Use of Fuzzy C-Mean and Fuzzy Min-Max Neural Network in Lung Cancer Detection

Kiran S. Darne, Suja S. Panicker

Abstract*— Lung cancer is a disease characterized by uncontrolled cell growth in tissues of the lung and is the most common fatal malignancy in both men and women. Early detection and treatment of lung cancer can greatly improve the survival rate of patient. Artificial Neural Network (ANN), Fuzzy C-Mean (FcM) and Fuzzy Min-Max Neural network (FMNN) are useful in medical diagnosis because of several advantages. Like ANN has fault tolerance, flexibility, non linearity, while FcM gives best result for overlapped data set, data point may belong to more then one cluster center and always converges .and , also, FMNN has advantages like online adaptation, non-linear separability, less training time, soft and hard decision. In this work, we propose to use FcM and FMNN on standard datasets, to detect lung cancer.

Index Terms*— Classification, Clustering, Fuzzy System, FCM, FMNN.

I. INTRODUCTION

Cancer is a leading cause of death worldwide. Lung cancer is a type of cancer that is considered as one of the most leading causes of death globally. [8] The early detection of the lung cancer is a challenging problem, due to the structure of the cancer cells. [2] Lung cancer is one of the most common cancers. Early detection and early diagnosis can significantly increase the survival rate. [3]

Lung cancer is a disease characterized by uncontrolled cell growth in tissues of the lung and is the most common fatal malignancy in both men and women. Lung cancer is considered to be as the main cause of cancer death worldwide, and it is difficult to detect in its early stages because symptoms appear only at advanced stages causing the mortality rate to be the highest among all other types of cancer. More people die because of lung cancer than any other types of cancer such as: breast, colon, and prostate cancers. There is significant evidence indicating that the early detection of lung cancer will decrease the mortality rate. As indicated by the latest statistics provided by World Health Organization, around 7.6 million deaths worldwide occur each year because of this type of cancer. Furthermore, mortality from cancer are expected to continue rising, to become around 17 million worldwide in 2030. [2]

Diagnosis and Prognosis are important terms related to disorders. Medical Diagnosis refers both to the process of attempting to determine or identify a possible disease or disorder and to the opinion reached by this process. From the point of view of statistics the diagnostic procedure involves classification tests. Prognosis is the projected course a disease will take, and can only be done after a diagnosis is made of particular condition.

For a person diagnosed with cancer, the prognosis may be grim, predicting death in only a few short months or weeks. Without a full knowledge of a particular condition it is not possible to make an accurate prognosis.

In the medical fraternity, several techniques such as Chest Radiograph (x-ray), Computed Tomography (CT), Magnetic Resonance Imaging (MRI scan) and Sputum Cytology exist to diagnose Lung Cancer.

From Machine Learning and Data Mining the following tools have been employed in successful classification - Artificial Neural Network, Decision tree, Bayesian Network, association rule mining, Clustering, Classification. With its ability to perform complex computations, and reproduce some of the flexibility and power of the human brain by artificial means; Artificial Neural Networks (ANN) have been used in various applications for successful detection of cancers. Such computations are performed by a dense mesh of computing nodes (neurons) and weighted connections, which operate collectively and simultaneously on most or all data and inputs. Neurons usually operate in parallel and are configured in regular architectures. They are often organized in layers, and feedback connections both within the layer and toward adjacent layers are allowed. [1]

II. RELATED WORK

In our previous work [1], we have surveyed the various Neural Networks that have been used in successful classification of medical data for various disorders. Examples include -Feed Forward Neural Network, Radial Basis Function (RBF) Network, Kohonen self-organizing network, Fuzzy Neural Network, Probabilistic Neural Network.

For the successful detection of Lung Cancers, various methods have been surveyed. They are as follows:

To detect the lung cancer in its early stages Hopfield Neural Network (HNN) and Fuzzy C-Mean (FCM) clustering algorithm were used for segmenting sputum color images. HNN showed better classification result than FCM, and it succeeded in extracting the nuclei and cytoplasm regions. However FCM failed in detecting the nuclei, instead it detected only part of it; and was not sensitive to intensity variations. [2]

With the development of medical technology, the medical images play a more and more important role in diagnosis. X-ray, CT and MRI images have widely been applied to diagnosis. [13]

ANN has the potential to improve the diagnostic accuracy. Biochemical Diagnosis, Imaging Diagnosis and Cytology Histology diagnosis are the three main methods for lung cancer diagnosis. Imaging diagnosis includes X-ray imaging, CT, MRI, angiography and interventional radiology.

Manuscript received on July, 2013.

Kiran S. Darne, Computer Engineering, Pune University/ MIT College, Pune, India.

Suja S. Panicker, Computer Engineering, Pune University/ MIT College, Pune, India.

CT is an important medical imaging method employing tomography, which is one of the most widely used for the diagnosis of lung cancer. The result of ANN was compared with those of logistic regression by ROC curve analysis. The diagnostic accuracy of ANN and logistic regression among all samples of the training group and test group were 96.6% and 84.6%. [3]

In [4] a Radial Basis Neural Network (RBFN) was used for lung cancer screening. Because of its learning characteristics it was selected to train the samples and then extract the internal relation between the pathogenic factors and inducing lung cancer, and eventually it generates empirical function and forecasts the new samples. The training function adopted Linear Least Square method (LLS) and the Gradient Descent hybrid learning algorithm to optimize the training process and the screening results. The accuracy of lung cancer identification was 95.32%. [4]

In order to identify characteristics of patient segments where average survival is significantly higher/lower than average survival across the entire dataset, Association Rule Mining techniques were used for the identification of hotspots from lung cancer data. Automated association rule mining techniques resulted in hundreds of rules, from which many redundant rules were manually removed based on domain knowledge. The resulting rules conformed to existing biomedical knowledge and provided interesting insights into lung cancer survival. The Hotspot algorithm is an association rule mining algorithm which is directed by a target attribute, which means that the consequent is fixed to the target attribute. It can be used for segmentation with both nominal and numeric targets. [5]

Computer Aided Diagnosis (CAD) are procedures in medicine that assist doctors in the interpretation of medical images. Imaging techniques in X-ray, MRI, and Ultrasound diagnostics yield a great deal of information, which the radiologist has to analyze and evaluate comprehensively in a short time. A relatively young interdisciplinary technology, CAD combines the elements of Artificial Intelligence and Digital Image Processing with radiological image processing.[6] and [7] use CAD for lung cancer.

The design and development of a two stage CAD system that can automatically detect and diagnose histological images such as CT scan of lung with a nodule into cancerous or noncancerous nodule was done in [6]. In the first stage the input image is pre-processed and the cancerous nodule region is segmented. Second stage involves diagnosis of the nodes based on fuzzy system and the grey level of the nodule region. While maintaining a high degree of true-positive diagnosis, this proposed method attained an accuracy of 90% and also high detection sensitivity and specificity, which meets the basic requirements of clinical diagnosis [6]

Machine Learning techniques were utilized to develop a CAD system, which consisted of Feature Extraction phase, Feature Selection phase and Classification phase. Different wavelets functions have been used in Feature Extraction/Selection, to find the one that produces the highest accuracy. Clustering K-nearest-neighbor algorithm has been developed utilized for classification. Testing was done using Japanese Society of Radiological Technology's standard dataset of lung cancer. Of the 154 nodule regions (abnormal) and 92 non-nodule regions (normal), accuracy levels of over 96% was achieved for classification.[7]

A new CAD scheme was proposed in [8] used for detecting lung nodules Firstly, the lung region is segmented from the CT data using methods such as adaptive threshold algorithm. An active contour model was built to segment and remove lung vessel accurately in the lung region. The suspicious nodules were detected and omitted renal vessel was filtered using a selective shape filter. Finally nodule features were extracted and rule-based classifier was used to distinguish true or false positive nodules. Overall detection rate was 85%, false positives at approximately 2/scan. These results indicate that this scheme can help radiologist to improve the diagnosis efficiency. [8]

The lung cancer and control group were detected using laser induced auto-fluorescence spectra of human serum in. Fluorescence is caused by absorbing electromagnetic radiation and emitting light with usually a longer wavelength, thus fluorescence of serum reflected the chemical component change. Advantages of fluorescence detection are its sensitivity, celerity, and safety. Shortcomings of this approach include low sensitivity of fluorescence and low concentration of serum. The intensity ratio of the three wavelength position of fluorescence spectroscopy of lung cancer groups and controls were used for Linear Discriminant Analysis (LDA). A total diagnosis accuracy of 83.3% was obtained. [9]

A nonlinear anomaly detector called Kernel RX-algorithm was applied to CT images for malignant nodule detection, which is very similar to anomaly detection in military imaging applications wherein RX-algorithm has been successfully applied. In [9] the original RX-algorithm was modified so that it can be applied to anomaly detection in CT images. Moreover, this kernel trick was used to map the data to a high dimensional space to obtain a kernelized RX-algorithm that outperforms the original RX algorithm. Results have suggested that the proposed approach could be an efficient technique for early lung cancer detection. [10]

Despite various efforts to develop new predictive models for early detection of tumor local failure in locally advanced Non-Small Cell Lung Cancer (NSCLC), many patients still suffer from a high local failure rate after radiotherapy. A graphical Bayesian network framework was used for predicting such local failures. Testing was done using a dataset of locally advanced NSCLC patients treated with radiotherapy. Experimental results demonstrate that this is an efficient method to develop predictive models of local failure in these patients and to interpret relationships among the different variables. The combined model of physical and biological factors outperformed individual physical and biological models, achieving accuracy of 87.78%, Matthew's correlation coefficient (r) of 0.74, and Spearman's rank correlation coefficient of 0.75 on leave-one-out cross-validation analysis. [11]

A revised Group Method of Data Handling (GMDH)-type neural network algorithm was used for medical image diagnosis and was applied to medical image diagnosis of lung cancer. In this algorithm, the knowledge base for medical image diagnosis is used for organizing the neural network architecture for medical image diagnosis, and this revised GMDH-type algorithm can identify the characteristics of the medical images accurately. The optimum neural network architecture fitting the complexity of the medical images is automatically organized so as to minimize the prediction error criterion defined as Prediction Sum of Squares (PSS), and it is shown that the revised GMDH-type neural network can be easily applied to the medical image diagnosis. With the conventional Neural Network, many different output images were obtained for various structural parameters and several iterative calculations of Back propagation are needed to find more accurate neural network architecture. Results indicate that the revised GMDH-type NN algorithm was accurate, and hence a useful method for the medical image diagnosis of the lung cancer. [12]

Estimation of the volume of the lungs and the viable lung tissue is an important step in the management of patients with severe pulmonary disease. The presence of gross pathology makes it impossible to perform lung segmentation automatically and reliably in CT scans of such patients. A statistical classifier classified these regions beforehand and the user corrects any errors until the segmentation of an entire slice is correct. The resulting lung segmentations showed a large overlap and a small average boundary distance when compared to completely manual delineations of the lung borders. [13]

A Parallel Immune Algorithm (IA) has been used for the detection of lung cancer in chest X-ray images based on object shared space. Template matching method is combined to the algorithm and Java Spaces is used as object shared space. Results show that the algorithm is efficient for detecting suspected lung cancer.[14]

III. PROPOSED WORK

Based on the above survey, in our current work we propose to employ Fuzzy Min-Max Classifier and Fuzzy c Means to the standard datasets of Lung Cancer, and thereby determine better solutions to the problem.

In the previous work [17] aimed at solving the Fault Diagnosis problem on the two standard datasets - UCI and NASA, a high classification accuracy was yielded by FMN ; which has been a major motivation to use FMN for Lung Cancer Detection problem.

Fuzzy Min-Max classification Neural Network is a supervised NN classifier that forms hyperbox fuzzy sets for learning and classification. Learning is an expansion/contraction process of the hyperboxes. An ordered pair is selected from the Training Data set, and the task is to find a hyperbox belonging to the same class that can expand (if necessary) to include the input pattern. If no such hyperbox is found, a new hyperbox is formed for that class label and is added to the Neural Network. Whether a hyperbox can expand, is determined by the Expansion Criteria. If this criterion is satisfied, the min and max points are updated. Whenever a hyperbox expands, or whenever a new hyperbox is created, it could lead to overlap. Hence a hyperbox Overlap Test is carried out for each dimension of input, to determine the possible overlap between hyperboxes belonging to different classes. Of all dimensions, the dimension that gives the minimum overlap value is determined; the latter is eliminated using a Contraction process. [15] Fig.2 shows the three layer architecture.

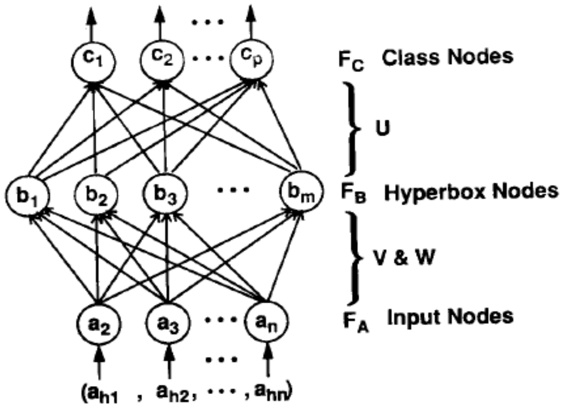

Fig2. Architecture of Fuzzy Min Max NN

Fuzzy C-Mean clustering is a method of clustering which allow one piece of data to belong to two or more clusters. Membership is assigned to each data point corresponding to each cluster center on the basis of distance between the cluster center and the data point. After each iteration, membership and cluster center are updated. Main objective of Fuzzy c- Mean algorithm is to minimize the objective function. [16]

FCM with its advantages such as – giving best result for overlapped data sets, allowing one piece of data to belong to two or more clusters[16] and FMN with advantages such as online adaptation, Non-linear separability, Less training time, Soft and hard decisions [15]; shall be applied to the standard Lung Cancer datasets separately and the results shall be noted.

Table I shows summary of the survey work done.

Table I: Performance of various methods proposed for lung cancer diagnosis

Sr. No.	Author & Year Published	Methodology	Datasets	Performance
1	Fatma Taher & Rachid Sammouda [2011]	HNN & FCM	100 sputum color images	HNN segmentation results were more accurate than FCM
2	Yongjun WU, Na Wang, Hongsheng ZHANG, Lijuan Qin, Zhen YAN, Yiming WU [2010]	CAD scheme of the CT and ANN	117 CT images of pulmonary nodules (58 benign and 59malignant)	ANN showed 96.6 % accuracy & Logistic Regression gave 84.6% accuracy.

3	Wang Tao, Lv Jianping and Liu Bingxin, [2011]	RBFN, LLS, gradient descent	5,000 valid data as samples	95.32% accuracy
4	Ankit Agrawal and Alok Choudhary [2011]	Association rule mining & hotspots	A subset of 13 patient attributes from the SEER data	Survival time of patients increased.
5	S. Aravind Kumar, Dr. J.Ramesh, Dr. P.T.Vanathi, Dr.K.Gunavathi [2011]	CAD, Segmentation, Fuzzy Systems	40 clinical cases containing 685 slice images.	90% classification accuracy
6	Xiaozhou Li, Rong Wang and Ming Lei [2011]	Fluorescence Spectroscopy	36 serums	83.3% classification accuracy
7	Hamada R. H. AI-Abs, Brahim Belhaouari Samir, Khaled Bashir Shaban, and Suziah Sulaiman [2012].	Machine Learning techniques.	247 chest Radiographs (154 images containing nodules and 93 normal images)	96% classification accuracy
8	Aminmohammad Roozgard, Samuel Cheng, and Hong Liu [2012]	kernel RX-algorithm	CT images from of size 512 x 512	Proved efficient.
9	Jung Hun Oh, Jeffrey Craft, Rawan Al-Lozi, Manushka Vaidya [2010]	Graphical Bayesian network framework	The dataset consisted of physical variables and four biomarker proteins	87.78% accuracy.
10	Tadashi Kondo, Junji Ueno and Shoichiro Takao [2012]	GMDH	medical images	Diagnosis efficiency can be improved.
11	Thessa T.J.P. Kockelkorn, Eva M. van Rikxoort, Jan C. Grutters and Bram van Ginneken [2010]	Computed tomography, lung segmentation.	12 thoracic CT scans from lung transplantation and ILD patients.	Useful when automatic segmentation methods fail.
12	Peng Gang, Yang Xiong, Liu Li [2011]	Immune algorithm	Chest X-ray medical images	Efficient for detecting suspected lung cancer.
13	Jia Tong, Wei Ying, Wu Cheng Dong [2010]	CAD, lung vessel Segmentation	90 thoracic CT scans	Overall detection rate of 85%

The main idea is to compare their results, and hence suggest better solution to detect lung cancer. It is proposed to test the above algorithms on standard datasets, however we are limiting to only numerical data.

In our current work, Lung cancer dataset from mldata repository [22] will be used for testing. It contains 137 instances and 8 attributes including – Treatment, Cell type, Status, Kernofsky, Months, Age, Therapy, Survival. CLASS

This dataset has been normalized using Min Max method and the same shall be used to further generate the Training & Testing Datasets.

IV. CONCLUSION

We have done a detailed survey about several methods used in the past for Lung Cancer detection, and a novel way for detection using Fuzz and Fuzzy c Mean is proposed in our current work. Both FMN and Fcm have yielded good in previous classification occur independent works, hence we aim to study the independent results of FMN and fcm on the standard datasets; after which a improved version should be available.

FUTURE SCOPE

Present work is restricted to numerical data. Use of Lung Cancer Image Datasets can be used in the future to test the results.

Abbreviation:

HNN- Hopfield Neural Network
FCM- Fuzzy C-Mean
ANN- Artificial Neural Network
CT- Computed Tomography
RBFN- Radial Basis Neural Network
CAD- Computer Aided Diagnosis
GMDH- Group Method of Data Handling

REFERENCES

1. Kiran S. Darne & Suja S. Panicker, "Use of Artificial Neural Networks in Medical Data Classification: A Review", International Conference on Advanced Computer Sciences, Communication and Information Technologies (ICACSIT), 2012.
2. Fatma Taher and Rachid Sammouda, "Lung Cancer Detection by Using Artificial Neural Network and Fuzzy Clustering Methods", IEEE GCC conference and exhibition, 2011.
3. Yongjun WU, Na Wang, Hongsheng ZHANG, Lijuan Qin, Zhen YAN and Yiming WU, "Application of Artificial Neural Networks in the Diagnosis of Lung Cancer by Computed Tomography", IEEE conference on Natural Computation, 2010.
4. Wang Tao, Lv Jianping and Liu Bingxin, "Research of Lung Cancer Screening Algorithm Based On RBF Neural Network", International Conference on Computer and Management (CAMAN), 2011.
5. Ankit Agrawal and Alok Choudhary, "Identifying HotSpots in Lung Cancer Data Using Association Rule Mining", 11th IEEE International Conference on Data Mining Workshops, 2011.

6. S.Aravind Kumar, Dr.J.Ramesh, Dr.P.T.Vanathi, Dr.K.Gunavathi, "Robust and Automated Lung Nodule Diagnosis from CT Images Based On Fuzzy Systems" International Conference on Process Automation, Control and Computing (PACC), 2011.
7. Hamada R. H. AI-Abs, Brahim Belhaouari Samir, Khaled Bashir Shaban, and Suziah Sulaiman, "Computer Aided Diagnosis System based on Machine Learning Techniques for Lung Cancer" International Conference on Computer & Information Science (ICCIS), 2012.
8. Jia Tong, Wei Ying, Wu Cheng Dong, "A Lung Cancer Lesions Detection Scheme Based on CT Image", 2nd International Conference on Signal Processing Systems (ICSPS),2010.
9. Xiaozhou Li, Rong Wang and Ming Lei, "Analysis on data fordetection of Lung cancer using serum auto-fluorescence", International Symposium on IT in Medicine and Education (ITME), 2011.
10. Aminmohammad Roozgard, Samuel Cheng, and Hong Liu, "Malignant Nodule Detection on Lung CT Scan Images with Kernel RX –algorithm", Proceedings of the IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI 2012), 2012.
11. Jung Hun Oh, Jeffrey Craft, Rawan Al-Lozi, Manushka Vaidya, "Predicting local failure in lung cancer using Bayesian networks" Ninth International Conference on Machine Learning and Applications,2010.
12. Tadashi Kondo, Junji Ueno and Shoichiro Takao, "Medical Image Diagnosis of Lung Cancer by Hybrid Multi-layered GMDH-type Neural Network Using Knowledge Base", Proceedings of ICME International Conference on Complex Medical Engineering,2012.
13. Thessa T.J.P. Kockelkorn, Eva M. van Rikxoort, Jan C.Grutters and Bram van Ginneken, "Interactive Lung Segmentation in CT Scans with Severe Abnormalities"2010.
14. PENG Gang, YANG Xiong, LIU Li , "Parallel Immune Algorithm for Lung Cancer Detection in X-Ray Images Based on Object Shared Space" 12th International Conference on Parallel and Distributed Computing, Applications and Technologies, 2011.
15. Patrick K. Simpson, "Fuzzy Min-Max Neural Networks Parts 1: Classification", IEEE Transaction on Neural Network, Vol 3, No.5, 1992.
16. James C. Bezdek, Robert Ehrlich and William Full , "FCM: The Fuzzy C-Means Clustering Algorithm" IEEE Transactions on Computers and Geosciences, Vol 10, No. 2-3, 1984.
17. Suja S.Panicker, P.S. Dhabe, M Dhore "Fault Diagnosis using Fuzzy Min Max Network Classifier", International Journal of Artificial Intelligent Systems and Machine Learning, July 2010.
18. http://www.cancer.gov
19. https://sites.google.com/site/dataclusteri ngalgorithms/home
20. http://www.learnartificialneuralnetworks.com/
21. http://www.topnews.in/health/diseases/lung-cancer?page=2
22. mldata.org/repository/Data/viewslug/datasets-numeric-veteran/