Classification of Lung Diseases using Optimization Techniques

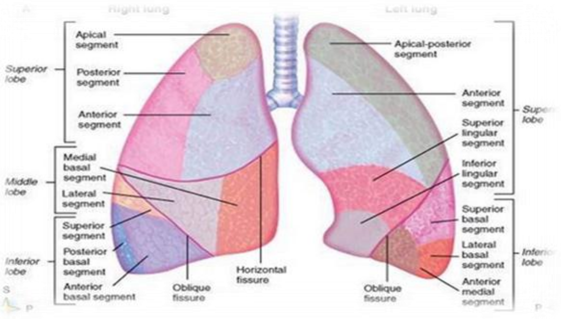
Tejinder Kaur[1] Er. Neelakshi Gupta[2]
[1,2]Department of Electronics & Communication Engineering
[1,2]Amritsar College of Engineering & Technology

Abstract—The lungs are part of complex apparatus, expanding and relaxing thousands of times each day to bring in oxygen and expel carbon dioxide. Lung disease can results from problems in any part of this system. In this paper, an automated approach for feature selection and classification of the lung diseases using CT images is presented. The lung CT image is engaged as the input. Filters are used to remove unwanted noise and enhance the quality of image. New feature selection technique that is hybridization of genetic and PSO is used to select features after extracting features using MAD technique. Once features have been selected then they are classified using MLP-NN classifier. In this paper optimization techniques such as genetic/Particle Swarm Optimization are used for feature selection process.

Key words: Classification, Genetic Algorithm, CT, Particle Swarm Optimization, Feature Extraction

I. INTRODUCTION

The lungs are part of complex apparatus, expanding and relaxing thousands of times each day to bring in oxygen and expel carbon dioxide. Lung disease can results from problems in any part of this system. Diseases that affect airways include: Chronic obstructive pulmonary disease, Chronic bronchitis, Emphysema Cystic fibrosis and many more. To cure the diseases it is very important to recognize the diseases accurately. Accomplished by rush of new development of high technology and use of various imaging modalities such as computed tomography (CT), magnetic resonance imaging (MRI), doppler ultrasound, and various imaging techniques based on nuclear emission PET (position emission tomography), SPECT(single photon emission computed tomography), more challenges arise; for example how to process and analyze a significant volume of images so that high quality information can be produced for disease diagnosis and treatment.



In this paper, an optimization technique for classification of the lung diseases using CT images is presented. The extracted features of lung ct images are selected by the using hybrid genetic\particle swarm optimization algorithm method, the classifiers are used to classify the datasets in to relevant datasets and the performance measures are evaluated for the datasets. The paper is organized as Section 2 deal with the literature review. Section 3 explains the proposed method in steps and in last step conclusion of the paper is presented.

II. LITERATURE REVIEW

Neha Sharma et al. (2015) [1] defined a new image based feature extraction technique for classification of lung CT images. A novel hybrid based method was obtained by combining the Gabor filter and Walsh Hadamard transform features using Median Absolute Deviation (MAD) technique. Thus it include the advantage of combined models. The proposed system comprised of three stages. In the initial stage, the images were given as input and features were obtained by applying novel fusion based feature wrenched method, followed by second stage, in which features were selected by applying Genetic Algorithm which selected the top ranked features. In the final stage, categeorisers namely, decision tree, memory based classification (KNN) , multilayer perceptron neural network (MLP-NN) were employed to perform classification of the lung diseases.

Sunanda Biradar et al. (2015) [2] proposed a method for developing a Computer Aided Diagnosis (CAD) system for detecting lung cancer by analyzing the Computer tomography (CT) images of lungs. The images were collected from the LIDC dataset and enhanced to increase the contract of images by Median filter. After enhancement morphological segmentation was used to segment lungs region from the CT images. Then segmented images were used to identify and classify the cancerous and noncancerous nodules using the support vector machine (SVM) classifier. The SVM polynomial has given the 96.6% accuracy and SVM quadratic function had given the 92% accuracy

Ms I. Christa Mary et al. (2014) [3] provided a comprehensive survey of Computer Aided Diagnosis system (CAD) for early detection of lung diseases. Lung diseases are mainly occurred as airway diseases, lung tissue disease and lung classification disease. In the computerized detection and classification of lung image five methodologies were used. These were pre-processing, segmentation, feature extraction, feature selection and classification. The comparisons with these methods were described in lab.

Fardin Ahmadizar et al. (2015) [4] proposed a method in which a new evolutionary-based algorithm had developed to simultaneously evolve the topology and the connections weights of ANNs by means of a a new combination of grammatical evolution(GE) and genetic algorithm(GA). GE had adopted to design the network topology while GA was incorporated for better weight adaption . The proposed algorithm needs to invest a minimal expert's effort for customization and was capable of generating any feedforward ANN with one hidden layer. Moreover, due to the fact that the generalization ability of an ANN may decrease because of overfitting problems, the algorithm utilizes a novel adaptive penalty apporvh to simplify ANNs generated through the evolution process. As a result, it produced much simpler ANNs that has better generalization ability and are easy to implement. This prpoposed method has been tested on some real world classification datasets, and results are statistically compared against existind methods. The results indicated that these algorithms outperrorms the other methods and provided the best overall performance in terms of the classification accuracy and number of hidden neurons. The results also presents the contribution of the proposed penalty approch in the simplicity and generalization ability of the generated networks.

III. PROPOSED METHODOLOGY

A. *Pre-Processing*

Guided filter and morphological filter are used to enhance the image quality and to remove noises. Guided filter is used as an edge preserving smoothing operators like popular "bilateral filter" but has better performance near edges. Guided filer is also a more genetic concept beyond smoothing: it can transfer the structures of guidance images to filtering output enabling new filtering applications like de-hazing and guided feathering. The dilation and erosion process is done for morphologically smoothing of images. Erosion involves the alteration of pixels at the edges of region i.e., exchanging binary 1 value to 0.While dilation is reverse process with regions growing out from their boundaries. The Erosion followed by dilation is known as Opening operation which suppresses the bright details smaller where dilation followed by erosion is closing operation which suppresses the dark details are computed. Here in this work the darker details are suppressed where the dilation is performed followed by erosion. The pre-processed image is obtained by applying the guided filter and morphological smoothening.

B. *Feature Extraction*

The main purpose of feature extraction is to reduce the original dataset by measuring certain features that distinguish one of interest from another one. Texture features and Pixel coefficients values are extracted by taking each sub-image from top left corner. Texture analysis is quantitative method that can be used to quantify and detect structural abnormalities. The Gabor Filter will get the orientation as input , set with the direction give the strong response for the location the target images. So as we change the orientation, that will allow the highlight of edges in more better way. Image size 256×256 pixel is divided into sub-image of 12 blocks. The output of Gabor filter is Gabor coefficients of whole image taken. The features extracted by Gabor filter are sum, min, max,mean,median. The Walsh Hadamard transform (WHT) is a non-sinusoidal, orthogonal technique deintegrate a signal into basic function set that are Walsh functions with square and rectangular waves with the value of $+1$ or -1. The output of WHT function is the pixel co-efficients on the whole image. Extracted features from WHT are sum, min, max,mean,median The features extracted from WHT and Gabor Filter are fused using template level fusion. WHT extracts the features from the Frequency domain and Gabor filter extracted features comparable to orientation selectivity, spatial localization and spatial frequency. Features extracted are then fused using MAD technique. MAD uses median for the deviation scores, and is more robust than the standard deviation as a measure of the dispersion. Main advantage of MAD is that it can be covered into values that close to the standard deviation.

C. *Feature Selection*

In machine learning and statistics, feature selection is also known as variable selection, attribute selection and variable subset selection, is a process of selecting relevant features for model constructions. Feature selection is a technique of choosing minimized applicable features which improves the classification accuracy. Feature selection methods are of three types: Filter method, Wrapper method and Hybrid method. In this paper we are applying hybrid method as we are using hybridization of two algorithms namely, Genetic Algorithm and Particle Swarm Optimization.

1) *Genetic Algorithm*

Genetic algorithm (GA) is a randomized search and optimization technique guided by the principles of evolution and natural genetics, having a large amount of implicit parallelism. GAs perform search in complex, large and multimodal landscapes, and provide near-optimal solutions for objective or fitness function of an optimization problem. An implementation of a genetic algorithm begins with a random population of chromosomes. One can evaluates these data structures and apply reproduction operators in such a way that the chromosomes which provide a better solution to the objective problem are given more chances to reproduce themselves than those chromosomes which gives poorer solutions.

2) *Particle Swarm Optimization*

Particle swarm optimization (PSO) is just a population based stochastic optimization technique developed by Dr. Eberhart and Dr. Kennedy in 1995, inspired by social behavior of bird flocking or fish schooling.

PSO shares many similarities with evolutionary computation techniques such as for example Genetic Algorithms (GA). The machine is initialized with a population of random solutions and pursuit of optima by updating generations. However, unlike GA, PSO does not have any evolution operators such as for example crossover and mutation. In PSO, the potential solutions, called particles, fly through the situation space by following the present optimum particles. Each particle monitors its coordinates in the situation space which are associated with the most effective solution (fitness) it has achieved so far. (The fitness value can be stored.) This value is known as *pbest* .Another "best" value that is tracked by the particle swarm optimizer is the greatest value, obtained up to now by any particle in the neighbors of the particle. This location is known as *lbest*. Each time a particle takes all the people as its topological neighbors, the most effective value is an international best and is known as *gbest*.

The particle swarm optimization concept and *lbest* locations (local version of PSO). Acceleration is weighted by a random term, with separate random numbers being generated for acceleration toward *pbest* and *lbest* location includes, at every time step, changing the velocity of (accelerating) each particle toward its *p*bests.

In past several years, PSO has been successfully applied in lots of research and application areas. It is demonstrated that PSO gets better results in a quicker, cheaper way compared with other methods.

Another reason that PSO is attractive is that there are few parameters to adjust. One version, with slight variations, is useful in a wide variety of applications. Particle swarm optimization has been useful for approaches that can be utilized across a wide selection of applications, in addition to for specific applications dedicated to a particular requirement.

D. Classification

Image classification analyzes the numerical properties of various image features and organizes the data into categories. Classification algorithms mainly employs into two phases of processing such as training and testing. In initial training phase, the characteristic properties of typical image features are isolated based on the training class, is created. In the subsequent testing phase, these feature-space partitions are used for classification of image features. There are various classification methods such as decision tree induction, Bayesian networks, KNN , MLP-NN. In this thesis work we are using MLP-NN classification method.

1) Multilayer Perceptron Neural Network

A multilayer perceptron is a feed forward artificial neural network model that maps set of input data onto a set of appropriate outputs. MLP has a hidden layer between input and output layers. MLP can implement non linear discriminants (for classification) and non linear regression functions (for regression). The biological neuron's functions are transformed by enumerating a differentiable nonlinear functions for MLP's each artificial neuron. .

IV. CONCLUSION AND FUTURE WORK

In this paper an optimization techniques for classification of lung diseases has been proposed. To realize an automatic and unsupervised classification, we implemented a framework that utilizes a combination of various techniques and composed of algorithms such as feature selection technique that uses hybrid technique with the combination of two algorithms that are genetic and Particle Swarm Optimization and for classification Multilayer Perceptron – Neural Network is used. The combination of these techniques provides effective methods for classification of lung diseases. This work has potential for further development because of its simplicity that will motivate to classify the types if lung diseases.

REFERENCES

[1] Neha Sharma, Anil Kumar, "Review and Analysis of Fusion model for the Classification of Lung Cancer diseases using Genetic Algorithm", IRJET June2015.

[2] Sunanda Biradar, Kavya Agalatakatti, "Lung Cancer identification using CT images", IJECS July2015.

[3] Ms.I.Christa Mary, Ms. J. Preethi, "A Survey on Computerized, Quantification and Classification of Lung Disease", IJAICT Nov2014.

[4] Fardin Ahmadizar, Khabat Soltanian, Fardin Akhlaghian, Loannis Tsoulos," Artificial Neural Network development by means of a novel combination of Grammatical Evolution and Genetic Algorithm", Volume 39, Elseiver March 2015.

[5] Ms. Swati P. Tidke, Prof. Vrishali A. Chakkarwar, "Classification of Lung Tumor using SVM", Volume 2 Issue. 5, IJECR Sep 2012.

[6] Chesner Desir, Caroline Petitjean, Laurent Heutt, Luc Thiberville, Mathieu Salaun, "An SVM-based distal lung image classification using texture discriptors", Elseiver2012.

[7] Gurpreet Kaur, Harpreet Singh," An Intelligent System for Lung Cancer Diagnosis Using Fusion of Support Vector Machines and Back Propagation Neural Network", International Journal of Science and Research (IJSR), 2013.

[8] C.Bhuvaneswari1, P.Aruna, D.Loganathan, "Advanced Segmentation Techniques using Genetic Algorithm for Recognition of Lung Diseases from CT Scans of Thorax", IJERA Aug2013.

[9] Bahriye Akay," A Study on Partile Swarm Optimization and Artifical Bee Colony algorithms for Multilevel Thresholding", Elseiver June2013.

[10] Mahdi Setayesh, Mengjie Zhang, Mark Johnston," A Novel Particle Optimization Approch to Detecting Continuous, Thin and Smooth Edges in noisy images", Volume 246, Elseiver 10 Oct 2013.

[11] C.Bhuvaneswari1, P.Aruna, D.Loganathan," Classification of Lung Diseases by Image Processing Techniques Using Computed Tomography Images", International Journal of Advanced Computer Research Volume-4 Number-1, 14 March-2014.

[12] C. Gunavathi, K. Premalatha," Performance Analysis of Genetic Algorithm with kNN and SVM for Feature Selection in Tumor Classification", International Journal of Computer, Control, Quantum and Information Engineering , 2014.

[13] G.Vijaya, A.Suhasini, R.Priya, "Automatic Detection of Lung Cancer in CT Images", IJRET May2014.

[14] Xiabi Liu, Ling Ma, Li Song, Yanfeng Zhao, Xinming Zhao, Chunwu Zhou, "Recognizing Common CT Imaging Signs of Lung Diseases through a New Feature Selection Method based on Fisher Criterion and Genetic Optimization"IEEE 2014.

[15] Kun-Huang Chen, Kung-Jeng Wang, Kung-Min Wang, Melani-Adrian Angelia, "Applying Particle Swarm Optimization-based decision tree classifier for cancer classification on gene expression data", Elseiver2014.

[16] Rina D.Rudyanto, Sjoerd Kerkstra,Eva Mvan Rik Xoart," Comparing algorithms for automated vessel segmentation in Computed Tomography scans of lung", Volume 18, Elseiver Oct 2014.

[17] Ling Ma, Xiabijiu, Li Song, Chunwu Zhou, Yankeng Zhao,Xinming Zhao ,"A new classifier Fusion method based on Historical and on-line classification reliability for recognizing common CT imaging signs of Lung diseases", Volume 40, Elseiver2015.

[18] Nali, Yuanxiang Li ," Image Restoration using Improved Particle Swarm Optimization", IEEE 2011.

[19] Shengjun Zhou, Yuanzhi Cheng, Shinichi Tamura," Automated lung segmentation and smoothing techniques for inclusion of juxtapleural nodules and pulmonary vessels on chest CT images", Volume 13, Elseiver 2014.

[20] Mahdi Setayesh, Mengjie Zhang, Mark Johnston," A Novel Particle Optimization Approch to Detecting Continuous, Thin and Smooth Edges in noisy images", Volume 246, Elseiver 10 Oct 2013.