Lung Cancer Detection Using SVM Algorithm

Mrs. Nandhini A[1], Muthu Sahin S H[2]

[1]Assistant Professor (SG), Department of Computer Applications, Nehru college of management, Coimbatore, Tamilnadu, India.

[2] II MCA, Department of Computer Applications, Nehru college of management, Coimbatore, Tamilnadu, India.

Abstract: *In this study investigates the formation of Lung cancer detection using the techniques image processing. The oncologist considered for blood test result. The system formed can take any of medical image within the three choices consisting of CT, MRI and Ultrasound image. This study explores the application of Support Vector Machine (SVM) classification algorithms for the detection of lung cancer, this improves its ability to handle high-dimensional data and provide robust classification result. The model proposed here is developed using PSO, genetic optimization and SVM algorithm used for feature selection and classification. This paper uses image processing to detect lung cancer and separates the lung image into its parts(segmentation)Find important features in the image (features extraction). Choose the most important features to use (features selection). The computer format accepts any clinical image within three choices of MRI, CT and Ultrasound image as input. SVM tool works very well for early detection and treatment of this lung cancer.*

Keywords: *Support Vector Machine, Lung Cancer, Image processing, Early Detection.*

1. Introduction:

Cancer is one of the main causes of non- accidental death. It has been established that lung cancer ranks first globally in terms of cancer-related deaths affecting both men and women. If individuals receive an early diagnosis, the death rate can be decreased since doctors can then provide the appropriate treatment within the allotted period. Cancer arises when a cluster of uncontrollably growing cells lose their equilibrium and develop into malignant tumors that spread across the surroundings tissues. Small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC) are two categories of cancer. The diagnosis and course of treatment for small cell and non- small cell lung cancer differ. The different computer-aided diagnosis techniques for lung cancer detection include image processing, pattern identification, and artificial neutral network. Images from CT, MRI, and ultrasound scans are used in this work. Creates cross-sectional pictures of particular scanned object region using a combination of computer-processed comprises several X-ray pictures obtained at various angles. The purpose of this work is to design a system that can output the desired image given any one of the three images as an input. The employed algorithms are effective in terms of accuracy, specificity, and sensitivity. The steps in the suggested model are as follows: as: Gathering a set of lung image data, preparing, identifying edges, morphological processing, and segmenting MRI and CT scans. There are parts that go over each step.

2. Methodology:

The techniques employed to create the suggested model are explained in this section. This allows for two photos. Finally, a comparison is made between each image. The following lung cancer of classification given steps figure 1.

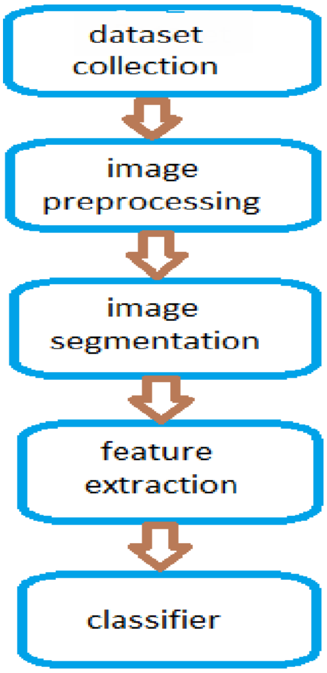


Figure 1. Proposed System

2.1. Data Collection:

This is the initial stage, where we keep all the CT scans. From the database containing the computed tomography picture of individuals with cancer and those without, two types of images are extracted.

2.2. Image Preprocessing:

This stage's primary goals are to improve the image and reduce or eliminate noise and undesired aberration. There are two steps here: Both image improvement and smoothing. A certain amount of noise is added to the CT picture when it is extracted from the machine noise reduction is achieved through image smoothing. Here, undesired or distortions or noise are removed using the median filter. Since it keeps the edges intact and doesn't distort the image, median filtering is a useful technique. Contrast adjustment is used to improve the image by lessening the impact of contrast changes on the image's quality. Through the transformation of new values to input pixel values, it improves visual contrast. As a result, default saturation of low and high data occurs.

2.3. Image segmentation:

For the purpose of identifying probable malignant lesions or other regions of interest (ROIs) in the lung images, segmentation is essential. Typical methods of segmentation consist of:

- Thresholding: A straightforward technique for segmenting an image in which pixel values are compared to a threshold.
- Techniques like region expanding, splitting, and merging are example of region-based segmentation.
- Edge Detection: To determine the boundaries of lesions, use edge- detection algorithm (such as Canny or Sobel).
- Morphological procedures: To enhance segmentation outcomes, use procedure like dilation and erosion.

2.4. Feature extraction:

We must compute the tumor's characteristics from the segmented image. This is a critical phase in the entire process when we will compute the features' area, perimeter, and eccentricity. These characteristics are all geometrical in nature.

1. Area: This indicates the number of pixels that are obtained at a high value and recorded as 1 in lung nodules. It is a scalar quantity, meaning that adding up the values is what it entails.

2. Perimeter: This second scalar quantity indicates how many pixels have a high value at the lung nodule's borders, meaning that they are summated at the boundary where their pixel value is 1.

3. Eccentricity: This is the ellipse's major axis length divided by the distance between its foci. Its value range from 0 to 1.

2.5. Classification:

At this stage, the type of cancer must be identified-benign or malignant. The SVM classifier is employed in this instance. SVM classifiers are models for supervised learning that are recognized by their patterns. Whenever there are precisely two classes in our data, support vector machines (SVM) can be applied. The optimal hyperplane is identified for data, classification; it yields two distinct classes with disparate data points, In the order to choose the optimal classifier hyper plane with the biggest margin between the two classes.

Particle Swarm Optimization Method:

This research employs the Particle Swarm Optimization method, a multilevel threshold approach for image segmentation. The particle swarm optimization principle is used to address the threshold problem. By determining the proper threshold values, the PSO algorithm can be utilized to get an appropriate division of the target image in line with a fitness function. The experimental results, which were tested on all three photos, demonstrated the efficacy of the technique. A population-based optimization system called particle swarm optimization (PSO) is based on the social behaviour of flocks of birds. This method starts with a collection of random particles and iteratively update generation is search of optimism.

Genetic Optimization Method:

Genetic algorithms fall within the broader category of evolutionary algorithms, which use strategies including inheritance, mutation, selection, and crossover that are modelled after natural evolution to produce answer to optimization issues.

Support Vector Machine:

In this research, supervised learning models with related learning algorithms-also referred to as support vector machines-are utilized to evaluate data and identify patterns for classification purposes. It is a non- probabilistic binary linear classifier since it starts with a collection of input data and predicts, for each input, the specified class out of the two options. SVM use kernel functions such as polynomial, BF, quadratic, and Multi-Layer Perception (MLP) to transfer the provided data into a different space.

SVM Algorithm Works:

Even in situation when the data is not otherwise linearly separable, SVM categorizes data points by mapping them to a high-dimensional feature space. After identifying a separator between the categories, the data is altered so that the separator can be represented as a hyperplane. After them, the group to which a new record should belong can be predicted using the features of fresh

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 09 | Sept - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

data.

The sum of the margins between two classes is the aim of SVM. The ideas presented are further clarified in the graphic below.

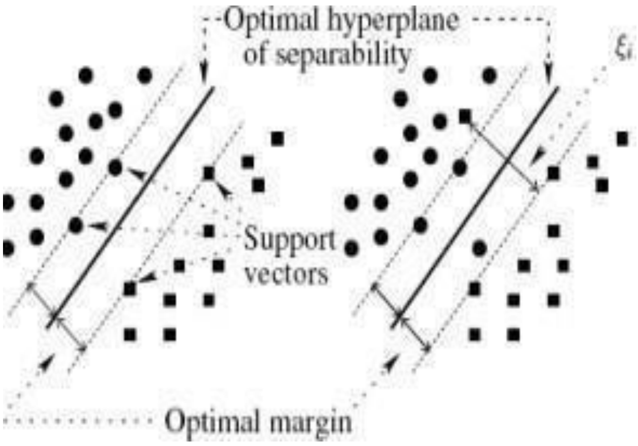


Figure 2. SVM Hyperplane

Regression and classification can be worked on in a variety of methods. SVM is renowned for its ability to solve problems, including those with highlighted aspects and unsolvable problems. This is a cursory process used to separate datasets, extract features from known sets, and create a decision plane to partition the sets.

The outcome of building a line that is closer to the training data in terms of distance. For binary classifications, the linear plane produces superior outcomes. To handle multi-class instances with non-linear planes, distinct toolbox kernels are utilized.

SVMs are also used to categorize cancer stages. A binary SVM classifier is trained for any category, regardless of whether each document in the training set is in that category or not. A medical report might not have only one stage.

The toolkit's functions are used by the SVMs in their operation. The computer is trained using an enormous array of outcomes at various levels. Following analysis of the test results, it provides a score that serves as a cutoff point to determine whether any updates are made that pertain to the same class.

The validation portion of the code is run in order to increase the amount of information while still producing meaningful outputs on the same data. First, the data is split up into smaller sets, and the machine that was trained on the other sets is used to extract the result for one of the smaller sets.

3. Results:

The Lung CT, MRI, and ultrasound pictures used in this work came from a specialized medical imaging facility. The Gabor filter is used to enhance the image. Following the enhancing stage, the images underwent layer separation and were subsequently transformed into grayscale images. The lung region, or ROI, was retrieved through the use of the Super Pixel Segmentation pictures are displayed in Figures 2,3, and 4, respectively.

The photos are subjected to additional feature extraction and feature classification stages in MATLAB's GUI, or graphics user interface, which yields a normal or abnormal result.

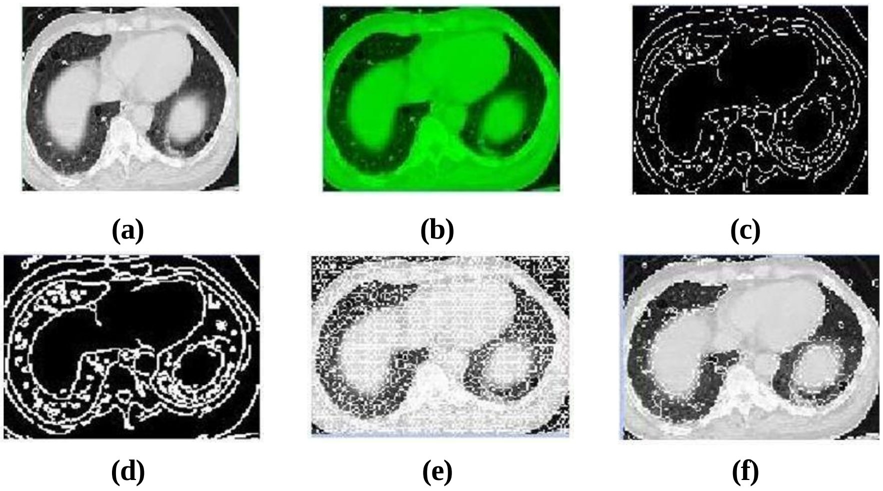


Figure 3. CT image segmentation procedures the following processing: a) input denoised image; b) green layer separation; c) gray level intensity; d) edge detection; e) morphological processing; and f) segmentation.

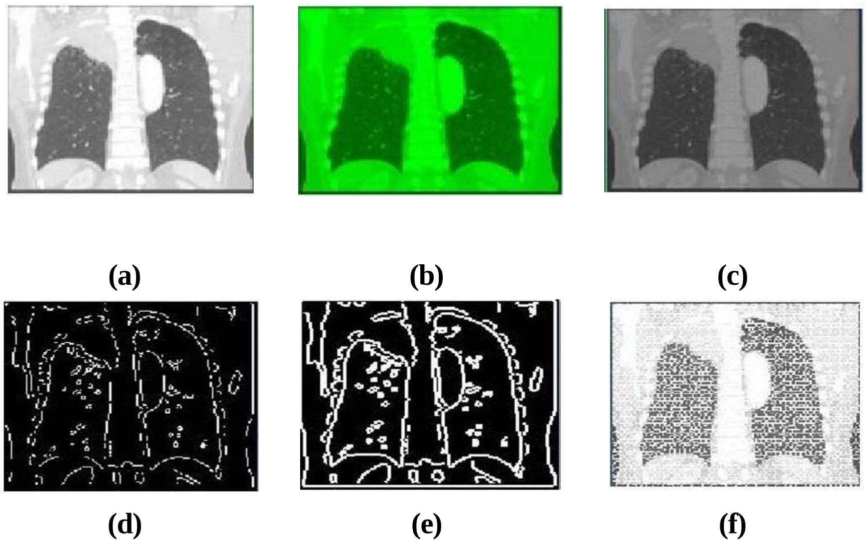
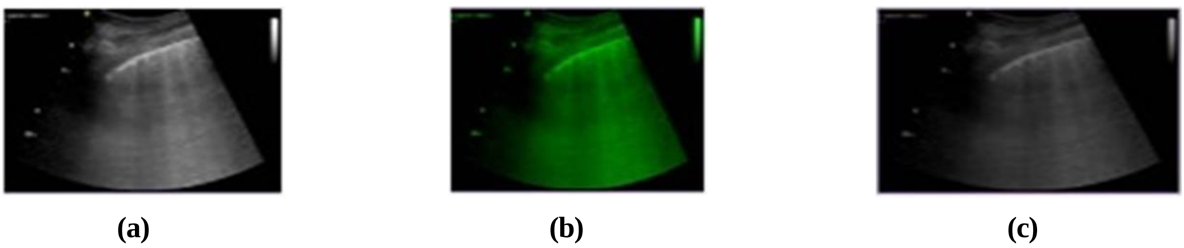
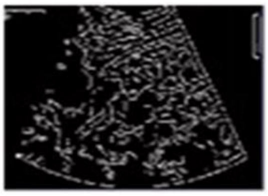
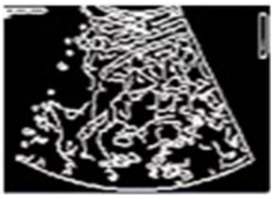


Figure 4. MRI image segmentation steps a) The denoised image input; b) The separation of green layers; c) The intensity of gary levels; d) edge detection; e) morphological processing; and f) segmentation.

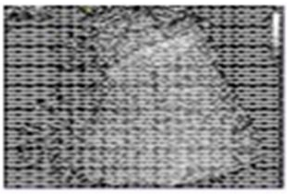




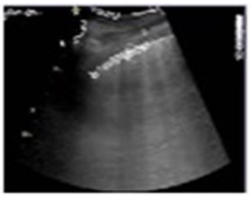
(d)



(e)



(f)



(g)

Figure 5. Procedures for Ultrasound Image Segmentation a) The denoised picture input; b) The separation of green layers; c) The strength of gray levels; d) edge detection; e) morphological processing; f) segmentation; and g) super pixel segmentation.

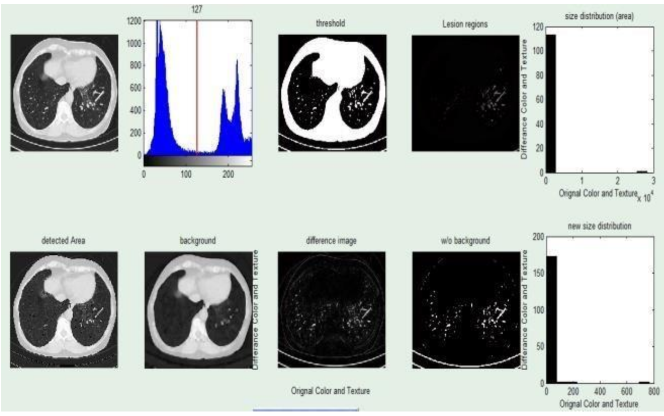


Figure 6. CT Image Feature Extraction

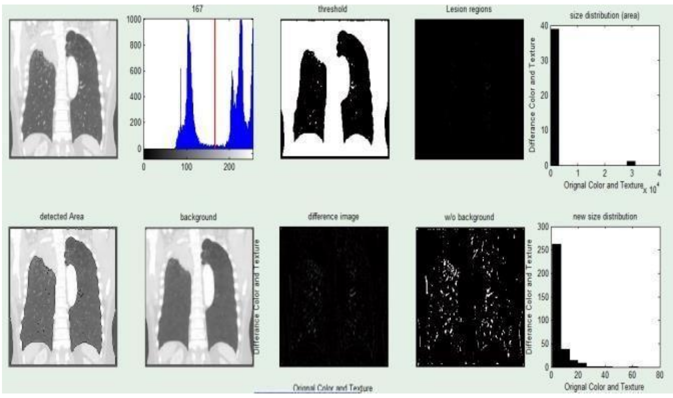


Figure 7. MRI Image Feature Extraction

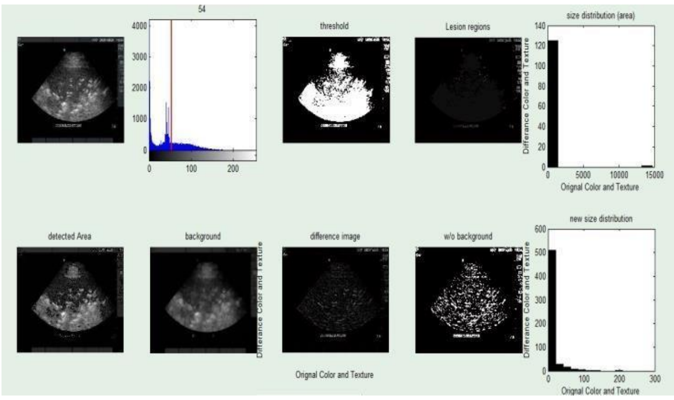


Figure 8. Ultrasound Image Feature Extraction

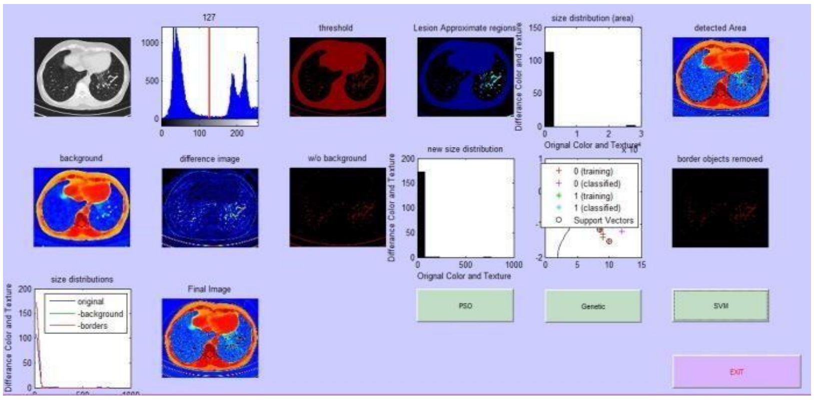


Figure 9. CT Image Feature Selection

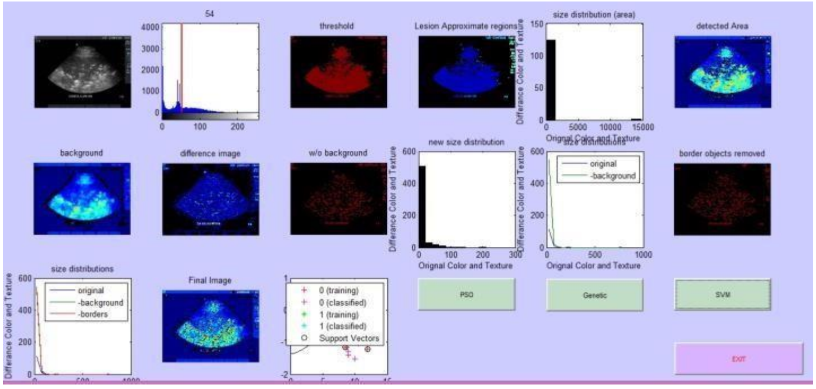


Figure 10. Ultrasound Image Feature Selection

4. Conclusion:

This survey examines the primary imaging modalities for cancer detection using image processing applied to CT, MRI, and ultrasound images. Our approach for segmenting CT, MRI, and ultrasound images was suggested. Studying the required features thar were retrieved for the two photos allows for the accurate identification of this system. We also employed feature selection using the PSO, GO, and SVM algorithms, which resulted in an accuracy of almost 89.5% and a decrease in false positives.

Future Scope: The method employed reduces system complexity and improves diagnosis confidence. For the two photos, the Gabor filter has been employed to reduce noise. After applying a canny filter

for feature selection and classification, producing an abnormal or normal nodule as the outcome. Cancerous nodules are referred to as abnormal nodules. In addition, we can use the Pearson's and Spearman's algorithms for classification in order to identity the areas in all three photos that are most susceptible to cancer.

Reference:

1. A Cancer Journal for Clinicians, 55 (2005), 10-30; American Cancer Society, Cancer Statistics, CA.
2. International Transport on Computing Sciences, Anita Chaudhary, Sonit Sukhraj Sing, 4, 2012.
3. International Conference on Computational Techniques and Artificial Intelligence (ICCTAI'2011), 17-19 November 2011, 872-880; Disha Shama, Gaga deep Jindal.
4. Information Conference on Biomedical Engineering, Egypt, 2002; EI-Baz A, Farag A.A., Falk R, Rocco R.L.
5. Computer-aided diagnosis in chest radiography: a survey, Gineken BV, Romeny BM, and Viergever MA, IEEE, Transactions on Medical Image, 20(12), 2001, 1228-1241.
6. In the IEEE Transactions on Medical Imaging, 20(12), 2001, Gineken BV, Romeny BM, and Viergever MA.
7. First IEEE International Conference on Power Electronics, Intelligent Control and Energy Systems, IEEE, 2016. Gawade Prathamesh Pratap, R.P. Chauhan, Detection of Lung Cancer Cells using Image Processing Techniques.
8. Moffy Crispin Vas and Amita Dessai: Using Image Processing Techniques to Classify Cancerous and Non-

Cancerous Lung Cancer Nodules International Conference on Engineering and Management Academic Research (ICAREM-17).

9. Aditya Tiwari, Shubhangi Khoragade, and First IEEE, 2016. Automatic Identification of Major Lung Diseases Using Chest Radiographs and Classification by Artificial Neural Network.
10. Dr. Anuradha Thakare, Pooja R. Katra, the 2nd International Conference for Convergence in Technology (12CT) 2017 explored the use of image processing and data classification techniques for the detection of lung cancer stages.
11. Neural Information Processing, IEEE, Lin D, Yan C, 2004.
12. IEEE Journal of Biomedical, 2013, 519-524; Tao Suna, Jingjing Wanga, Xia Li Pingxin Lvc, Fen Liua, Yanxia, Luoa, Qi Gaoa, Huping Zhua, Xiuha Guo.
13. Schwartz L.H., Jiang L, Zhao B, Gamsu G, Ginsberg M.S., Journal of Applied Clinical Medical Physics, 4, 2003.