Indian Journal of Science and Technology, *Vol 9(48), DOI: 10.17485/ijst/2016/v9i48/106499, December 2016*
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Identification of Cancer in CT Images based on SVM and PSO using Gene Selection Algorithm

Suhas A. Bhyratae[1], Shanthi Mahesh[2], A. S. Harshith[1] and Harshitha C. Shekar[1]

[1]Department of ISE, Atria Institute of Technology, ASKB Campus, 1st Main Road, AGS Colony, Anand Nagar, Bengaluru - 560024, Karnataka, India

[2]Information Science and Engineering, Atria Institute of Technology, ASKB Campus, 1st Main Road, AGS Colony, Anand Nagar, Bengaluru - 560024, Karnataka, India

Abstract

Objectives: The objective of the proposing system is to classify Computed Tomographyimages into two classifications like cancernodes and without cancernodes for lung pictures using SVM and PSO classifiers. **Methods/Analysis:** Computed tomography (CT) cancerous lung images are used. Techniques employed include the usage of support vector machine and particle swarm optimization classifiers. Proposed techniques include histogram equalization, ROI masking using thresholding, wavelet transform, gray level co-occurrence matrix. Hence resulting in results in high accuracy, also yields better result in terms of the recognition accuracy. **Findings:** SVM shows higher accuracy rate of identifying the cancerous nodules than PSO for different iterations held. These comparison results between SVM and PSO bring out the efficiency of comparing between the two classifications algorithms. In the existing system 1.afusion of classification techniques was used to determine the lung nodes in the CT image and feature extraction based technique was implemented on the node. 2. As the second step, the features were computed based on its texture to distinguish the blood vessels. The proposed system is implemented to classify by combining of SVM and PSO. Hence leading to Low performance and accuracy in classification. With the results obtained in the proposed system it differs in various aspects like the lung nodes are well identified, better result obtained due to comparisons hence highlights the fact that proposed system had far more advantages than the existing system.

Keywords: Computed Tomography, Gene Selection, Particle Swam Optimization, Support Vector Machine

1. Introduction

The uncommon growing of cells inside the human body mentions cancer tends to increase in unrestrained fashion.

This research aims to understand, examine and identify the best algorithm that yields high accuracy in identifying the cancer image. To achieve efficiency in gene selection algorithms from thousands of applicant genes that can donate in the identifying cancers, the main objective is to develop a classifier and to emerge a unique method with combination of particle swarm optimization (PSO)and support vector machine (SVM).

The performance of this work with other well-known standard classification methods will be compared. It is been conclude that the proposed method outclasses the well with other standard classifiers for input datasets based on computational analysis. The inordinate importance in texture classification, medical image data processing, soft computing and data mining techniques goes to Feature collection. A good feature collection method in classification investigates the structures of samples increases the processing speed, accuracy prediction and avoids complexity. Here in task, PSO implements a feature selection and SVM[1] assist PSO as a fitness function for the classification.

2. Materials and Methods

The materials include the CT images that are taken as inputs. Methods used here are histogram equalization, ROI masking using thresholding, wavelet transform, gray level co-occurrence matrix, PSO and SVM. The input is

*Author for correspondence

the CT image of the lung which is further decomposed from unsigned integer 16 to unsigned integer 8. This is further pre-processed to remove the noisy data. Layer background removal/foreground extraction is done to obtain the region of interest (ROI). This image is then divided into 4 images in 4 quadrants by wavelet transform technique then feature extraction is done by GLCM technique. After which the trained features help in enhancing the image and it's brightness by depicting the layers of the scanned image. The resultant image is then given as an input to perform classification using SVM and PSO. Finally the output obtained is the accuracy rate that depicts which algorithm yields higher accuracy.

Implementation includes the major steps: Pre-processing, ROI Extraction, Transformation and Feature Extraction**.** There are several ways of encoding the information in an image: Binary image, Gray scale image, Indexed image, True colour or RGB image. Next is testing to discover errors.

Black-box technique is used in this system. It is testing technique can be applied to the structure of module without any prior knowledge of the internal mechanisms. This technique takes inputs and gives outputs without seeing by what means the software mechanism.

2.1 Top-Level Design

The proposed block diagram ia as shown in Figure 1.

2.1.1 Level 1 DFD for Pre-processing/ROI-Extraction

It includes the following processes:-

1. The CT images gotten from the repositories having many sub-areas and each sub-area having certain differences within an image.
2. Few CT images have diverse forms and might having the cancer data in it because of the abnormal growth in the cells.

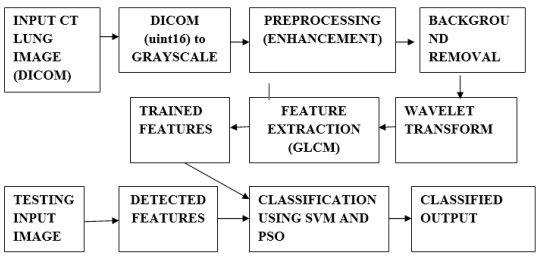


Figure 1. Proposed Block Diagram.

3. An area of interest in each CT image was mined to eradicate the variances in the CT with and without cancer nodes to get good classification results.
4. The Hough transform is used to compute the ROI to estimate all the CT images to its border, leaving some region in the inside which is relevant information and creating it black outside.
5. In the CT images the unique task is to extract ROI as a pre-processing.

2.1.2 Level 1 DFD for Wavelet Transform

It includes the following processes:-

1. Next task after preprocessing is to segment for separation of the ROI from other parts of the body as well as cells from the scan.
2. Segmentation step is not necessary in proposed system.
3. In proposed system, in the place of segmentation the ROI is extracted from images are converted to the transform domain by spatial domain.

Discrete Wavelet Transform

Here a CT image is altered using a multi-scale tool called DWT (Discrete Wavelet Transform).The Discrete Wavelet Transform tool which is useful on discrete data to get a multi- gauge picture.

1. To arrest important data of a CT image in a small information is used to deal for representation by showing its capability.
2. The Discrete Wavelet Transform tool allows a ordered breakdown of an feedback symbol to series of low resolution referential symbol.
3. The upper frequency mechanisms permits illustration of the image in a more closed manner and the image energy greatness is in focusing of a minor set of coefficients are proposed by Discrete Wavelet Transform.
4. In order to hunt a wavelet families this will capably signify a sign of attention in a large variation of applications.

2.1.3 Level 1 DFD for Feature Extraction

It includes the following processes:-

1. Image texture offers a vital basis of data in the current condition of the human health of a detected region.
2. The extraction of texture information of the lung nodes has been done by the GLCM (grey level co-occurrence matrix).

3. The spatial dependence of grey level values can be taken within an image is done by GLCM which is used more in texture analysis method as well as in biological imaging that shows the capability.
4. The computed features were: autocorrelation, entropy, measure of correlation, contrast, dissimilarity), energy, variance and inverse difference moment.

2.1.4 Level 1 DFD for Feature Selection

1. In pattern recognition design cycle, the most crucial step is article or attributes selection since it has to search automatically the best subset of attributes.
2. After the finding of all computational texture features, an examination need to be done to measure the importance of each feature in each wavelet sub-band.
3. The main intention was to decrease the feature set and this work has to be carried out automatically by choosing the best feature.

2.1.5 Level 1 DFD for Classification

1. To categorize any object of the class in the pattern classification[3] should be define.
2. Here, the classification step was completed using a SVM.
3. To train the SVM classifier, a combination of two different features from the set of 11 feature vector was tested.
4. To obtain the best two significant features out of all 11 features, each scale and filter was tested.
5. The selection was done by a status of the features from autonomously evaluating criteria of the total value of two-sample t-test with mutual variance.

3. Results and Discussion

The CT images are referred to as samples. The number of sample to be used are not constrained to any particular images. By taking the obtained CT images as input the analysis is performed. Whereas the number of iterations is changed as per the requirement. In this system the iteration value is performed for 100,200,300,400,500 iterations. Keeping the expected accuracy to be 100% when the iteration is performed it is clearly observed that the accuracy obtained for SVM is more for both cancerous and non-cancerous images. SVM yields accuracy always

Table 1. Testing Table for cancerous images

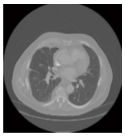
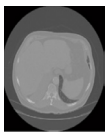
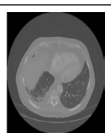
INPUT		EXPECTED OUTPUT		OBSERVED OUTPUT	
		ACCURACY		ACCURACY	
		PSO	SVM	PSO	SVM
CANCEROUS IMAGES		100%	100%	77.7778%	95.1613%
		100%	100%	77.7778%	96.1613%

Table 2. Testing Table for Non-cancerous images

INPUT		EXPECTED OUTPUT		OBSERVED OUTPUT	
		ACCURACY		ACCURACY	
		PSO	SVM	PSO	SVM
NON-CANCEROUS IMAGES		100%	100%	77.7778%	98.3871%
		100%	100%	77.7778%	96.7742%

greater than PSO and also nearing to 100%. The software used is MATLAB 7.14 Version.

In this section we are also discussing on Accuracy of PSO and SVM for cancerous images shown in Table 1. As well Table 2 shows the Accuracy of PSO and SVM for Non-cancerous images.

4. Conclusion

This draws the conclusion that the SVM classifier yields better result when compared to PSO classifier. With the accuracy rate of SVM classifier nearly equal to 100%. This shows that the SVM classifier has higher accuracy.

5. Acknowledgement

This opportunity is used to express deep sense of gratitude to Prof. Dr. Shanthi Mahesh and Prof. Suhas A Bhyratae., Dept. of ISE for providing us all the necessary facilities for successful completion of this work.

6. References

1. Pashaei E. A novel gene selection algorithm for cancer identification based on support vector machine and particle swarm optimization. Dept of Comput Eng. Yildiz Tech Univ, Istanbul, Turkey.
2. Tu C-J. Feature Selection using PSO-SVM. IAENG International Journal of Computer Science. 2007; 1–6.
3. Brown MP. Knowledge-based analysis of microarray gene expression data by using support vector machines. Proc Natl Acad Sci. 97:262–7.