

An Early Diagnosis System for Predicting Lung Cancer Risk Using Adaptive Neuro Fuzzy Inference System and Linear Discriminant Analysis

Mustain Billah* and Nazrul Islam

Department of Information and Communication Technology, Mawlana Bhashani Science and Technology University, Tangail, Bangladesh

*Corresponding author: Mustain Billah Department of Information and Communication Technology Mawlana Bhashani Science and Technology University Tangail, Bangladesh, Tel: +8801929175180; E-mail: mustainbillahx@gmail.com

Rec date: May 18, 2016; Acc date: June 16, 2016; Pub date: June 23, 2016

Copyright: © 2016 Billah M et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Citation: Billah M et al. Zika virus disease (ZIKV) Infographics, language-tailored: Innovations and Good Practices in Integrated Disease Surveillance Programme Kashmir, India 2016. J Mol Pathol Epidemiol. 2016, 1:1.

Abstract

Lung cancer is the number one cause of cancer deaths in both men and women worldwide. The general prognosis of lung cancer is poor because doctors tend not to and the disease until it is at an advanced stage. But it is very necessary to diagnose the disease early for taking preventive steps. In this paper, an Adaptive Neuro Fuzzy Inference System (ANFIS) and Linear Discriminant Analysis (LDA) based lung cancer diagnosis system is proposed. This diagnosis system has mainly two steps: Feature extraction-reduction and classification. First, lung cancer historical data sets are collected from different hospitals. They are then preprocessed. To reduce the lung cancer features dimensionality, Linear Discriminant Analysis (LDA) is applied. Reduced features are then fed into AN-FIS classifier system. Classification accuracy, sensitivity and specificity analysis are performed for performance evaluation of proposed system. Obtained accuracy of about 95.4% shows that the proposed intelligent system has a good diagnosis performance and can be used as a promising tool for lung cancer diagnosis.

Keywords: Adaptive neuro fuzzy inference system; Linear discriminant analysis; Lung cancer; Sensitivity and specificity analysis; Classification accuracy

Introduction

Lung cancer is a type of cancer that begins in the lung. Actually it is the uncontrolled cell growth in tissues of the lung. If left untreated, this growth can spread beyond the lung into nearby tissue or other parts of the body. Lung cancer is the leading cause of cancer deaths in the United States, among both men and women. Lung cancer claims more lives each year than do colon, prostate, ovarian and breast cancers combined.

There are three main types of lung cancer Non-Small Cell Lung Cancer is the most common type of lung cancer. About 85% of lung cancers are non-small cell lung cancers. About 10%-15% of lung cancers are small cell lung cancers. This type of lung cancer tends to spread quickly. Fewer than 5% of lung cancers are lung

carcinoid tumors. They are also sometimes called lung neuroendocrine tumors. Most of these tumors grow slowly and rarely spread.

However, lung cancer is also the most preventable cancer. Lung cancer symptoms usually do not appear until the disease has progressed. So, cure rate and prognosis depend on the early detection and diagnosis of the disease. But this early diagnosis is not an easy task. Whenever, a patient goes to hospital, doctor asks about medical history and symptoms and arranges for some tests such as CT scan that takes X-Ray through sections of body.

A physician commonly takes decisions by evaluating the current test results of a patient or the physician compares the patient with other patients with the same condition by referring to the previous decisions. Therefore, it is very difficult for a physician to diagnose lung cancer. For this reason, an early diagnosis system for predicting lung cancer risk is proposed in this study to help the physician in the diagnosis of lung cancer diseases. Linear Discriminant Analysis (LDA) and Adaptive Neuro Fuzzy Inference System (ANFIS) is used in this study. ANFIS has been used for many feature selection purposes in many papers [1-7]. LDA-ANFIS based system has been used for various disease diagnoses such as hepatitis, diabetes but there is no work on lung cancer diagnosis using LDA-ANFIS method.

Various research works are found on lung cancer diagnosis [8-11]. A comparative analysis on these papers and proposed system is presented in this study also. In the paper [7] authors found out some features of lung cancer. Computer aided diagnosis system using image processing was used for lung nodule detection in [8]. [9] Proposed computerized scheme for automated detection of lung nodule.

The paper is organized as follows: section 2 gives a brief description of theory of pattern recognition, LDA and ANFIS. Proposed system is explained in section 3. In section 4 the obtained results are presented. Finally, in Section 5, the discussion and conclusion are presented.

Theory

Pattern recognition

It is generally easy for a person to differentiate the sound of a human voice, from that of a violin; a handwritten numeral "3," from an "8"; and the aroma of a rose, from that of an onion. However, it is difficult for a programmable computer to solve these kinds of perceptual problems. These problems are difficult because each pattern usually contains a large amount of information, and the recognition problems typically have an inconspicuous, high-dimensional, structure.

Pattern recognition [12] is the science of making inferences from perceptual data, using tools from statistics, probability, computational geometry, machine learning, signal processing, and algorithm design.

Pattern is defined as composite of features that are characteristic of an individual. In classification, a pattern is a pair of variables fix, wg where x is a collection of observations or features (feature vector) and w is the concept behind the observation (label). The quality of a feature vector is related to its ability to discriminate examples from different classes. Examples from the same class should have similar feature values and while examples from different classes having different feature values.

If the characteristics or attributes of a class are known, individual objects might be identified as belonging or not belonging to that class. The objects are assigned to classes by observing patterns of distinguishing characteristics and comparing them to a model member of each class. Pattern recognition involves the extraction of patterns from data, their analysis and, finally, the identification of the category (class) each of the patterns belongs to.

Linear discriminant analysis

Linear Discriminant Analysis (LDA) is a method of finding such a linear combination of variables which best separates two or more classes [13]. In itself LDA is not a classification algorithm, although it makes use of class labels. However, the LDA result is mostly used as part of a linear classifier. The other alternative use is making a dimension reduction before using nonlinear classification algorithms. LDA seeks to reduce dimensionality while preserving as much of the class discriminatory information as possible [14].

Adaptive neuro fuzzy inference system

ANFIS [15] derives its name from adaptive neuro-fuzzy inference system. It works similarly to that of neural networks. Using a given input/output data set, ANFIS constructs a fuzzy inference system (FIS) whose membership function parameters are tuned (adjusted) using either a back propagation algorithm alone or in combination with a least squares type of method. This adjustment allows the fuzzy systems to learn from the data they are modeling [16] (Figure 1).

Assume that the fuzzy inference system has two inputs x and y and one output f . A first-order Surgeon fuzzy model has rules as the following:

Rule1

If x is A_1 and y is B_1 , then $f_1 = p_1x + q_1y + r_1$

Rule2

If x is A_2 and y is B_2 , then $f_2 = p_2x + q_2y + r_2$

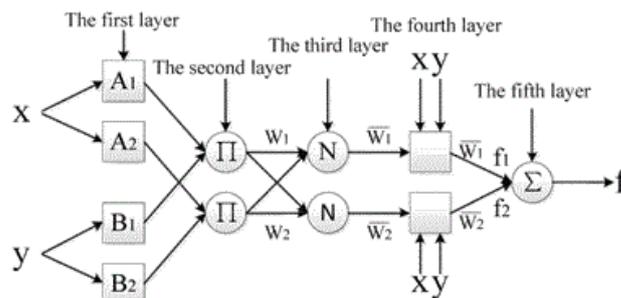


Figure 1: Adaptive Neuro Fuzzy Inference System with two inputs.

Here,

$$\text{output} = \frac{w_1.f_1 + w_2.f_2}{w_1 + w_2}$$

Proposed LDA-ANFIS early diagnosis system of lung cancer has mainly two stages: Feature Extraction and reduction using LDA and classification with ANFIS.

Feature extraction - reduction stage

First, clinical data sets of 1000 lung cancer patients diagnosed in different hospitals were extracted. Each patient record contains more than 20 features. As the dataset contains a large number of feature variables, Linear Discriminant Analysis (LDA) was conducted to this dataset and features are reduced to only 9 features Table 1.

In this lung cancer database, there are two classes: out of risk and very high risk.

Classification with ANFIS

In the classification stage, two class features obtained in the feature extraction and feature reduction stages are given to inputs of Adaptive Network based On Fuzzy Inference System (ANFIS) classifier.

In the structure of the ANFIS classifier, both artificial neural network and fuzzy logic are used. ANFIS classifier is formed of if-then rules, couples of input-output and learning algorithms of neural network. These are used in the training of ANFIS classifier. In this experimental study, the ANFIS classifier has nine inputs ($x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8,$ and x_9) and one output (y). The parameters of the ANFIS classifier used in this study can be given as below

The number of layers: 5 Input: 9

The number of rules: 512 Output: 1

Input membership functions: Bell-shaped Training parameters
learning rule: Hybrid Learning Algorithm

Sum-squared error: 0.0000001

Reaching Epochs number to sum-squared error: 2200

Table 1: Reduced features of lung cancer dataset using LD

No	Feature	Values
1	Hemoptysis	Yes,No
2	loss of weight	Yes,No
3	loss of appetite	Yes,No
4	Chest or rib pair	Yes,No
5	Fatigue	Yes,No
6	severe cough	Yes,No
7	finger clubbing	Yes,No
8	Thromocytosis	Yes,No
9	Abnormal spirometer	Yes,No

The results obtained using LDA-ANFIS

The correct diagnosis performance of LDA-ANFIS based lung cancer risk diagnosis system is estimated by using sensitivity and specificity analysis and classification accuracy performance evaluation methods, respectively.

In this experimental study, 800 samples were used for training and another 200 samples for testing the system. Figure 2 shows the confusion matrix of output class and target class. Here, '0' represents the class 'out of risk' and '1' represents the class 'high cancer risk'.

		Confusion Matrix		
Output Class	0	275 27.5%	21 2.1%	92.9% 7.1%
	1	25 2.5%	679 67.9%	96.4% 3.6%
		0	1	95.4% 4.6%
		Target		

Figure 2: Confusion matrix for two classes.

From this confusion matrix, sensitivity, specificity and accuracy can be measured. Here, TP: True Positive; TN: True Negative; FP: False Positive; FN: False Negative.

Sensitivity and specificity analysis

Sensitivity (also called the true positive rate, or the recall in some fields) measures the proportion of positives that are correctly identified.

TP

$$\text{Sensitivity} = \frac{TP}{TP + FN}(\%) \quad (2)$$

Specificity (also called the true negative rate) measures the proportion of negatives that are correctly identified.

TN

$$\text{Specificity} = \frac{TN}{FP + TN}(\%) \quad (3)$$

The obtained values of sensitivity and specificity of proposed LDA-ANFIS based early lung cancer risk diagnosis system are given in Table 2.

Table 2: Values of sensitivity and specificity

	Sensitivity	Specificity
PCA-AIRS(Polat and Gunes, 2006)[10]	100	94.44
LDA-ANFIS(Esin, Akif and Derya, 2009)[17]	96.66	91.66
Proposed LDA-ANFIS	96.44	92.90

Classification accuracy analysis

Accuracy of a measurement system is the degree of closeness of measurements of a quantity to that quantity's true value.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN}(\%)$$

The classification accuracy of proposed LDA-ANFIS based early lung cancer risk diagnosis system is given in Table 3. From the Table 3 above, it is clear that proposed system has higher accuracy than other methods [17,18].

Table 3: Classification Accuracy of different models

Used method	Authors Name	Classification Accuracy(%)
PCA-ANFIS[18]	Polat and Gunes	89.47
PCA-AIRS[10]	Polat and Gunes	94.12
LDA-ANFIS[17]	Esin, Akif and Derya	94.16
Proposed LDA-ANFIS	Used in this study	95.4

Discussion

In this study, an early diagnosis system of lung cancer risk prediction based on Adaptive Neuro Fuzzy Inference System (ANFIS) and Linear Discriminant Analysis (LDA) is proposed. Performance evaluation using sensitivity and specificity analysis and classification accuracy shows that, proposed LDA-ANFIS based automatic diagnosis system for lung cancer obtains very promising results in classifying the possible lung cancer patients. Using this system physicians can take rapid decision on lung

cancer risk prediction. Moreover, patients can also use proposed diagnosis system without any prior medical or clinical knowledge about lung cancer. Diagnostic clinics can also use this model. This model can be implemented in computer software and mobile apps for increasing its availability and usability.

Though proposed model overcomes the problem of previous models and out-performs them, the sensitivity and specificity for proposed model is less than optimal. To solve this problem in future, PCA-ANFIS based system will be proposed for increasing accuracy in predicting lung cancer risk.

References

1. Turabieh H (2016) GA-based Feature Selection with ANFIS Approach to Breast Cancer Recurrence. *International Journal of Computer Science Issues (IJCSI)* 13: 36.
2. Rezaei A, Noori L, Taghipour M (2016) The Use of ANFIS and RBF to Model and Predict the Inhibitory Concentration Values Determined by MTT Assay on Cancer Cell Lines, *International Journal of Information Technology and Computer Science (IJITCS)* 8: 28.
3. Cosma G, Acampora G, Brown D, Rees RC, Khan M, et al. (2016) Prediction of Pathological Stage in Patients with Prostate Cancer: A Neuro-Fuzzy Model. *PLoS one* 11: e0155856.
4. Kalaiselvi C, Nasira GM (2015) Prediction of Heart Diseases and Cancer in Diabetic Patients Using Data Mining Techniques. *Indian Journal of Science and Technology* 8: 14.
5. Trivedi R, Singh TN, Gupta N (2015) Prediction of blast-induced Flyrock in opencast mines using ANN and ANFIS. *Geotechnical and Geo-logical Engineering* 33: 875891.
6. Wang CY, Tsai JT, Fang C, Lee TF, Chou JH, et al. (2015) Predicting survival of individual patients with esophageal cancer by adaptive neuro-fuzzy inference system approach. *Applied Soft Computing* 35: 583590.
7. Shigematsu H, Lin L, Takahashi T, Nomura M, Suzuki M, et al. (2005) Clinical and bio-logical features associated with epidermal growth factor receptor gene mutations in lung cancers. *Journal of the National Cancer Institute* 97: 339346.
8. Gurcan MN, Sahiner B, Petrick N, Chan HP, Kazerooni, et al. (2002) Lung nodule detection on thoracic computed tomography images: preliminary evaluation of a computer-aided diagnosis system. *Medical Physics* 29: 25522558.
9. Arimura H, Katsuragawa S, Suzuki K, Shiraishi LiF, Sone J, et al. (2004) Computerized scheme for automated detection of lung nodules in low-dose computed tomography images for lung cancer screening 1. *Academic Radiology* 11: 617629.
10. Polat K, Gunes S (2008) Principles component analysis, fuzzy weighting pre-processing and artificial immune recognition system based diagnostic system for diagnosis of lung cancer. *Expert Systems with Applications* 34: 214-221.
11. Daliri MR (2012) A hybrid automatic system for the diagnosis of lung cancer based on genetic algorithm and fuzzy extreme learning machines. *Journal of medical systems* 36: 1001-1005.
12. Gibson W (2004) *Pattern Recognition* 2003.
13. Scholkopf B, Mullert KR (1999) Fisher discriminant analysis with kernels. *Neural networks for signal processing* 9: 1.
14. Tang F, Tao H (2007) Fast linear discriminant analysis using binary bases. *Pattern Recognition Letters* 28: 2209-2218.
15. Jang JSR (1993) ANFIS: adaptive-network-based fuzzy inference system. *Systems, Man and Cybernetics, IEEE Transactions on* 23: 665-685.
16. Jang JSR, Sun CT, Mizutani E (1997) Neuro-fuzzy and soft computing; a computational approach to learning and machine intelligence.
17. Dogantekin E, Dogantekin A, Avci D (2009) Automatic hepatitis diagnosis system based on linear discriminant analysis and adaptive network based on fuzzy inference system. *Expert Systems with Applications* 36: 11282-11286.
18. Polat K, Gne S (2007) An expert system approach based on principal component analysis and adaptive neuro-fuzzy inference system to diagnosis of diabetes disease. *Digital Signal Processing* 17: 702-710.