

# Pattern Analysis for Lung Cancer Cell Identification

Anita Kumar<sup>1</sup>, K.Mohamed Amanullah<sup>2</sup>

<sup>1</sup> Department of Computer Science

<sup>2</sup> Department of Computer Applications

<sup>1,2</sup> Bishop Heber College, Trichy

**Abstract-** Millions of death report caused by cancer in every year report has been adopted in World Health Organization (WHO) reports ungovernable cell development in the tissues of the lung is called as lung cancer. The early detection of lung cancer is a major challenging problem, due to the arrangement of the cancer cells, where huge number of the cells is overlapped with everyone. Classification is very important part of data analysis for Lung cancer prediction. Knowledge discovery and data mining have found numerous applications in business and scientific domain. The goal here is to identify characteristics of patient segments where average durability is significantly higher/ lower than average survival across the entire dataset by using the Artificial Bee Colony (ABC) algorithm; the dimensionality of the dataset is reduced in order to reduce the computation complexity.

**Keywords-** Classification, Feature Extraction, Lung Cancer, Artificial Bee Colony (ABC) Algorithm

## I. INTRODUCTION

Lung cancer is one of the [24] commonest cancers in the industrialized world and persons with this grave disease must deal not only with the physical effects but also with the psychosocial aspects. Lung cancer [25] is a disease of abnormal cells multiplying and growing into a tumor. The overall 5-year survival rate for lung cancer patients increases from 14 to 49% if the disease is detected in time. There is significant evidence indicating that the early detection of lung cancer will decrease mortality rate. There are many techniques to diagnose lung cancer, such as Chest Radiography (x-ray), enumerate Tomography (CT), Magnetic Resonance Imaging (MRI scan) and Sputum Cytology. However, most of these techniques are costly and time consuming. In other terms, most of these techniques are detecting the lung cancer in its advanced stages, where the patient's chance of surviving is less. Hence, there is a great need for a new technology to diagnose the lung cancer in its early stages. Image processing techniques provides a very good quality tool for improving the manual analysis.

In this modern age of fully automated trend of living, the automated diagnostic systems play an important and vital role. Automated diagnostic system models in Medical Image processing are one such field where numerous systems are

proposed and still many more under conceptual design due explosive growth of the technology today. From the previous decades, we have witnessed an explosive growth of Digital image processing for analysis of the data that can be captured by digital images and artificial neural networks are used to aggregate the analyzed data from these images to produce a diagnosis prediction with high accuracy instantaneously where digital images serve as tool for input data [20] [21]. Hence in the process of surgery these automated systems help the surgeon to identify the infected parts or tumors in case of cancerous growth of cells to be removed with high accuracy hence by increasing the probability of survival of a patient. In this one of such an automated system for cancer cell classification which helps as a tool assisting surgeon to differentiate cancerous cells from those normal cells i.e. percentage of carcinoma cells, instantaneously at the time of the surgery. Here the pathological images serve as input data. The analysis of all the pathological images is directly based on four steps: 1) image filtering and enhancement, 2) segmentation, 3) feature extraction 4) analysis of extracted features in the pattern of recognition system or classifier [21].

## II. LITERATURE SURVEY

In 2014, Chien-Hung Huang et al [7] has proposed differentially expressed genes are identified via an expression dataset generated from lung adenocarcinoma tumor and adjacent non-tumor tissues using protein synergy network, micro-array data analysis and cluster analysis with up and down regulated communities. Input taken from 18 patients cancerous and non-cancerous tissues with 41 samples and their findings of systematic strategy to discover potential drugs and target genes for lung cancer. From which eight drugs from drug bank and three drugs from NCBI. Very small number of patients considered for analysis and takes more time and cost to consume the entire process.

In 2014, Dr.D.P.Shukla et al [9] reviewed about various data mining techniques such as clustering, classification, regression, association rule mining, CART(Classification And Regression Tree) are widely used in Healthcare domain. These are the approaches used to improve the quality of prediction and diagnosis with respect to different kinds diseases such as cancer, cardio vascular abnormalities and others are using various algorithms for example genetic

algorithm, association rule mining, k-means clustering, Naïve Bayesian classification is applied in huge volume of medical data.

In 2009, Mao et al. [2] explores the application of Apriori-Gen algorithm on the disease association study which has a large data set and tries to find the association among multiple Single Nucleotide Polymorphisms (SNPs) that may be responsible for the disease. The goal of disease association study is to assess accumulated information targeted to find interaction of multi-SNPs which are associated to complex diseases with significantly high accuracy and statistical power. The proposed theory was also applied to check the genetic data of the type 2 diabetes. This will save a very large amount of money and time for diagnosis and can be done before the onset of diseases.

Fang et al. [26] have used a network-based biomarker detection approach coupled with gene set enrichment analysis to identify and validate genes associated with lung cancer and relating pathways. Results revealed a panel of new and unexpected genes with potential physiological functions in cigarette smoking or lung cancer using the coupled discovery strategy, in addition to those consistent with previous findings in cigarette smoking and lung cancer. Thus, it was indicated that disease-specific network biomarkers, interaction between genes/proteins, or cross-talking of pathways give many other specific values for the development of precision therapies for lung.

Polat et al. [27] have detected lung cancer using principles component analysis (PCA), fuzzy weighting preprocessing and artificial immune recognition system

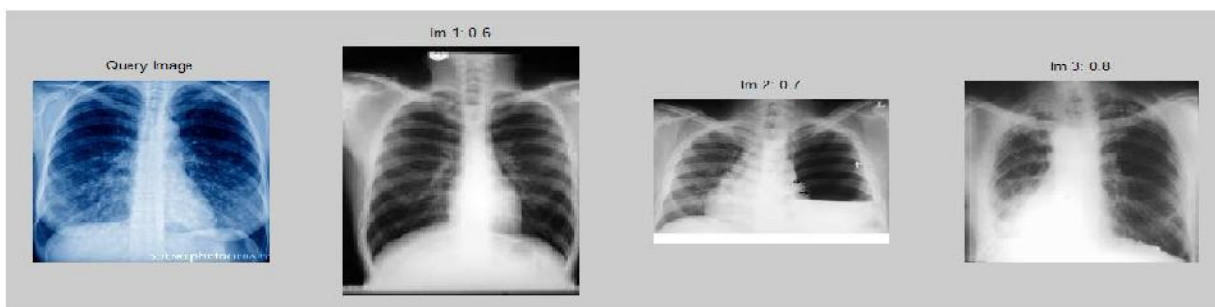
(AIRS). The system has three stages. First, dimensionality of lung cancer dataset that has 57 features was reduced to four features using principles component analysis. Second, a weighting scheme based on fuzzy weighting pre-processing was utilized as a pre-processing step before the important classifier. Third, artificial immune recognition system was used classifier. Experiments were done on the lung cancer dataset to diagnose lung cancer in a fully automatic manner. The obtained classification accuracy of system was 100% and it was very promising with regard to the other classification applications.

### III. METHODOLOGY

Each image sample is stored to a size of 512 X 512 pixels. Generally, the quality of image is affected by different artifacts due to non-uniform intensity, variations, motions, shift, as well as noise [3]. Thus, the pre-processing of image aims at selectively removing the redundancy present in scanned images without affecting the details which that play a key role in the diagnostic process. Hence, Histogram Equalization becomes the important step in preprocessing. So each image is preprocessed to improve its quality.

#### Classification

Classification is the process of classifying the cancerous images by extracting the features of the given image suffering from the cancer and these features are compared with the features of the given sample images. In this paper 35 sample images have been given for classification and the features of these images are compared with the given image and hence lung cancer is detected.



Classification of Cancerous Images

#### Dimensionality Reduction

The lung cancer dataset contains the information about the patients those who have lung cancer, other cancer diseases and no cancer disease and also the corresponding risk factors and symptoms of those patients. By using the ABC algorithm the size of the dataset is reduced in order to decrease the computational complexity.

#### Dataset

KDD refers to the non-trivial extraction of implicit, previously unknown and potentially useful information from data in databases.

| Age | Gender | Symptoms | Disease |
|-----|--------|----------|---------|
| 25  | Male   | Medium   | Yes     |
| 32  | Male   | High     | Yes     |
| 24  | Female | Medium   | Yes     |
| 44  | Female | High     | Yes     |
| 30  | Male   | Low      | No      |
| 21  | Female | Low      | No      |
| 18  | Male   | Low      | No      |
| 34  | Male   | Medium   | No      |

Memorize the best solution so far

Cycle=cycle+1

Until cycle=mcn

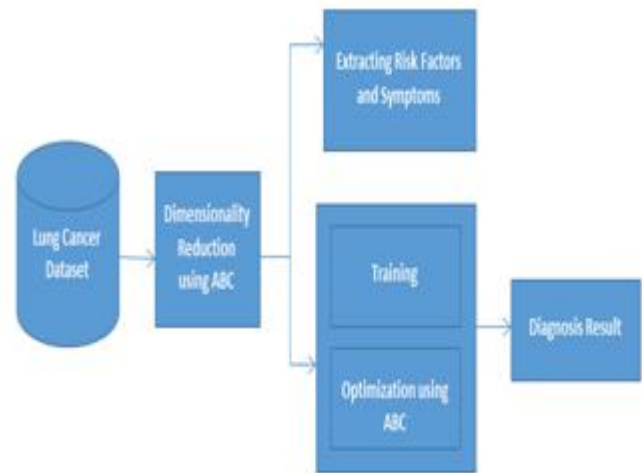
#### IV.PROPOSED TECHNIQUE

The performance of our proposed lung cancer pre-diagnosis system is analyzed by using the statistical measures which is given in [28].To carry out the performance examination process we make use of lung cancer dataset

#### Algorithm Specification

#### ABC ALGORITHM

The architecture of proposed lung cancer pre-diagnosis system is shown in Fig1. In the beginning, the dimensionality of the given dataset is reduced using Artificial Bee Colony (ABC) algorithm. After the dimensionality reduction, the reduced dataset are given as the input to the pre-diagnosis stage. Here, the high risk factors and symptoms which cause lung cancer are extracted from the dimensionality reduced dataset. These features are given as the input to the Feed Forward Back Propagation Neural Network (FFBNN).



#### Algorithm Specification

Load training samples

Generate the initial population

Evaluate the fitness

Set cycle to 1

Repeat

For each employed bee {

Produce new solution

Calculate the value of f

Apply greedy selection process}

Calculate the probability values

For each onlooker bee {

Select solution depending on probability

Produce new solution

Calculate the value of f

Apply greedy selection process}

If there abandoned solution

which has the patients list those who have lung cancer, other cancer and no cancer and also the corresponding risk factors and symptoms of those patients. By using the familiar ABC algorithm, the given lung cancer dataset size is reduced. The risk factors and the symptoms obtained from the reduced dataset are utilized to validate the performance of the result.

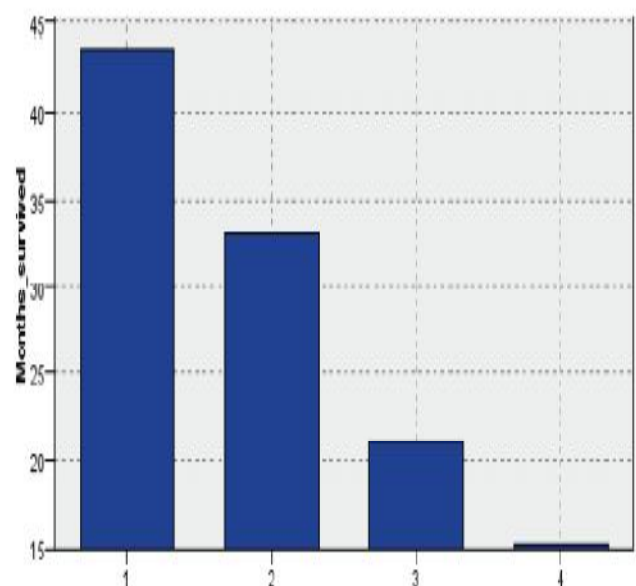


Fig 1: Survival time vs. Cancer grade

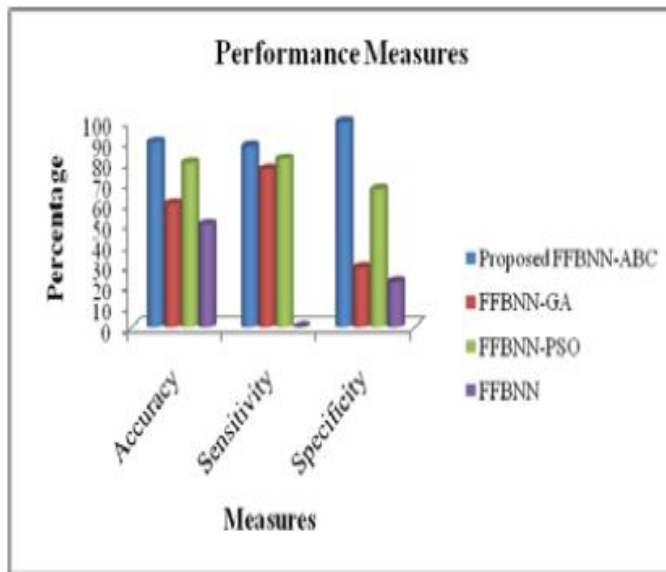


Fig 2: Performance Outcomes in Terms Of Accuracy, Sensitivity and Specificity

Occurrence of Lung cancer disease is high in India, especially in rural India, did not get noticed at the early stage, because of the lack of awareness utilizing such as accuracy, sensitivity and specificity are compared to FFBNN based to the PSO (Particle Swarm Optimization) , and Genetic Algorithm(GA) approaches are proposed in different category.

## V. CONCLUSION

The proposed system was implemented and a huge set of test data's were utilized to analyze the outcomes of the proposed lung cancer pre-diagnosis system. Thus the proposed lung cancer pre-diagnosis system offers a significant tempo of accuracy, sensitivity and specificity. We can say that proposed method more precisely diagnosis the lung cancer from the given test data by seeing the elevated rate of measurements. Therefore by utilizing the Feed Forward Back propagation Neural Network (FFBNN) and ABC techniques, our proposed lung cancer pre-diagnosis system proficiently diagnoses the lung cancer.

## REFERENCES

- [1] "An Approach for Discretization and Feature Selection of Continuous-Valued Attributes in Medical Images for Classification Learning".
- [2] Basavaraj .S. Anami<sup>1</sup> and Vishwanath.C.Burkpalli 2  
1. Principal, K.L.E.Institute of Technology, Hubli-580030, India 2.Research Scholar, Basaveshwar Engineering College, Bagalkot – 587102, India.
- [3] "Texture based Identification and Classification of Bulk Sugary Food Objects", ICGST-GVIP Journal, ISSN: 1687-398X, Volume 9, Issue 4, August 2009.
- [4] Bing Gong, School of Computer Science and Technology Heilongjiang University Harbin, China, "A Novel Learning Algorithm of Back-propagation Neural Network" ,2009 IITA International Conference on Control, Automation and Systems Engineering.
- [5] ZAKARIA SULIMAN ZUBI<sup>1</sup>, REMA ASHEIBANI SAAD, "Using Some Data Mining Techniques for Early Diagnosis of Lung Cancer", Recent Researches in Artificial Intelligence, Knowledge Engineering and Data Bases, LIBYA, 2011.
- [6] "Digital Image Processing" by Gonzalez & Woods 2nd edition
- [7] H. Tamura, S. Mori, and T. Yamawaki, "Texture features corresponding to visual perception," IEEE Trans. On Systems, Man, and Cybernetics, vol. Smc-8, No. 6, June 1978.
- [8] J. Smith and S.-F. Chang, "Transform features for texture classification and discrimination in large image database".IEEE Intl. Conf. on Image Proc., 1994.
- [9] Castleman K R. "Digital image processing". NJ: Prentice Hall, 1996.
- [10] Manjunath, B., Ma, W.: "Texture features for browsing and retrieval of image data". IEEE Trans on Pattern Analysis and Machine Intelligence 18 (1996) 837842
- [11] Alexander Suhre, A. Enis Cetin, TulinErsahin, Rengul Cetin-Atalay, "Classification of cell images using a generalized harris Corner Detector"
- [12] R. Adams and L. Bischof, "Seeded Region Growing", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 16, pp. 641-647, 1994.
- [13] R. Haralick, "Statistical and structural approaches to texture", Proceedings of the IEEE, vol. 67, pp. 786–804, 1979.

- [14] C. Harris and M.J. Stephens. “A combined corner and edge detector”. In *Alvey Vision Conference*, pages 147–152, 1988. D.
- [15] Ballard and C. Brown, “Computer Vision”, Prentice-Hall, Inc., 1982, Chap. 6. [19]. Davies, “Machine Vision: Theory”, Algorithms and Practicalities, Academic Press, 1990, Chap. 18.
- [16] A K Jain, “Fundamentals of Digital Image Processing”, Prentice-Hall, 1986, Chap. 9.
- [17] D. Vernon Zhi-Hua Zhou, Yuan Jiang, Yu-Bin Yang, Shi-Fu Chen , “Lung Cancer Cell Identification Based on Artificial Neural Network Ensembles”, National Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093, P.R.China. *Artificial Intelligence in Medicine*, 2002, vol.24, no.1, pp.25-36. @Elsevier.
- [18] Leonard Fass , “Imaging and cancer: A review”, GE Healthcare, 352 Buckingham Avenue, Slough, SL1 4ER, UK Imperial College Department of Bioengineering, London, UK.
- [19] Jingtangshan, P. R. China , “Application of Neural Networks in Medical Image Processing” ISBN 978-952-5726-09-1, Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10), 2-4, April. 2010, pp. 02.
- [20] Liu Lu and Liu Wanyu, "A Method of Pulmonary Nodules Detection with Support Vector Machines", In *IEEE Proceedings of the Eighth International Conference on Intelligent Systems Design and Applications*, Kaohsiung, pp. 32-35, 2008.
- [21] Ying-Wooi Wan and Guo, “Constructing gene-expression based survival prediction model for Non-Small Cell Lung Cancer (NSCLC) in all stages and early stages”, In *IEEE Proceedings of International Conference on Bioinformatics and Biomedicine Workshop*, Washington, pp. 338, 2009.
- [22] Xiaocong Fang, Michael Netzer, Christian Baumgartner, Chunxue Bai and Xiangdong Wang, “Genetic network and gene set enrichment analysis to identify biomarkers related to cigarette smoking and lung cancer”, *Cancer Treatment Reviews*, Vol. 39, No. 1, pp. 77–88, 2013.
- [23] Kemal Polat and Salih Gunes, “Principles component analysis, fuzzy weighting pre-processing and artificial immune recognition system based diagnostic system for diagnosis of lung cancer”, *Expert Systems with Applications*, Vol.34, No. 1, pp. 214–221, 2008.