

DOI: <https://doi.org/10.24297/ijct.v20i.8844>

A Novel Integrated Prognosis & Diagnosis System for Lung Cancer Disease Detection using Soft Computing Techniques

Dr. Nadiminti Venkata Ramana Murty¹, Prof.M.S.Prasad Babu²

¹ Associate Professor, Department of CSE, Engineering and Technology Program, Gayatri Vidya Parishad College for Degree and PG Courses (A), Rushikonda, Visakhapatnam, Andhra Pradesh

².Retired Senior Professor, Department of CS&SE, AU College of Engineering (A), Andhra University, Visakhapatnam

Abstract::

Nowadays, lung cancer is one of the ranking first causes of mortality worldwide among men and women. Although there are a lot of treatment options like surgery, radiotherapy, and chemotherapy, five-year survival rate for patients is quite low. However, survival rate may go up to 54% in case lung cancer is identified in an early stage. Therefore, early detection of lung cancer is vital to decrease lung cancer mortality. Medical Experts are continuously trying to find the best solution for the early prediction and diagnosis of Lung Cancer Disease; in this Research work, an attempt has been made to design and develop a novel integrated soft computing predictive system to handle various types of patients' clinical data to diagnose the lung cancer disease. Here data mining techniques are used to handle the numeric and textual data, image processing techniques are used to handle CT scan images, neural networks are used to train the lung cancer patient images, and fuzzy inference mechanism is used to predict the lung cancer stages. This integrated approach results in detection of lung cancer disease with Prognosis and suggesting diagnosis by the expert system for lung cancer disease. Even in cases of small-sized nodules (3–10 mm), the proposed system is able to determine the nodule type with 96% accuracy.

Index Terms: Data Mining, Fuzzy Inference Mechanism, Image Processing, Neural Networks, and Soft Computing,

1. INTRODUCTION

The novel model suggested in this paper is to diagnose the lung cancer diseases using the techniques of image segmentation and decision making. A Fuzzy Inference System (FIS) is developed and result is tested by using artificial neural network. The methodology proposed for lung nodule detection consists of the following major components as shown in the figure 1. The figure 1 consists of five major components: Knowledge Base, Inference Engine, Working Memory, Database and a User Interface.

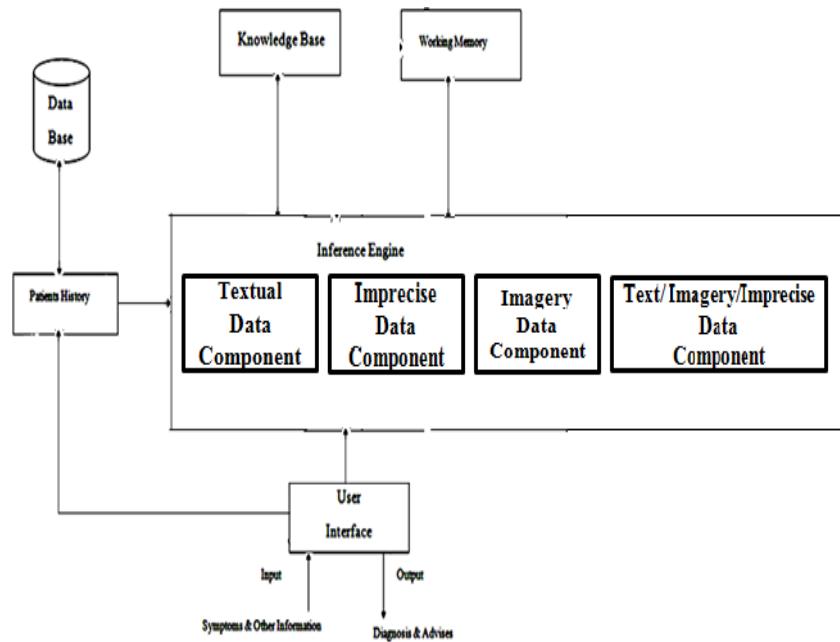


Fig.1. Architecture of a novel integrated predictive system for lung cancer diagnosis

The Major Components of the Architecture shown in the figure 1 are

1.1 Knowledge Base

Knowledge base consists of set of facts and rules. Here facts are true statements already tested by the system and rules are written by the user. An Input query given at the user interface can be solved by using the rules and facts in knowledge base.

1.2. Working Memory

Working memory consists of the data recently processed by the system. This memory is similar to a cache memory which contains the information about the frequently used rules and facts. This is also a very useful block in prediction system.

1.3 Database

In most of the prediction system designs a separate database will be maintained apart from the existing knowledge base to store the history of old records and other valuable information. In this predictive system design, the database is exclusively used to store the old patient records of lung cancer disease like CT Images, Test and Diagnostic Reports etc..

1.4 User Interface

User interface is the key component in the system through which user can interact with the predictive system. Here symptoms and patient test reports are given as input to the system and diagnosis and advice given by the expert are taken as output.

1.5. Inference Engine

Inference engine is used to infer the data from the knowledge base. The inference engine for the proposed integrated system contains four major independent components: Numeric Data Component, Image Data Component, Imprecise data components and training data component.

2. METHODOLOGY

In this novel predictive system design, the inference mechanism methodology is shown in the figure 2. The inference engine of the predictive system contains four major components, each of which are handling different types of feature of the lung cancer patient records. The components used in this methodology are explained

hereunder.

The four major components are

- i) Numerical/Categorical Data Component
- ii) Imprecise Data Component
- iii) Imagery Data Component
- iv) Text/Imagery Data Component

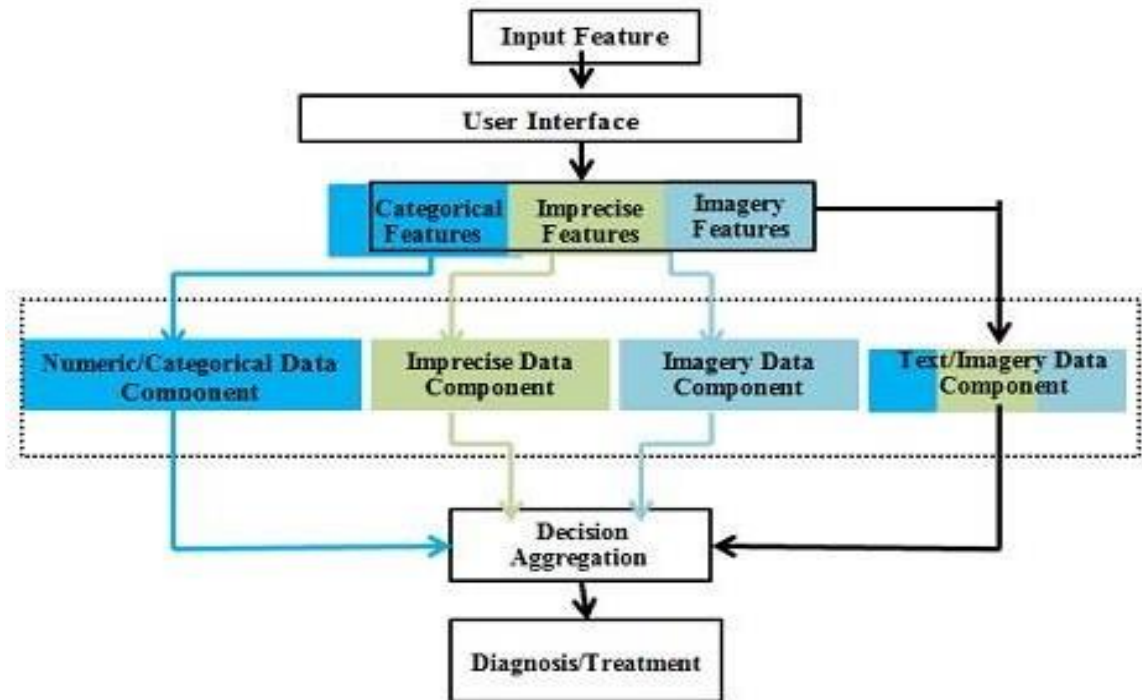


Fig.2 Inference mechanism of Predictive System for Lung Cancer Diagnosis

2.1. Features of Lung Cancer Dataset

In this experimentation, patient records of lung cancer disease are taken from the popular dataset of machine learning repository. The detailed features of the dataset are given Table 1

Table 1. Features of Lung Cancer Dataset

Lung Cancer Dataset Description				
Attribute No	Attribute Name	Type	Nominal Ranges	Description
1	HB	Numeric	1 to 4	Hemoglobin
2	Age (years)	Numeric	0 to 100	Age
3	RBC	Numeric	1 to 5	Red Blood Cells
4	NEU	Numeric	1 to 4	White Blood Cells
5	LENFO	Numeric	1 to 5	White Blood Cells Type
6	PLT	Numeric	1 to 5	Platelets
7	MPV	Numeric	1 to 5	Mean Platelet Volume
8	PLR	Numeric	1 to 5	Positive Likelihood Ratio

9	NLR	Numeric	1 to 5	Negative Likelihood Ratio
10	T	Imagery	T1,T2,T3,T4	Size and/or extension of the primary tumor
11	N	Imagery	{low, medium, high}	Degree of spread to regional lymph nodes
12	M	Imagery	{low, medium, high}	Presence of distant metastasis
13	T1	Imagery	< 3cm	Size of the Tumor
14	T2	Imagery	> 3 < = 7	Size of the Tumor
15	T3	Imagery	> 7	Size of the Tumor
16	T4	Imagery	> 8	Size of the Tumor
17	Gender	Numeric	{male, female}	Male or Female
18	Smoking	Imprecise	{low, medium, high}	Range of Smoking
19	Exposure to harmful chemicals	Imprecise	{low, medium, high}	Exposure Range
20	Hoarse Voice	Numeric	{yes, no}	Abnormal Change in the Voice
21	Dyspnea	Numeric	{yes, no}	Shortness of Breath
22	Unexpected Weight Loss	Imprecise	{low, medium, high}	Sudden Weight loss
23	Dysphagia	Numeric	{yes, no}	Difficulty or Discomfort in Swallowing
24	Family History	Numeric	{yes, no}	Histology of Cancer
25	Wheezing with Sound	Numeric	{yes, no}	Whistling sound while breathing
26	Hemoptysis	Numeric	{yes, no}	Coughing up blood
27	Pain in Chest or Bone	Imprecise	{low, medium, high}	Severity of Pain
28	Pneumonia Level	Imprecise	{low, medium, high}	Fever level

2.2 Numeric/Categorical Data Component

This component is used to handle numeric/categorical data of the liver patient records. The details are given Table1. These features are handled by Data mining techniques to generate inferencing rules and thereby to develop predictive system for determining whether the patient has cancer or not.

2.3. Imprecise Data Component

This component is used to handle categorical data of the lung cancer patient records. The details are given in Table 1. These features are handled by Fuzzy inferencing techniques to determine the stage of the cancer. i.e. one of the four stages.

2.4. Imagery Data Component



This component is used to handle imagery data of the lung cancer patient records. The details are given in Table 1. These features are handled by Image Processing techniques to detect the tumors and their sizes and thereby determining whether a patient has cancer or not.

2.5. Text/ Imagery Data Neural Network Component

This component is used to handle Text/Imagery data in the input feature dataset given in Table 1. These features are handled by Neural Networks to detect whether the patient has cancer or not.

3. SYSTEM DESIGN

The proposed system handles numeric/categorical/imagery features supplied by users to determine lung cancer malignancy as well as the necessary diagnosis if the patient has lung cancer. The detailed steps of the proposed methodology are mentioned hereunder.

Step 1: User logs into the system with user name and password.

Step 2: User enters the Lung Cancer patient record feature set which contains combination of Textual, Imprecise & Imagery features. The feature set of a patient's record is given in Table 1.

Step 3: Given feature set is segregated into three feature subsets and those subsets will be supplied to respective components as shown in figure 2.

The feature subsets are

- a) Textual feature set contains numeric/categorical features of the given feature set as mentioned in Table 1.
- b) The imprecise feature set contains fuzzy features of the given feature set as given in Table 1
- c) The imagery feature set contains CT scan image related data as mentioned in Table 1.

Step 4: Each component evaluates received feature subsets and generates results based on the inference mechanism discussed above.

Step 5: Decision Aggregation will be done by the system and suggests necessary diagnosis/treatment if the given.

4. IMPLEMENTATION

The Methodology of the research work is implemented as mentioned in this section.

The trace of implementation steps are mentioned below.

Step 1: User logged into the system with user name= xxxx and password = ****

Step 2: User Inputs the Lung Cancer patient record which is a combination of Numeric and Categorical Data. The cancer patient record details contains the features F1, F2, F28, as given in Table 1.

Step 3: These features are handled by individual data components in the inference engine.

Step 3(a) : The Numerical data in the patient records described in Table 1 is handled by Numeric/Categorical data component. Data Mining concepts are used in this data component.

For example :

Input :

If (Age is Medium) and (Gender = Male) and (Hemoptysis is yes) and (pain in chest or bone is high) and (shortness of breath is yes) and (unexpected weight loss is high)

Output

lung cancer = yes

The categorical feature set described in Table 2 of a lung cancer patient record is handled by Numeric/Categorical Data component. Data mining concepts are used in this data component.



For example :

Input :

If (Family History is yes) and (Exposure to Harmful Chemicals is high) and (pain in chest or bone is medium) and (shortness of breath is yes) and (unexpected weight loss is medium)

Output

lung cancer = yes

Step 3(b) : The imprecise data features of the patient records described in Table 1

For example :

Input :

If (Family History is yes) and (Hemoptysis is yes) and (pain in chest or bone is high) and (Wheezing with sound is yes) and (Dysphagia is yes) and (unexpected weight loss is high)

Output:

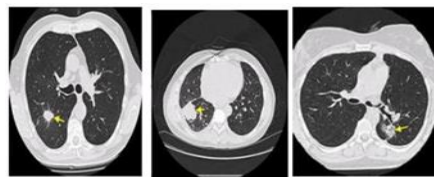
lung cancer = yes and Stage = 3.

Step 3(c) : The imagery data described in table 4 is handled by Imagery data

For example :

The input & output images are represented in figures 3(a) & (b).

Input: Image data



Output: Image Data

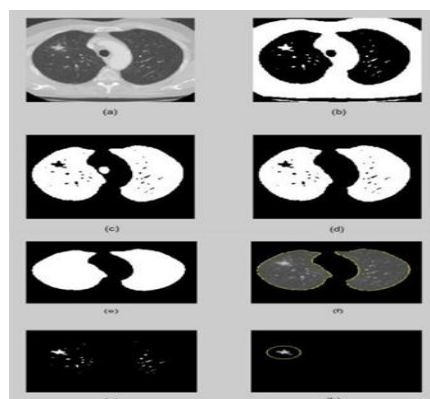


Fig: 3(a) & 3(b)

Step 3(d) : The Textual/imagery data is handled by Text/ Imagery data.

Example 1 : Textual Data

Input :

If (Exposure to Harmful chemicals is low) and (Hemoptysis is low) and (pain in is low)

Output:



lung cancer = no.

Example 2: Imagery Data, Input and Output are shown in the figures 4(a) &(b).

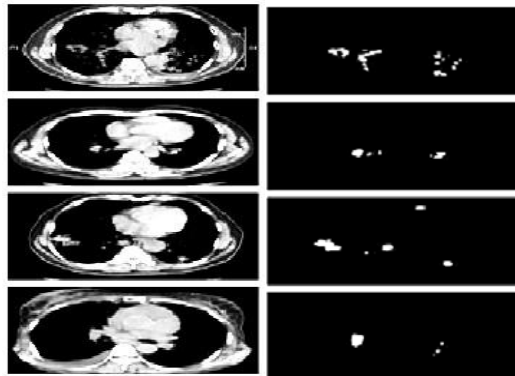


Fig 4(a) & 4(b)

Step 4: Decision Aggregation will be performed by the system based on the results obtained from individual components.

Step 5: Finally the system suggests necessary Diagnosis/Treatment. The figure. 5 shows a typical treatment suggested by the system

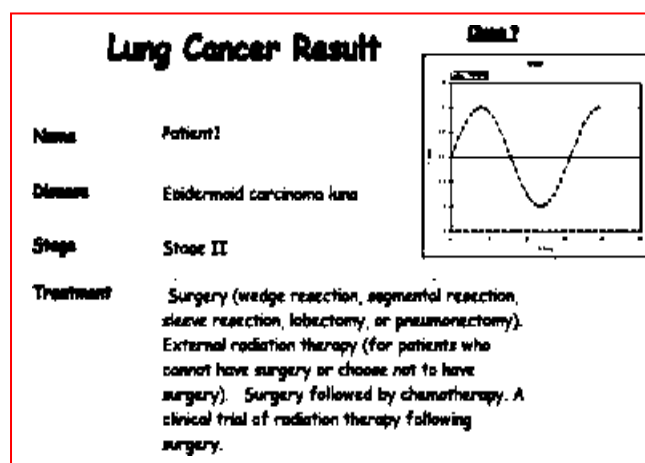


Fig.5. Suggested treatment

5. CONCLUSIONS

Finally a Novel Integrated Predictive System is developed using Soft computing techniques for Prognosis and Diagnosis of Lung Cancer Disease which is helpful for both patients and doctors to detect the stage of the Lung Cancer using all sorts of datasets like Imagery dataset, Categorical dataset and Numerical datasets. It is also observed that the results of the system are more accurate and promising when compared with other Lung Cancer Analyst systems.

6. FUTURE ENHANCEMENTS

This research work can be developed as a smart device application. In this case, Graphical User Interface (GUI) is required to enable the user to get contact with the system easily. Using smart devices provide the users with the ability to use the system anywhere, which makes the process of diagnosis faster and easier.

7. REFERENCES:

1. S. Saleh, N. Kalyankar, and S. Khamitkar, "Image Segmentation via using Edge Detection", International Journal on Computer Science and Engineering(IJCSE), vol. 2, no. Three, pp. 804-807, 2010.
2. S.K. Vijai Anand, "Segmentation coupled Textural Feature Classification for Lung Tumor Prediction", Proceedings of the International Conference on Computing, Communications and Networking Technologies ICCCT, 2010, pp. 518-524.
3. Mohammed Al-Zoube & Baha Khasawneh, (2010), "A Data Mashup for Dynamic Composition of Adaptive Courses", The International Arab Journal of Information Technology, Vol. 7, No. 2.
4. Hamdan, H. and Garibaldi, J.M., "Adaptive neuro-fuzzy inference system (anfis) in modelling breast cancer survival", in Fuzzy Systems (FUZZ), International Conference on, IEEE, (2010), 1-8.
5. J. C. Obi and A. A. Imainvan, "Decision Support System for the Intelligent Identification of Alzheimer using Neuro Fuzzy logic," International Journal on Soft Computing, DOI : 10.5121/ijsc.2011.2203, vol. 2, no. 2, (2011), pp. 25-38.
6. K. Lavanya, M. A. S. Durai, N. C. Sriman and N. Iyengar, "Fuzzy Rule Based Inference System for Detection and Diagnosis of Lung Cancer," International Journal of Latest Trends in Computing ISSN: 2045-5364, vol. 2, (2011), pp. 165-171.
7. Ada and Rajneet Kaur, " Early Detection and Prediction of Lung Cancer Survival using Neural Network Classifier", international journal of application or innovation in engineering and management, Volume 2, Issue 6, June 2013 ISSN 2319 – 4847.
8. Atiyeh Hashemi, Abdol Hamid Pilevar, Reza Rafeh, "Mass Detection in Lung CT Images Using Region Growing Segmentation and Decision Making Based on Fuzzy Inference System and Artificial Neural Network", I.J. Image, Graphics and Signal Processing, 2013, pp. 16-24.
9. Ahmed S. Moussa, Sara A.Yones, "Tumor Volume Fuzzification for Intelligent Cancer Staging", Applied Soft Computing, Volume 35, October 2015.
10. P. Nithya, B. Umamaheshwari, R. Deepa, Detection Of Lung Cancer Using Data Mining Classification Techniques, IJARCSSE, Vol. 5, Issue 7, July 2015.
11. Eman Magdy, Nourhan Zayed, and Mahmoud Fakhr, " Automatic Classification of Normal and Cancer Lung CT Images Using Multiscale AM-FM Features", International Journal of Biomedical Imaging, Volume 2015, Article ID 230830, 7 pages.
12. S. Vijayarani and S. Sudha, "An Efficient Clustering Algorithm for Predicting Diseases from Hemogram Blood Test Samples", Indian Journal of Science and Technology, vol.8, pp. 1–8, Aug. 2015.
13. Al-Hamadani, B. and T. Shubita (2015). "Novel Hyper-Active Algorithm to Estimate Missing Microarray Attributes." Computer and Information Science 8(3):10 <http://www.airccj.org/CSCP/vol6/csit65201.pdf> , Computer Science & Information Technology (CS & IT), pp. 01–09, 2016. © CS & IT-CSCP 2016 DOI : 10.5121/csit.2016.60501
14. J.Thomas, R.Theresa Princy, " Human Heart Disease Prediction system using data mining techniques", IEEE 10.1109/ICCPCT.2016.7530265, 4 Aug 2016 <https://www.hindawi.com/journals/ijbi/2013/942353/>.