IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
PP 41-46

Enhanced Edge Detection Based Support Vector Machine towards the Prediction of Lung Cancer

P.Dhanalakshmi[1], Dr.G.Sathyavathy[2]
[1](Research Scholar, Department of Computer Science, Sri Ramakrishna College of Arts and science for Women, Coimbatore-46 Email :dhanalakshmi.8030@gmail.com)
[1](Assistant Professor, Department of Computer Science, Sri Ramakrishna College of Arts and science for Women, Coimbatore-46. Email :satyacs@srcw.org)

Abstract: *Image Processing is a process of performing operations on an image with an intention of getting an improved image. It is also used to mining to get the hidden information in an image. Edge detection is the method of finding the boundaries of object in the image. Recently image processing techniques are being used in the prediction of diseases, where it helps to increase the classification accuracy. In this paper a novel edge detection method based on least square support vector machines is proposed to detect edges with the target of predicting the lung cancer. This paper uses sensitivity, specificity, accuracy and F1-measure as the metric tomeasure the performance. The results show that the proposed methods outperforms than other algorithm.*
Keywords: *Edge Detection, classification, SVM, image processing, accuracy.*

I. INTRODUCTION:

Image Processing is a process of performing operations on an image with an intention of getting an improved image. It is also used to mining to get the hidden information in an image. Image processing is considered as a kind of signal processing, where the input is given an image and output might be image or characteristics/features related to the image. Currently, image processing is getting developed quickly. It frames center research territory inside computer science and software engineering disciplines as well.
The three basic steps of image processing are:

- Import the image using tools.
- Analyze and manipulate the image.
- Export the image based on alteration or analysis.

Medical image classification has become more and more important for disease diagnosis and treatment. Through medical image feature extraction and classification, medical images can be distinguished as normal or abnormal. Besides, when doctors visit medical imaging library, they often need to get a certain type of image for narrowing the scope of search, which also requires medical image classification. The conventional medical image classification algorithms are usually based on images' basic features, such as color, texture and shape features. However, these low-level features cannot reflect the underlying information of images well, thus always leads to the Semantic gap problem between low-level features and high-level semantic information. Therefore, the semantic gap problem is one of the biggest challenges for medical image classification.

The need for edge detection in image processing is getting increased in a significant manner. Edge detection is connected to the research areas related to sensing and is utilized to save the critical basic properties, recognition of pattern, image classification and so on. Though there exist multiple proposals in this thrust area, still it doesn't meet the research objective of detecting the edge with increased accuracy. For this problem, this research work utilizes the concept of least square support vector machines to detect edges with the target of predicting the lung cancer.

The rest of the paper is organized as follows: A review of previous related work is presented in Section 2. In Section 3, proposed method is presented with contourlet HMT and LS-SVM Classifier. In Section 4, discusses the dataset and performance metrics used. In Section 5, we describe results by comparing with multi-scale non-negative sparse coding [Ruijie Zhang et al.,2017]. Finally, in Section 6, we draw conclusions with future enhancement.

II. LITERATURE REVIEW:

J.Guo et al., 2017 made an analysis on segmentation framework and proposed a fully automatic framework to segment Branch retinal artery occlusion (BRAO) regions based on 3D spectral-domain optical coherence tomography (SD-OCT) images. Kumar et al.,2017 proposed a method for classifying medical images that uses an ensemble of different convolutional neural network (CNN) architectures, where the CNNs were a state-of-the-art image classification technique that learns the optimal image features for a given classification task. S.Roychowdhury and M.Bihis, 2016 proposed a workflow for medical researchers/practitioners by automating generalized workflow which focuses on data inferencing rather than dealing with the intricate details of predictive modeling, such as feature and model selection. I.Diamant et al.,2017 presented a novel variant of the bag-of-visual-words (BoVW) method for automated medical image classification, where it improves the BoVW model by learning a task-driven dictionary of the matched visual words as per the task using a mutual information-based criterion.M. Toutain et al.,2016 used the framework of partial difference equations on weighted graphs as a methodology to address the problem of computer-aided cytology, where it was introduced to perform image smoothing and filtering, this framework has been extended to address segmentation and semi-supervised clustering of any discrete domain that can be represented by a graph of arbitrary topology. C.Kuo et al., 2014 proposed an automatic method for selecting the radial basis function (RBF) parameter (i.e., σ) for a support vector machine (SVM), and also authors have proposed a kernel-based feature selection method with a criterion that was an integration of the previous work and the linear combination of features.

A.Olaode et al.,2014 studied the problems emerge in image categorization and proposed an unsupervised image categorization model in which the semantic content of images were discovered by utilizing Probabilistic Latent Semantic Analysis by clustering the images into unique groups based on semantic content similarities using K-means algorithm.Mathan Kumar &R.PushpaLakshmi, 2018 proposed a approach for image retrieval based on classification, Multiple Kernel SIFT (MKSIFT) that extracts the features from the pre-processed input image, where it utilizes the steps of SIFT to extract the feature points. Khatami et al.,2017 proposed a deep convolutional neural network was proposed to classify the information of radiology images, but millions of data were needed by deep networks whereas the shortage of balanced large datasets in medical domain motivated the authors to trust on even the second prediction category rather than just the best one. Hence the best predicted categories were considered for a query test, followed by a similarity-based search technique. Kleyko et al.,2017proposed a model for medical image classification based on the usage modalities of principles of hyper-dimensional computing and reservoir computing. It was demonstrated that the highest classification accuracy of the proposed method was better than the existing.

III. PROPOSED WORK:

3.1. Contourlet HMT

The primary disadvantage of wavelets is its confined capacity to catch directional data. Do et al., 2006 [18] proposed a transformation of contourletwhich consolidates a Laplacian Pyramid (LP) with a Directional Filter Bank (DFB). In that LP was utilized to occupy the pixel breaks and connection pixelbreaks straight through a DFB. DFB breaks down each accurateLP subordinate band into two directional subordinate bands. Contourlet HMT effectively enhances edges in bidirectional area which proficiently portraying smooth shapes by directional and expanded basis components. Each coefficient in the common measure has four youngsters in the following higher subordinate band, where a new 4 subordinate bands will rise. In this way, the contourlet coefficient can be denoted in a quadtree structure.Individual contourlet coefficient can be displayed by a combination of distribution by Gaussian as being in either in high state or low state. We acquire coefficients which are not fine and coefficients which are well fine.

3.2. Least Square Support Vector Machine (LSSVM)

SVM is benchmark algorithm for machine-learning that began in statistical learning theory [J.A.K. Suykens and J. Vandewalle, 1999]. SVM receives a supplementary risk minimization rule that can conquer the contradiction between underfitting and overfitting. Also, it has a solid improvement to diminish the impact of noisewith the target of increasing the accuracy.SVMoverwhelms the issues of measuring through a nonlineartransform and matrix kernel function, abstaining from including computational multifaceted nature for a high dimensional space. LSSVM [T. Van Gestel et al.,2002] Endeavour's to deal with the arrangement of linear conditions as opposed to curved quadratic programming, which is performed by utilizing standard SVM to diminish computational unpredictability and increment arrangement speed. LSSVM Endeavour's to diminish the least square error on samples test while in the meantime expanding the edge between two classes. Broad experimental correlations exhibit that LSSVM gets better execution on an assortment of classification issues.
Consider an arrangement of N training cases; LSSVM takes the accompanying structure:

$$y = \frac{\text{[illegible]}}{\text{[illegible]}} \qquad (1)$$

where w is the weight, b is the bias and the nonlinear mapping $\varphi(x)$ plots the input data into a high feature space. LSSVM has the capacity to solvetheoptimization issues thorough the following:

$$Min\left[\frac{\text{[illegible]}}{\text{[illegible]}}w^T w + \frac{\text{[illegible]}}{\text{[illegible]}}\sum (o_i - y_i)\right] \qquad (2)$$

where C is the deal parameter between the terms in four sides, subject to

$$o_i = \frac{\text{[illegible]}}{\text{[illegible]}} \qquad (3)$$

By utilizing theLagrangian function [20], it's possible to derive the equations as:

$$w^T = \qquad (4)$$

$$o = \qquad (5)$$

$$\alpha_i = \sum C(y_j - o_i) \qquad (6)$$

$$O_i = \qquad (7)$$

Here, $\alpha = (\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_i)^T$, where$\alpha_i \in R$are Lagrangian. In this manner, Mercer's condition [N. Aronszajn, 2005] can be utilized as the kernel function. We have utilized a function of Gaussian, which will be very useful to perform tedious nonlinear connection between the samples of input and output. It is chosen as the kernel function, whereit can be denoted as

$$k(x_i, x_j) = \exp\left(\right) \times \varphi(x_i) \qquad (8)$$

So the solution of $f(x)$ is the following:

$$f(x) = sgn\left(\right)$$

$$f(x) == sgn\left(\right) \qquad (9)$$

3.3. Edge detection scheme using LSSVM classification

In this part, we enhance the contourlet HMT coefficients for the detection of unique edges and diminishing noise in boisterous pictures by using BayesShrink in a contourlet HMT space. Contourlet HMT is a multidirectional and multi-scale advancement that has improved bidirectional recurrence localization. By making use of using the LS-SVM, it is recommended that classifying contourlet HMT coefficients into two classes by comparing to hard and soft regions. In this manner, contourlet HMT coefficients at each scale comparing to surface and edges can be utilized for those to smooth regions, and distinctive edge esteems can be set in the BayesShrink limit, denoising and using angles to identify edges. This proposed plan can conquer the bidirectional confinement of the existing techniques and catch the bidirectional edges of a picture. Additionally, the edge data in various coefficients and directions can coordinate with each other; it is intertwined with an edge combination weighted in normal.

3.4 Classification of contourlet HMT coefficients using LSSVM

This research work considers the information of the input image as $f_{i,j}(XY)$, where the value of i and j ranges from 1 to N. N is considered as percentage of noise added to actual image, which ought to be zero-mean white added substance Gaussian noise with deviation 0.1. As a second level, contourlet HMT decomposition is carried on the info noisy image for J level; next, it is estimated to get a better coefficient A_i and coarser coefficientsD_i, where $i = 1,2,\ldots,J$ signifies the decomposition level. Finally, the subordinate point of contourlet HMT does not contain any confined coefficients since it has hign dimensional normality. Consequently, element vectors are build with the objective of contourlet HMT coefficients holding coarser coefficients $D_i, i = 1,2,\ldots,J$ andD_i^1 . We utilize a mapping of binary$I_i[x,y]$to assistin choosing coefficients and ought to be discarded for most extreme reconstructedimage quality by considering the coefficients of HMT (C_{xy}). In addition, coefficients of $N_i$contourlet HMT having maximum supporting values are chosen as criteria1 (C1) feature vector and those having the minimum supporting value are chosen as criteria2 (C2) feature vector.Engage with LSSVM method. C1 and C2 feature vectors are considered for training, and O1 and O2 are assumed as the training objective. At that specific pixel, we get the preparation tests $\phi = \{C1, C2, O1, O2\}$.The coefficients of classifying contourlet HMT in coarser coefficients are considered in two classes which are noise coefficient (0) and edge coefficient (1).

3.5Denoising of noisy contourlet HMT coefficients

Elimination of noisy coefficient begins with utilization of BayesShrinkrule [Shi Y et al.,2017] using a Bayesian numerical structure for images to deliver coefficient subordinate limits, which are about ideal for comfortablethresholding [Z. Wang et al.,2004]. Thereshold ofBayesShrink edge $(TB)$for every noisy coefficientwhich is classified is depicted as below:

$$TB = \qquad (10)$$

where$\hat{\sigma}^2$is the change in Gaussian noise, and $\hat{\sigma}_x$ is the noise's standard deviation. The change Gaussian noise is evaluated as the median deviation after the conversion.

$$\hat{\sigma} = \quad , Y_{ij} \in D_i \qquad (11)$$

Approximate the standard deviation of the image is

$$\hat{\sigma}_x = \sqrt{\max(\hat{\sigma}_Y^2 - \hat{\sigma}^2} \qquad (12)$$

where

$$\hat{\sigma}_Y^2 = \sum Y_{ij}^2 \qquad (13)$$

Additionally, $\hat{\sigma}_V^2$ is aapproximation of the fluctuation in the perceptions; that is the number of the contourlet HMT coefficients on the subordinate point J. In this way, denoising limit for each noise coefficient's class are predicted.

3.6Edge Detection

Once after eliminating the noise from the class denoisedcoefficients, the edges can be detected by utilizing the gradients. The difference of angle between the gradient coefficients is calculated using:

$$Gradient\ (G) = \sqrt{|R_1|^2 + |R_2|^2} \qquad (14)$$

where$R_1\ and R_2$ denotes to halfway subsidiaries of the vertical and horizontal directions.

.

Itacquires a most extreme G and determined the limit along with the direction of G. On the off chance that a state of every segment fulfils two conditions:

- ✓ Greceives the maximum value in local.
- ✓ $G \geq T$, subordinate points in this area are the edge points.

Canny algorithm is utilized to find the edges by horizontal, vertical and diagonal, where the canny algorithm uses quad filters for edge detection.

3.7Integration of edge information

Weighted average technique [Guang-Xin LI et al.,2005] is utilized to integrate better and worst coefficients so far estimated. The entire pixels of the edge are not detected in the process of edge detection. The maximum part of the pixels are detected, nevertheless, still there may be couple of directional edge pixels not detected. In this way, information of the edge is detected from different better coefficients and it has an unequivocal level correspondingly. One edge image is formed by utilizing the edges. At long last, we movein the inverse method utilizing contourlet HMT coefficients to get the yield the final image.

IV. DATASET AND PERFORMANCE METRICS:

4.1 ADL Dataset

ADL dataset [Liu D et al.,2016] is used in this research work. ADL dataset contains bovine histopathology imagesof lung, which was kindly published by Animal Diagnostics Lab, Penn-sylvania State University. 50 randomly selected images are used for this research work with the target of classifying the affected lungs related to cancer.

4.2 Performance Metrics

The official performance metric mean classification accuracy, specificity, sensitivity and F-Measure [Tang J et al.,2012] are used to evaluate theperformance of these algorithms. Their definitions are as follows.

$$Sensitivity\ = \frac{[illegible]}{[illegible]} \times 100$$

$$Specificity\ = \frac{[illegible]}{[illegible]} \times 100$$

$$Accuracy\ = \frac{[illegible]}{[illegible]} \times 100$$

$$F - Measure\ = \frac{[illegible]}{[illegible]} \times 100$$

Moreover, the receiver operating characteristic (ROC) curve isemployed. Besides, to make the experimental results more convincing, we evaluate the dataset using 10-fold cross-validation. That isthe dataset is split into 10 approximately equally sized partitions.All the algorithms run 10 times, each time using nine partitions asthe training set and the other partition as the test set. The averageof the repeated experiments is more convincing.

V. RESULTS AND DISCUSSION

From the figure Fig.1 and Fig.2 it is evident that theresult of EEDSVM is consistent to those of MSNNSC. The sensitivity, specificity, classification accuracy, and F-Measure are 95.55, 60, 92, and 95.5 respectively. EEDSVM significantly outperforms in predicting the lung cancer than other algorithm MSNNSC. Table values of Fig.1 and Fig.2 is shown in Table 1.

Table 1: Classification results of MSNNSC and EEDSVM on ADL dataset (%).

Method	Specificity	Sensitivity	Accuracy	F-Measure
MSNNSC	93.18	50	88	93
EEDSVM	95.55	60	92	95.5

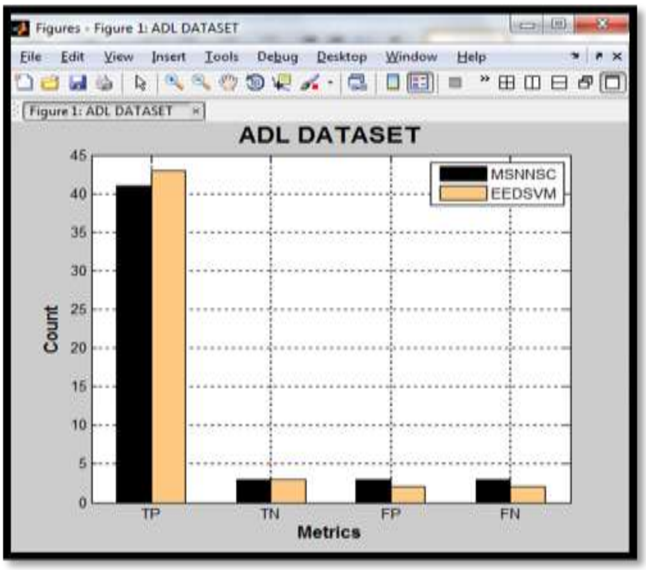

Fig.1 TP TN FP FN Analysis

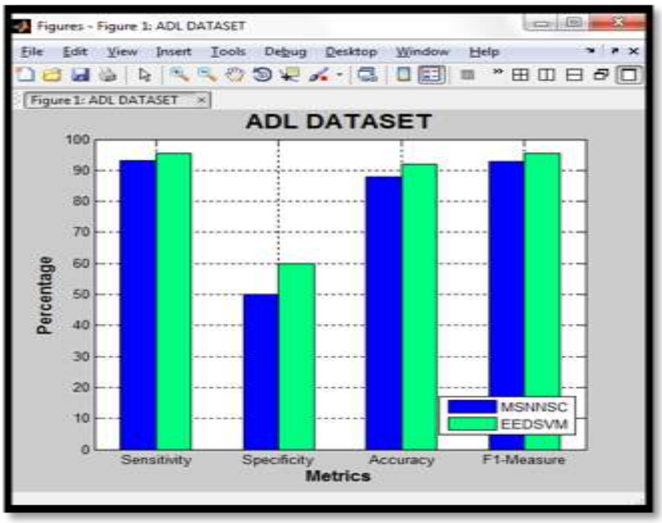

Fig.2 Classification Result

VI. CONCLUSION:

In this paper a novel edge detection method based on least square support vector machines is proposed to detect edges with the target of predicting the lung cancer. Experimental resultsdemonstrate that our algorithm performs superior to other related algorithm and which can efficiently predict the lung cancer.In the future, we will study medical image classification based onsome other classification algorithms with the concept of edge detection.

REFERENCES:

[1]. A. Olaode, G. Naghdy and C. A. Todd, "Unsupervised Image Classification by Probabilistic Latent Semantic Analysis for the Annotation of Images," 2014 International Conference on Digital Image Computing: Techniques and Applications (DICTA), Wollongong, NSW, 2014, pp. 1-8.
[2]. C. Kuo, H. H. Ho, C. H. Li, C. C. Hung and J. S. Taur, "A Kernel-Based Feature Selection Method for SVM With RBF Kernel for Hyperspectral Image Classification," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 7, no. 1, pp. 317-326, Jan. 2014.
[3]. Diamant, E. Klang, M. Amitai, E. Konen, J. Goldberger and H. Greenspan, "Task-Driven Dictionary Learning Based on Mutual Information for Medical Image Classification," in IEEE Transactions on Biomedical Engineering, vol. 64, no. 6, pp. 1380-1392, June 2017.
[4]. Guang-Xin LI, Ke Wang, Li-Bao Zhang, The fast algorithm of weighted multi-resolution image fusion, China J. Image Graph. 10 (12) (2005)1529–1536.
[5]. J. Guo, W. Zhu, F. Shi, D. Xiang, H. Chen and X. Chen, "A Framework for Classification and Segmentation of Branch Retinal Artery Occlusion in SD-OCT," in IEEE Transactions on Image Processing, vol. 26, no. 7, pp. 3518-3527, July 2017.
[6]. J.A.K. Suykens, J. Vandewalle, Least squares support vector machine classifiers, Neural Process. Lett. 9 (3) (1999) 293–300.

[7]. Khatami, M. Babaie, A. Khosravi, H. R. Tizhoosh, S. M. Salaken and S. Nahavandi, "A deep-structural medical image classification for a Radon-based image retrieval," 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE), Windsor, ON, 2017, pp. 1-4.

[8]. Kleyko, S. Khan, E. Osipov and S. P. Yong, "Modality classification of medical images with distributed representations based on cellular automata reservoir computing," 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), Melbourne, VIC, 2017, pp. 1053-1056.

[9]. Kumar, J. Kim, D. Lyndon, M. Fulham and D. Feng, "An Ensemble of Fine-Tuned Convolutional Neural Networks for Medical Image Classification," in IEEE Journal of Biomedical and Health Informatics, vol. 21, no. 1, pp. 31-40, Jan. 2017.

[10]. Liu D, Wang S, Huang D, Deng G, Zeng F, Chen H. Medical image classificationusing spatial adjacent histogram based on adaptive local binary patterns[J].ComputBiol Med 2016;72:185–200.

[11]. M. Toutain, A. Elmoataz, X. Desquesnes and J. H. Pruvot, "A Unified Geometric Model for Virtual Slide Image Processing and Classification," in IEEE Journal of Selected Topics in Signal Processing, vol. 10, no. 1, pp. 151-160, Feb. 2016.

[12]. Mathan Kumar & R. PushpaLakshmi, Multiple kernel scale invariant feature transform and cross indexing for image search and retrieval, The Imaging Science Journal, volume 66, issue 2, 2018, Pages:84-97.

[13]. N. Aronszajn, Theory of reproducing kernels, Trans. Am. Math. Soc. 686 (1950)337–404.[23] E.J. Balster, Y.F. Zheng, R.L. Ewing, Feature-based wavelet shrinkage algorithm for image denoising, IEEE Trans. Image Process. 14 (12) (2005) 2024–2039.

[14]. S. Roychowdhury and M. Bihis, "AG-MIC: Azure-Based Generalized Flow for Medical Image Classification," in IEEE Access, vol. 4, pp. 5243-5257, 2016.

[15]. T. Van Gestel, J. Suykens, G. Lanckriet, A. Lambrechts, B. De Moor, J. Vandewalle, Bayesian framework for least squares support vector machine classifiers, Gaus-sian processes and kernel fisher discriminant analysis, Neural Comput. 15 (5)(2002) 1115–1148.

[16]. Tang J, Zha ZJ, Tao D, Chua TS. Semantic-gap-oriented active learning formultilabel image annotation[J]. IEEE Trans Image Process2012;21(4):2354–60.

[17]. Z. Wang, A.C. Bovik, H.R. Sheikh, et al., Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (2004) 600–612.