Original Article

# Design A Hybrid Model For Lung Cancer Classification By Applying Svm Classifier With Ffbpnn On Computed Tomography Images

**Vikrant Shokeen[1], Anshu Sharma[2], Dr. Pankaj Nanglia[3], Dr. Paramjit Singh[4], Dr. Archna Sharma[5], Dr. Sushil Bansal[6]**

[1]Research Scholar, Maharaja Agrasen University. Email: shokeen18@gmail.com
[2]Research Scholar, Maharaja Agrasen University. Email: anshukaushik10@gmail.com
[3]Associate Professor, Maharaja Agrasen University. Email: nanglia.pankaj@gmail.com
[4]Assistant Professor, Maharaja Agrasen University. Email: ppparamjitsingh@gmail.com
[5]Assistant Professor, Delhi Institute of Pharmaceutical Sciences and Research (DIPSAR). Email: archanakuk@gmail.com
[6]Professor, Maharaja Agrasen University. Email: sk93recj@gmail.com
**DOI:** 10.47750/pnr.2022.13.S08.223

## Abstract

The Current Study elaborates the classification of Lung Cancer by traversing the potential consumption of a Hybrid Model in which ROI (Region of interest) is recognised through different segmentation techniques by extracting key points through different features extraction and optimization techniques are implemented to the dataset of CT scan images of 800 pictures. A total of 800 patients were included in this retrospective study and divided into the training set and test set with a ratio of 8:2 randomly. The research work developed a Hybrid Model by using distinct Optimization techniques i.e. Particle Swarm Optimization (PSO). Artificial BEE colony (ABC) and features extraction method namely SIFT (Scale Invariant Feature Transform, Speed Up Robust Feature (SURF), in context of the least execution time with least mean square error rate with support vector machine (SVM) classifier. Additionally, working of the Hybrid Model has been gauged in respect of parameter Accuracy, Error rate, Precision, Recall, and Execution Time. The overall Accuracy of the hybrid model is 99.56% while recall value is 89%, F-measure 93.44% have been obtained for the Hybrid Model.

**KEYWORDS:** Lung Cancer, Segmentation, Feature Extraction, SIFT, PCA, SURF, PSO, and SVM classifier.

## INTRODUCTION

Lung cancer is one of the most hazardous diseases, which can be commonly exhibited nowadays. The Death Rate is increased by 10 to 15% all over the world and more than 12,000 deaths are counted in India per year [1]. This perilous disease took up momentum very quickly that execute the diagnosis of disease at the initial stage by the Analysts in healthcare centres. Many patients have been recovered from this disease at the initial stage by the dominant remedy. Now as per Doctors' concerns "Automating the complete detection" process is the only way to diagnose lung cancer. A large number of efforts have been done by medical experts to take this analysis further to the upcoming phase [2].

Lung cancer can be categorized into two stages as follow:

- Training Stage
- Classification Stage

**LITERATURE REVIEW: -**

The selection process adopted in this review relies on adhering the specifications of the preferred reporting items for systematic reviews and meta-analyses which is a practical and efficient approach for writing survey studies. Concretely, a search was performed for the last seven years (January 2015–February 2022) [3]. This section emphasized on the latest trends for the detection of Lung Cancer while having a sufficient number of studies to be discussed. In this context various studies have been conducted and some of them are closely associated with the current scenario proposed in this section. Here dataset of Pictures, merits, demerits, and the conducted method accuracy implementation time and lapse count are talked through to take this research to new heights.

A study has been conducted by Shubham Dodia et.al (2022) "Recent Advancement in Deep Learning based Lung Cancer Detection: A systematic review". This paper discusses an overview of lung cancer, along with publicly available benchmark data sets for research purposes. Recent research performed in medical image analysis of lung cancer using deep learning algorithms is compared using various technical aspects such as efficiency, advantages, and limitations. These discussed approaches provide insight into techniques that can be used to perform the detection and classification of lung cancer. Numerous techniques adapted in the acquisition of the images, extraction of relevant features, segmentation of region affected, selection of optimal features, and classification are also discussed [3].

A study is directed by Pankaj et. al (2020) named "A hybrid algorithm for long cancer classification using SVM and Neural networks". In this a collection of 500 pictures are taken into consideration 75% of data is utilized for the training purpose and the rest 25% is utilized for the testing purpose. for the feature extraction of Optimized Features by applying different feature and segmentation techniques by using SVM classifier with Feed-Forward Back Propagation Neural Network (FFBPNN) is implemented to the dataset this model is untitled as a hybrid model, with such a high 98.08% classification accuracy [4].

Likewise, Paramjit et.al (2019) proposed a study named "Improved lung cancer segmentation using k-means and cuckoo search". In it, the database of CT scan pictures is taken and applied Artificial Bee colony (ABC) and cuckoo search methods. Here SVM works as a cross-validator to reveal the accuracy, precision, recall, and F-measure [5]. The study shows that increment in accuracy by 5%, Precision by 6% recall by 3%, and F-measure by 4% are gained through ABC and a boost of 10% in accuracy,11% in precision 12% in recall and 11% in F-measure has been observed through cuckoo search.

Another study by Nanglia et.al (2019) titled "Detection &classification of lung cancer at an early stage by applying feature extraction optimization and neural networks on hybrid structure". In this study, Dr. Nanglia took 400 CT scan pictures for the examination of the best feature extraction technique and best optimization technique. Additionally, across validator, support vectors machine (SVM) is gone through for better categorization. In this, it was observed that the hybrid model figured out that an execution time of 1.94 seconds. which is the smallest and minimal fault rate of 29.25 is perceived from surf and GA algorithm techniques. Finally, complete classification accuracy for the hybrid model was measured and seemed to be 99.67% by taking this peculiar hybrid model [6-8].

One more important work regarding lung cancer is done by Nanglia et. al (2018) called "comparative investigation of different feature extraction techniques for lung cancer detection system". This review conveys the relative examination among these feature extraction methods SIFT, SURF, and Principle component analysis (PCA). This learning discloses that the average accomplished time is 448 seconds and an average error rate of 25.704 is found for the surf methods. Therefore, this research finalizes that the SURF method is the outstanding approach concerning the time execution and error rate [9-11].

## IMPLEMENTATION PROCESS: -

In this section execution procedure of the hybrid model discussed. In addition, the dataset identification is also reviewed in this stance. The execution moves that are applied to obtain this special hybrid model for spotting lung cancer as soon as possible are categorized below: -

- Distribution and enhancement of the images.
- Relative examination of these images.

- Appointment of the reasonable key characteristics of extraction methods.
- The relative examination took place between the feature extraction methods and implementation best feature extraction method on the hybrid model.
- Tutoring and categorization have been done with the help of a classifier support vector machine.
- Consequences were measured concerning parameters, accuracy, sensitivity, F-measure, Precision, execution time, and recall [12-16].

## DATASET:

In this study, the suggested hybrid model dictates ELCAP lung images dataset primitively mapped out and evolved by Cornell University. The first edition of the dataset was built in 2019 which had 1000 CT scan images (documented) of 1mm slice thickness [17-21].

## RESEARCH METHODOLOGY:

This research technique is used for the categorization process (Train samples [ cancerous, non-cancerous], Test set)

- Segmentation and Optimization of the images.
- Extract key points for cancerous and non-cancerous images
- Train Optimized valid key features
- Classify and diagnose
- End

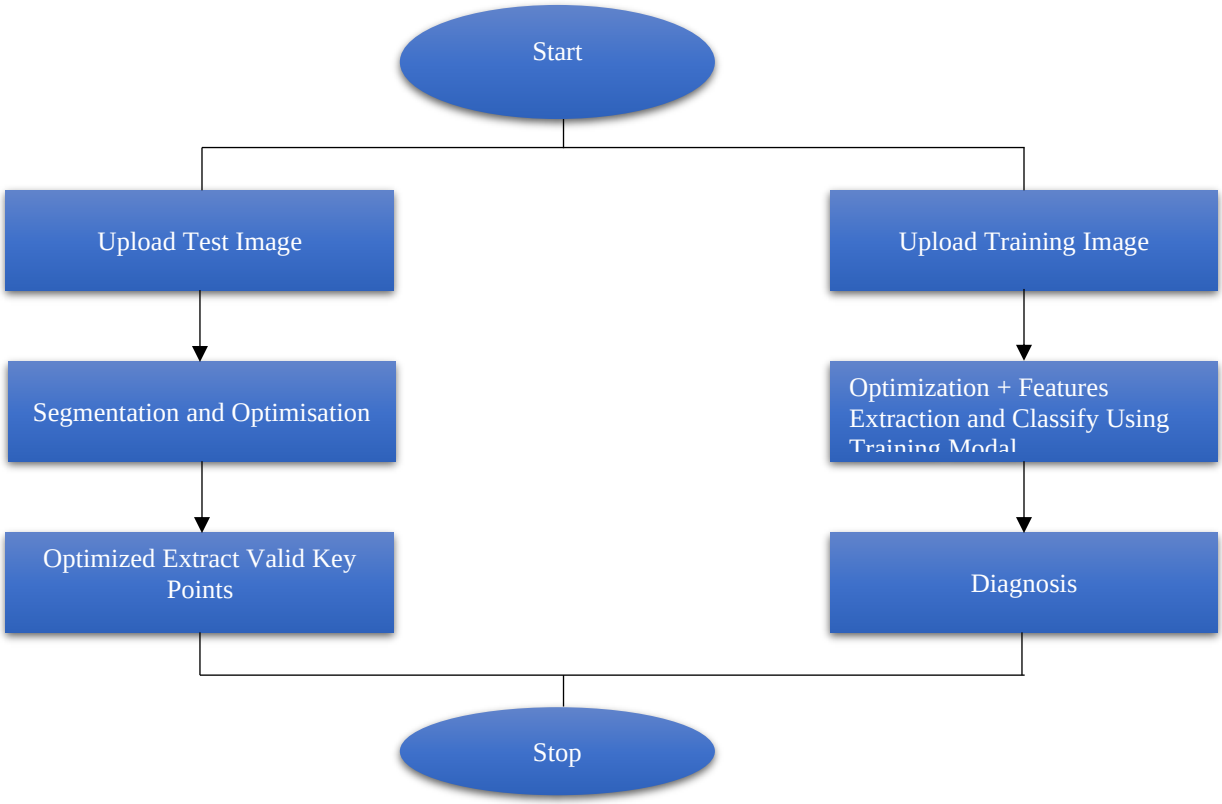



**Figure 1. Architecture diagram for Lung Cancer Classification**

## EXPERIMENTAL RESULTS AND CONCLUSION

- **SEGMENTATION AND OPTIMIZATION**

Multiple segmentation methods are put into the images to search out the Region of Interest (ROI) in this condition artificial bee colony (ABC), PSO (Particle swarm optimization). Techniques board of GUI which was generated in Matlab 16. It is mentioned that there are two panels namely the testing & model training panels. Segmentation takes place in both parts consequently [21-24].

- **COMPARATIVE ANALYSIS OF SEGMENTATION AND OPTIMIZATION**

Relative comparison is taken out among these methods in the context of parameter, accuracy, sensitivity, F – measure and precision. In the first occurrence, the average mean of accuracy for the PSO is 98.90 the average mean of sensitivity is found at 95, the average mean of F-measure is studied at 97 and the average mean precision value is provided 94. Identically the average mean of all these values are taken into consideration and observed relatively in the artificial BEE colony (ABC) case and exhibited in figure 3. The average mean of accuracy for the ABC is 97.87, the average mean of sensitivity is calculated at 88, average mean of precision value is found to be 94. So in the context of the above-calculated values, it is clearly said that PSO is the foremost optimization technique and in this hybrid model, the PSO method seems to give that outcome[24-28].

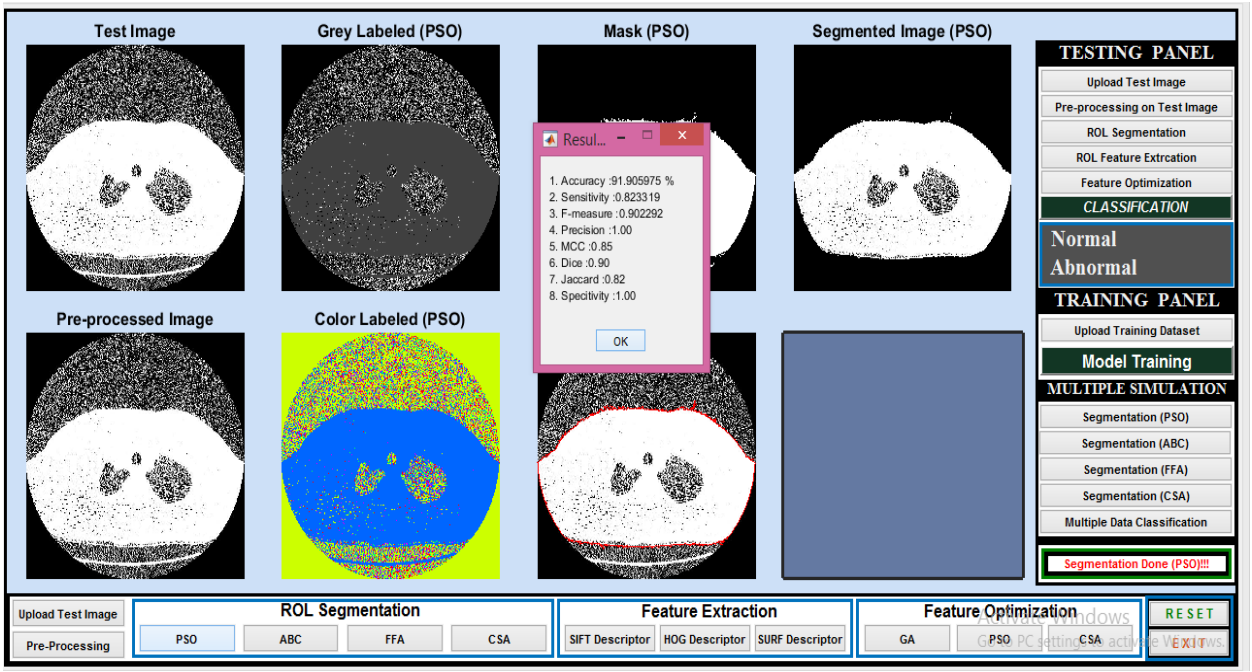



**Figure 2 Various Feature Extraction Techniques implemented in Hybrid Mode**

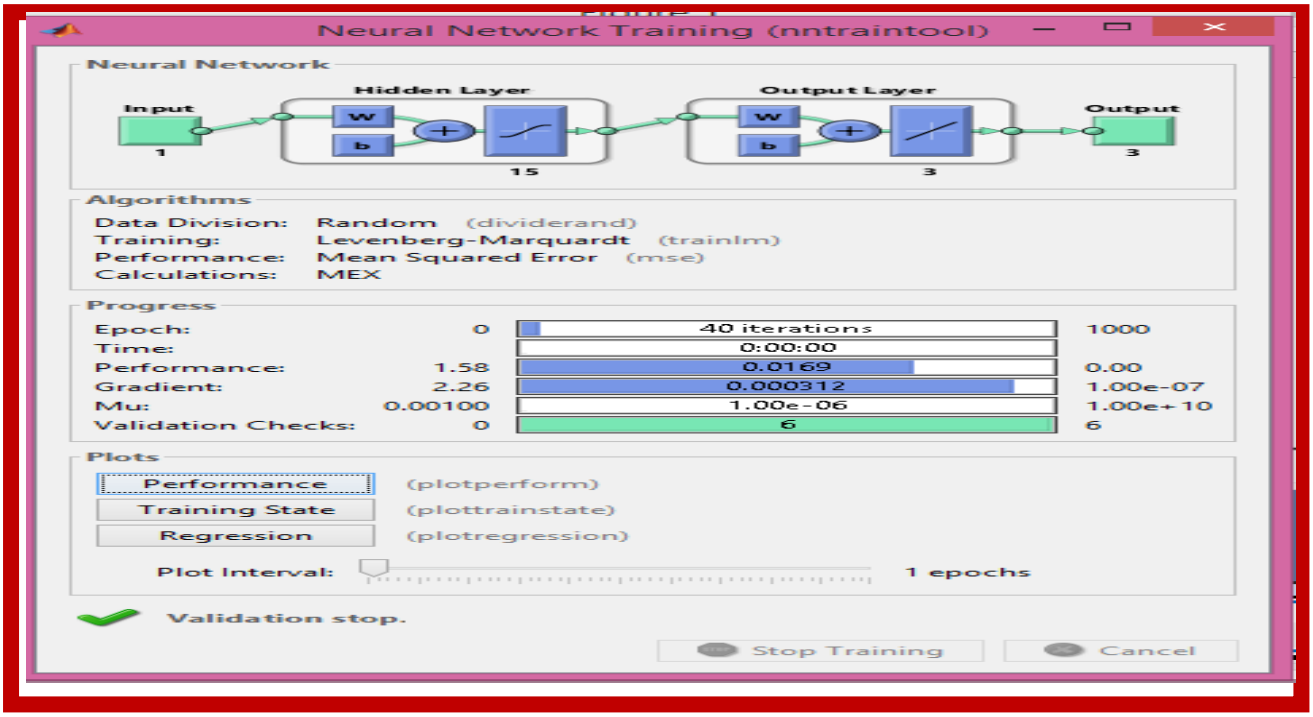



**Figure 3 Implementation of Neural Network by using Levenberg-Marquardt Model**

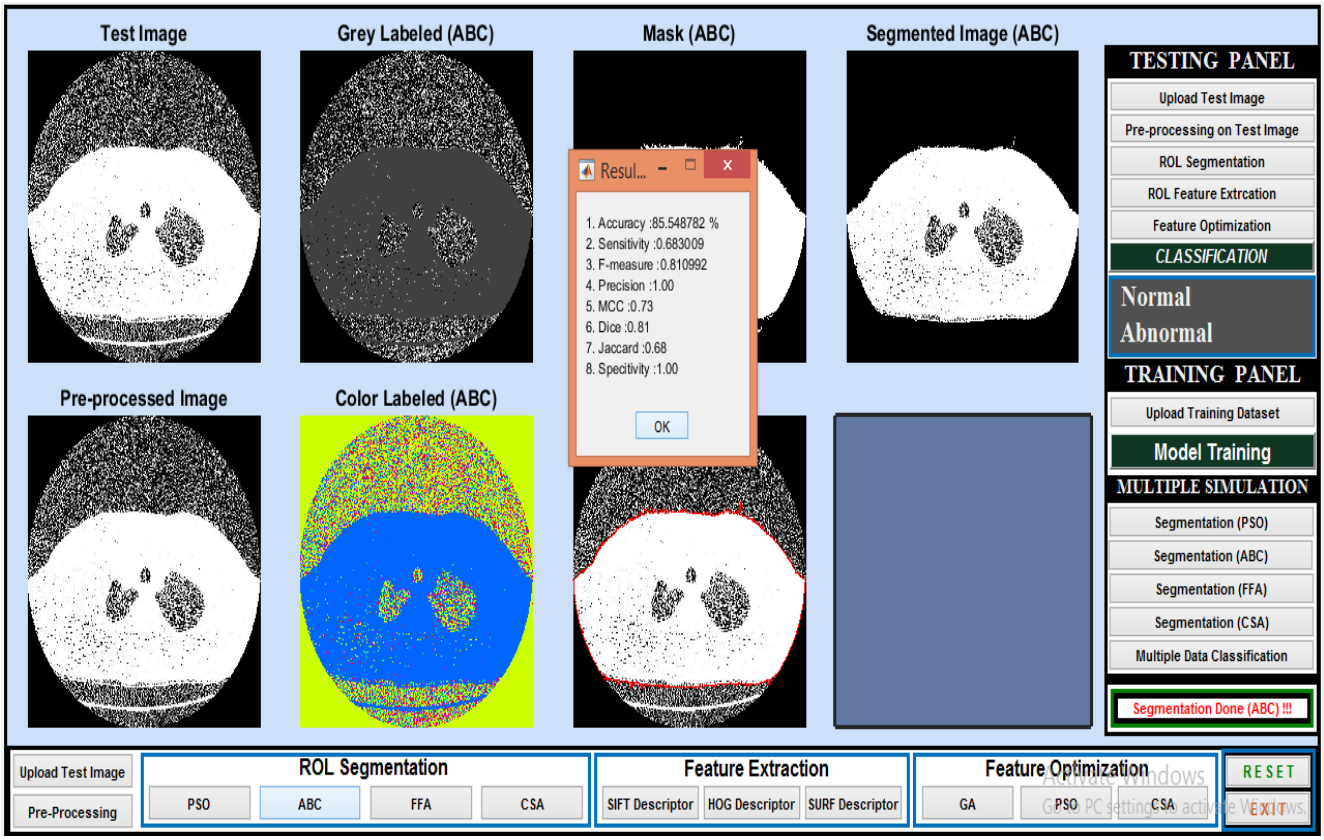
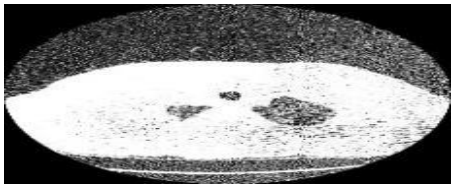
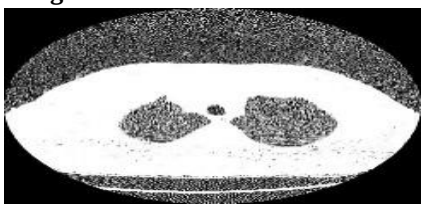
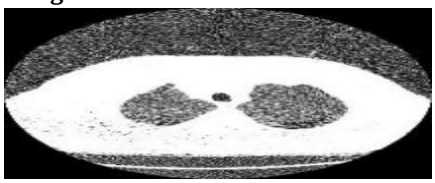
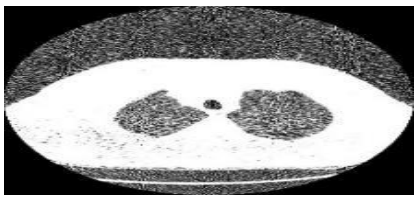
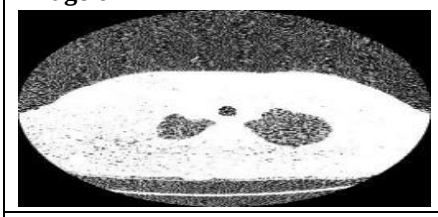
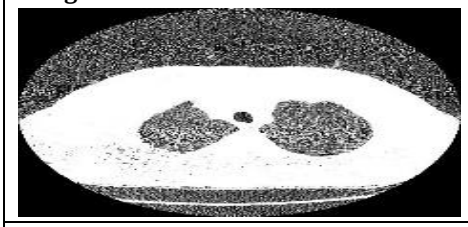
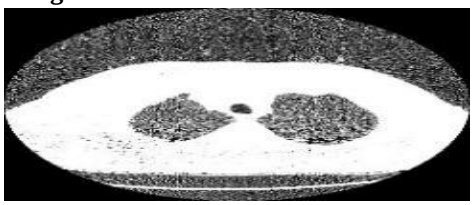
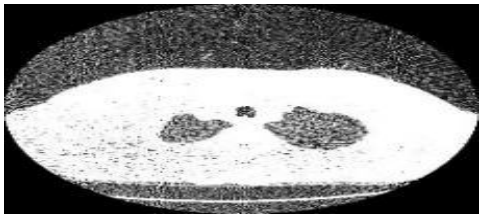



**Figure 4 Utilization of different Feature Optimization Techniques**

**Table 1: Comparative analysis among PSO and ABC**

| | Parameters | POS | ABC |
|---|---|---|---|
| **Image 1**  | Accuracy | 98.32 | 95.36 |
| | Sensitivity | 0.98 | 0.90 |
| | F- Measure | 0.98 | 0.95 |
| | Precision | 0.97 | 0.96 |
| **Image 2**  | Accuracy | 97.45 | 97.05 |
| | Sensitivity | 0.95 | 0.93 |
| | F- Measure | 0.98 | 0.97 |
| | Precision | 0.96 | 0.97 |
| **Image 3**  | Accuracy | 95.68 | 94.70 |
| | Sensitivity | 0.97 | 0.95 |
| | F- Measure | 0.96 | 0.97 |
| | Precision | 0.91 | 0.92 |
| **Image 4** | Accuracy | 98.25 | 96.10 |
| | Sensitivity | 0.98 | 0.95 |
| | F- Measure | 0.97 | 0.96 |

| | | | |
|---|---|---|---|
|  | Precision | 0.96 | 0.97 |
| **Image 5**  | Accuracy | 99.10 | 82.46 |
| | Sensitivity | 0.97 | 0.62 |
| | F- Measure | 0.99 | 0.73 |
| | Precision | 0.92 | 0.94 |
| **Image 6**  | Accuracy | 92.75 | 86.68 |
| | Sensitivity | 0.86 | 0.90 |
| | F- Measure | 0.91 | 0.82 |
| | Precision | 0.93 | 0.88 |
| **Image 7**  | Accuracy | 99.15 | 96.65 |
| | Sensitivity | 0.98 | 0.92 |
| | F- Measure | 0.99 | 0.97 |
| | Precision | 0.93 | 0.94 |
| **Image 8**  | Accuracy | 98.52 | 96.85 |
| | Sensitivity | 0.97 | 0.94 |
| | F- Measure | 0.98 | 0.97 |
| | Precision | 0.94 | 0.98 |
| **Image 9**  | Accuracy | 98.36 | 97.54 |
| | Sensitivity | 0.96 | 0.95 |
| | F- Measure | 0.98 | 0.97 |
| | Precision | 0.96 | 0.97 |
| **Image 10**  | Accuracy | 98.18 | 94.78 |
| | Sensitivity | 0.95 | 0.87 |
| | F- Measure | 0.97 | 0.94 |
| | Precision | 0.93 | 0.95 |

## TRAINING AND CLASSIFICATION USING SVM CLASSIFIER

In this section, the best classifier after deploying the best feature extraction technique SURF and best optimization technique on the hybrid model is Support Vector Machine (SVM) [28-30]. Two kernel functions (i.e. linear and polynomial) are performed by the SVM classifier as shown in the figure. The polynomial kernel function reduces the data size and difficulties of the network and increases the valid feature and speed of computation in pre-processing of the image. Thus the only selected kernel support vectors are to be passed through the neural networks in the hybrid model. The hybrid structure has both training and classification modules as mentioned above.

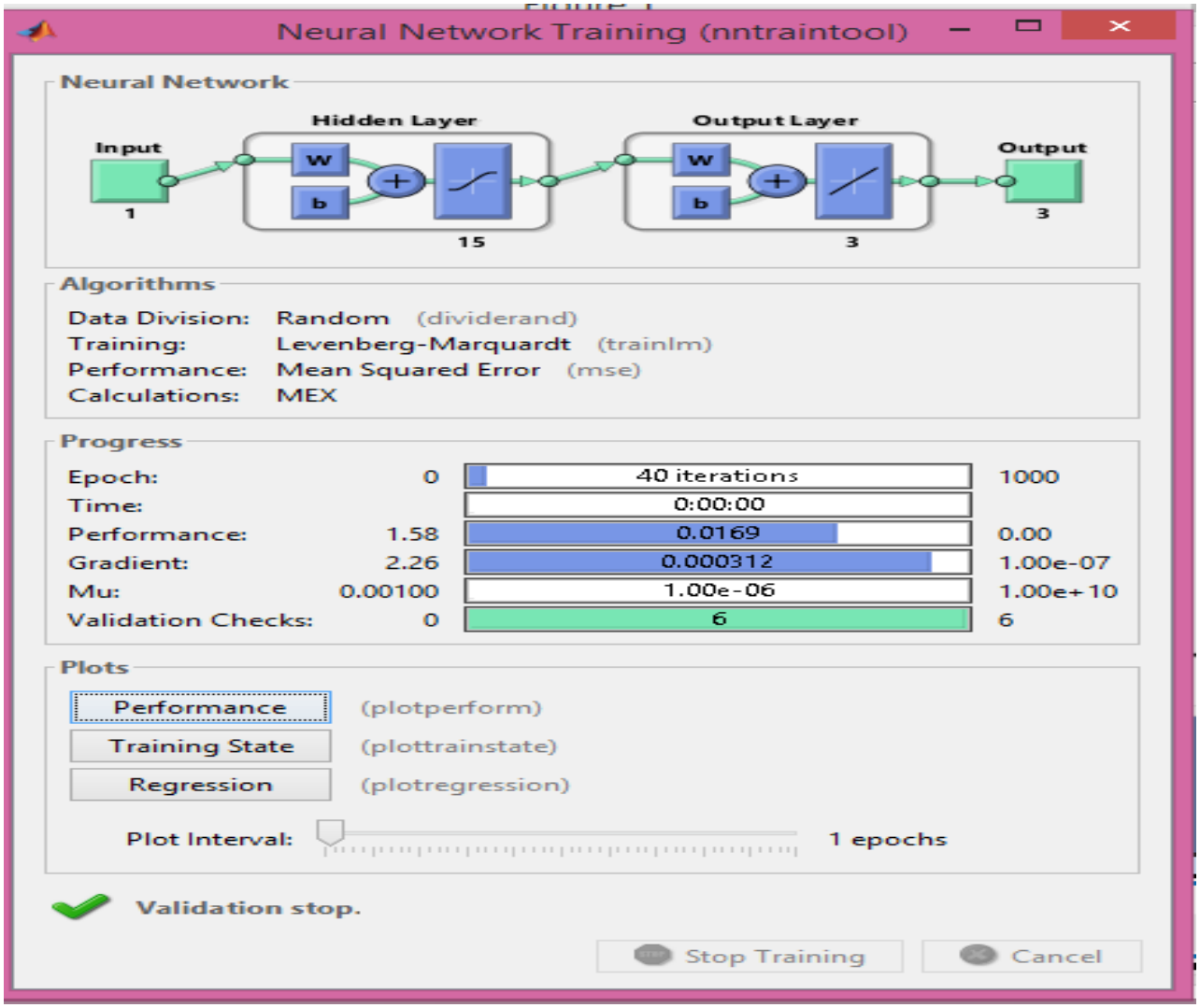



**Figure 5 Utilization of SVM classifier with Feed Forward Back Propagation Neural Network**

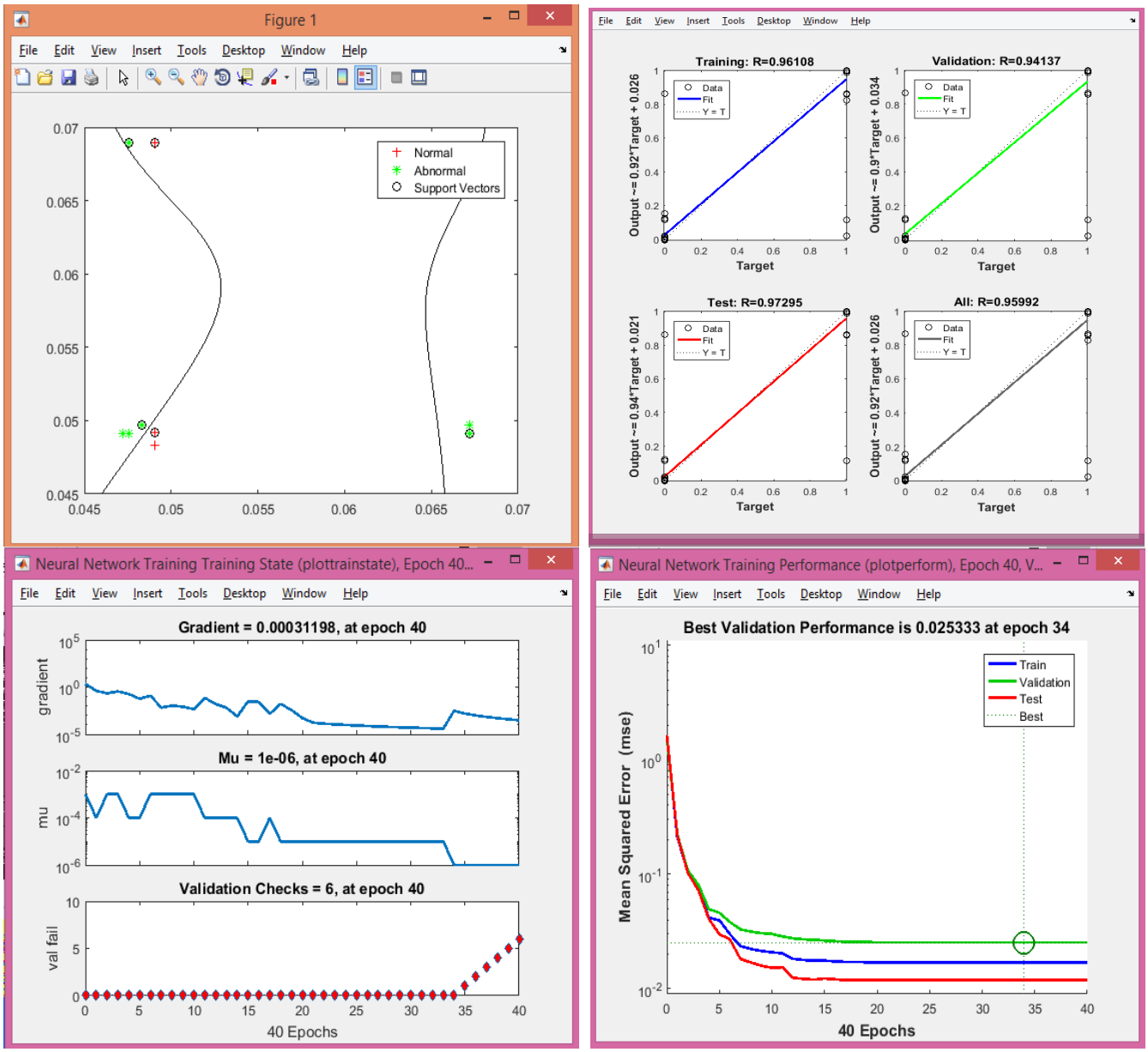



**Table 3: Performance of Hybrid Model in the terms of different parameters**

| | Accuracy | Error | Precision | Recall | F-Measure | Execution Time |
|---|---|---|---|---|---|---|
| Image 1 | 99.74 | 0.26 | 0.98 | 0.88 | 0.93 | 0.0395 |
| Image 2 | 99.79 | 0.21 | 0.99 | 0.89 | 0.94 | 0.0007 |
| Image 3 | 98.91 | 1.09 | 0.99 | 0.88 | 0.93 | 0.0007 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Image 4 | 99.61 | 0.39 | 0.99 | 0.91 | 0.95 | 0.0008 |
| Image 5 | 99.71 | 0.29 | 0.98 | 0.87 | 0.92 | 0.0030 |
| Image 6 | 99.81 | 0.19 | 0.99 | 0.89 | 0.94 | 0.0002 |
| Image 7 | 99.4 | 0.6 | 0.99 | 0.89 | 0.94 | 0.0016 |
| Image 8 | 99.61 | 0.39 | 0.99 | 0.89 | 0.93 | 0.0013 |
| Image 9 | 99.65 | 0.35 | 0.99 | 0.9 | 0.94 | 0.0012 |
| Image 10 | 99.38 | 0.62 | 0.99 | 0.89 | 0.94 | 0.0013 |

## CONCLUSION: -

The proposed Hybrid Model generated very high accuracy, sensitivity, specificity, and precision and compared with other state of art methods. Because of its simplicity and possible global search capabilities, evolutionary algorithms have shown tremendous promise in the area of feature selection in the latest years. 'Relative review of feature vector' and 'evaluated features in this research work' have been presented with full analysis with respect to mean square error and minimum execution time. A detailed explanation is given of various optimization techniques (i.e. Feature vector) that are performing and sorted by SVM and able to achieve good accuracy of 99.03% approximately. This research work also concluded that Support vector machine with polynomial kernel property by applying Feed Forward Back Propagation Neural Network on CT scan images produced tremendous results in the terms of parameter like Accuracy, MSE Precision, Recall and F-Measure and execution time as shown in the table 3. Further, classification accuracy elucidate that the features were optimized correctly. This current research work provides many opportunities to upcoming scholars. Different Types of Lung Cancers or different stages of Lung cancers can also be analysed accurately through the same designed model Hybrid Model.

## 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 (AICR) 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 CT scan Images". International Research
12. Journal of Engineering and Technology (IRJET) e-ISSN, 2395-0056.
13. 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.
14. 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.
15. Malik, Mayank Kumar, Pankaj Bhatt, Tarun Kumar, Jaspal Singh, Vipin Kumar, Abdul Faruk, Shivkanya Fuloria, Neeraj Kumar Fuloria, Vetriselvan Subrimanyan, and Sunil Kumar. 2022. "Significance of Chemically Derivatized Starch as Drug Carrier in Developing Novel Drug Delivery Devices." The Natural Products Journal 12. https://doi.org/10.2174/2210315512666220819112334.

16. Burger, W., Burge, M. J., Burge, M. J., & Burge, M. J. (2009). Principles of digital image processing (p. 221).London: Springer.
17. Pang, Y., Li, W., Yuan, Y., & Pan, J. (2012). Fully affine invariant SURF for image
18. matching. Neurocomputing, 85, 6-10.Huijuan, Z., & Qiong, H. (2011, September). Fast image matching based on improved SURF algorithm.
19. In Electronics, Communications, and Control (ICECC), 2011 International Conference on (pp. 1460-1463). IEEE.
20. Pankaj. 2021a. "Anti-Cancer Cyclodextrin Nanocapsules Based Formulation Development for Lung Chemotherapy." Journal of Pharmaceutical Research International, 54–63. https://doi.org/10.9734/jpri/2020/v32i3931024.
21. Chand, Sumanta Prasad, Sandip Debnath, Mehdi Rahimi, Mohd Shaikhul Ashraf, Pankaj Bhatt, and Saima Ahmed Rahin. 2022. "Contextualization of Trait Nexus and Gene Action for Quantitative and Qualitative Characteristics in Indian Mustard." Journal of Food Quality 2022: 1–24. https://doi.org/10.1155/2022/4387318.

22. 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.
23. Bhatt, Pankaj, Vipin Kumar, Richa Goel, Somesh Kumar Sharma, Shikha Kaushik, Shivani Sharma, Alankar Shrivastava, and Mulugeta Tesema. 2022. "Structural Modifications and Strategies for Native Starch for Applications in Advanced Drug Delivery." BioMed Research International 2022: 1–14. https://doi.org/10.1155/2022/2188940.

24. 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.
25. Al-Snafi, Ali Esmail, Suruchi Singh, Pankaj Bhatt, and Vipin Kumar. 2022. "A Review on Prescription and Non-Prescription Appetite Suppressants and Evidence-Based Method to Treat Overweight and Obesity." GSC Biological and Pharmaceutical Sciences 19 (3): 148–55. https://doi.org/10.30574/gscbps.2022.19.3.0231.

26. Predicting outcomes of non-small cell lung cancer using CT image features. IEEE Access, 2, 1418-1426.
27. Ahamed, Saahirabanu, Pankaj Bhatt, S. J. Sultanuddin, Ranjan Walia, M. Akiful Haque, and S. B. InayathAhamed. 2022. "An Intelligent IoT Enabled Health Care Surveillance Using Machine Learning." In 2022 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI). IEEE.
28. Hawkins, S. H., Korecki, J. N., Balagurunathan, Y., Gu, Y., Kumar, V., Basu, S., ... & Gillies, R. J. (2014).Predicting outcomes of non-small cell lung cancer using CT image features. IEEE Access, 2, 1418-1426.
29. 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.