INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

*Corresponding author.

Tel: +91-9751995269
drsanjith@gmail.com

An unconventional SVM classification using Chaos Pso optimization for lung cancer discovery

C Thinkal Dayana¹, K S Mithra¹, S Sanjith²*

1 Assistant Professor, Department of Computer Science, St Alphonsa College of Arts and Science, Karinkal, 629157, India
2 Assistant Professor & Head, Department of Computer Science, St Alphonsa College of Arts and Science, Karinkal, 629157, India. Tel.: +91-9751995269

Abstract

Objectives: The main purpose of this work is to detect the cancer region and to classify the particular region based on Support Vector Machine (SVM) classifier. **Methods:** Optimization technique is used after classifying the cancerous region in order to improve the accuracy of the Lung cancer CT images. The proposed method is improved using a novel Chaos Particle Swarm Optimization (CPSO) technique. The MATLAB is used to optimize the technique. **Findings:** The achieved accuracy of SVM classifier using CPSO is 97.4% which is higher when compared to PSO, Genetic algorithm which yields an accuracy 89.5% and genetic optimization for feature selection and ANN for lung cancer classification which obtains 95.87% accuracy.

Keywords: Chaos Particle Swarm Optimization; SVM; CT image; classification

1 Introduction

Lung cancer is the most common disorders which affects the people and can reduce the survival rate. The medical professionals are able to diagnose the diseases with the help of CT images. To reduce the mortality rate of cancer patients, initial identification of lung cancer with proper treatment is needed. But the initial identification is not an easy task which requires some fundamentals image processing steps followed by Optimization technique[1]. Image processing plays a vital role in disease diagnosis from medical imaging[2]. In this paper a novel approach for lung cancer classification is proposed, in which Chaos PSO optimization is utilized. The simulated results have been obtained based on the performance measures. In[3] proposed the modified bacterial foraging optimization technique for lung cancer classification which evaluates the performance parameter. It shows the simulated result for back propagation neural network is better than SVM method. MLP-NN lung disease classification technique using PSO algorithm which improves the feature selection and also analyzed the accuracy, bit error rate[4–8]. In[9] studied various ways to detect lung cancerous region from the CT image based on PSO and Genetic algorithm. The obtained accuracy of this method is 89.5%. A new idea for identifying the lung cancer status which uses genetic optimization for feature selection and ANN for lung cancer classification. This technique provides the accuracy

of 95.87% [10]. From the literature review it is observed that the accuracy is improved using Optimization techniques. Hence, the proposed method uses Support Vector Machine using Chaos PSO Optimization techniques is applied on lung CT images to improve the accuracy of simulated results. The primary contribution of this paper is to detect the lung cancer region using SVM classifier with novel CPSO algorithm in order to improve the accuracy.

This study is structured as further as follows: Section II gives the detailed explanation of Proposed system. Section III shows the simulated results and Finally, the conclusion of this paper is explained in Section IV.

2 Materials and Methods

The basic steps required for image processing with Optimization techniques are used in the proposed method along with the steps required for the implementation of the system is mentioned in Figure 1.
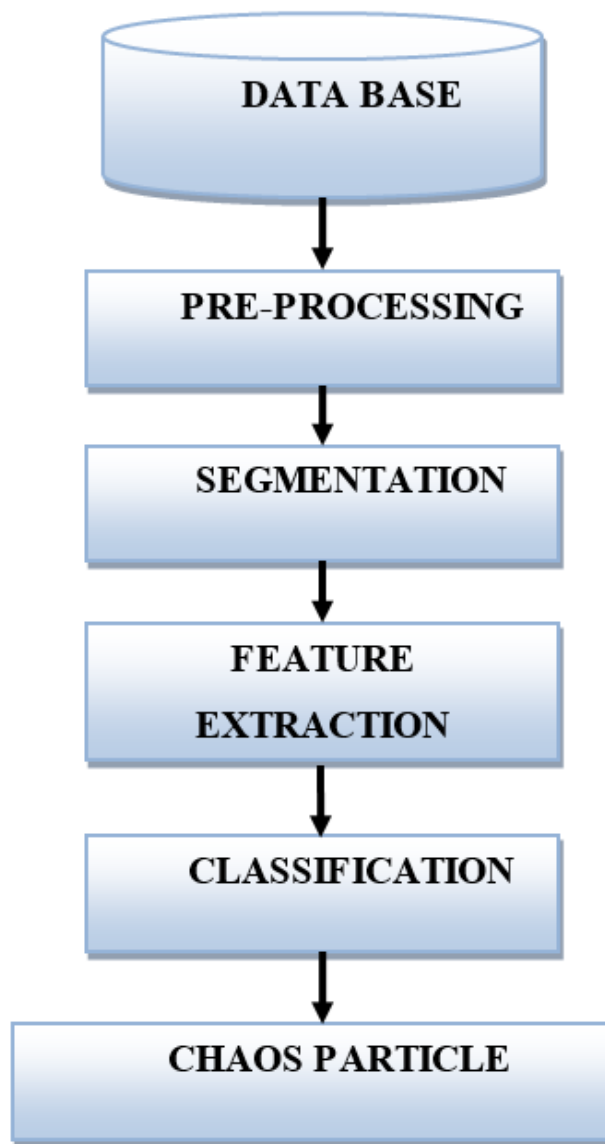

Fig 1. Block diagram of the proposed system

Image acquisition

The capturing of input image from the database is done through image acquisition process, the image obtained from the database will be in DICOM format. The images obtained for proposed research is from Lung Image Database Consortium (LIDC) datasets. This dataset contains data of 50 patients.

Image pre-processing

Generally, CT images will contain a superior amount of noises, so a noise removal process is mandatory for further processing. Image preprocessing step is used to perform the background noise removal. The median filter is used to suppress the salt and pepper noise from the image. Due to its de-noising power and computational efficiency characteristics median based filters are supposed to act as backbone filters[11].

Image segmentation

Segmentation is the process used to separation the medical images into various regions. The aim of segmentation is to shifting the representation of lung image into meaningful, which smoothen the analyzing process of the particular image without any hitches. Adaptive global thresholding is the simplest process used to segment the lung images. It separates the cancer region and non-cancer region very efficiently from the CT lung images[12,13]. This method is mainly applicable were the single value threshold is not working accurately, at that time pixel value depends on its position of lung image.

Feature extraction

Feature extraction is mainly performed for the goal of dimensionality reduction. In this research we used Gray Level Co-occurrence Matrix, In which the total number of rows and columns are similar to the gray levels, G_0. The Figure 2 shows the representation of Gray Level Co-occurrence Matrix (GLCM)[9].
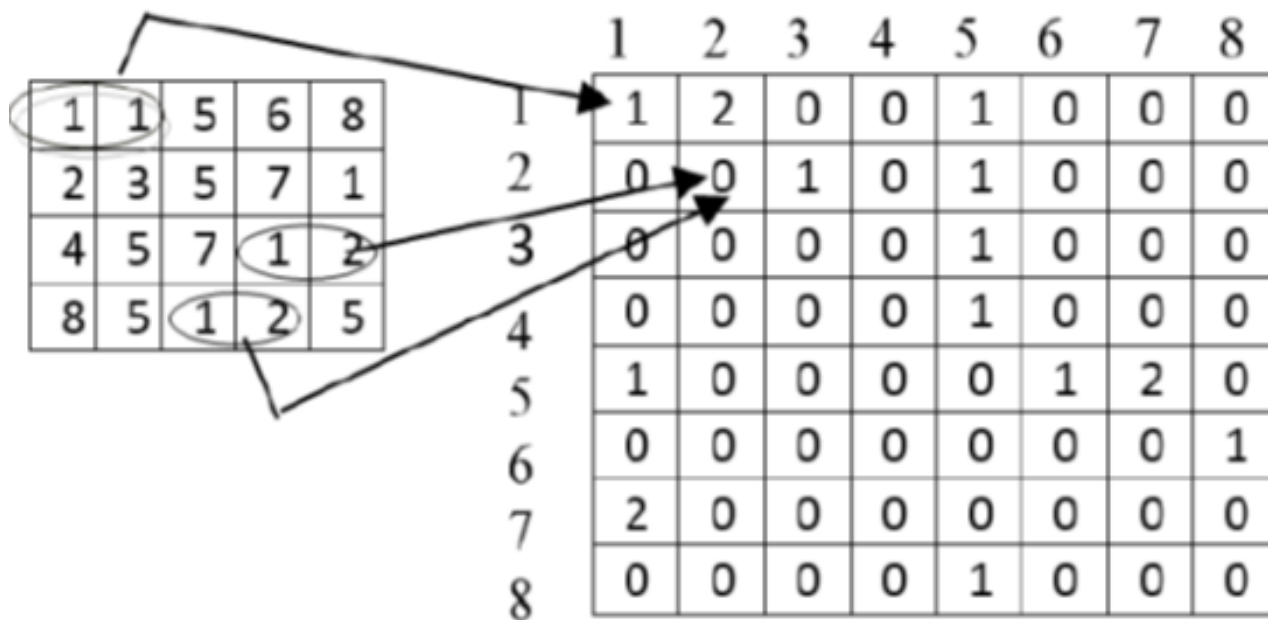

Fig 2. Gray Level Co-occurrence Matrix

Image classification

After the reduction features of GLCM technique the obtained result is provided as the input for SVM classifier[14]. Support Vector Machines is a binary classification method used to analyze data and recognize the patterns that receipts a group of

inputs and predicts the input between two choices present, which makes SVM as non-probabilistic binary linear classifier. It is said to be supervised learning models that uses kernel function as shown in Figure 3.
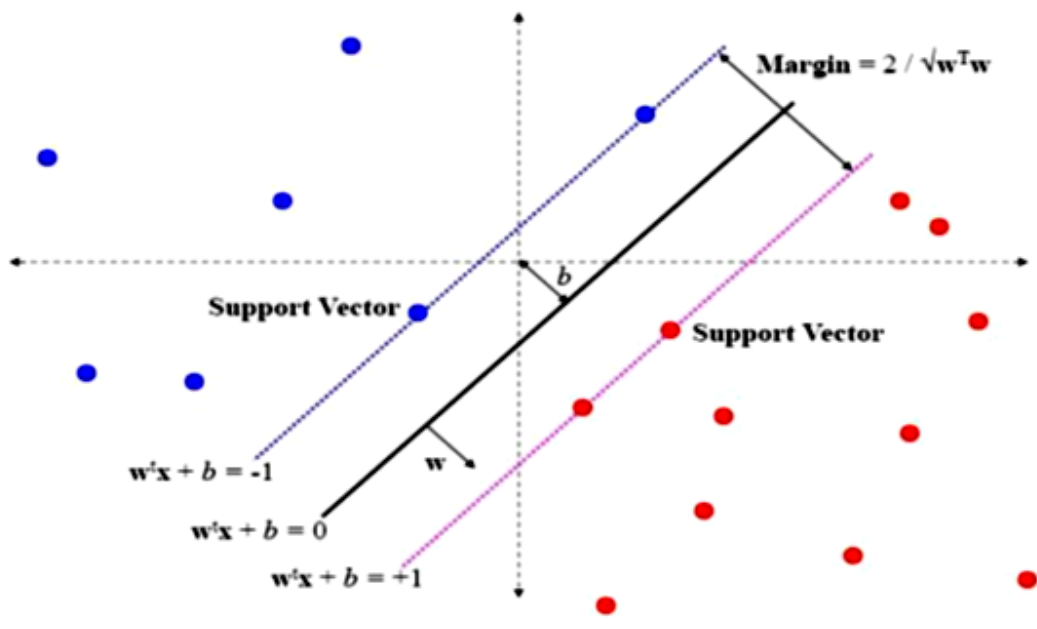

Fig 3. SVM classifier

Chaos particle swarm optimization

This paper uses CPSO algorithm which is a population-based algorithm for solving the thresholding problem. It is a multilevel thresholding technique which accurately finds the threshold value with a fitness function by dividing the target image. So that the optimization problem can be solved. This optimization looks the updated iterations after defining with a set of random particles[15]. The experimental result shows the improved accuracy than the existing technique.

$$min f(x_0, x_{1,.........}x_m) \; a_k \leq x_k \leq b_k$$

3 Experimental Results

In this research the simulation result yielded by using MATLAB. The inputs for this experiment are the images taken from LIDC dataset. The proposed system uses totally 50 number of images. In the dataset 75% data is used for training and 25% used for testing process.

Initially, the input image is preprocessed using median filter. After that the output of preprocessed image is segmented to isolate the cancerous region as shown in Figure 4. The GLCM feature extracted values for contrast, correlation, energy, mean and homogeneity is shown in Table 1 All of the extracted features were besed on pixel information of segmented image region. The mean feature is proportionl to the brightness of the image[16,17].

Table 1. GLCM feature extraction

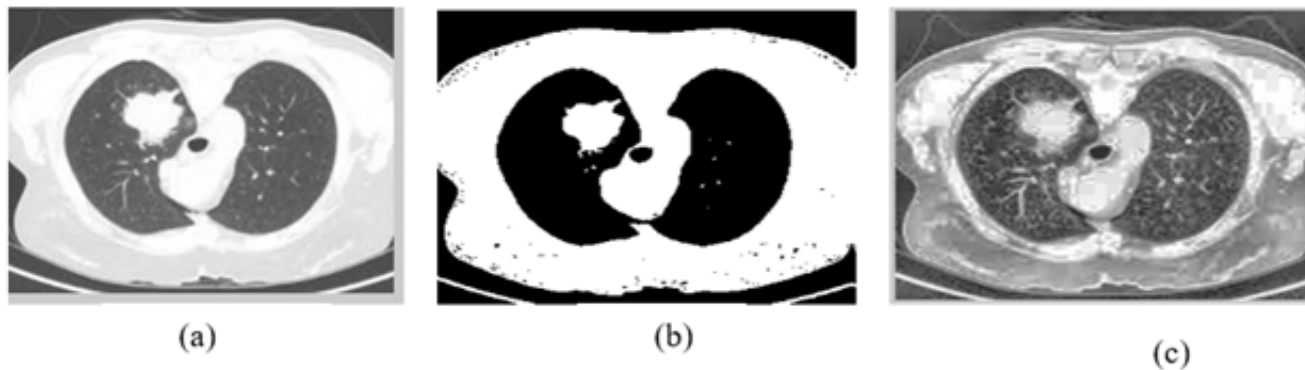| GLCM features	Sample
Contrast	0.0036
Correlation	0.9487
Energy	0.9466
Homogeneity	0.9974
Mean	0.6

Fig 4. (a) Input Image (b) Preprocessed Image (c) Segmented Image

SVM classifier predicts the output as whether normal or abnormal as shown in Figure 5 and this method helps the medical professionals to find the lung cancer at the early stage to avoid the mortality rate in the Globe.

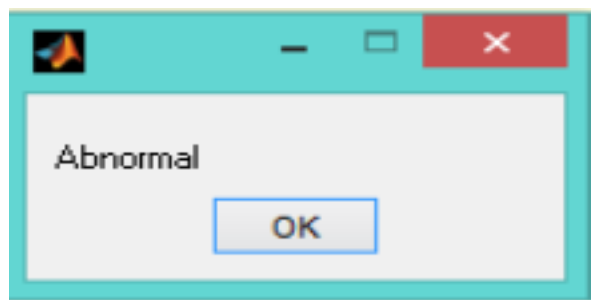**Fig 5.** SVM Classifier output

Figure 6 shows the performance analysis of SVM using Genetic Algorithm, SVM using PSO and SVM using Chaos PSO are compared the parameters such as accuracy, sensitivity and specificity. Performance is measured and proved that the proposed system achieves the better accuracy of 97%.
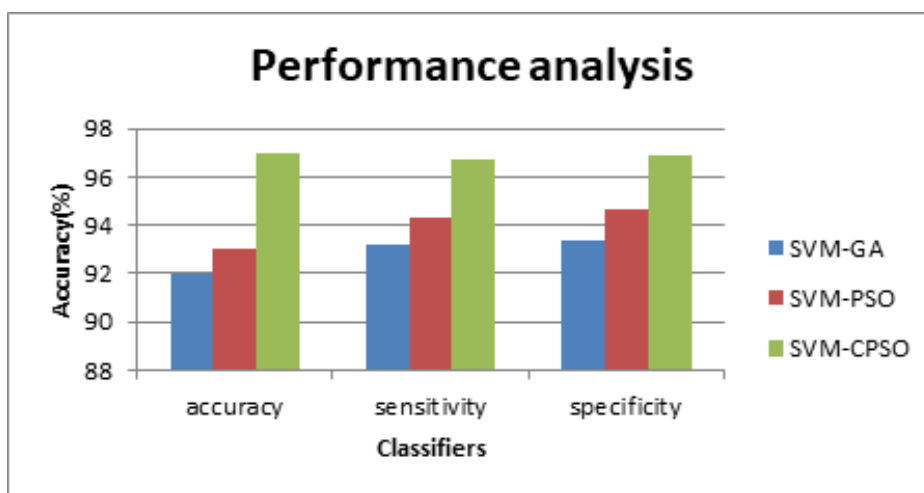

Fig 6. Performance analysis of various optimization techniques with SVM classifier

4 Discussions

Experimentation were carried out from real time datasets obtained from the reputed databases. The results achieved by the research is related with few similar works carried out using SVM classifiers the chosen few related research and the obtained results are discussed below.

In [18] totally 11 patients are considered with 1 normal case, 2 gentle cases and 8 malicious cases about 3278 sectional images. 50% of dataset are measured for training and 50% as testing phases. The performance results of the three kernels are measured using the parameters sensitivity and specificity and obtained approximately 82.2% of accuracy.

A method for segmentation [19] of MRI, CT and Ultrasound images. For identification of cancer cell is done by studying the necessary features extracted for the two images. Ultrasound images as well to detect the validity of this system. We used feature selection as well by the use of PSO, Genetic Optimization and SVM algorithm giving an accuracy of about 89.5% with reduction in false positive.

The Lung Image Database Consortium dataset (LIDC) has been used for training and testing purpose in [20] using classification effected by means of an SVM. A Receiver Operating Characteristics (ROC) curve is used to analyze the performance of the system. Overall, the system has accuracy of 95.16%, sensitivity of 98.21% and specificity of 78.69%. SVM provides the accuracy of 92.5% in [21].

The achieved accuracy of SVM classifier using CPSO is 97.4% which is higher when compared to PSO, Genetic algorithm which yields an accuracy 89.5% and genetic optimization for feature selection and ANN for lung cancer classification which obtains 95.85% accuracy. While comparing the proposed method with the existing SVM-PSO and Genetic algorithm for accuracy, the proposed method yields a higher value than the existing methods.

The accuracy of system can be improved if training is performed by using a very large image database. The different basic image processing techniques are used for prediction purpose. The proposed system helps physician to extract he tumor region and evaluate whether the tumor is benign or malignant.

5 Conclusion

The proposed research is to detect the lung cancer in early stage with the help of an CT image and this method works by classifying the normal and abnormal CT lung images from the given set of inputs. The classification process is carried out by Support Vector Machine classifier. The simulation is undergone by using MATLAB. Initially, the images were preprocessed by median filter, followed by adaptive global thresholding segmentation. The segmented output is processed using GLCM method for further classifying the images. As per the outcomes of the simulation process the obtained accuracy of SVM is 97%. The accuracy can be further improved by using SVM with Chaos PSO which yields higher accuracy than the other optimization techniques.

The proposed Chaos PSO along with SVM is compared with two existing methods namely SVM with Genetic algorithm and SVM with PSO in which the proposed method performed well in terms of accuracy of 97.6% in which other two methods obtained only 89.3% and 95.85% respectively.

References

1) Arul C, Jaison B, K A. A discrete wavelet based feature extraction and hybrid classification technique for Microarray data analysis. *TSWJ*. 2014. Available from: https://www.hindawi.com/journals/tswj/2014/195470.

2) de Pinho Pinheiro CA, Nedjah N, de Macedo Mourelle L. Detection and classification of pulmonary nodules using deep learning and swarm intelligence. *Multimedia Tools and Applications*. 2020;79:15437–15465. Available from: https://dx.doi.org/10.1007/s11042-019-7473-z.

3) Mithra KS, Emmanuel WRS. Automatic Methods for Mycobacterium Detection on Stained Sputum Smear Images: a Survey. *Pattern Recognition and Image Analysis*. 2018;28(2):310–320. Available from: https://dx.doi.org/10.1134/s105466181802013x.

4) Babita R, Kumar A, Ravneet G, K. A Modified approach for Lung cancer detection using Bacterial Foraging Optimization Algorithm". *International journal of scientific research Engineering& Technology*. 2016;5:39–42. Available from: https://www.scribd.com/document/322070115.

5) Tajinder K, Neelakshi G. Classification of Lung Diseases using Particle Swarm Optimization". *International journal of advanced research in Electronics and Communication Engineering*. 2015;4:2440–2446. Available from: http://ijarece.org/?page_id=1829.

6) Al-Hegami AS, Bin-Ghodel ASH. A Particle Swarm based Approach for Classification of Cancer based on CT Scan. *International Journal of Computer Applications*. 2019;178(12):26–31. Available from: https://doi.org/10.5120/ijca2019918862.

7) K P, S P. A Novel Hybrid Segmentation Method with Particle Swarm Optimization and Fuzzy C-Mean Based On Partitioning the Image for Detecting Lung Cancer. Preprints . 2019. Available from: https://doi.org/10.20944/preprints201906.0195.v1.

8) Rostami M, Forouzandeh S, Berahmand K, Soltani M. Integration of multi-objective PSO based feature selection and node centrality for medical datasets. *Genomics*. 2020;112(6):4370–4384. Available from: https://dx.doi.org/10.1016/j.ygeno.2020.07.027.

9) Asuntha A, Neelima S, S A. PSO, Genetic optimization and SVM algorithm used for lung cancer detection. *Journal of Chemical and Pharmaceutical research*. 2016;p. 351–359. Available from: https://www.jocpr.com/abstract/pso-genetic-optimization-and-svm-algorithm-used-for-lung-cancer-detection-5133.html.

10) Yuvarani, Sakthi. Optimization techniques for Lung cancer analysis- A survey". *International journal of pure and applied Mathematics*. 2018;118:207–214. Available from: https://acadpubl.eu/jsi/2018-118-7-9/articles/8/28.

11) Priyanka, Kalsh A, Sharma K, Saini K. GLCM and its features". *International Journal of Advanced Research in Electronics and Communication Engineering*. 2015;4:1–8. Available from: http://ijarece.org/wp-content/uploads/2015/08/IJARECE-VOL-4-ISSUE-8-2180-2182.

12) Swati P, Chakkarwar VA. Classification of lung tumour using SVM". *International Journal of Computational Engineering Research*. 2012;2:1254–1257. Available from: http://www.ijceronline.com/papers/Vol2_issue5/N02512541257.

13) Liu P, Xie M, Bian J, Li H, Song L. A Hybrid PSO–SVM Model Based on Safety Risk Prediction for the Design Process in Metro Station Construction. *International Journal of Environmental Research and Public Health*. 2020;17(5):1714–1714. Available from: https://dx.doi.org/10.3390/ijerph17051714.

14) Simone L, Jakobovic D, Picek S. Classification of cancer data, analyzing gene expression data using a Fuzzy decision tree algorithm. Springer. 2018. Available from: https://link.springer.com/chapter/10.1007/978-3-319-65455-3_13.

15) Moghadas-Dastjerdi H, Ahmadzadeh M, Samani A. Towards computer based lung disease diagnosis using accurate lung air segmentation of CT images in exhalation and inhalation phases. *Expert Systems with Applications*. 2017;71:396–403. Available from: https://dx.doi.org/10.1016/j.eswa.2016.11.013.

16) Priyadharshini J, Shirdhinkar M. Detection of Lung cancer using SVM Classification". *International research journal of Engineering and Technology*. 2017;4:378–381. Available from: https://ieeexplore.ieee.org/abstract/document/8878774.

17) Mithra KS, Emmanuel S, R W. FHDT: Fuzzy and Hyco-entropy-based Decision Tree Classifier for Tuberculosis Diagnosis from Sputum Images. 2018. Available from: https://link.springer.com/article/10.1007/s12046-018-0878-y.

18) Parveen SS, Kavitha C. Classification of Lung Cancer Nodules using SVM Kernels. *International Journal of Computer Applications*. 2014;95(25):25–28. Available from: https://dx.doi.org/10.5120/16751-7013.

19) Asuntha A, Brindha A, Indirani S, Srinivasan A. Lung cancer detection Using SVM algorithm and optimization techniques. *J Chem Pharm Sci*. 2016;9(4):3198–3203.

20) Mohamad MS, Omatu S, Deris S, Yoshioka M. A Constraint and Rule in an Enhancement of Binary Particle Swarm Optimization to Select Informative Genes for Cancer Classification. In: Pacific-Asia Conference on Knowledge Discovery and Data Mining. Springer. 2013;p. 168–178.

21) Tidke PS, Chakkarwar VA. Classification of lung tumor using SVM. *International Journal Of Computational Engineering Research*. 2012;2(5).