Innovative Knowledge Based System – Decision Support System for Diagnosis of Carcinogenesis

N.P. Gopalan [#1], A. Malathi [#2]

[#1]Professor, National Institute of Technology, Trichirappalli, INDIA

[#2]Assistant Professor, Department of Computer Science, M S Ramaiah College of Arts, Science and Commerce, Bangalore, INDIA

E-Mail : npgopalan@nitt.edu, malathi_cs@msrcasc.edu.in

Abstract - Medical diagnosis is one of the most important significant applications of computer science. Knowledge based system is to produce intelligent machine or software which emulate human being's intelligence. Integration of artificial Intelligence and knowledge based systems leads to a new technology. Application of knowledge based system has been aptly applied to make diagnosis process easier and faster. Earlier diagnosis of carcinogenesis saves enormous lives. The goal of this paper is to evaluate the risk factor of lung cancer disease using knowledge based systems - artificial neural network, fuzzy and expert system and the implementation of ANN using back propagation , fuzzy system with linguistic variables ,membership function along with using expert system rules and image processing. . In this paper the diagnosis systems are demonstrated using neuro fuzzy, expert systems and image processing.

Keywords -Knowledge based system, artificial neural networks, back propagation, expert system, fuzzy logic, membership function, linguistic variables, decision support, inference engine, image processing.

I. INTRODUCTION

Knowledge-based systems (KBSs) is a computer program that uses the heuristic human reasoning through various procedures, mechanisms and techniques, in order to solve complex problems that do not have a traditional algorithmic approach. 'Knowledge based systems' are used in implementing various applications by adopting the techniques of artificial intelligence. In recent years knowledge based systems are used widely used in diagnosing especially in carcinogenesis. Diagnosing the disease correctly is difficult and it is very important in the field of medicine. The more research works on this topic are being done by various laboratories, organization, etc,. Various applications have been developed using knowledge based systems. It adopts various artificial intelligence techniques. In this paper application of knowledge based system in diagnosing lung cancer disease, carcinogenesis (the creation of cancer disease), is the process that normal cells are converted into cancer cells. It falls under three categories : neural networks, fuzzy logic and expert system.In 1970s an article by Gorry (1973), interest increased in the potential use of AI techniques has been increased in medical systems. Paulo J. Lisboa and Azzam F.G. Taktak (2006) , in the paper systematic review had done on artificial neural networks especially in decision support. Street W.N(1998), implemented neural model to diagnose breast cancer .The patients data sets are used in artificial neural networks model. The neural network model predicts by applying the probabilities at various intervals of time for every patient and an attempt is made to diagnose the disease with 'ok' , 'poor' and etc. Seker .H. , Odetao M.,Petroric D and Naguib R.N.G(2003) had developed an artificial neural network model using backpropagation algorithm with around 50 inputs and had achieved 95% of accuracy .G. Wilym s. Lodwick, M.D., Richard Conners and Charles A. Harlow (1979), have implemented a model to diagnose the carcinogenesis using neural and the neural model had applied to diagnose breast cancer. Zahan, S. Michael, C. Nikolakeas, S (1997), developed a hierarchical model using fuzzy approach in medical diagnosis. The model varies from other fuzzy model approaches by minimum two characteristics. Initially it performs by considering output universe of discourse, the degree of a certain diagnosis and not the diagnosis itself and also it offers specific techniques of dealing with a large variety of uncertainties that are involved by a diagnosis. Pietro Torasso(1985) , discussed the major characteristics of the medical expert systems particularly in the medical field especially in evaluating the liver function. An expert system can be defined as Russell, S. and P. Norvig (2002), an intelligent software that uses knowledge base and inference rules to find solution to solve the very difficult problem require human expertise. John, R.I (2005) describes a fuzzy logic approach in medical diagnosis . A normal view of medical diagnosis has been introduced in clinical or medical settings . The relevant possible uses of fuzzy cognitive maps are showed. David E.Smith(1992) , ES continue to evolve for specific applications in medical diagnosis.The author showed in his study use of expert systems in medical schools and within the professional medical community**.** An expert system is a software that emulates the knowledge of a human expert on complex problems . and uses this knowledge to solve problems in a fashion similar to the expert. Expert systems are the major practical

application of artificial intelligence. Suapang.P(2010) developed a software using Borland Delphi 6.0 to implement medical Image Processing.

II .ARTIFICIAL NEURAL NETWORK MODEL

Artificial neural network (ANN) is an approach that involves machine learning, models human brain and contains a number of artificial neurons . Neural networks are used in medical diagnosis especially in diagnosing carcinogenesis .The capability of the physician to efficiently handle and cure the disease is explicitly based on physician's capability to diagnose cancer at the starting stages. Neural network is like human brain connected with neurons and made of different layers. The ANN model has three layers with feed forward neural network model, which is one of the commonly used model. The first level or layer, called input layer takes input and put into internal layers are known as hidden layers. The outer layer , output layer takes the output from inner layers , after performing the required process gives it to outer world is known as output layer. The proposed model uses the feed forward network with three layers. The input layer has twelve nodes , the first eleven nodes specify the input the twelfth node is additional node assigned with -1.The hidden layer has five nodes and output layer has two nodes. All the nodes are connected with each other. Each link assigned with a weight. The back propagation algorithm is used in network model by changing the weight values to perform the network learning. The weight values of every node were initialized randomly between -1 and +1. The weights are trained in network model in such a way that to minimize the error value [Xin] which specifies the mean square error between proposed output values and actual output values.

ei(n) = di(n) – yi(n). The neural model error is defined as the sum of the squared errors of the output neurons:

$$E(n) = \sum ek2(n)$$

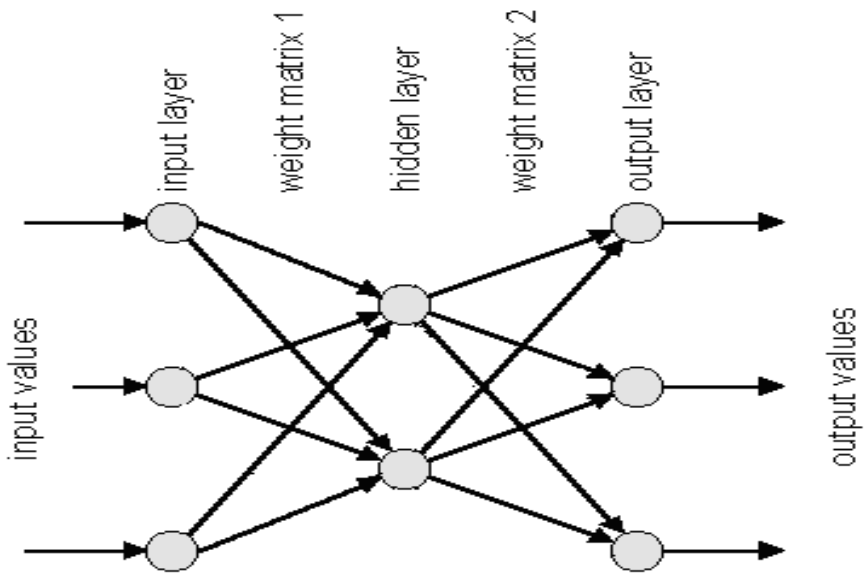


Fig.1 represents the typical neural network.

The total mean squared error value is the average of neural network errors of the training values. In back propagation algorithm the number of iterations are based on the sum of squares of errors of the output layer values for all data that are training in an epoch is less than 0.01 , i.e. threshold value.

$$EAV = \frac{1}{N} \sum_{n=1}^{N} E(n))$$

A. Experimental Results

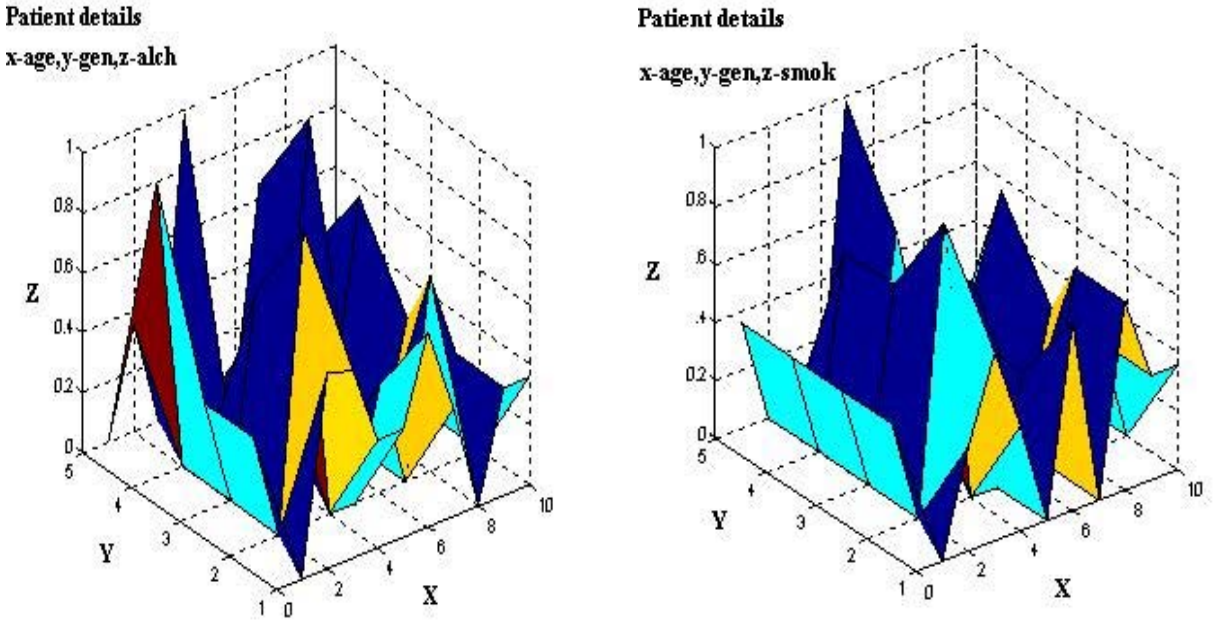
ANN constructed to perform a specific application, that the neural model should be trained . A neural network model has to be trained rather than programmed. The 'learning' in artificial neural network models is accomplished by examples. This 'learning' is called as 'training' in network model, since the training is done by changing the weight values in the model through various iterations. The total iterations of the neural model

and the time of convergence depends on the initial weight value. Nominal variables denote the input values of the input layer . A bi-state variables are denoted by the transformation and represented by numeric values .

For e.g if age >= 34 the first three nodes takes the values {0,1,0}, if age < 34 the next three nodes takes the values {1,0,0} and if age >=54 the next three nodes i.e. node7, node8 and node9 takes the values {0,0,1}. Each layer is a basically a function which takes some variables (in the form of vector *u*) and transforms it to another variable(another vector *v*) by multiplying it with coefficients and adding some biases *b*. These coefficient is known as weight matrix *w*. Size of the v vector is known as *v*-size of the layer.

v=sum(w.*u)+b

The hundred cancer patients especially lung cancer data had been collected from various research hospitals(KMS, MSR) and trained with the neural model.The results obtained depicts around 90% of accuracy. Then extended to 800 patients. The ANN is implemented using Matlab. The 3D graph is depicted below:



ANN gives around 90% accuracy. Using back propagation algorithm the results are found to be better. Neural networks are superior in terms of classification accuracy to multilayer perceptron trained with back propagation algorithm.

III. FUZZY EXPERT SYSTEM – PROPOSED MODEL

Fuzzy Expert system is highly used in medical diagnostic problems due to its inherent imprecision and uncertainty in the medical data. Fuzzy applications are widely used in medical diagnosis. It is based on imprecision, specifically, the way people make decisions on imprecise and non-numerical values. Expert System is a branch of Artificial Intelligence . Expert systems are used in many areas such as educational software, decision support systems etc.Fuzzy Expert system software replicate the thinking process of a human and would make logical decisions accordingly especially in the field of medical diagnosis.

The proposed fuzzy expert system consists of two stages of lung cancer risk, first stage cancer risk (FSCR) and second stage cancer risk (SSCR). The inputs for FSCR are demographic data such as age, gender, smoking, alcohol and the inputs for SSCR :output of FSCR , cough , vomiting , loss of weight and chest pain. The SSCR and FSCR are shown in figure 3 & 4.

A. Linguistic Variables

Linguistic variables are used to implement a fuzzy expert system. Linguistic variables can be defined by words, such as ‘low’, ‘high’ , ‘young," ,etc, by giving the values in its range. FES rules are described with linguistic variables (values) using IF – THEN rules. It contains two parts; The first part is antecedent block that lies between the IF and THEN and the second part is a consequent block that follows THEN. The membership function is the basis of a fuzzy logic controller .They convert the non-fuzzy values data in to fuzzy data. Membership functions contain set of fuzzy variables. For example, the gender values can be denoted using fuzzy variables as "Female” and "Male”. The input values or features are divided into fuzzy linguistic values or levels , for example age is described as young, middle and old. The output risk level is classified into three

linguistic levels namely low, medium and high. Adjustments in FES controller are handled by ES rules . Fuzzy controller consists of rules or inference engine rules, membership functions and defuzzification function.

The FE System – Factors

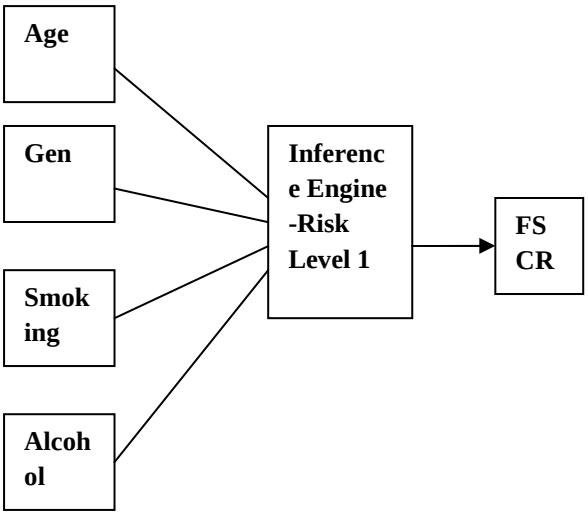


Fig 3 - Model of FSCR

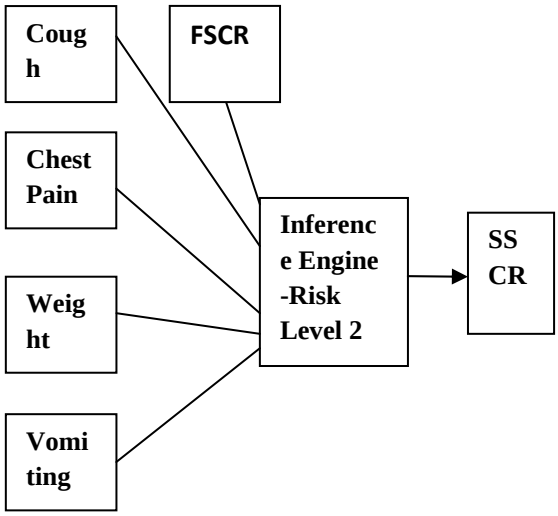


Fig 4 - Model of SSCR

B. Fuzzification and Defuzzification

The fuzzification transforms crisp values into the linguistic terms by fuzzy membership functions. The membership function is used to associate a grade to each linguistic term. Fuzzification is done by the factors and the membership functions. These formulas are framed with the help of the experts and various literature reviews. And the membership functions and formulas are derived as follows.

$$\text{Age (A)} = \begin{cases} 1 & ; 84 < a \\ a & ; 0 \le a \le 84 \end{cases}$$

$$\text{Smok (S)} = \begin{cases} 1 & ; 25 < s \\ s & ; 0 \le s \le 25 \end{cases}$$

$$\text{Alco (Al)} = \begin{cases} 1 & ; 5.5 \le al \\ al & ; 0 \le al < 5.5 \end{cases} \qquad \text{------ (I)}$$

In the set of derived rules (i), (ii) and (iii) the linguistic variables are Age, Smok and Alco. For example, for the linguistic variable the linguistic expressions can be obtained as follows:

$$\mu_{Young}(A) = \begin{cases} a / 28 & ; 0 \le a \le 3.5 \\ 0 & ; \text{otherwise} \end{cases}$$

$$\mu_{middle}(A) = \begin{cases} 0. & ; a \le 3.5 \\ a/10 & ; 3.5 < a \le 6.5 \\ 0 & ; \text{otherwise} \end{cases}$$

$$\mu_{Old}(A) = \begin{cases} 0 & ; a \le 8.5 \\ a / 9 & ; 8.5 < a \end{cases} \qquad \text{---------- (II)}$$

The weight of the age is defined as .

$$W_{age} = \begin{cases} -1 & ; \text{if } 0 \le \mu_{age} \le 0.4 \\ 0.1 & ; \text{if } 0.4 < \mu_{age} \le 0.6 \\ 0.3 & ; \text{if } 0.6 < \mu_{age} \le 1 \end{cases} \qquad \text{---------- (III)}$$

Similarly, The suitable linguistic expressions for the other input variables such as smoking, alcohol are determined. The output factor of FLLCR is calculated using the linguistic expressions (I).The possible output factors are Low, Middle and high, which can be expressed using formulas .

Defuzzification is a process that converts quantifiable value into fuzzy value. The membership function degrees of the fuzzy sets are converted into a specific real value or decision value.

C. First Stage Cancer Risk (FSCR)

The first stage cancer risk has been found from the history of patient, such as gender, age factor, smoking parameter and alcohol parameter are described as linguistic variables.The proposed Fuzzy Expert system (FSCR) has fifty four rules that frames a total of fifty four pairs of output , which deals the risk factors of lung cancer .This helps in developing fuzzy system that can precisely finds the first stage cancer risk factor level of the lung cancer patient. The developed FES rules to find FSCR are given the tabular form . Fifty four rules are framed . The rules are ,

Table 1 : FSCR rules

Rules	Age	Gender	Smoking	Alcohol	FSCR
R1	Y	F	No	No	L
R2	Y	F	No	O	L
---------- ---------- ----------	----------	------- ----------	------	----------	------
R53	O	F	R	No	M
R54	O	M	R	R	H

For example, Rule 1 is explained below:

Rule 1: If the patient if female , age is young i.e. less than 35 and no smoking habit then the FSCR is low.

D. Second Stage Cancer Risk (SSCR)

The second stage cancer risk factor is based on FSCR output of and the patient clinical or medical symptoms, such as weight loss, cough , etc,. The SSCR framed with forty eighty rules . The developed FES rules to find SSCR are shown in the Table 2. The fuzzy rules are ,

Table 2 : SSCR fuzzy rules

Rule No	FSCR	Cough	Chest Pain	Vomitting	SSCR
R1	L	Y	N	N	L
R2	M	N	N	Y	L
------		----	----	----	-----
R47	H	Y	Y	N	M
R54	H	Y	Y	Y	H

For example, Rule 1 is explained below:

Rule 1: If FSCR is low and the patient has the symptoms such as cough or cold , chest pain, weight loss and vomiting sensation =yes then Lung Cancer Risk = Medium .Since the age is young the cancer risk becomes low.

E. FES -Experiment Results

The FSCR outputs are high, medium and low which are based on the input parameters and weighs.The FSCR is low if the weighted value is <= 0.45, medium if 0.46 to 0.65 and high if weighted valus >=0.65. By applying the input parameters as FSCR, loss of weight, chest pain, cough, and vomiting, output of SSCR is calculated. The output factor of SSCR is calculated using the linguistic expressions and member functions (II) The SSCR outputs are low, medium and high which are based on the average .The SCRS is low if value is <= 0.45, medium if 0.46 to 0.65 and high if >=0.65 .

For example, Age -30; Gender- F;Smoking –No ; Alcohol –No , then the values are 0.107142857, 0.4 , 0 and 0.001 .By applying fuzzy Rule 1 (table 1) the value of FSCR is 0.00867143 . FSCR is applied with the clinical symptoms . For example the clinical symptoms are , cough –y , chest pain –y , weight loss– y and vomiting sensation – y , By applying fuzzy Rule 1 (table 2) the value of SSCR 0.523469.

IV. CONCLUSION

The most common and deadly disease is 'lung cancer' .Detection of cancer in the early stage is the key of its cure. More than 40% cancer deaths are preventable if it diagnosed at early stage. Developing countries are facing the shortage of medical experts in medical field. The major purpose of the KBS is to provide expert advice from the developed system especially in diagnosing lung cancer. The developed KBS would act like doctor in a box.

REFERENCES

[1] Antognetti, P. and Milutinovic, V. "Neural Networks: Concepts, Applications, and Implementations (Eds.)" PH, Englewood Cliffs, NJ.,1991.

[2] David E. Smith, Expert systems for medical diagnosis: A study in technology transfer ,The Journal of Technology Transfer , Volume 17, Issue 4, pp 45-53

[3] Gorry G.A -Computer assisted clinical decision making method, Med.12:45-51

[4] G Wilym s(1). Lodwick,M.D.,Richard Conners(2) ,Ph.D and Charles A. Harlow(2),Ph.D, Dept. of Radiology ,Univ. of Missouri , Colunbia (@)Dep of Electrical Engg. ,Louisiana state univ.,Louisiana -IEEE 1979

[5] John RI, Innocent PR. - Modeling uncertainty in clinical diagnosis using fuzzy logic,Centre for Computational Intelligence, School of Computing, De Montfort University, Leicester, 2005 Dec;35(6):1340-50.

[6] Liverpool, UK, Department of Clinical Engineering, Royal Liverpool University Hospital, Liverpool, UK ,Neural Networks-Volume 19 Issue 4 (May 2006)

[7] M Fathitorbaghan, D Meyer-MEDUSA – A Fuzzy Expert System for medical diagnosis of acute abdominal pain-Methods of Information in Medicine, Volume: 33, Issue: 5 Pages: 522-529.

[8] Paulo J. Lisboa , Azzam F. G. Taktak-"The use of artificial neural networks in decision support in cancer" A systematic review- - School of Computing and Mathematical Science, Liverpool John Moores University,

[9] Pietro Torasso- Knowledge based expert systems for medical diagnosis, Statistics in Medicine Volume 4, Issue 3, pages 317–325, July/September 1985

[10] Russell, S. and P. Norvig, 2002. Artificial Intelligence: A Modern Approach, Prentice Hall, II Edition

[11] Seker H., Odetayo M., Petrovic D., and Naguib R.N.G.- "A Fuzzy Logic Based Method for Prognostic Decision Making in Breast and Prostate Cancers", IEEE Trans. On Information Technology in Biomedicine, 2003. page 256

[12] Street W.N– "Artificial Neural Network model for prognostic prediction" ,Proceedings of 15th International conference on ML -Fuzzy logic based method for prognostic decision making in breast and prostrate,1998. cancers-Biomedicine (IEEE transactions) 2003.

[13] Suapang, P &Yimmun, S-Medical image processing and analysis for nuclear medicine diagnosis in IEEE conference on Control Automation and Systems (ICCAS), 2010, Page(s):2448 – 2451, Print ISBN:978-1-4244-7453-0

[14] Michael, C., Zahan, S. Nikolakeas, S-A fuzzy hierarchical approach to medical diagnosis-Tech. Univ. of Cluj-Napoca;This paper appears in: Fuzzy Systems, Proceedings of the Sixth IEEE International Conference (1997), Volume: 1, On page(s): 319-324 .