ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

A Hybrid Approach Based on Texture Feature Analysis in CT Images

Pooja M, Veena Devi

M. Tech Student (Communication Systems), Department of ECE, R.V College of Engineering, Mysore Road, Bengaluru, India

Assistant Professor, Department of ECE, R.V College of Engineering, Mysore Road, Bengaluru, India

ABSTRACT: Lung cancer is the supreme prevalent cancer and is the leading cause of cancer deaths worldwide. The global survival rate of lung cancer patients is merely 14%. Lives of cancer patients can be saved if the cancer is detected in the initial stages. They are many types of formats of medical data available one among them is Computed Tomography (CT) which is the preferred imaging modality in cancer detection with improved diagnostic accuracy. Even thoughCT is advantageous over other modalities, visual inspection of these images may be an error prone task, as it is difficult to distinguish between background tissues and lung nodules and subject to inter and intra observer variability. Therefore, computational systems are essential to assist radiologists in the elucidation of images and accurate diagnosis. This paper aims at developing a hybrid methodology for automatic detection of lung cancer with greater enhanced accuracy. Image pre-processing methods such as grey level concurrency matrix (GLCM) fused with Principle Component Analysis(PCA) and Support Vector Machine(SVM) were performed to remove over-fitting and to increase the classification accuracy. Lung region of interest (ROI) were extracted from images using morphological operators. GLCM statistical texture features were preferred as they extract more texture information from the cancer regions than the visual assessment. The proposed method was carried out using images of lung cancer patients and implemented using MATLAB. The performance of the proposed methodology is shown.The proposed method provides better classification and cancer detection.

KEYWORDS: gray scale, ROI, GLCM, SVM, PCA, Fuzzy c-means

I. INTRODUCTION

Disease after heart diseases in the world Cancer is the second most common disease. The overall 5-year survival rate of lung cancer patients is only 14%, but detecting at an early stage increases the survival rate as high as 60 – 70%. This proposed system is developed to detect lung cancer automatically as interpretation of a larger image dataset by radiologists is time consuming and laborious which may lead to misclassification.The fuzzy c means was a novel concept for the classification of normality and abnormality in cancer detection. However this source of classification will not remember the previous feature values thus making a way to over-fitting thus reducing the classification accuracy. In order to avoid this we propose a hybrid approach that would enhance the classification procedure by fusing the SVM and principal component analysis for better classification thus avoiding over-fitting. SVM is mainly considered because SVM deals better than as compared to any other classifiers for large database. Initially this SVM-PCA classifier is trained for the set of images with lung cancer whose features are calculated using GLCM to train the SVM. Medical images will have more texture features which are extracted using GLCM; this is a textures feature extraction method with best efficiency. Another approach would be using the concept of principal component analysis for classification. Once the training process is done query image is taken. Depending upon the texture feature of the query image classifier will do the classification and detects the lung cancer. The automatic detection of lung is necessary as large amount of human errors will occur during classification of lung cancer by radiologist [1-4].

II. PROPOSED SYSTEM

In this paper, we will describe in detail about the construction of GLCM matrix along various possible directions and extracting their features. We have also commented about the preferred direction of GLCM processing and the

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

importance of SVM and PCA Further, we showed that the use of a PCA-SVM is far more optimal than the naïve classifier for classifying the lung cancer images.

A. Gray Level Concurrency Matrix

The GLCM, which is a square matrix, can reveal certain properties about the spatial distribution of the gray-levels in the texture image. It was defined by Haralick et al.[5-9] in 1973. It shows how often a pixel value known as the reference pixel with the specific intensity value occurs in a specific relationship to a pixel value known as the neighbor pixel with the intensity value j. So, each element of(i,j) of the matrix is the number of occurrences of the pair of pixel with value i and a pixel with value j which are at a distance d relative to each other. The spatial relationship between two neighboring pixels can be specified in many ways with different offsets and angles, the default one being between a pixel and its immediate neighbor to its right. In the present work, four possible spatial relationships $0^{\circ}, 45^{\circ}, 90^{\circ}, 135^{\circ}$ were specified and implemented.

Mathematically, for a given image I of size$K \times K$, the elements of a $G \times G$ gray-level co-occurrence matrix MCO for a displacement vector $d = (dx, dy)$ is defined as

$$M_{ij} = \sum_{x=1}^{k} \sum_{y=1}^{k} \begin{cases} 1, if\ I(x,y) = i\ and\ I(x+d_x, y+d_y) = j \\ 0 \end{cases} \quad (1)$$

Each element of the GLCM is the number of times that two pixels with gray tone i *and* j are neighborhood in distance d and direction θ. Hence, these matrices are symmetric in nature and the co-occurring pairs obtained by choosing θ equal to 0° would be similar to those obtained by choosing θ equal to 180°. This concept extends to $45^{\circ}, 90^{\circ}$ *and* 135° as well. With all these considerations, the GLCM matrix is calculated for each of the four possible angles.

B. PCA + SVM

SVM belongs to kernel methods. Kernel algorithms map data from an original space into a higher dimensional feature space using non-linear mapping. An original algorithm from the original space is used in the feature space. Although the high-dimensional space increases the difficulty of the problem (curse of dimensionality), a trick for computing the scalar products in the feature space exists. Computation of the scalar product between two feature space vectors can be done using kernel functions. But our scope of work would fuse

Hence the main intention of PCA would obviously be the dimensionality reduction of the high dimensional data while reaching the robustness. It transfers set of the identical data into the set of uncorrelated dataset. This usage helps in the redundancy elimination of the data in the image. This way of approximation is applied to the multispectral image with the help of Eigen values and eigen vectors of the corresponding matrix or correlation matrix in the multispectral satellite images. These Eigen vectors would be like a kernel which would convolve on the image and slides over it. Next, after convoluting individual bands are extracted which are synonymously called as the principle components.This principle component is called the first component which is replaced with the Image which gives the new first principle component. The new component thus obtained along with other principal components obtained this way is transformed with the assistance of inverse PCA to form the fused image. The PCA technique is often a robust and simple technique for fusing images and also it may not be the best approach to fuse for high resolution images and low resolution multispectral images. The reason is that it may distort multispectral characteristics of the data [11].To explain it more precisely consider a set of M*N features formed in terms of matrix with class labels +1 and -1 in an binary SVM which contains a matrix containing the features of training images.

The other variant attempted was to apply PCA to the training data before giving it as the input to the SVM routine. Since each image has N features, the covariance matrix construction in MATLAB is not possible due to memory limitations. So, the training set is downsized and converted to the dimensions of (N-a) samples. The application of PCA has its own advantage and disadvantage. PCA aims to represent the data effectively along the directions of maximum scatter. The good thingis, the original image will have a lot of redundant information and these points will be discarded by PCA. But, the undesirable thing is that, the two classes may be well separated in their original dimension, but the PCA in the event of projecting the samples in the scatter directions, may make the samples from two classes to overlap and thereby present a difficult mixture of training samples to the SVM routine.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

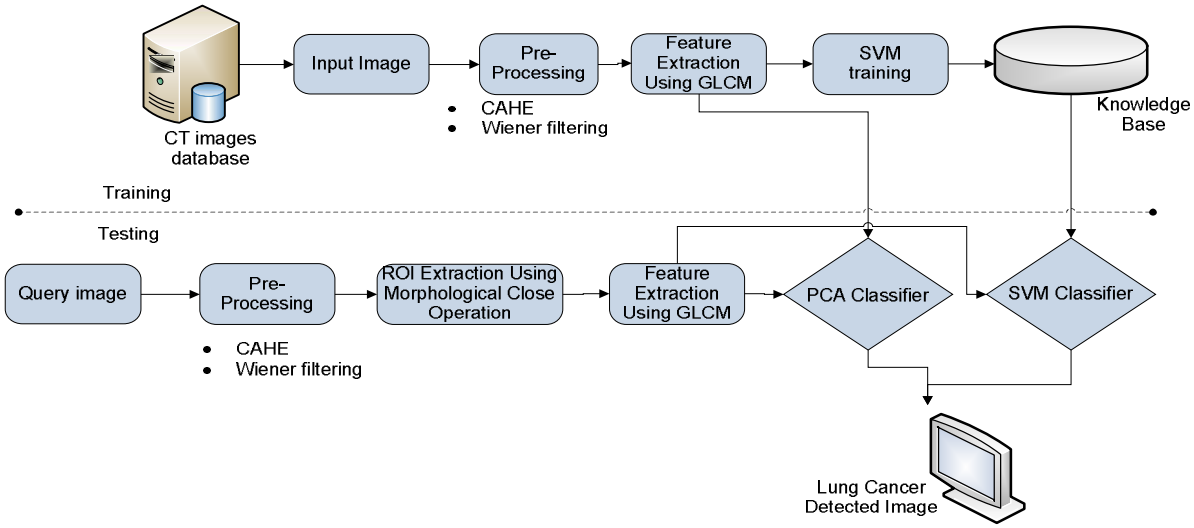


Figure 1A: Block diagram of the PCA-SVM system

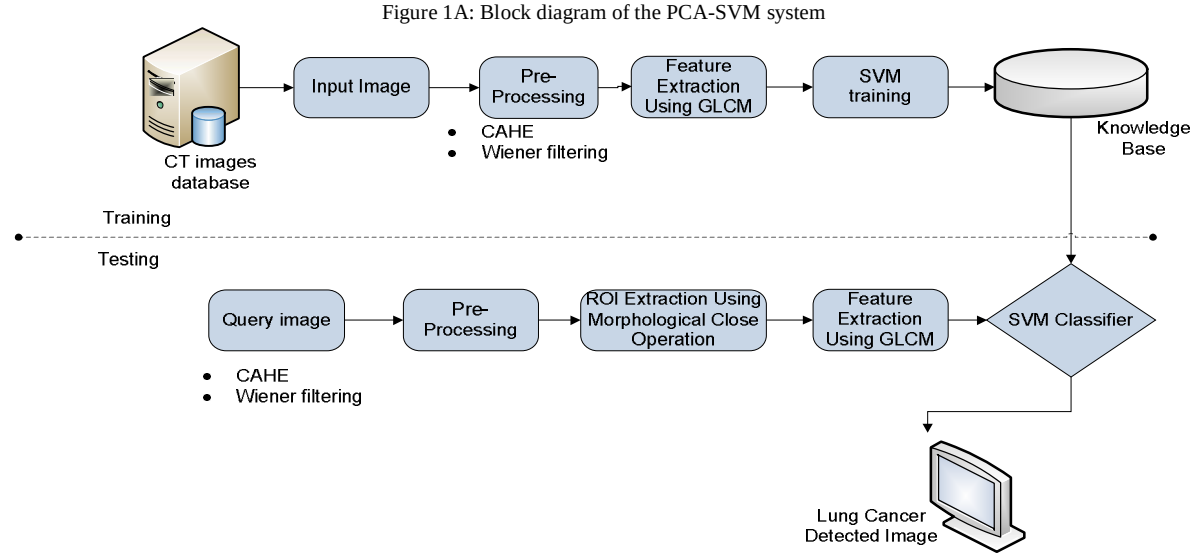


Figure 1b: Block diagram of the SVM system

C. SUPPORT VECTOR MACHINE (SVM)

Support vector machines (SVMs) are a set of supervised learning methods used or classification, regression and outlier's detection.SVMs (Support Vector Machines) are a useful technique for data mining. Although SVM is considered easier and convenient to use than Neural Networks, users not familiar with it often get unsatisfactory results at first.

a. SVM TRAINING

Our search for training (Extracting cancer) the knowledge base cancer detection is based on the insight that features represent the core of the classification on profiles. Following the occurrence of tumor region led us to many examples of cancer.

In contrast, there were other more serious cases where the tumor seemed particularly unclassifiable, for most of these cases we use support vector machines as a trivial method. Hence features from the GLCM are trained in the svm to form a structure with features and the relevant parameters like hyper-plane, bias line, group etc.

b. SVM PREDICTION

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

Although users do not need to understand the underlying theory behind SVM, we introduce the basics necessary for explaining our procedure. A prediction task usually involves separating data into training and testing sets. Each instance in the training set contains the class labels and several the features or observed variables). The goal of SVM is to produce a model (based on the training data) which predicts the target values of the test data given only the test data attributes.

Theoretically it is quite cumbersome to understand SVM without the proper fundamentals about the prediction of the events. Hence we have explained the fundamental aspects in the introduction.The linear SVM model for any given population of a dependent and independent variable can be formulated as:

$$y = \beta_0 + \beta_1 x + \varepsilon \quad (2)$$

Where

y - Dependent variable, x - independent variable, β_0 - the intercept of the line, β_1- the slope of the line, ε - statistical noise or error

The linear SVM model provides how well the data points lie on the line. The line should be drawn in such a way that almost all data points should lay on this line. By looking at the regression line and arrangement of data points on the line specifies how well the regression model fits into the dataset.

The slope of a line is a number that describes both the direction and the steepness of the line. If we find that the slope of the regression line is significantly different from zero, we will conclude that there is a significant relationship between the independent and dependent variables.

The y-intercept is the place where the regression line $y = \beta_0 + \beta_1 x + \varepsilon$ crosses the y-axis (where x = 0), and is denoted by β_0. The slope of a regression line is used with a t-statistic to test the significance of a linear relationship between x and y.

In regression analysis, the (Gaussian) radial basis function kernel, or RBF kernel, is a popular kernel function used in various learning algorithms. In particular, it is commonly used in vector classification. The widely used Radial Basis Function (RBF) kernel is known to perform well on a large variety of problems. RBF network can be used to find a set of weights for a curve fitting problem. The weights are in higher dimensional space than the original data [10].

Alternatively, it could also be implemented using the adjustable parameter sigma plays a major role in the performance of the kernel, and should be carefully tuned to the problem at hand. If overestimated, the exponential will behave almost linearly and the higher-dimensional projection will start to lose its non-linear power. In the other hand, if underestimated, the function will lack regularization and the decision boundary will be highly sensitive to noise in training data.

There are three basic classes of radial basis functions.

a. $\phi(r) = (r^2 + c^2)^{1/2}$, for some c>0 and r∈ R

b. $\phi(r) = \frac{1}{\phi(r)=(r^2+c^2)^{1/2}}$, for some c>0 and r∈ R

c. $k(x,y) = \exp\left(-\frac{\|x-y\|^2}{2\sigma^2}\right)$ (3)

d. $k(x,y) = \exp\left(-\gamma\|x-y\|^2\right)$ (4)

Gaussian functions are probably the most used. In general, the selection depends on the application. The above analogies prescribe to describe the data of a kernel trick used as a part of the regression and non-linear regression analysis. In the context of this our system would predominantly exploit the kernel on the regression systems to quantify the prediction analysis and their error rate. The error rate is a measure of the accuracy of predictions. Recall that the regression line is the line that minimizes the sum of squared deviations of prediction. In our context we embody the non linear regression line patterns in order to differentiate the unethical words based on the database for the prediction.

D. FLOWCHART OF THE PROPOSED SYSTEM

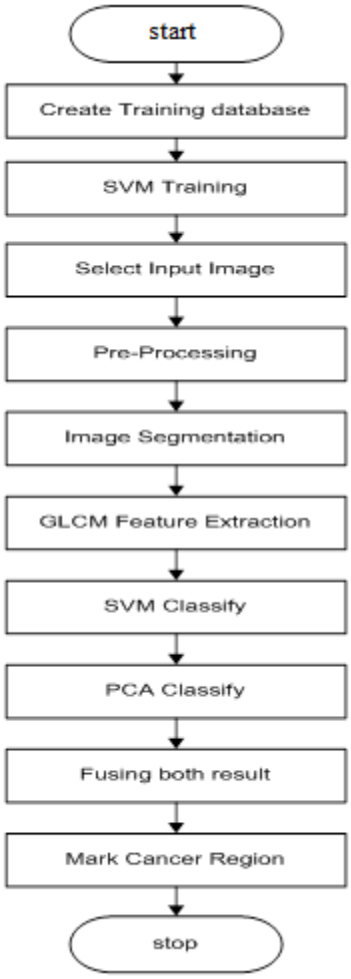
The data flow diagram illustrates the implementation of the data flow diagram of a lung cancer detection system under various phases of traversal. Initially image segmentation operations are performed to form a well component connected image and to excel the boundaries of the face. The next cycle of GLCM will detach the phase at different textures by establishing the values of statistical features, thus dividing into SVM and PCA classifier to train the features. Then, fuse them together to perform classification to mark the severity of the cancer into normal and abnormal.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016



III. RESULTS AND DISCUSSIONS

Experimental result analysis is the process of analyzing the output of experiments carried on the system. The verified application has been experimented with various inputs and the results are analyzed for its performance and accuracy. The detail analysis of the experiments and results obtained is explained below.

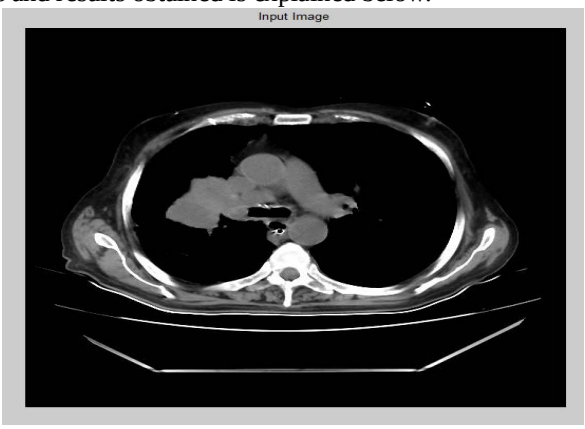


Figure 2: input image

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

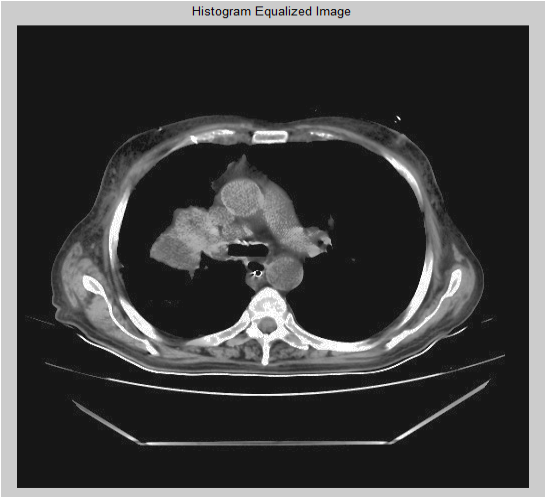


Figure 3: histogram equalized image

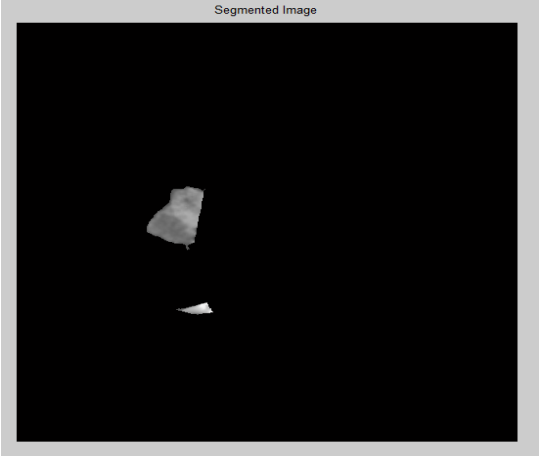


Figure 4: segmented image

Figure 5: lung cancer detection image

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

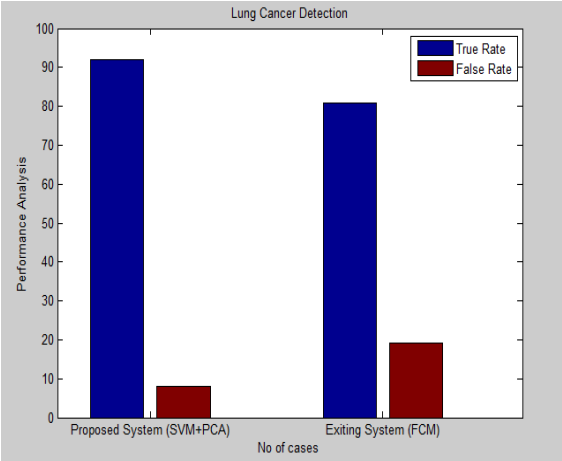


Figure 6: Performance Analysis Graph

IV. CONCLUSION

From the present work, it can be concluded that gray level co-occurrence matrix finds to be best when simulated with SVM and PCA. This conclusion is justified, as a pixel is more likely to beover-fitted to a closely located pixel than the one located far away. θ = (0°, and 90°) by fusing together SVM-PCA shows comparable results for GLCMprocessing with fusing together PCA and SVM

ACKNOWLEDGMENT

The authors are very grateful to VEENA DEVI S V ,Asst. Professor Dept. of ECE,RVCE for interesting discussions regarding this work.

REFERENCES

[1] J. Sklansky, “Image segmentation and feature extraction”, IEEE transaction on systems, Man and Cybernetics, 8(4):237-247, 1978
[2] A. Rosenfeld, Editor: Digital Picture Analysis, Springer Verlag, Berlin 1976
[3] R.M. Haralick, K. Shanmugam, I. Dinstein, “Textural Features for Image Classification”, IEEE Trans. on Systems, Man and Cybernetics (1973)610 – 621.
[4] T. Ojala and M. Pietikäinen, “Texture Classification, Machine Vision and Media Processing Unit”, University of Oulu, Finland. Available at
[5] A. Eleyan and H. Demirel, “Co-occurrence matrix and its statistical features as a new approach for face recognition”, Turk J ElecEng& Comp Sci., Vol.19, No.1, pp. 97-107, 2011
[6] R.A.Braga, C.M.B.Nobre, A.G. Costa, T.Safadi and F.M. da Costa, “Evaluation of activity through dynamic laser speckle using the absolute value of the differences”, Optics Communication, Elsevier, Vol. 284, pp. 646–650, 2011
[7] I. Sobel, “An Isotropic 3×3 Gradient Operator”, Machine Vision for Three Dimensional Scenes, Freeman, H., Academic Press, NY, 376-379, 1990.
[8] J. Prewitt, “Object Enhancement And Extraction, in Picture Processing and Psychopictorics, B. Lipkin and A. Rosenfeld, Academic Pres, 1970.
[9] L.G. Roberts, “Machine Perception of Three-Dimensional Solids”, in optical and Electro-Optical Information Processing, J.T. Tippett et al., MIT Press, Cambridge, 159-197, 1965.
[10] ShervanFekriErshad, “Texture Classification Approach Based on Combination of Edge & Co-occurrence and Local Binary Pattern”, Int'l Conf. IP, Comp. Vision, and Pattern Recognition, 626-629, 2011
[11] J. Matthews. C. Cortes and V. Vapnik, Support vector networks, Machine Learning 20 (1995) 273–297.
[12] Ada, RajneetKaur, Feature Extraction and Principal Component Analysis for Lung Cancer Detection in CT scan Images.