INTERNATIONAL JOURNAL FOR TRENDS IN ENGINEERING & TECHNOLOGY
VOLUME 3 ISSUE 3 – MARCH 2015 – ISSN: 2349 – 9303

Lung Disease Classification Using Support Vector Machine

D.KAVINYA [1]
[1]K.S.Rangasamy College of Technology, Tiruchengode, TamilNadu.
kavinyacs@gmail.com

DR. A.GNANA BASKARAN[2]
[2]K.S.Rangasamy College of Technology, Tiruchengode, TamilNadu.

Abstract— Classification plays a vital role in disease detection and diagnosis. Classification of lung diseases is an important part for disease diagnosis. It assists diagnosis of disease with greater efficiency. Here Computed tomography (CT) images of lung diseases are classified. In this data mining classification algorithm, support vector machine (SVM) is to be optimized using ant colony optimization (ACO) algorithm. The feature of the CT scan image is extracted using wavelet transformation and the moment invariants, it is believed that it will provide a better output for classification. The Further principle component analysis also provides reduced dimensionality of image it is an added advantage for efficient classification. This optimized svm will provide a better classification accuracy.

Index Terms— Ant colony optimization (ACO), Computed tomography (CT), Principle component analysis (PCA), Support vector machine (SVM), Moment Invariant (MI).

1 INTRODUCTION

Nowadays computed tomography (CT) images are widely used for disease diagnosis and the radiation emission by the CT scans are lower. It helps doctors for disease detection. Efficiently physicians can detect the abnormalities in the body and locate the changes. Using different gray levels of the computed tomography images they can be distinguished. Lung diseases can be caused due to infections, medications and several other disorders. Computed tomography images are widely used in disease diagnosis. The CT scan images are called tomograms having more detail than standard X-rays. A CT scan produces images of structures within the body like internal organs, blood vessels, bones and tumours (Kalender 2011).

Medical images is pervasive for patients disease detection and it provides efficient standard for The Digital Imaging and Communications in Medicine (DICOM) efficient standards [2,15,16] allows storing textual descriptions, known as metadata, along with the images in terms of speed and resolution the CT images plays an efficient tool for diagnosis.

The computerized tomography has been found to be the most reliable method of early detection of tumors [3, 4]. Chiou et al. [12] have proposed the application of neural network based hybrid system for lung nodule detection, based on artificial neural network architectures [7]. C. Bhuvaneswari et al. (2014) [1] proposed A new fusion model of classification of the lung diseases using genetic algorithm. It consists of phases, namely: Pre-processing, Feature extraction, Feature selection, Classifiers, Performance evaluation for various classifiers. Emphysema, Bronchitis, Pleural effusion and normal lung images were considered for classification of that work. For Pre-processing Median filter and morphological smoothening filter techniques are applied to remove the noise from the images and enhance the image. For feature extraction novel Fusion Median Absolute Deviation (MAD) technique was used which combines Gabor feature (Texture Co-Efficient) and Walsh Hadamard Transform (Pixel Co-Efficient). For feature selection Genetic algorithm is used.

The proposed work is done as follows, the input to the system is CT lung images. The Pre-processing phase includes Denoising and Edge Enhancement. In Denoising, the random noise is removed by discriminating regions of different densities. In Edge Enhancement, Laplacian Filters are used. The feature extraction technique is carried out using wavelet transforms. MIs are used as features for shape recognition and classification. Feature selection module has been performed by Sequential Backward Floating Selection. The classification of the images is done through the optimized svm [14] and the performance measures, classification accuracy of each classifiers are discussed in this work.

2 MATERIALS AND METHODS

CT images of lung diseases such as cancer, pneumonia, normal lung were considered for classification in this work.

3 PROPOSED WORK

The proposed work is represented in various steps as follows.

3.1 Preprocessing

Image Preprocessing is the first phase which generates preprocessed lung volume. Pre-processing is used to enhance the marks and edges that have to be matched. This process includes Denoising and Edge Enhancement. In Denoising, the random noise was removed by Adaptive Weiner Filter by discriminating regions of different densities. In Edge Enhancement, Laplacian Filters were used for the enhancement of the edges. Pre-processing includes Image Denoising and Edge Enhancement. Image denoising is done to remove the noise from the query image. Noise is removed using Weiner filter.

3.2 Feature Extraction

Majority of the medical images are generally in gray levels with few modalities containing color images. These images of different categories can be distinguished by using their shape and texture characteristics.

3.2.1 Coiflet Wavelet Function

Daubechies (1992) designed Coiflets which are discrete wavelets having scaling functions with vanishing moments. Coiflet wavelets are near symmetric with the wavelet functions having N/3 vanishing moments and N/3-1 scaling functions. The wavelet function ψ has 2N moments equal to 0 and the scaling function ψ has 2N-1 moments equal to 0. The support of length of both functions is 6N-1 (Beylkin et al 1991).

The coifN ψ and φ has more symmetric than the dbNs. Regarding support length, coifN has to be compared to db3N or sym3N. With regard to the number of vanishing moments of ψ, it is more similar to db2N or sym2N.

If *s* is a sufficiently regular continuous time signal, for large *j* the Coefficient (s, ϕj, k) is approximated by $2^{-j/2}$ s (2^{-j} k).

If s is a polynomial of degree d, $d \leq N - 1$, then the approximation becomes equality. This property is useful when calculating the difference between an expansion over the $\phi_{j,k}$ scaling function of a given signal and its sampled version.

In Coiflet wavelet function, there are 2N moments equal to 0 and the scaling function has 2N-1 moments equal to 0. The above two functions have a support of length 6N-1. Both Daubechies and Coiflets wavelet families have orthogonal and bi-orthogonal properties.

3.2.2 Moment Invariants (MI)

MIs was frequently used as features for remote sensing, image processing, shape recognition and classification [11]. The Moments provide an object's characteristics that uniquely represent its shape. Invariant shape recognition is through classification in multidimensional moment invariant feature space. Many techniques were developed to derive invariant features from the moments for object recognition/representation.

Conventionally, the moment invariants are computed based on information from both shape boundary and its interior region. The Moments used to construct MIs are defined in continuous the discrete form implementation is for practical purpose. Given a function f(x, y), these regular moments are defined (Keyes 2006) by:

$$pq = \iint x\ p\ yq\ f(x,y)dxdy$$

Is a two-dimensional moment of function f(x,y). The order of moment is (p + q) where p and q are both natural numbers. In digital form implementation this becomes:

$$M_{pq} = \sum_x \sum_y x^p\ y^q\ \ f(x,y)$$

3.3 Feature Reduction

Feature reduction is performed using principle component analysis, which is the general technique which uses sophisticated underlying mathematical principles to transform a number of possible correlated variables into a smaller number of variables named as principal components.

In general terms, PCA uses a vector space transform for reducing the dimensionality of large data sets. By using mathematical projection, original data set, which may have involved many variables, that often, be interpreted in just a few variables. It is therefore the case that an examination of the reduced dimension data set will allow the user for spotting trends, patterns and outliers in the data, far more easily than would have been possible without performing the principal component analysis (Jolliffe 2005).

PCA computes new variables called principal components which are obtained as linear combinations of original variables.

3.4 OPTIMIZING SUPPORT VECTOR MACHINE USING ANT COLONY OPTIMIZATION

In general svm provides better classification whereas if an efficient optimization algorithm is used with svm its performance can be increased further [5, 8, 9, 12, and 13]. Here ant colony optimization algorithm is used to enhance the classification of svm.

3.4.1 Support Vector Machines

SVM is a popular technique for numerical data classification [10]. This type of learning algorithm was introduced in the year of 1990s. SVM maps feature non-linearly into n dimensional feature space when provided with a feature set capable of being represented in space. When a kernel is introduced with high computation in the SVM algorithm, the inputs are in the form of scalar products, the classification is then achieved. The issue is translated into a convex quadratic optimization problem and a clear answer being attained by convexity (Boser et al 1992). In SVM, an attribute is a predictor variable and a feature is a transformed attribute. A set of features describing an example is in the form of a vector. The feature vectors define the hyper plane. The optimal hyper plane is constructed by the SVM with the aim to separate vector clusters with a class of attributes on one side of the plane and on the other with different attributes. The margin represents the distance between hyper plane and Support Vectors (SV). SVM analysis tries for positioning the margin in such a way that space between it and support vectors are increased. The Figure shows a simplified SVM process overview.

SVM can be used to classify elements in a certain feature space. Svm work in a two-step process which is as follows:

- Training: With the representative sample data the support vectors are generated.
- Regression/classification: It is used for unknown data in the feature space.

3.4.2 Ant Colony Optimization

The ant colony optimization algorithm (ACO) is a probabilistic technique for solving computational problems which can be reduced to finding good paths through graphs [6].

The general algorithm is relatively simple and based on the set of ants, each making a single possible round-trips along the cities. In every stage, the ant chooses to move from one city to another according to some rules:

- It must visit each city exactly once;
- The distance city has less chance of being chosen
- The more intense the pheromone trail laid out on an edge between two cities, the greater will be the probability that that edge will be chosen
- Having completed its journey, the ant deposits more pheromones on all of the edges it has traversed, if the journey is short
- After each iteration, trails of pheromones will evaporate.

4. CONCLUSION

In this work performance of support vector machine will be increased using ant colony optimization algorithm for disease detection by using the lung CT images. In this feature extraction is to be done using wavelet transformation and the moment invariants, it is believed that it will provide a better output for classification. Further principle component analysis also provides reduced dimensionality of image it is an added advantage for an efficient classification. Classification plays a major role in disease detection, the classification algorithm will be optimized such that it provides an efficient classification of lung diseases.

REFERENCES

[1] C. Bhuvaneswari, P. Aruna, D. Loganathan : "A new fusion model for classification of the lung diseases using genetic algorithm" Egyptian Informatics Journal (2014) 15, 69–77

[2] Chiou YS, Lure YF, Freedman MT, Fritz S. Application of neural network based hybrid system for lung nodule detection. In: Computer-based medical systems, proceedings of sixth annual IEEE symposium, IEEE; 1993. p. 211–216

[3] Graham R, Perriss R, Scarsbrook A. Dicom demystified: a review of digital file formats and their use in radiological practice. Clin Radiol 2005:1133–40

[4] Graham R, Perriss R, Scarsbrook A. Dicom demystified: a review of digital file formats and their use in radiological practice. Clin Radiol 2005;60:1133–40

[5] Karaboga, D. and B. Basturk, On the performance of artificial bee colony (ABC) algorithm," Applied Soft Computing, Vol. 8, No. 1, 687{697, 2008

[6] Karaboga, N., A. Kalinli, and D. Karaboga, \Designing digital IIR filters using ant colony optimisation algorithm," Engineering Applications of Artificial Intelligence, Vol. 17, No. 3, 301{309, 2004.

[7] Lin D, Yan C. Lung nodules identification rules extraction with neural fuzzy network. Neural Inform Process IEEE 2002; 4:2049–53.

[8] Haupt RL, Haupt SE. Practcal genetic algorithms. second ed. New Jersey, United States: John Wiley & Sons; 2004.

[9] Jiang, H, Tang, F & Zhang, X 2010, 'Liver cancer identification based on PSO-SVM model', Proceedings of the eleventh IEEE International Conference on Control Automation Robotics & Vision, pp. 2519-2523.

[10] Jrad .N, Congedo.M, Phlypo, R, Rousseau, S, Flamary, R, Yger, F & Rakotomamonjy, A 2011, 'SW-SVM: sensor weighting support vector machines for EEG-based brain–computer interfaces', Journal of neural engineering, vol.8, no.5.

[11] Keyes, L 2006, 'Using Moment Invariants for Classifying Shapes On large Scale Maps' , Journal of Computers, Environment and Urban systems, vol.25, pp.119–130.

[12] George, G & Raj, VC 2011, 'Review on feature selection techniques and the impact of SVM for cancer classification using gene expression profile'. International Journal of Computer Science & Engineering Survey, vol.2, no.3, pp.16-27

[13] Gao. F 2013, 'Application of SVM Optimization Based on GA in Electronic Sphygmomanometer Data Fusion', International Journal of Computer Science Issues, vol. 10, no. 1, pp.1-6.

[14] Sonali Jain(2013) "A Machine Learning Approach: SVM for Image Classification in CBIR" International journal of application in engineering and management, Volume 2, Issue 4, April 2013

[15] Tianxia Gong et al(2007), "Classification of CT Brain Images of Head Trauma" pp. 401–408, Springer-Verlag Berlin Heidelberg 2007

[16] V.Krishnaiah, Dr. G. Narsimha, Dr .N. Subhash Chandra (2013), "Diagnosis of Lung Cancer Prediction System Using Data Mining Classification Techniques", International Journal of Computer Science and Information Technologies, Vol. 4