

Classification of Stages of Lung Cancer using Genetic Candidate Group Search Approach

Manasee Kurkure¹, Anuradha Thakare²

Department of Computer Engineering, PCCOE Pune-44.

Abstract: Cancer is one of the most commonly affected diseases in the progressing countries. Early diagnosis of cancer plays a significant role in curing cancer patients. Thousands of people every year die due to lung cancer, so there is a need for accurate prediction of this disease. In this paper, a Genetic Candidate Group search algorithm 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, to overcome this difficulty this hybrid method is combined with Naïve Bayes Classification and different stages of cancer images fast and accurate. The number of images are tested and the results are obtained. The accuracy of 82 percentage is achieved in classification.

Keywords: Data Clustering, Genetic Candidate Group search Algorithm(GCGS), Naïve Bayes Classification, Optimization.

I. Introduction

Cancer is a disease, involving abnormal cell growth. In cancer, cells divide and grow uncontrollably, forming malignant tumors, and affect nearby organs within the body. These tumors can grow and affect almost all the body functionality. 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. Thus the process of classification of lung cancers is always challenging.

A. Genetic Algorithm

A genetic algorithm (or GA) is a search technique used in to find exact or approximate solutions to optimize the problems. Genetic algorithms are a particular class of Evolutionary algorithms that use techniques inspired by biology such as inheritance, mutation, selection, and crossover.

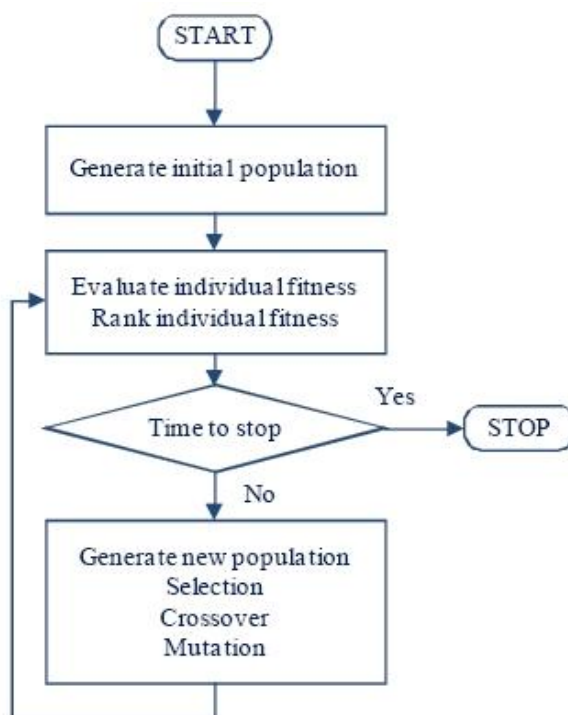


Fig 1: Genetic Algorithm

The optimization is done through the exchange of genes between parents. A simple GA consists of five steps:

1. Start with a randomly generated population of M chromosomes, where M is the size of population, l – length of chromosome x .
2. Calculate the fitness value of function $\varphi(x)$ of each chromosome x in the population.
3. Repeat until M offspring's are created:
 - 3.1 Probabilistically select a pair of chromosomes from current population using the value of the fitness function.
 - 3.2 Produce an offspring y using crossover and mutation operators, where $I = 1, 2, \dots, N$.
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. [15]

Thus GA is applied to find the optimum feature subset through the evolution process. The proposed fitness function of the GA is used to test the performance of an individual.

B. Naive Bayes Algorithm

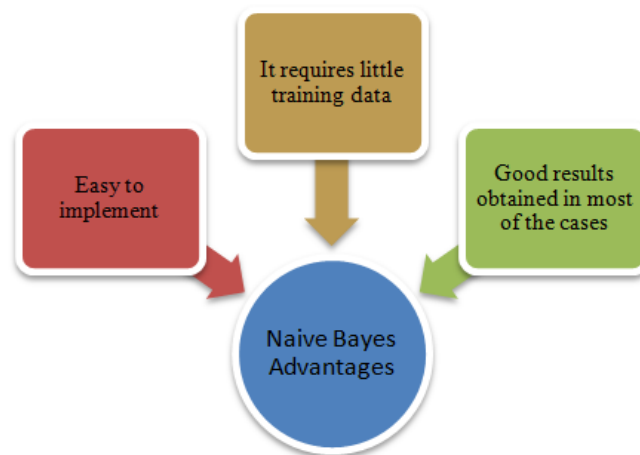


Fig 2: Advantages of Naive Bayes

The Naïve Bayesian classifier is a supervised learning method. The Naïve Bayesian classification system is based on Bayes rule. In some domains its performance is performs good as compare to that of neural network and decision tree learning.

II. 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-neighbors are chosen at each iteration for classification by using GA, the test samples are classified with these neighbors and the accuracy is calculated for different number of K values,hence the computation time of K-NN is reduced from the obtained results with this method. In paper the segmentation process starts by detecting the lung edge using canny edge detection filters. To improve the edge detection, an Euler number method is applied in this paper.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 extended the bit string chromosome of the traditional GA for a bit-array chromosome, which conforms closely with a logical edge representation. They introduced problem space reduction ranking selection, dynamic operator rates, and stochastic evaluation on the operators. The GA was compared to the SA approach using ideal and actual medical images from different modalities including MRI, CT, and ultrasound. 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.

III. Proposed Work

In the proposed work, first take the images as input. Apply image processing techniques for pre-processing and feature extraction. After feature extraction of that particular image Naïve Bayes is applied on that image and further optimization is made by GCGS Algorithm. The following figure 3 shows the proposed workflow of the system.

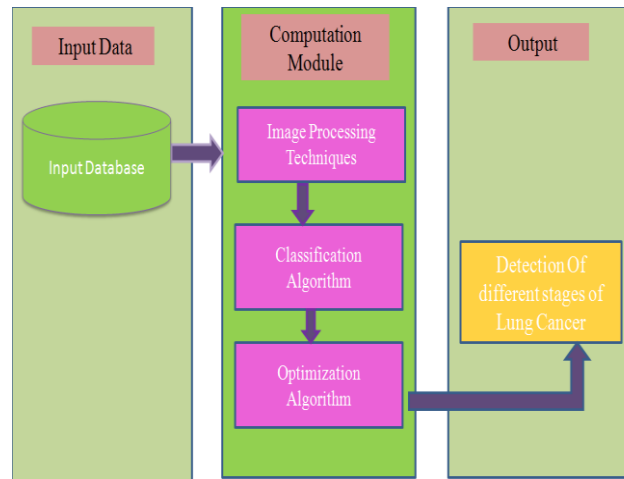


Fig 3: Proposed System Model

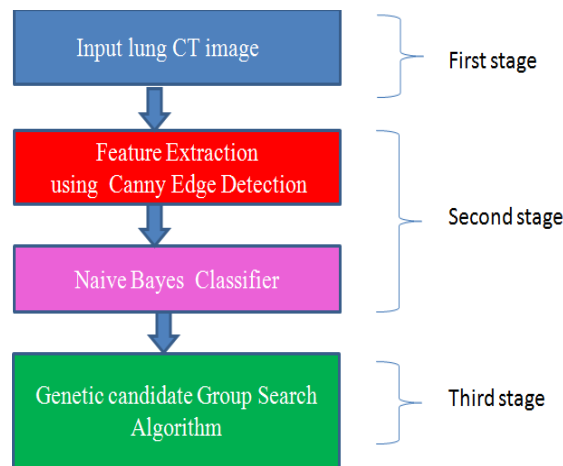


Fig.4: Workflow of Proposed Model

In the above fig 2, we will first take image as an input. Then we will use the Canny detection method to preprocess the image by detecting only strong edges. From that detected image we will generate feature vectors. The output of Naïve Bayes classification with the help of GCGS algorithm provides accurate and fast results, whether the given image is normal or cancerous.

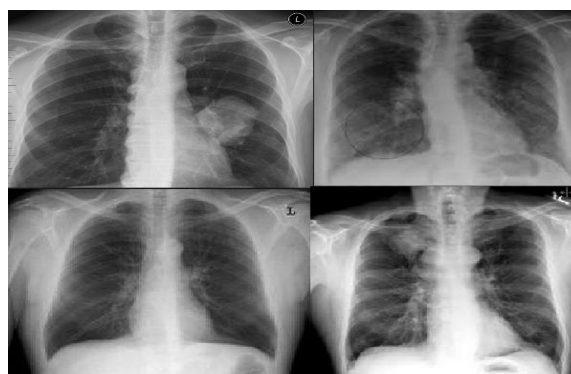


Fig 5: Lung Cancer Dataset

A. Following steps used in the proposed Algorithm:

Input: image i_1 from test DB and set of images for train

DB List $L\{img\}$

Output : Reational probable images

1. Read image from user and convert into buffer image
2. for each(k into L)
3. Calculate pixel weight = distance between $L[k]$ and i_1 .
4. Define threshold as T .
5. Apply filtration

For each(p up to W)

Foe each(p_1 onto H)

$Diff = W[p] * H[p_1]$

If($weight > T$ or $Diff > T$)

Add $L[i]$ into output List

6. end for

7. End procedure.

Genetic Algorithm for filtering the accuracy.

Input: Chromosome with n features (output image List $L[]$ from NB algorithm 1 image features), number of generations, population size, crossover probability (P_c), mutation probability (P_m).

Output: Single image with detection

1. Initialize the population randomly with chromosomes.
2. Initialize N (total number of records in the training set).
3. for each chromosome in the new population
4. Apply Crossover to best selected chromosome.
5. Apply Mutation for each chromosome to generate new population.
6. Calculate fitness = $f(D \cup I \cup A)$
7. End for
8. Select 50% best fit chromosome and remove worse fit chromosome.
9. Go to step no 3.
10. Stop

B. Mathematical Model

Let S be a system

$S = \{S_1, S_2, S_3, S_4\}$ where

S_1 represents User registration system.

S_2 represents feature extraction module.

S_3 represents database image distance calculation Module.

S_4 represents result matching and detection using NB and GCGSA.

$$I = \{U_1, U_2, U_3, \dots, U_n\}$$

$$G = \sqrt{Gx^2 + Gy^2}$$

$$f = \{f_1, f_2, f_3, \dots, f_n\}$$

Let c be the class label

$$N = \{\text{cancerous, no cancerous}\}$$

$$P(c|f) = \frac{P(f|c)P(c)}{P(f)}$$

Let ,

D_t = distance from database

D_n = distance from new image.

Feature = $\{F_{i1}, F_{i2}, F_{i2}, \dots, F_{in}\}$

Retrieve all images from database,

Where,

Foreach (Distance $_1$ =from database each user image);

Crossover= $\{\text{Feature}_i, \dots, \text{feature}_{i+1}\}$

Mutation= $\{\text{Feature}(i), \dots, \text{Feature}(n)\}$

Where,

Fitness = $f(D \cup I \cup A)$

$D = D_t - D_n$
 $I = 4 * \text{Area} * \pi / \text{perimeter}^2$
A= number of pixels
Threshold $t = 0.35$
Classify image base on current score.
N= {cancerous, Non cancerous}

IV. Results And Discussions

The results of proposed Lung Cancer Detector are presents in this section. Initially the images are taken as input and feature vectors are generated from segmented images by applying histogram and gray scale techniques. The stepwise execution is presented in figures from Fig 6 to Fig 9. The final output from the test image is given to GCGS algorithm and the given image is cancerous or non cancerous is predicted.

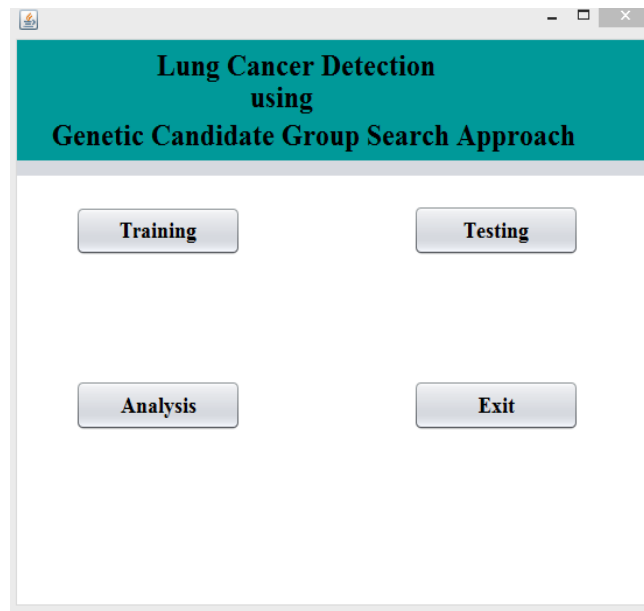


Fig 6: Main Form

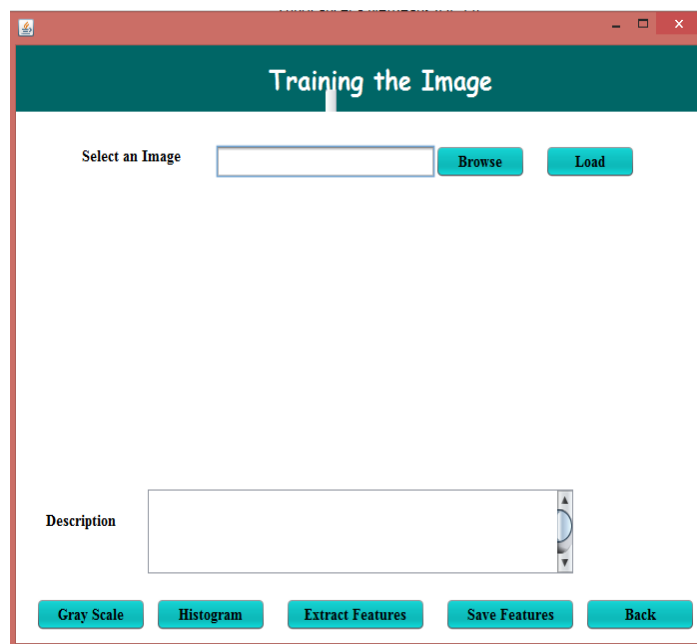


Fig 7: Training Phase

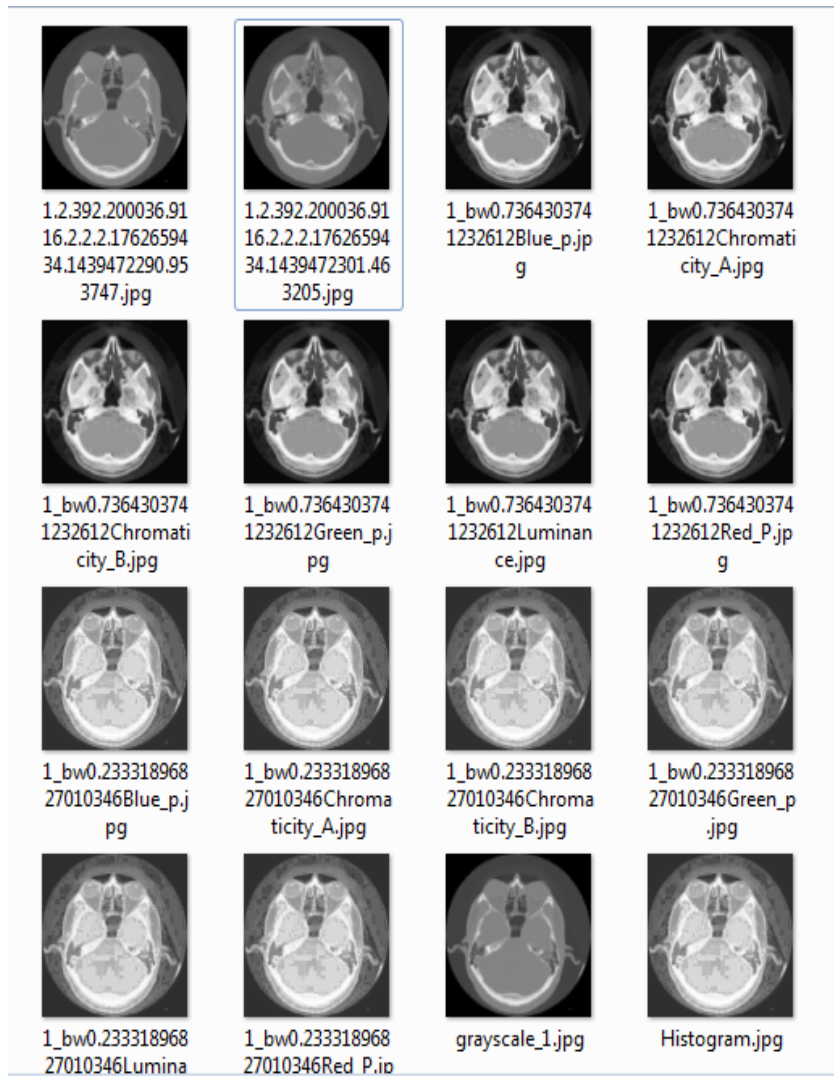


Fig 8: Features Generation of the image.

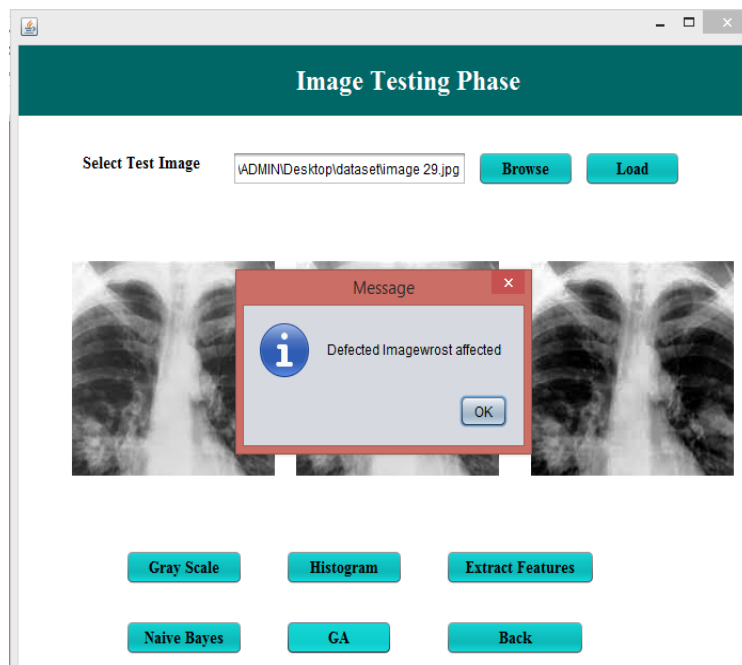


Fig 9: Final Result using GCGS

V. Conclusion

To improve the classification performance of the large amount of data Genetic Candidate Group Search Algorithm (GCGS) is proposed. For this optimization to be achieved classification, i.e. Naive Bayes classifier is applied on a number of images and 82% accuracy is obtained. In this way different stages of lung cancer are predicted.

References

- [1] P. Bhuvaneshwari, 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 Gudmundsson, 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 Transaction, 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):25122524, August 2008.
- [13] Ghassan Mahmoud husien Amer Dr. Ahmed Mohamed Abushaala, "Edge Detection Methods", 978-1-4799-8172-4/15/\$31.00 ©2015 IEEE
- [14] Nisha1, Rajesh Mehra2, Lalita Sharma, Comparative Analysis of Canny and Prewitt Edge Detection Techniques used in Image Processing, International Journal of Engineering Trends and Technology (IJETT) – Volume 28 Number 1 - October 2015.
- [15] Shanta H Biradar, "Reboost Image Segmentation Using Genetic Algorithm", International Journal of Computer Science Trends and Technology (IJCST) – Volume 3 Issue 3, May-June 2015