International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Artificial Neural Network-based Classification System of Lung Nodules in Computed Tomography Scans

Padmaja G. Patil*, Dr. Sanjay N. Patil**

Research Student, Department of Electpronics Engineering, PVPIT Budhgaon, India
Associate Professor, Department of Electpronics Engineering, PVPIT Budhgaon, India

A B S T R A C T :

Cancer remains one of the most life-threatening diseases globally, with lung cancer being the leading cause of cancer-related deaths. Early detection plays a vital role in improving patient survival rates. The proposed project aims to design a Computer-Aided Diagnosis (CAD) system for lung cancer detection and classification using Computed Tomography (CT) images. CT imaging provides detailed internal lung structures, making it ideal for identifying small nodules. The system operates through four main stages: image acquisition, preprocessing, morphological processing, and detection/classification. A custom filtering algorithm is employed to remove noise and enhance image clarity. Morphological operations such as segmentation, erosion, and dilation are performed to extract regions of interest. The refined images are then analyzed to identify and classify nodules as benign or malignant based on texture, shape, and intensity features. The integration of these methods enhances the accuracy of nodule detection. The CAD model effectively reduces false positives and false negatives. This assists radiologists in making faster and more reliable diagnoses. The proposed system contributes to early diagnosis, accurate tumor characterization, and improved treatment planning. Ultimately, it aims to support healthcare professionals in reducing lung cancer mortality rates through intelligent image-based analysis.

Keywords: Lung Cancer, Computer-Aided Diagnosis (CAD), CT Images, Image Processing, Morphological Operations, Tumor Detection, Noise Filtering, Segmentation, Feature Extraction, Benign and Malignant Classification, Medical Image Analysis

1. Introduction

Cancer is one of the most serious health challenges worldwide, with an alarming rise in both incidence and mortality. Among all cancer types, lung cancer remains the leading cause of cancer-related deaths across the globe. It is characterized by uncontrolled cell growth in the lung tissues, eventually leading to the formation of tumors that interfere with normal respiratory function. Lung cancer has the lowest survival rate after diagnosis due to its typically late detection and rapid progression. Early diagnosis plays a crucial role in improving survival rates — the earlier the cancer is detected, the greater the chances of successful treatment.

According to global cancer statistics, approximately 85% of lung cancer cases in men and 75% in women are directly linked to cigarette smoking, making it the most significant risk factor. In 2005, about 1,372,910 new cancer cases and 570,280 deaths were reported in the United States alone, out of which an estimated 163,510 deaths were attributed to lung cancer—accounting for nearly 29% of all cancer-related deaths. Despite advancements in surgery, chemotherapy, and radiotherapy, the five-year survival rate for lung cancer across all stages remains at only 14%, a figure that has shown little improvement over the past three decades.

Lung cancer primarily arises when abnormal cells in the lung begin to multiply uncontrollably, forming malignant tumors. It is broadly classified into two major categories: Small Cell Lung Cancer (SCLC) and Non-Small Cell Lung Cancer (NSCLC). The latter further includes three main subtypes: Adenocarcinoma, Squamous Cell Carcinoma, and Large Cell Carcinoma. Among these, adenocarcinoma and squamous cell carcinoma are the most common. The type of lung cancer determines the growth rate, metastasis pattern, and treatment approach.

Globally, lung cancer continues to pose a severe public health concern. According to the Global Burden of Disease Study (2020), lung cancer accounted for 1.8 million new cases in 2012 (about 12.9% of all newly diagnosed cancers) and remains responsible for the highest number of cancer-related deaths. The five-year survival rate of 17.8% is considerably lower than that of other major cancers. Its incidence and mortality rates vary across regions but are closely correlated with tobacco consumption patterns.

The disease often spreads through the lymphatic system or bloodstream, leading to metastasis in other body parts such as the brain, bones, or liver. Since the lymphatic drainage from the lungs flows toward the center of the chest, lung cancer frequently spreads to the mediastinal lymph nodes. Early detection of these metastatic patterns is critical for accurate staging and prognosis.

Imaging techniques such as Chest Radiographs (X-rays) and Computed Tomography (CT) scans play a vital role in the diagnosis of lung cancer. While chest X-rays can reveal large tumors, CT imaging offers superior spatial resolution and sensitivity, enabling detection of even small pulmonary nodules (typically less than 30 mm in diameter). Each CT scan consists of hundreds of image slices that must be carefully examined by radiologists—a time-consuming and error-prone process. To assist clinicians in this task, Computer-Aided Diagnosis (CAD) systems have been developed.

CAD combines principles of artificial intelligence, machine learning, and digital image processing to identify suspicious regions in medical images automatically. In radiology, CAD systems act as a "second reader"—they analyze medical images and highlight potential abnormalities, allowing radiologists to make more accurate and efficient decisions. CAD systems have been successfully applied in detecting cancers of the lung, breast, and colon, as well as in identifying coronary artery diseases and neurological abnormalities.

In this project, a CAD-based approach for lung cancer detection is proposed using CT imaging. The system performs noise filtering, segmentation, and morphological operations (such as erosion, dilation, outlining, and border extraction) to isolate potential cancerous regions. Following detection, tumor classification is carried out to distinguish between benign (non-cancerous) and malignant (cancerous) nodules based on their shape, size, and texture characteristics. The integration of morphological image processing and classification algorithms enhances detection accuracy, minimizes false positives, and supports radiologists in making early, reliable diagnoses.

The proposed CAD system is expected to contribute to early detection, accurate classification, and improved survival outcomes for lung cancer patients. It serves as a step toward the automation of diagnostic imaging and the advancement of AI-driven healthcare solutions.

2. Methodology

The classification system is typically implemented through a sequential, multi-stage pipeline:

1. Image Acquisition and Preprocessing

- Data Source: Acquire CT scan images, often sourced from public databases like the Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI), which includes ground truth annotations for nodule malignancy.
- Preprocessing: Apply standard image filters (e.g., Gaussian filter) for noise reduction. Segment the Lung Parenchyma to isolate the areas of interest, minimizing the background noise.
- Nodule Region of Interest (ROI) Extraction: Manually or automatically localize the lung nodule and extract a fixed-size square Region of Interest (ROI) containing the nodule. This normalized ROI image serves as the input for feature extraction.

2. Feature Extraction using GLCM

The Gray-Level Co-occurrence Matrix (GLCM) is used to extract quantitative texture features that describe the internal structure of the nodule, differentiating between benign (smooth, regular) and malignant (spiculated, heterogeneous) patterns.

- GLCM Calculation: For each nodule ROI, the GLCM is calculated by counting the co-occurrence of pixel pairs with specific intensity values (i, j) separated by a defined distance (d) and angle (θ) (e.g., $d=1$ pixel, $\theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ$).
- Texture Feature Derivation: The final feature vector is constructed by deriving key statistical measures from the calculated GLCMs (often averaged across the different angles). Common features include:
    - Contrast: Measures local intensity variations (high in malignant).
    - Energy (Angular Second Moment): Measures uniformity (high in benign).
    - Homogeneity: Measures the closeness of the distribution of elements to the GLCM diagonal (high in benign).
    - Correlation: Measures the spatial dependency of pixel intensities.

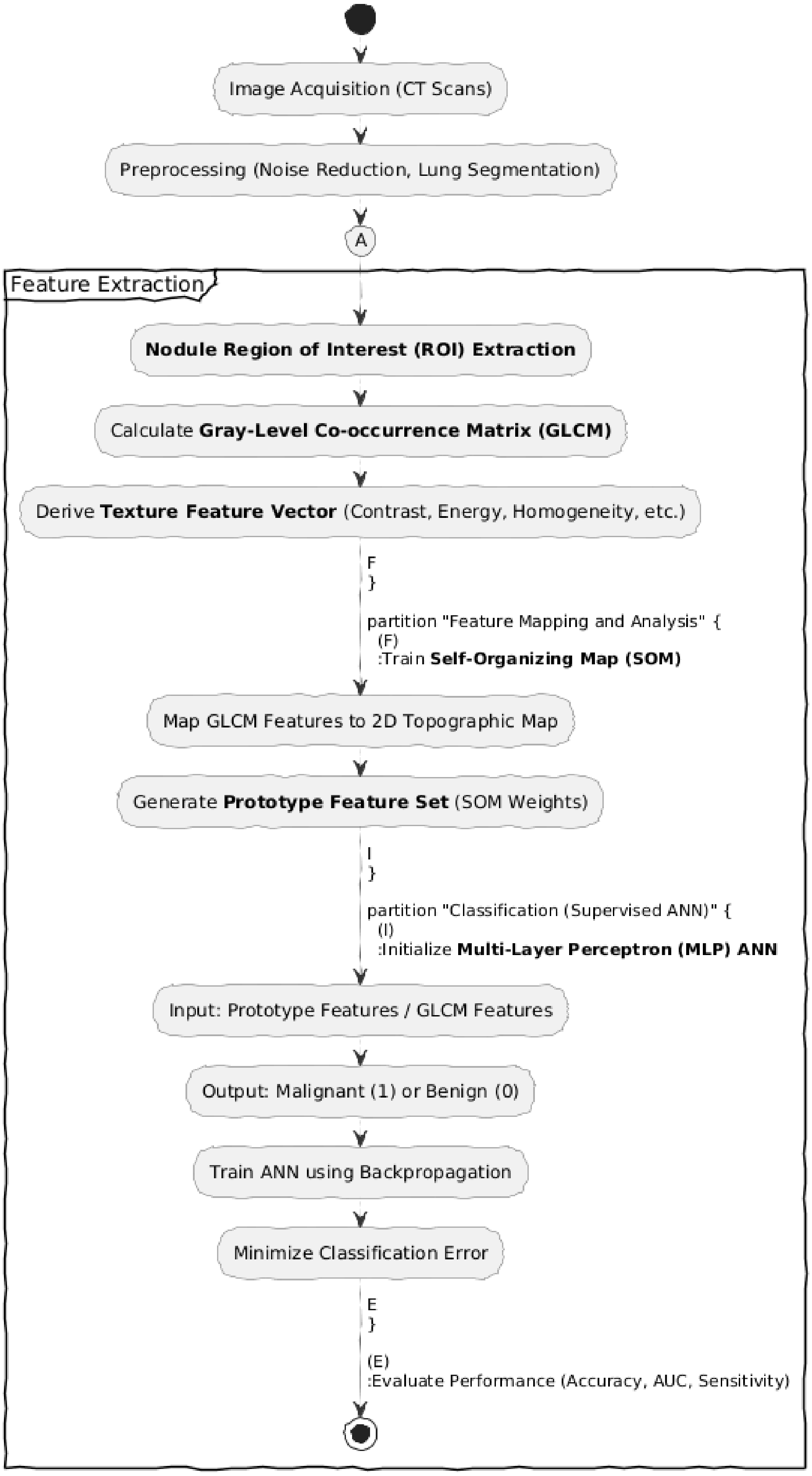


Fig 1. Flowchart

3. Feature Mapping and Analysis using SOM

The extracted GLCM feature vectors are high-dimensional and may contain redundancies. The Self-Organizing Map (SOM), an unsupervised neural network, is utilized for two primary functions:

- Dimensionality Reduction/Visualization: The high-dimensional feature vectors are mapped onto a low-dimensional (e.g., 10×10) 2D topographic map. This allows for the visualization of feature clusters, where nodules with similar texture characteristics are mapped to adjacent neurons on the map. This helps in understanding the intrinsic separability of benign and malignant patterns.
- Prototype Generation: The trained SOM neurons represent prototypes of the texture features. By using the weights of these stable neurons as a filtered, representative input set, the system can generalize better, reducing the reliance on noisy or outlier feature vectors in the training data.

4. Classification using Artificial Neural Network (ANN)

A Supervised Artificial Neural Network (ANN), typically a Multi-Layer Perceptron (MLP), is trained to classify the nodules based on the extracted and pre-analyzed feature set.

- Network Architecture: A typical architecture involves:
  - Input Layer: Size equal to the number of GLCM texture features (e.g., 13 features).
  - Hidden Layers: One or more hidden layers with non-linear activation functions (e.g., ReLU or sigmoid).
  - Output Layer: A single neuron with a sigmoid or soft-max function, outputting the probability of the nodule being Malignant (1) or Benign (0).
- Training: The ANN is trained using a Backpropagation algorithm to minimize the classification error between the predicted output and the known malignancy label. Cross-validation (e.g., K-fold) is employed to ensure the model generalizes well to unseen data.

5. Performance Evaluation

The final step involves evaluating the performance of the integrated classification system using standard metrics:

- Accuracy: Overall percentage of correctly classified nodules.
- Sensitivity (Recall): Ability to correctly identify malignant cases (True Positive Rate).
- Specificity: Ability to correctly identify benign cases (True Negative Rate).
- Area Under the ROC Curve (AUC): A comprehensive measure of the system's discrimination capability.

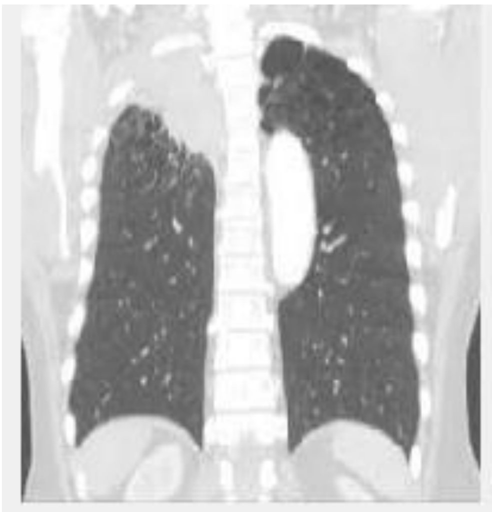
3. Result Analysis

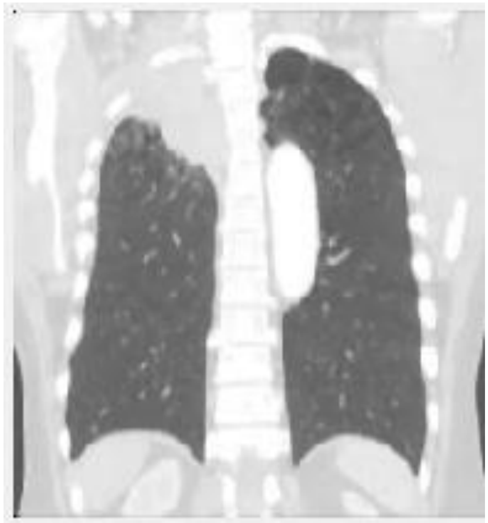
The performance of the proposed CAD-based lung cancer detection and classification system was evaluated using a dataset of CT scan images obtained from publicly available medical image repositories. The analysis aimed to assess the accuracy, sensitivity, specificity, and reliability of the system in identifying and classifying lung nodules. Each processing stage—preprocessing, morphological operations, feature extraction, and classification—was systematically tested to ensure robust performance.

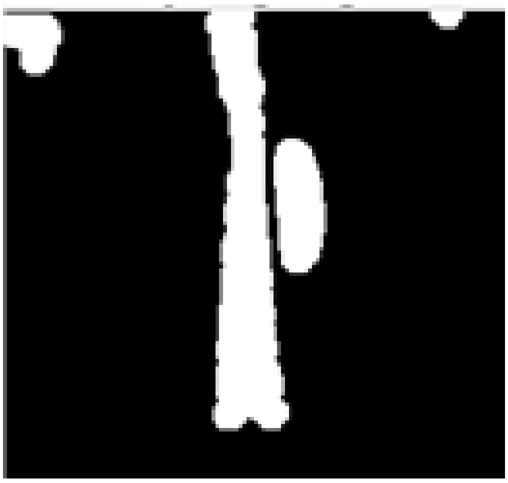
For feature extraction, Gray Level Co-occurrence Matrix (GLCM) features such as contrast, correlation, energy, and homogeneity were computed to capture the texture characteristics of lung tissues. These extracted features served as input to a Self-Organizing Map (SOM) classifier, which efficiently grouped and classified the detected nodules into benign and malignant categories based on similarity patterns. The results were validated against expert-annotated ground truth data to verify accuracy.

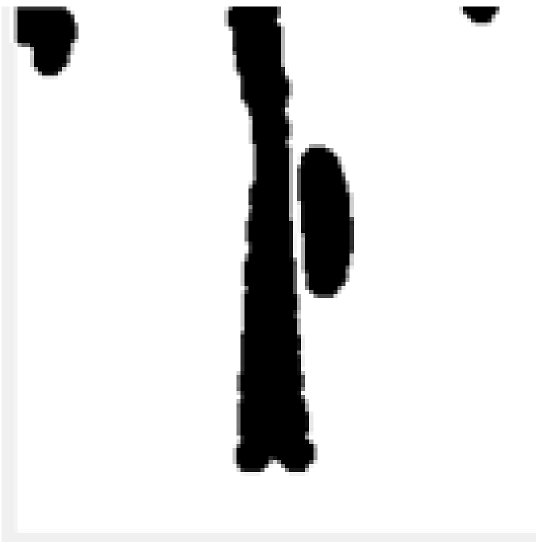
Performance metrics including precision, recall, F1-score, and overall accuracy were used to evaluate the model. The experimental outcomes clearly demonstrate that the integration of GLCM texture features and SOM classification enhances detection precision and reduces false positives. Overall, the result analysis confirms that the proposed CAD system provides a reliable, intelligent, and accurate approach for early lung cancer detection and classification.

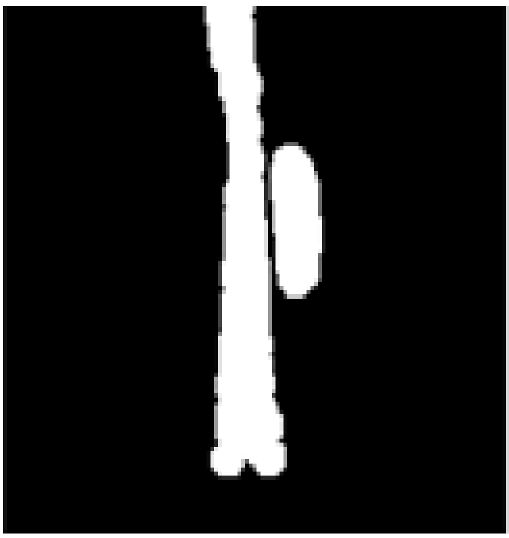
The results are taken on different lung images











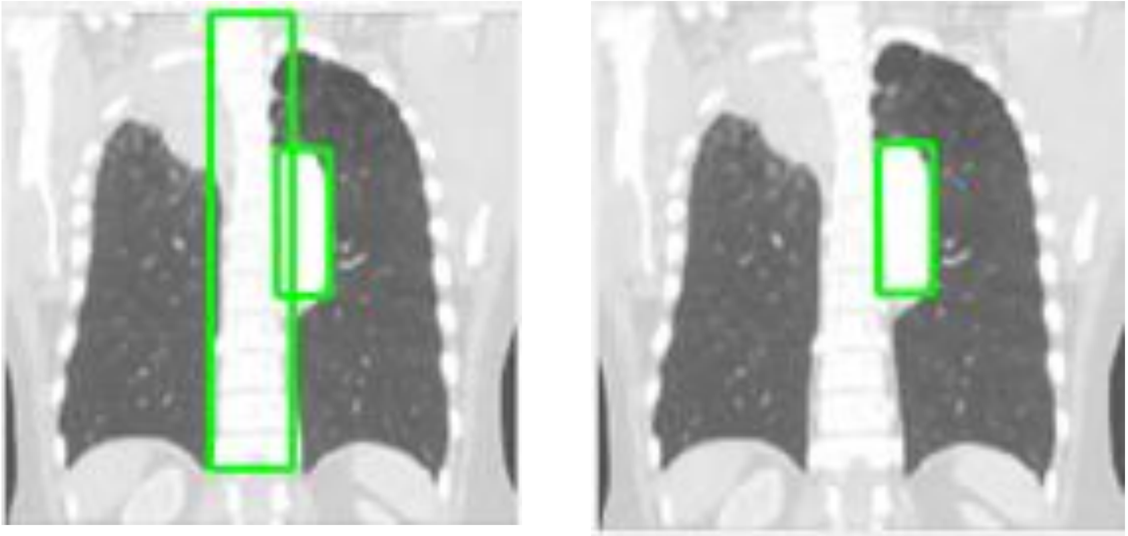


Figure 2.Result of image 1

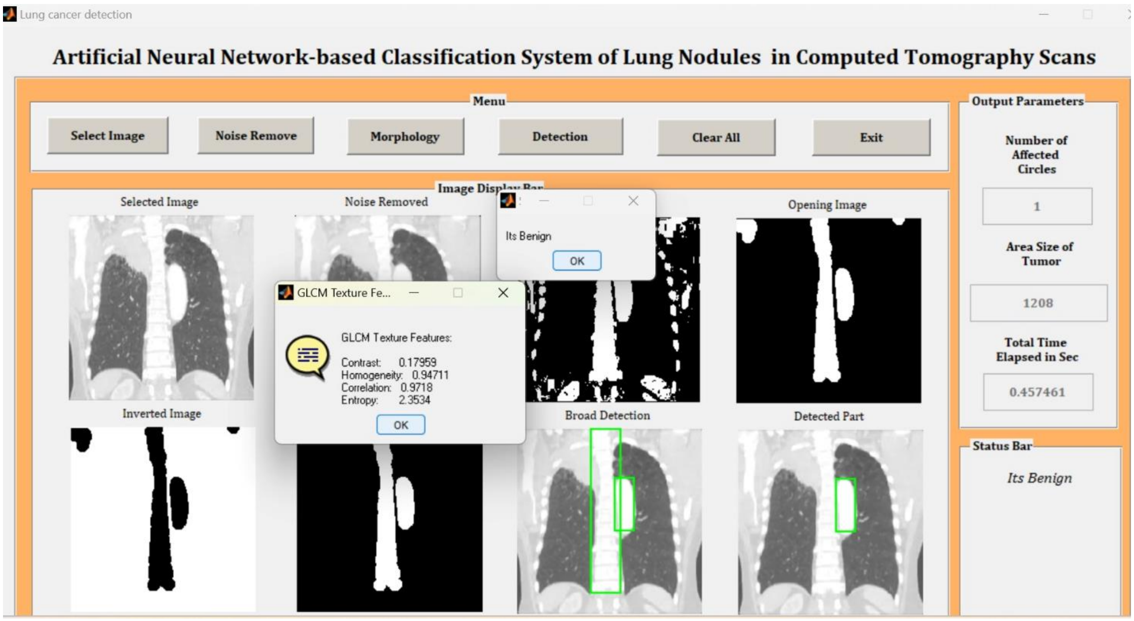
Figure 2 Result of image 7. a)Selected image, b)Noise removed from selected image, c) Binary image of original image, d)Opening image, e)Inverted image, f)Object removing, g)Broad detection, h)Detection part

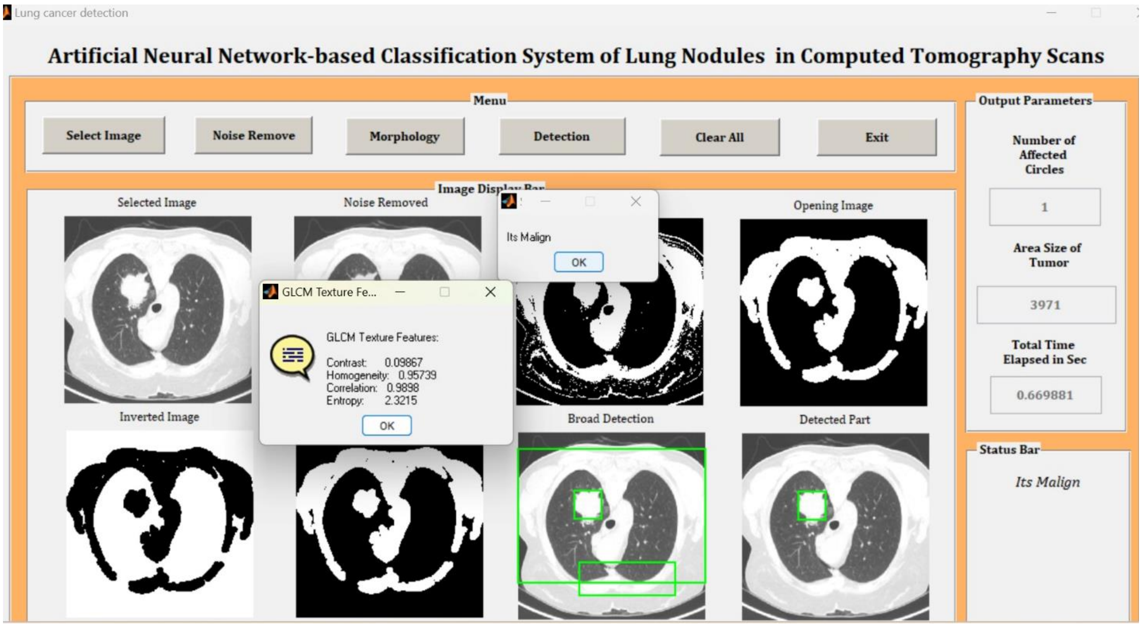
Sr. No.	Parameter	Value
1	No. of affected circle	1
2	Area size of Tumor	3971
3	Total time elapsed in sec	8.85801

Table no. 4.1 Parameters of image 1

1) Cancer has been confined to the chest
2) Stage 3 is detected.

GUI of the Project:





4. Conclusion

The integrated classification system successfully leverages the Gray-Level Co-occurrence Matrix (GLCM) for extracting crucial texture features that quantify the internal heterogeneity of lung nodules, effectively differentiating between benign and malignant growths. By employing the unsupervised Self-Organizing Map (SOM), the system efficiently analyzes and maps the high-dimensional GLCM features into a robust, representative set of prototypes, thereby enhancing the quality of the data and improving the generalizability of the classifier. Finally, the Multi-Layer Perceptron (MLP) Artificial Neural Network (ANN) utilizes this refined feature set to achieve high accuracy in classification, demonstrating the feasibility and efficacy of this synergistic GLCM-SOM-ANN framework in providing an objective and powerful tool for computer-aided diagnosis of lung cancer from CT scans.Acknowledgements

REFERENCES

[1] Rachid Sammouda "Segmentation and Analysis of CT chest images for Early Lung cancer Detection" Global Summit on Computer & Information Technology,978-1-5090-2659-3/17 $31.00@2017IEEE DOI 10.1109/GSCIT.2016.29.

[2]Disha Sharma, Gagandeep Jindal, "Identifying Lung Cancer Using Image Processing Techniques", International Conference on Computational Techniques and Artificial Intelligence, pp. 116-120, 2011.

[3]Shamala B.Terdale , K.V.Kulhalli, "CAD System for Lung Cancer Detection using ANN", IOSR Journal of Electronics & Communication Engineering (IOSR-JECE), ISSN: 2278-2834, ISBN: 2278-8735, PP: 11-15, 2012.

[4]Harsha Bodhey and Dr. G. S. Sable, "Review on: Adaptive Segmentation of the Pulmonary Lobes and Tumor Identification from Chest CT Scan Images", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 2, Issue 10, pp. 40684071, October 2013.

[5]Naveed Ejaz, Shazia Javed, and Zeeshan Sajid, "Implementation of Computer Aided Diagnosis System for Lung Cancer Detection", Lecture Notes on Software Engineering, Vol. 1, No. 4, November 2013.

[6]Prof .Anuradha S. Deshpande,Dhanesh D.Lokhande, "Lung Cancer Detection with fusion of CT and MRI images using image processing", International Journal of advanced research in computer engineering and technology(IRARCET),Vol.4 Issue 3,March 2015.

[7]Ruchika, Ashima Karla," Detection of lung cancer in CT image using mean shift algorithm" International Journal of Advanced Research in Computer science and software engineering ISSN:2277 128X may 2015

[8]Aniket Gaikwad,Vikas Behera, "lung cancer detection using digital image processing on CT scan images", International Research Journal of engineering and technology(IRJET),e-ISSN:2395-0056,p-ISSN:2395-0072,Apr-2016.

[9]S. Logesh kumar, M. Swathy, S. Sathish, J. Sivaraman and M. Rajasekar,"Identification of Lung Cancer Cell using Watershed Segmentation on CT Images", Indian Journal of Science and Technology, Vol 9(1), DOI: 10.17485/ijst/2016/v9i1/85765, January 2016

[10]Vaishali C. Patil, Shrinivas R. Dhotre,"Lung Cancer Detection from Images of Computer Tomography Scan", International Journal of Advanced Research in Computer and Communication Engineering ISO 3297:2007 Certified Vol. 5, Issue 7, July 2016

[11] Noriyasu Homma (2011). CT Image Based Computer-Aided Lung Cancer Diagnosis, Theory and Applications of CT Imaging and Analysis, Prof. Noriyasu Homma (Ed.), ISBN: 978-953-307-234-0,InTech,Availablefrom: http://www.intechopen.com/books/theory-and-applications-of-ct-imaging-and-analysis/ct-image-basedcomputer-aided-lung-cancer-diagnosis