Lung Anomaly Detection System (LADS) Using SVM based on Firefly Algorithm

Rasha Abdul Razak A P[1], Harish Binu K P[2]

[1]PG Student, Computer Science and Engineering, MEA Engineering College, Kerala, India

[2]Assistant Professor, Computer Science and Engineering, MEA Engineering College, Kerala, India

Abstract: *Computed Tomography is one of the most popular technique which is used for lung anomaly detection. In this paper, an automatic lung anomaly detection system is developed by using SVM and Firefly optimization algorithm. For that features from the CT images are selected from various types of feature selection techniques such as Local Binary Pattern (LBP), Discrete Wavelet Transform (DWT) and Histogram of Oriented Gradients (HOG). Then the selected feature set are optimized by firefly optimization, to find out an optimal feature subset from candidate features. Then the feature set is filtered by a guided filter, which is mainly used in medical applications to improve the accuracy of our system. The filtered features are then classified by using a commonly used classifier SVM (Support Vector Machine) which classify the lung disease from the CT image. This paper further compare LADS with FIG method and LADS shows better recognition performance.*

Keywords: Computed Tomography, DWT, FIG, HOG, LADS, LBP, SVM

1. Introduction

Lung diseases are the anomalies that affect the lungs and which is the most common medical condition found in the world nowadays. Air pollution, smoking, genetics and infections are the factors that affecting lung disorders. Computed Tomography (CT) scan plays an important role in the diagnosis of lung diseases. The difficulty of recognizing the lung lesions for radiologists leads to the research interest in Computer Aided Diagnosis (CAD) and Content Based Medical Image Retrieval (CBMIR) based on CT scan. The computer should have the abilities such as classifying and detecting CT findings of lung lesions to support CAD and CBMIR applications. The CT findings or CT features denotes what radiologists see in CT scan, for detecting diseases.

The term lung disease refers to many disorders affecting the lungs, such as asthma, COPD infections like pneumonia and tuberculosis and many breathing problem. Some lung disease can lead to respiratory failure. It is important to pay attention to the symptoms as they could be first sign of respiratory failure.

Computed Tomography (CT) is one of method which is used for lung anomaly detection. A CT scanner takes multiple X-rays of the chest and a computer create detailed images of the lungs and surrounding sections. CT scanning is fast, painless, noninvasive and accurate. Because of its ability to detect very small nodules in the lung, CT scan is especially effective for diagnosing lung cancer at its earliest, most curable stage. Here an automatic lung anomaly system is developed by using lung CT images. We have been witnessing the enormous increase in CT images of the human lungs, which should be read in time. This challenge plus the difficulty of recognizing subtle lesions even for radiologists promote the research interests in the computer-aided diagnosis (CAD) and the content-based medical image retrieval (CBMIR) based on thoracic CT scans. To support CAD and CBMIR applications, computer should have the abilities of classifying, detecting, and quantifying CT findings of lung lesions and the CT findings denote what radiologists see in CT scans for diagnosing diseases, which are also often called CT features or CT manifestation [1]. This work focus on the problem of automatic classification of CT findings of lung lesions in CT scans and then classify CT Findings of lung lesions under the ignorance of underlying diseases.

CT image of lungs includes a slice of the ribcage, which contain a large number of structures, such as, arteries, blood vessels, respiratory vessels, parenchyma and pulmonary pleural, each with its own specific information. Thus, for pulmonary disease diagnosis and analysis, it is necessary to select features from lung structures

The various algorithms used for this work are various feature selection algorithms such as LBP, HOG, DWT. Firefly algorithm is the optimization technique used and guided filtering technique is used for filtering the image. The most commonly used classifier SVM is used to classify the lung anomalies.

This paper mainly focuses on the commonly seen lung anomalies such as tuberculosis, bronchitis, calcification and cavity and vaculous.

2. Related Works

There is an extensive literature on object detection, but here we mention just a few relevant papers on lung anomaly detection system [4,11,2,10]. Dehmeshki et al proposes a shape-based genetic algorithm template-matching (GATM) method for the detection of nodules with spherical elements. The main motive of this system is to develop a novel nodule detection method based on the geometric shape features incorporated to the GATM model[4] .Parveen and Kavitha et al introduces a new system named FSFCM (Firefly Search with Fuzzy C-Means) for segmentation of CT lung nodules[14].Abbas et al proposes a novel segmentationalgorithm for lungs based on high-resolution computed tomography (HRCT) scan images.Zhu et al [21]

proposes method to find and select texture features of solitary pulmonary nodules (SPNs) detected by computed tomography (CT) and then evaluate the performance of support vector machine (SVM).

Liu et al [1] develops a feature selection method based on FIsher criterion and genetic optimization, called FIG, to tackle the CISL recognition problem. In this FIG, the Fisher criterion is applied to evaluate feature subsets which is based on genetic optimization algorithm is developed to find out an optimal feature subset from the candidate features. In contrast, our system provides a better recognition system than that of FIG. This paper further compare the accuracy against this FIG method.

3. Proposed System

This work focus on the problem of automatic classification of CT findings of lung lesions in CT scans and then classify CT Findings of lung lesions under the ignorance of underlying diseases. The various algorithms used for this work are various feature selection algorithms such as LBP, HOG, DWT. Firefly algorithm is the optimization technique used and guided filtering technique is used for filtering the image. The most commonly used classifier SVM is used to classify the lung anomalies. This system consists of three stages.
1) Image preprocessing.
   - Feature selection.
   - Filtering
2) Optimization.
3) Classification and training.
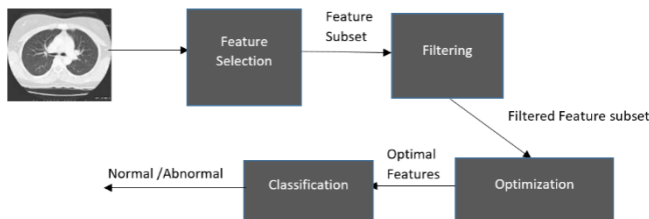
Figure 3.1: Outline of LADS

Figure 3.1 shows the outline of LADS and its modules are explained below.

3.1. Feature Selection Methods

In machine learning, feature selection, is the process of selecting a subset of relevant features (variables, predictors). The feature selection techniques used in LADS are
1) LBP
2) HOG
3) DWT

LBP (Local Binary Pattern)
Local binary patterns (LBP) is used for feature selection in computer vision. Huang et al [5] claims that Local Binary Patterns (LBP) is a non-parametric descriptor whose aim is to efficiently summarize the local structures of images.

The LBP feature vector, is created in the following manner:

1) Divide the input image into cells for example 16x16 pixels for each cell).
2) For each pixels in the cell, the pixel is compared with each of its 8 neighbors (on its left-top, right-top, left-middle, left-bottom, etc.
3) Write "0" when the center pixel value is greater than that of the neighbor value, otherwise, write "1" which gives an 8-digit binary number and then which is converted to decimal number. The derived binary number is referred to as Local Binary Patterns or LBP codes.
4) Now compute the histogram over the cell, by frequency of each number occurring. Frequency means each combination of which pixels are smaller and which are greater than the center. This histogram can be seen as a 256-dimensional feature vector.
5) Normalize the histogram computed.
6) Concatenate the normalized histograms of cells which gives a feature vector for the entire window.

HOG (Histogram of Oriented Gradients)
The histogram of oriented gradients (HOG) (Dalal and Triggs) [6] is a feature descriptor which is used in computer vision and image processing for object detection. The basic idea behind the histogram of oriented gradients descriptor is that shape within an image local object appearance and can be described by the distribution of intensity gradients and edge directions. Here the image is divided into small regions called cells, and for the pixels within each cell, a histogram of gradient directions is computed. Descriptor which is formed is the concatenation of these histograms. In order to attain accuracy, the local histograms can be contrast-normalized by calculating a measure of the intensity across a larger region of the image, called a block, and then using this value to normalize all cells within the block.

Discrete Wavelet Transform
Kociołek et al. says that the discrete wavelet transform (DWT) is a linear transformation that operates on a data vector whose length is an integer power of two, transforming it into a numerically different vector of the same length. DWT is a tool which separates data into different frequency components, and then learns each component with resolution matched to its scale.

The important feature of DWT is its multiscale representation of function. A given function can be analyzed at various levels of resolution by using the wavelets. It is also invertible and can be orthogonal. Discrete wavelet transform is used to compute characteristics from a signal on various scales proceeding by a high pass and low pass filtering. The basic feature selection procedure consists of
1) Decompose the signals using DWT into N levels by filtering and decimation to obtain the approximation and detailed coefficients.
2) Extracting the features obtained from the DWT coefficients.

Using wavelet transform, we can suppress noise which are out of frequency band of the signal.
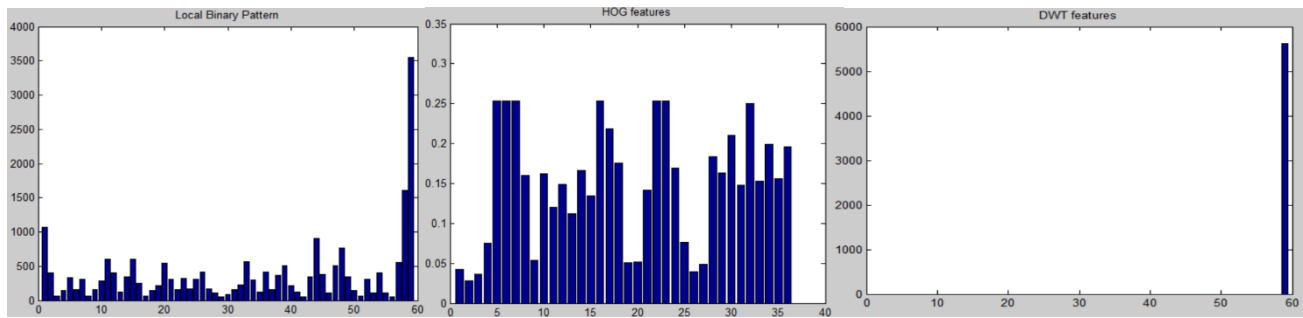
Figure 3.2: Features selected using LBP, HOG , DWT

3.2 Filtering Technique

Image filtering plays a crucial role in image processing. Filtering is an important part of digital image processing because it has a variety of applications. Here we use guided filter for this system.

Guided Filter

Guided filter is derived from a local linear model such that guided filter generates the filtering output by considering the content of a guidance image, which can be the filtering image itself or another different image. The guided filter can be perform as an edge preserving smoothing operator like the bilateral filter but has better performance near the edges. Guided filter will enhance the sharpness and reduce the noise of blurred, noisy images in relatively less computational time. The computational complexity is independent of its kernel size. This filter has an edge preserving smoothing property and has an O(N) time (in number of pixels N) algorithm for both gray scale and color images [8].
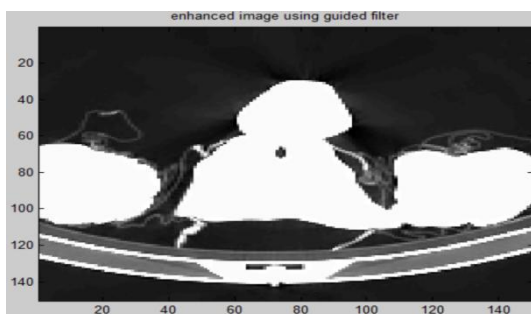
Figure 3.3: Filtered Image

Guided filter is a linear translation variant filter [3] such that the output pixel of the image I is computed as the linear transform of the guidance image G, where G can be the input image itself or another image in a local window w_k centered at pixel k[3].

In case of HD filtering, guided filter adopts fast and non approximation characteristics of linear time algorithm. Hence, it is used as one of the fastest edge preserving filters. Guided filter is fast and accurate filter. Guided Filter gives smoothen image and enhanced image. Guided Filter algorithm can be simulated in MATLAB.

3.3 Optimization Technique

The main purpose of optimization in image processing is for compensating the missed or corrupted image data and to find good correspondences between input images. Here we use firefly optimization technique for optimizing the feature subset obtained after the feature selection method.

Firefly optimization Algorithm

In order to explain firefly algorithm we want to follow some idealized rules in Yang, 2009.They are

1) Fireflies are unisex so that one firefly will be attracted to other fireflies without considering their sex;
2) Attractiveness is directly proportional to their brightness, thus for any two flashing fireflies, the less brighter one will move towards the brighter and they both decrease as their distance increases. If there is no brighter one than a particular firefly, it will move randomly;
3) The brightness of a firefly is affected by the landscape of the objective function. Other forms of brightness can be defined in a similar way to the fitness function in genetic algorithms.

3.4 Classification and Training

For classification and training the most commonly used classifier SVM is used.

SVM (Support Vector Machine)

Support vector machine is based on structural risk minimization principle. SVM performs very well in high-dimensional spaces and also on both training and testing data. It does not suffer the small size of training dataset since the decision surface of SVM-based classifier is evaluated by the inner product of training data. The basic principle behind SVM classifier is to construct a hyperplane such that it maximizes the margin between negative and positive examples. The hyperplane is determined by the examples that are closest to the decision surface and the decision surface is computed by the inner product of training data, which enables to map the input vectors through function into feature space. The feature space is defined by kernel K(x, y). For noise and to avoid overfitting, slack variables are introduced.
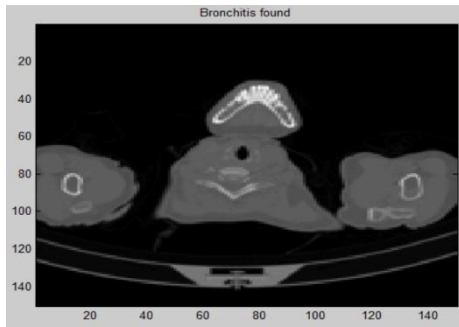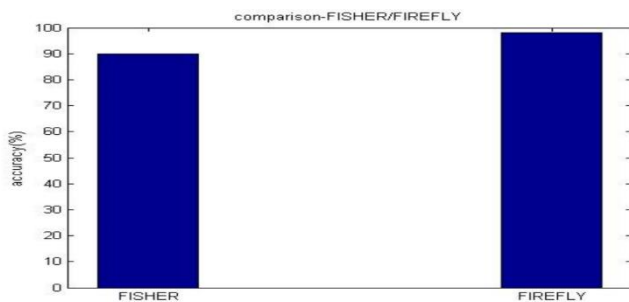
Figure 3.5: Output of LADS

4. Experimental Results

In this work we evaluated the results of both LADS and FIG system. The experiments are conducted to the data set collected from Moulana hospital, kerala, India. The result is evaluated against the accuracy of the system.

The accuracy of FIG System is about 90% and whereas that of LADS is 98.5%.

Fig 4.1: Comparison Graph

From the graph 4.1 it is clear that LADS has a better accuracy than that of FIG. Thegraph is plotted from the results of the experiments that has applied to dataset.

5. Conclusion

Proposed a new feature selection method based on SVM and firefly optimization for recognizing CISLs. The main contributions of this work are summarizedas follows.

- The problem of recognizing CISLs in lung CT images is put forward, which is important for the CAD and the CBMIR based on thoracic CT scans. To my knowledge, this problem has not received much attention of researchers. The previous works on lung tissue classification mainly concern about how to distinguish abnormal tissues from normal ones or identify among different visual patterns of a specific lung disease.
- As demonstrated by the experimental results, our LADS method can bring more effective recognition results at the satisfactory computation costs, compared with single type of features and the full set of original features. Furthermore, it brought slightly better recognition performance and much better computation efficiency than the commonly used genetic feature selection method based on classification accuracy rate.
- Here some image preprocessing steps are added to further improve the performance of CISL recognizer. Afilter is used to filter the blood vessels to get rid of the confusion

between vessels and CISLs and also enhance the regions wrapping CISLs to make the visual appearance of CISLs clearer and thus increase the possibility of correct classification.

References

[1] Liu, X., Ma, L., Song, L., Zhao, Y., Zhao, X., and Zhou, C. (2015). Recognizing common CT imaging signs of lung diseases through a new feature selection method based on fisher criterion and genetic optimization. *IEEE journal of biomedical and health informatics*,19(2):635–647.

[2] Abbas, Q., Khan, M. T. A., Farooq, A., and Celebi, M. E. (2013). Segmentation of lungs in hrct scan images using particle swarm optimization. *Int. J. Innov. Comput. Inf. Control*, 9(5):2155–2165**.**

[3] Nikhil Dhengre1, Kishor P. Upla2, Ronak D. Trivedi3 "Multimodal Biomedical Image Fusion: Use of logGabor and Guided filters with Non-subsampled Contourlet Transform "

[4] Dehmeshki, J., Ye, X., Lin, X., Valdivieso, M., and Amin, H. (2007). Automated detection of lung nodules in ct images using shape-based genetic algorithm. *Computerized Medical Imaging and Graphics*, 31(6):408–417.

[5] Huang, D., Shan, C., Ardabilian, M., Wang, Y., and Chen, L. (2011). Local binary patterns and its application to facial image analysis: a survey. *IEEE Transactions onSystems, Man, and Cybernetics, Part C (Applications and Reviews)*, 41(6):765–781.

[6] Dalal, N. and Triggs, B. (2005). Histograms of oriented gradients for human detection.In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer SocietyConference on*, volume 1, pages 886–893. IEEE

[7] Kociołek, M., Materka, A., Strzelecki, M., and Szczypinski, P. (2001). Discrete wavelet transform-derived features for digital image texture analysis. In *International Conference on Signals and Electronic Systems*, pages 99–104**.**

[8] Pham, C. C. and Jeon, J. W. (2014). Efficient image sharpening and denoising using adaptive guided image filtering. *IET Image Processing*, 9(1):71–79.

[9] Yang, X.-S. (2009). Firefly algorithms for multimodal optimization. In *International symposium on stochastic algorithms*, pages 169–178. Springer.

[10] Parveen, S. S. and Kavitha, C. (2015). Segmentation of ct lung nodules using fcm with firefly search algorithm. In *Innovations in Information, Embedded and CommunicationSystems (ICIIECS), 2015 International Conference on*, pages 1–6. IEEE.

[11] Zhu, Y., Tan, Y., Hua, Y., Wang, M., Zhang, G., and Zhang, J. (2010). Feature selection and performance evaluation of support vector machine (svm)-based classifier for differentiating benign and malignant pulmonary nodules by computed tomography. *Journal of digital imaging*, 23(1):51–65.

[12] Rasha Abdul Razak A P and Harish BinuKP, "A Survey on Various Image FilteringTechniques for Image Sharpening", International Journal of Science and Research (IJSR),ISSN 23197064, Volume 6 Issue 4 , April 2017

Author Profile

Rasha Abdul Razak A P received her B.Tech degree in ComputerScience from MEA Engineering College, Perinthalmanna in 2011. Right now she is pursuing her M.Tech degree in Computer Science at MEA Engineering College, Perinthalmanna from 2015 to 2017. Her research interests lie in Image Processing.

Harish Binu K P, Assistant Professor, MEA Engineering College,Perinthalmanna. Received the B.Tech degree in Computer Science & Engineering from NSS Engineering College, Palakkad, University of Calicut and M.tech degree in Computer Science from College of Engineering, Cochin university of Science and Technology. His research interests include Image Processing and Digital design.