www.ijecs.in
International Journal Of Engineering And Computer Science ISSN: 2319-7242
Volume 5 Issue 5 May 2016, Page No. 16736-16739

Introducing automated system for Lung Cancer Detection using Evolutionary Approach

Manasee Kurkure, Anuradha Thakare
Department of Computer Engineering
PCCOE Pune-44
manasi.k17@gmail.com
Department of Computer
Engineering
PCCOE Pune-44
adthakare@yahoo.com

Abstract— Cancer is one of the most commonly affected diseases in the developed countries. Early diagnosis plays a significant role in curing cancer patients. Every year, thousands of people die due to Lung cancer. In this paper, a novel Candidate group search algorithm based on evolutionary approach is proposed. This optimization algorithm allows assistant doctors to identify the nodules present in the lungs at the early stages. As manual interpretations are time consuming and very critical. Genetic algorithm (GA) helps in identifying genes that help to classify patient lung cancer status with notable predictive performance.

Keywords— Data Clustering, Genetic Algorithm, Naïve Bayes Classification, Candidate Group Search algorithm, Canny Edge Detection Algorithm, Optimization.

1. INTRODUCTION

Cancer which is clinically referred as a malevolent neoplasm is a extensive group of diseases, involving unregulated cell growth. In cancer, cells subdivide and grow hysterically, forming malignant tumors, and invade nearby parts within the body. These tumors can grow and hinder the digestive, nervous, and circulatory systems and releases hormones that may alter the body functionality. There are about 200 different known cancers that are shown in human. Each of these are characterized by the type of the cell that is first affected. Surgical removal of lung cancer still remains the gold standard in preventing lung cancer. Early diagnosis of lung cancer is therefore important to prevent the spread of the cancer. Treatment of lung cancer also varies depending on the type tumor present. Classification of different tumor types is thus important to ensure higher survival rates. However, classification of lung cancers is challenging. Currently, cancer classification is based on subjective interpretation of histopathological and clinical data. Classification also depends on the site of origin of the tumor. Clinical information may be incomplete at times and the wide classes of most tumors lack morphologic features which are essential in classification .The primary cause of Lung Cancer is tobacco consumption. More than 60 % of new lung cancers occur in never smokers or former smoker. Lung Cancer is rare below age 40,which rates increasing until age 80 ,after which the rate tapers off. Acute bronchitis, asthma, Chronic Obstructive Pulmonary Disease (COPD), chronic bronchitis Emphysema, Acute respiratory distress syndrome (ARDS) and Lung cancer. As per World cancer report 2014 lung cancer is the most common cause of cancer-related death in men and women, and was responsible for 1.56 million deaths anually. The major causes of the lung diseases are smoking, inhaling the drugs, smoke and allergic materials. The computed tomography (CT) images assists in detecting the extreme of the lung diseases. For the analysis of the proposed method CT image is sufficient also the visibility of soft tissue is better. There are several types of lung cancer, and these are divided into two main groups: Small cell lung cancer and non-small cell lung cancer which has three subtypes: Carcinoma, Adeno carcinoma and Squamous cell carcinomas[1].

A. CANNY EDGE DETECTION ALGORITHM

In this proposed work for the noise removal and contrast enhancement the images are pre-processed to obtain accurate enhanced images. Canny filter is used in Feature extraction. The output values of Canny filter are given to Naive Bayes Classification which is optimized by Genetic Candidate Group Search Algorithm (GCGS).The purpose of edge detection in general is to reduce the amount of data in an image, restoring the structural properties to be used for further image processing. Several algorithms exists, and this technique focuses on a particular one developed by John F. Canny(JFC) in 1986 [11]. Even though it is not new, it has become one of the standard edge detection methods and it is still used in research [10] [12].

The aim of JFC was to develop an algorithm that gives optimal results with respect to the following conditions:

1. Detection: The probability of detecting real edge points should be more while the probability of falsely detecting non-edge points should be less. This corresponds to maximizing the signal-to-noise ratio.
2. Localization: The detected edges should be as close as much as it can to the real edges.
3. Number of responses: One real edge should not result in more than one detected edge.

B. GENETIC ALGORITHM

GAs were proven to be the most powerful optimization technique where the problem is a large solution space. This explains the increasing popularity of GAs applications in image processing and other fields. They are applied where exhaustive search for solution is expensive in terms of computation time. Applications of GAs for image processing extend from detecting edges to making complex decisions or classifying extracted features. Genetic algorithms are based on natural selection discovered by Charles Darwin. They use

natural selection method of fittest individuals for optimization problem solver. Optimization is performed through natural exchange of genetic material between parents. Offsprings are formed from parent genes. Fitness of offsprings is evaluated. The fittest individuals are allowed to used only. In computer world, genetic material is replaced by strings of bits and natural selection replaced by fitness function. Matting of parents is represented by cross-over and mutation operations. A simple GA consists of five steps :

1. Start with a randomly generated population of N chromosomes, where N is the size of population, l – length of chromosome x.
2. Calculate the fitness value of function φ(x) of each chromosome x in the population.
3. Repeat until N offspring's are created:
   3.1. Probabilistically select a pair of chromosomes from current population using value of fitness function.
   3.2. Produce an offspring y_i using crossover and mutation operators, where i = 1, 2, …, N.
4. Replace current population with newly created one.
5. Go to step 2.

In some cases of GA, the whole population is formed for strings having the same length.

C. NAIVE BAYES ALGORITHM

The Naïve Bayesian classifier is a straightforward and frequently used method for supervised learning. The Naïve Bayesian classification system is based on Bayes rule and works as follows. One highly practical Bayesian learning method is the naive Bayes learner, often called the naive Bayes classifier. In some domains its performance has been shown to be comparable to that of neural network and decision tree learning. This section introduces the naive Bayes classifier. The naive Bayes classifier applies to learning tasks where each instance x is described by a conjunction of attribute values and where the target function f (x) can take on any value from some finite set V. A set of training examples of the target function is provided, and a new instance is presented, described by the tuple of attribute values (The learner is asked to predict the target value, or classification, for this new instance. The difference between the naive Bayes learning method and other learning methods we have considered is that there is no explicit search through the space of possible hypotheses (in this case, the space of possible hypotheses is the space of possible values that can be assigned to the various $p(v_i)$ and $p(\frac{a_i}{v_i})$ terms. Instead, the hypothesis is formed without searching, simply by counting the frequency of various data combinations within the training.

2. RELATED WORK

In paper to combat the limitations of traditional K-NN, a novel method to improve the classification performance of K-NN using Genetic Algorithm (GA) is done [1]. The proposed G-KNN classifier is applied for classification and similar k-neighbours are chosen at each iteration for classification by using GA, the test samples are classified with these neighbours and the accuracy is calculated for different number of K values to obtain high accuracy, hence the computation time of K-NN is reduced from the obtained results in this method. The MATLAB image processing toolbox based implementation is done on the CT lung images and the classifications of these images are carried out. The k value**,** execution time and accuracy is calculated and tabulated. Such early detection might be helpful for physicians.

In paper the segmentation process starts by detecting the lung edge using canny edge detection filters[2]. To improve the edge detection, Euler number method is applied in this paper. Later, morphology method is used to make the lung edge better so that the final output of lung region can be generated. After implementing the segmentation task, the output in the form of lung region mask is compared to the GT image to check their similarity. In the evaluation, the Jacquard Similarity Coefficient is used to calculate the similarity. The value derived from the test is moderately high although it cannot exceed other prior researchers score.

In this paper, they have investigated the application of GA to edge detection of medical images using cost minimization to accurately localize thin, continuous edges[3]. They have based the optimization on cost evaluations and transformations defined by Tan *et al.* [3], [14], where SA was used for the optimization. They extended the bit string chromosome of the traditional GA to a bit-array chromosome, which conforms closely with a logical edge representation. They introduced problem space reduction with dependent regions. To increase the performance of the traditional GA, they added reduced surrogate crossover, ranking selection, dynamic operator rates, and stochastic evaluation on the operators. The GA improved the Pratt figure of merit from 0.77–0.85for ideal images. For actual images, the value of the cost function was used for quantitative comparison. For MR images, the GA improved the cost function value by 18%. Similar results were obtained for other modalities. The detected edges were thin, continuous, and well localized. Most of the basic edge features were detected.

In the paper the GA-based feature selection method is proposed to determine the optimal feature subset where samples belonging to different classes are well discriminated and the features are the least redundant[4].A novel fitness function of the GA algorithm is provided to evaluate the candidate feature subsets, and the MI is used to compute the correlation information among attributes.

Sr.No.	Paper	Techniques Used	Performance
1.	Detection of Cancer in Lung With K-NN Classification Using Genetic Algorithm	KNN, Genetic Algorithm	Performance measures like the classification rate and the false positive rates are analyzed
2.	Image Segmentation for Lung Region in Chest X-ray Images using Edge Detection and Morphology.	Canny Edge Detection , Euler Method,	Segmentation is done on various edge detection technology

DOI: 10.18535/ijecs/v5i5.69

3.	Edge Detection in Medical Images Using a Genetic Algorithm	Edge detection algorithms, genetic algorithms	Here the application of GA's to edge detection of medical images using cost minimization to accurately localize thin, continuous edges.
4.	An Intelligent System for Lung Cancer Diagnosis Using a New Genetic Algorithm Based Feature Selection Method.	Genetic algorithm , Feature selection, Machine learning	A novel fitness function of the GA algorithm is provided to evaluate the candidate feature subsets, and the MI is used to compute the correlation information among attributes

Table 1. Survey Table

3. PROPOSED WORK

In the proposed work we first take the image as input. We will apply image processing techniques like Canny edge detection algorithm for feature extraction. In the last stage classification which is achieved, and optimization by Genetic Candidate Group Search Algorithm is used to detect the stages of cancer.

The following is the graphical representation of the module.

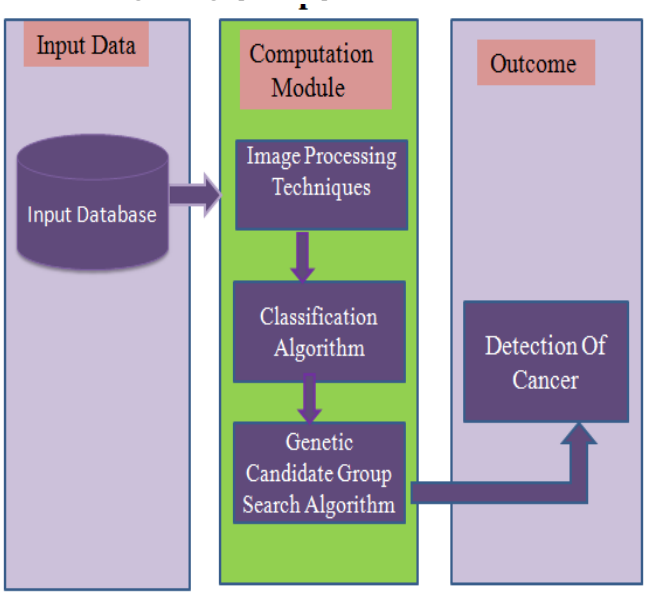

Fig.1. Proposed Model

Fig.2. Analysis of PR and PSNR values

The above intermediate results are shown in the paper. Accuracy to find edges are measured with respect to three parameters as shown in fig 2.We can see that the PSNR ratio is highest in Canny i.e. higher value of PSNR is good means the ratio of signal to noise is higher and execution time required is also less. So we can say that Canny works better specially good in medical domain rather than Sobel and prewitt.

6. Conclusion

The limitations of Candidate Group Search Algorithm, are overcome by a novel method to improve the classification performance by combining Genetic Algorithm (GA). The proposed Naive Bayes classifier is applied for classification and the accuracy will be calculated for different number of values using novel algorithm. The gradient-based approaches such as the Prewitt filter is very sensitive to noise. So Canny edge detection algorithm is used as it is less sensitive to noise but are computationally more expensive compared to Robert's operator Sobel, and Prewitt operator. From the above analysis

Sr. No.	Edge Detection Techniques	Performance Ratio	Execution Time (sec)	Peak Signal Noise Ratio
1	Canny	2.76	34.4	7.973
2	Sobel	2.05	36.8	7.535
3	Prewitt	2.1	37	7.265

we can say that Canny edge detection method is the best.

REFERENCES

[1] P . Bhuvaneswari , Dr. A. Brintha Therese, "Detection of Cancer in Lung With K-NN Classification Using Genetic Algorithm", Science Direct 2nd International Conference on Nanomaterials and Technologies. (CNT 2014)

[2] Anisha P R, Kishor Kumar Reddy C, **"**A Pragmatic approach for Detecting Liver Cancer using Image Processing and Data Mining Technique" SPACES-2015.

[3] Mohd Nizam Saad, Noraidah Sahari@ Ashaari, Hamzaini Abdul Hamid, " Image Segmentation for Lung Region in Chest X-ray Images using Edge Detection and Morphology", 2014 IEEE International Conference on Control System, Computing and Engineering, 28 - 30 November 2014, Penang, Malaysia..

[4] Mark´us Gud□mundsson, Essam A. El-Kwae, and Mansur R. Kabuka*, **"**Edge Detection in Medical Images Using a Genetic Algorithm",IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. 17, NO. 3, JUNE 1998 .

[5] Chunhong Lu & Zhaomin Zhu & Xiaofeng Gu, "An Intelligent System for Lung Cancer Diagnosis Using a New Genetic Algorithm Based Feature Selection Method", Springer Science+Business Media New York 2014.

[6] G.T. Shrivakshan1, Dr.C. Chandrasekar, "A Comparison of various Edge Detection Techniques used in Image Processing" IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 5, No 1, September 2012.0

[7] Shweta Kharya, Shika Agrawal, "Naive Bayes Classifiers: A Probabilistic Detection Model for Breast Cancer", International Journal of Computer Applications (0975 – 8887) Volume 92 – No.10, April 2014

[8] Sukhmeet Kaur, Kiran Jyoti,Kalyan Roy, Joydeep Mukherjee, "Image Similarity Measure using Color Histogram,Color Coherence, Vector, and Sobel Method," Acm Computing Surveys, International Journal of Science and Research (IJSR), India Online ISSN: 2319-7064.

[9] Shweta Kharya, Shika Agrawal ,Sunita Soni, " Naive Bayes Classifiers: A Probabilistic Detection Model for Breast Cancer " International Journal of Computer Applications (0975 – 8887) Volume 92 – No.10, April 2014.

[10] Sergei Azernikov. Sweeping solids on manifolds. In Symposium on Solid and Physical Modeling, pages 249–255, 2008.

[11] John Canny. A computational approach to edge detection. Pattern Analysis and Machine Intelligence, IEEE Transactions on, PAMI-8(6):679–698, Nov. 1986.

[12] F. Mai, Y. Hung, H. Zhong, and W. Sze. A hierarchical approach for fast and robust ellipse extraction. Pattern Recognition, 41(8):2512–2524, August 2008.