09	0.39	0.29	0.19	0.60	0.39	0.35	0.62
Precision	0.98	0.99	0.99	0.99	0.98	0.99	0.99	0.99	0.99	0.99
Recall	0.88	0.89	0.88	0.91	0.87	0.89	0.89	0.89	0.90	0.89
F-Measure	0.93	0.94	0.93	0.95	0.92	0.94	0.94	0.93	0.94	0.94
Execution Time	0.0395	0.00071	0.00072	0.0008	0.003	0.00015	0.00013	0.000133	0.00012	0.00012

REFERENCES

- 1.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, 2020.
- 2.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.
- 3.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.
- 4.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.
- 5.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.
- 6.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.
- 7.Yin, Y., Sedlaczek, O., Muller, B., Warth, A., Gonzalez-Vallinas, M., Grabe, N., ... & Drasdo, D. (2017). Tumorcell load and heterogeneity estimation from diffusion-weighted MRI calibrated with histological data: an example from lung cancer. IEEE Transactions on Medical Imaging.
- 8.Ma, L., Wang, D. D., Zou, B., & Yan, H. (2017). An Eigen-binding site based method for the analysis of antiEGFR drug resistance in lung cancer treatment. IEEE/ACM transactions on computational biology and bioinformatics, 14(5), 1187-1194.
- 9.Zhang, L., Zhou, W., Velculescu, V. E., Kern, S. E., Hruban, R. H., Hamilton, S. R., ... & Kinzler, K.

- W.(1997). Gene expression profiles in normal and cancer cells. *Science*, 276(5316), 1268-1272.
10. Fischer, S. (2017). Sniffing for Cancer: Nano Noses Hold Promise for Detecting Lung Cancer and Other Diseases. *IEEE pulse*, 8(4), 20-22.
11. Gaikwad, A., Inamdar, A., & Behera, V. (2016). Lung cancer detection using digital Image Processing On CTscan Images". *International Research Journal of Engineering and Technology (IRJET)* e-ISSN, 2395-0056.
12. Lambin, P., Rios-Velazquez, E., Leijenaar, R., Carvalho, S., van Stiphout, R. G., Granton, P., & Aerts, H. J. (2012). Radiomics: extracting more information from medical images using advanced feature analysis. *European journal of cancer*, 48(4), 441-446.
13. Peng, G., Tisch, U., Adams, O., Hakim, M., Shehada, N., Broza, Y. Y., & Haick, H. (2009). Diagnosing lung cancer in exhaled breath using gold nanoparticles. *Nature nanotechnology*, 4(10), 669-673.
14. Juan, L., & Gwun, O. (2009). A comparison of sift, pca-sift and surf. *International Journal of Image Processing (IJIP)*, 3(4), 143-152.
15. Burger, W., Burge, M. J., Burge, M. J., & Burge, M. J. (2009). *Principles of digital image processing* (p. 221). London: Springer.
16. Pang, Y., Li, W., Yuan, Y., & Pan, J. (2012). Fully affine invariant SURF for image matching. *Neurocomputing*, 85, 6-10. Huijuan, Z., & Qiong, H. (2011, September). Fast image matching based-on improved SURF algorithm.
17. In *Electronics, Communications and Control (ICECC)*, 2011 International Conference on (pp. 1460-1463). IEEE.
18. Whitley, D. (1994). A genetic algorithm tutorial. *Statistics and computing*, 4(2), 65-85.
19. Houck, C. R., Joines, J., & Kay, M. G. (1995). A genetic algorithm for function optimization: a Matlab implementation. *Ncsu-ie tr*, 95(09).
20. Karaboga, D., & Basturk, B. (2007). A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm. *Journal of global optimization*, 39(3), 459-471.
21. Fourie, P. C., & Groenwold, A. A. (2002). The particle swarm optimization algorithm in size and shape optimization. *Structural and Multidisciplinary Optimization*, 23(4), 259-267.
22. Narayanan, B. N., Hardie, R. C., Kebede, T. M., & Sprague, M. J. (2017). Optimized feature selection-based clustering approach for computer-aided detection of lung nodules in different modalities. *Pattern Analysis and Applications*, 1-13.
23. Hawkins, S. H., Korecki, J. N., Balagurunathan, Y., Gu, Y., Kumar, V., Basu, S., ... & Gillies, R. J. (2014). Predicting outcomes of nonsmall cell lung cancer using CT image features. *IEEE Access*, 2, 1418-1426.
24. Westaway, D. D., Toon, C. W., Farzin, M., Sioson, L., Watson, N., Brady, P. W., ... & Gill, A. J. (2013). The International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society grading system has limited prognostic significance in advanced resected pulmonary adenocarcinoma. *Pathology-Journal of the RCPA*, 45(6), 553-558.
25. Hawkins, S. H., Korecki, J. N., Balagurunathan, Y., Gu, Y., Kumar, V., Basu, S., ... & Gillies, R. J. (2014). Predicting outcomes of nonsmall cell lung cancer using CT image features. *IEEE Access*, 2, 1418-1426.
26. Kureshi, N., Abidi, S. S. R., & Blouin, C. (2016). A predictive model for personalized therapeutic interventions in non-small cell lung cancer. *IEEE journal of biomedical and health informatics*, 20(1), 424-431.