Design of an Intelligent based System for the Diagnosis of Lung Cancer

James, G. G.[1]
[1]Department of Computer Science,
Rhema University, Aba, Abia State

Chukwu, E. G.[2]
[2]Department of Computer Science,
Federal University of Technology, Ikot-Abasi, Nigeria

Ekwe, P. O.[3]
[3]Department of Computer Science,
Federal College of Agriculture Ishiagu, Ebonyi State

ASOGWA, E. C.[4]
[4]Department of Education Computer Science,
University of Nigeria Nsukka, Enugu State

DARLINGTON, C. H[5]
[5]Department of Computer Science,
Federal College of Agriculture Ishiagu, Ebonyi State

Abstract:- The search for quick detection and solution of Lung cancer is traced back to the Stone Age by medical experts and health workers. Often time, patients are confronted with the challenges of running through different medical outfits and spending a lot of money to fight the deadly disease and still ended up in their untimely grave. This is not because the medical experts could not handle the patient appropriately, but because the disease has gotten to the final or terminal stage, and at this point it is irreversible. The quest to reduce the challenges faced by patients and medical experts to get a timely solution to the diagnosis or detection of lung cancer in other to save the life of the patient has motivated this research. The work aims at designing and implementing an intelligent-based system for the diagnosis of lung cancer driven by fuzzy c-means algorithm.The methodology used was Object-Oriente Analysis and Design (OOAD). The fuzzy model was designed in three main stages includingfuzzification, FCM inference system, and defuzzification. The triangular membership (Tmf) function was used to map the input parameters to the output parameter. The FIS model was designed using Mamdani's inference mechanism. C# programming language was used on a windows 10 platform to integrate the Matlab FIS output for better performance and MySQL database 5.7.14 from WAMP server 3.0.6 was used as the back-end engine. To validate this work, data was captured directly from the 50 volunteers. The training, testing, and checking FCM KMSI values of 0.00123, 0.03123, and 0.06301 respectively were observed in the process at 150 epochs and an average error of 0.0002102 was observed at 300 epochs.

Keywords:- *Fuzzy Cluster Means Algorithm, Cancer Detection, Fuzzification, Intelligent System.*

I. INTRODUCTION

The 21st-century generation has noted from expert reports that approximately more than 70% of patients who died from Lung Cancer or other related cancerous disease are due to late, wrong, and incompetent diagnosis or detection. The terminality of this deadly ailment is not due to the inability of medical experts to handle the patient appropriately, but because the disease has gotten to the final or terminal stage and at this point seems to be irreversible. These caused patients to be confronted with the menace of running through different medical outfits and spending a lot of money to fight the deadly disease and still ending up in their untimely graves which makes lung cancer as well as other species of cancer be categorized into the family of terminal diseases.

However, the inability of medical experts to quickly detect and find solutions to cancerous Lung Tumor in other to save and or prolong the life of patients suffering from these ailments has motivated this research.

II. WEAKNESS OF THE PREVIOUS TECHNOLOGY

The system in Shah and Suralkar(2016) above suffers from the following limitations:

- Inability to provide more accuracy and save more time, as a result of confusion with the use of two almost similar algorithms.
- Inability to detect tumors in the lungs.

III. THE PROPOSED SYSTEM

The proposed system is a fuzzy cluster means an algorithm-based model for diagnosing and detecting of lung cancer. FL has the capability of handling imprecise, incomplete, and vague information, as well as the ability to represent partial truth, hence could handle the classification appropriately.

The essence of applying the FCM algorithm is to build an intelligent system that will cluster symptoms using unsupervised machine learning techniques to enhance the credibility of results presented to patients.

The system shall have four (4) input variables and a single output. The input variables to be considered are difficulty in breathing, severe cough, frequent vomiting of blood, and server dizziness respectively.

IV. STRENGTH OF THE PROPOSED TECHNOLOGY

The proposed system has the following benefits:

- Possession of an inbuilt intelligent diagnosis and detection model based on the Fuzzy c-means algorithm that provides the system with the ability to cluster symptoms using unsupervised machine learning techniques to enhance the credibility of results presented to patients
- Ability to reduce the challenges of wrong and late diagnosis of this lung tumor which is one of the major causes of untimely death in the world.
- Ability to comprehend the facility to diagnose the tumor even from day 1 of development.
- Ability to enhance timely information on the state of health of the patient's lung.

V. DESIGN METHODOLOGY

The method employed for this work is Object-Oriented Analysis and Design (OOAD). This is the procedure of identifying software engineering requirements and developing software specifications in terms of a software system's object model, which comprises interacting objects. The methodology can then be developed by making the utmost use of existing techniques in such a way as to satisfy the requirements (Nasrabadi and Feng, 2008). The major phases of software development using object–oriented methodology are object-oriented analysis, object-oriented design, and object-oriented implementation.

VI. DATABASE DESIGN

The sample symptoms of diseases in lung cancer were selected (i.e. difficulties in breathing, severe cough, coughing blood, and severe dizziness).This disease set is given in table 1.

Table 1: Symptoms of Diseases in lung cancer (* indicates overlapping symptoms of the disease)

BREATH-LESSNESS	SEVERE COUGH	COUGHING BLOOD	SEVERE DIZZINESS
No symptoms	*severe	Slight blood	*severe
normal	Slight	mild	Not all
			*slight

Table 2: A Frame for Holding Patient Symptoms (asterisks (*) indicates any float value between 0 and 1)

SLOT NAME	DATA TYPE	VALUE
di	Float	*
di+1	Float	*
…	…	….
dn	Float	*

The frame consists of disease as slots.

Lung cancer di di+1 . . . Dn

Frame 1: A frame for lung cancer

The slots di …dn, where n = 4, holds knowledge about symptoms for alcoholic cirrhosis. For example:

$d_1$holds value for breathlessness
$d_2$holds value for severe cough
$d_3$holds value for coughing blood
$d_4$holds value for severe dizziness

A. File Structure

The file structure shown in Table 3 represents the database for sharing knowledge in the knowledge base.

Table 3: File Structure for Knowledge Base

FIELD NAME	DATA TYPE	DESCRIPTION	LENGTH
d_i	Float	Holds value for breathlessness	4
d_{i+1}	Float	Holds value for severe cough	4
d_{i+2}	Float	Holds value for coughing blood	4
d_{i+3}	Float	Holds value for severe dizziness	4

VII. SYSTEM ARCHITECTURE FOR THE PROPOSED SYSTEM

Figure 5 shows the architecture diagram of the proposed system:

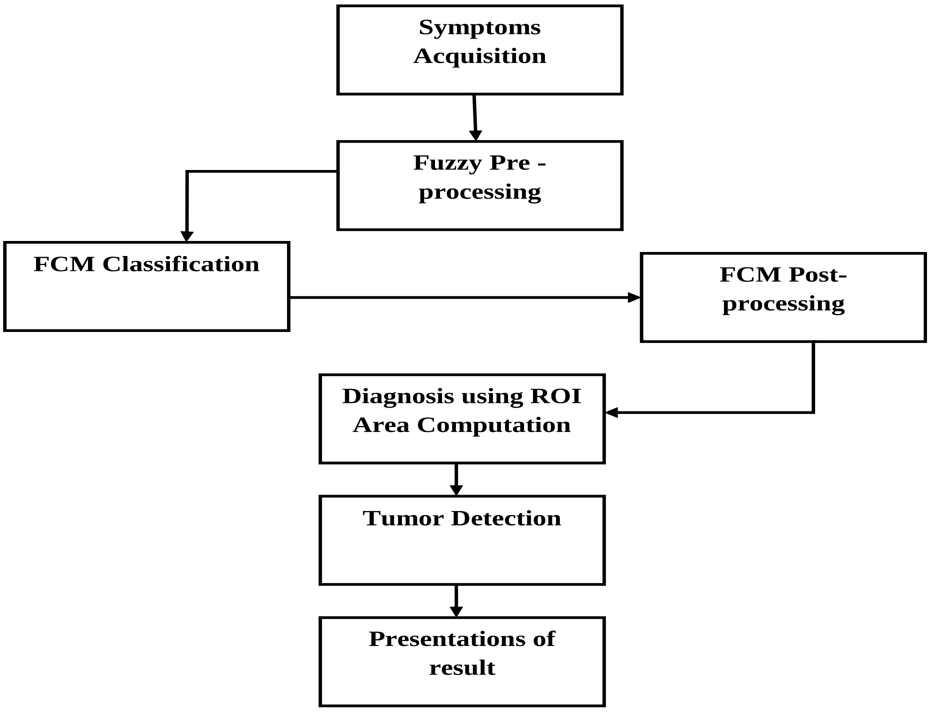

Fig. 1: Architecture of Proposed System (Adapted from Shah and Suralkar, 2016)

VIII. USE CASE DIAGRAM

The Use Case Diagram depicted in figure 2 shows the actors, the relationships and interaction within the system

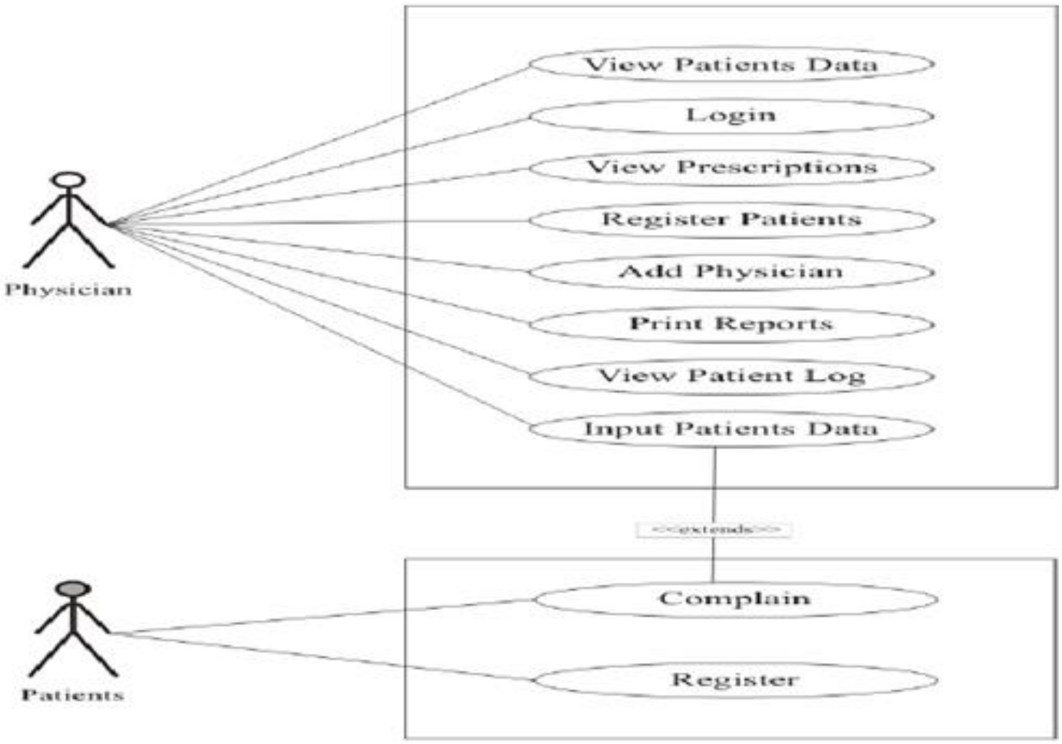

Fig. 2: Use Case Diagram

IX. EVALUATION OF THE FUZZY UNIT

A. Fuzzy logic model

In this work, a type-1 fuzzy logic model is used. This model is based on a triangular membership function that defines a degree of membership of all crisp values within the specified universe of discourse. The conceptual architecture of the fuzzy logic model used is presented in Figure 3.

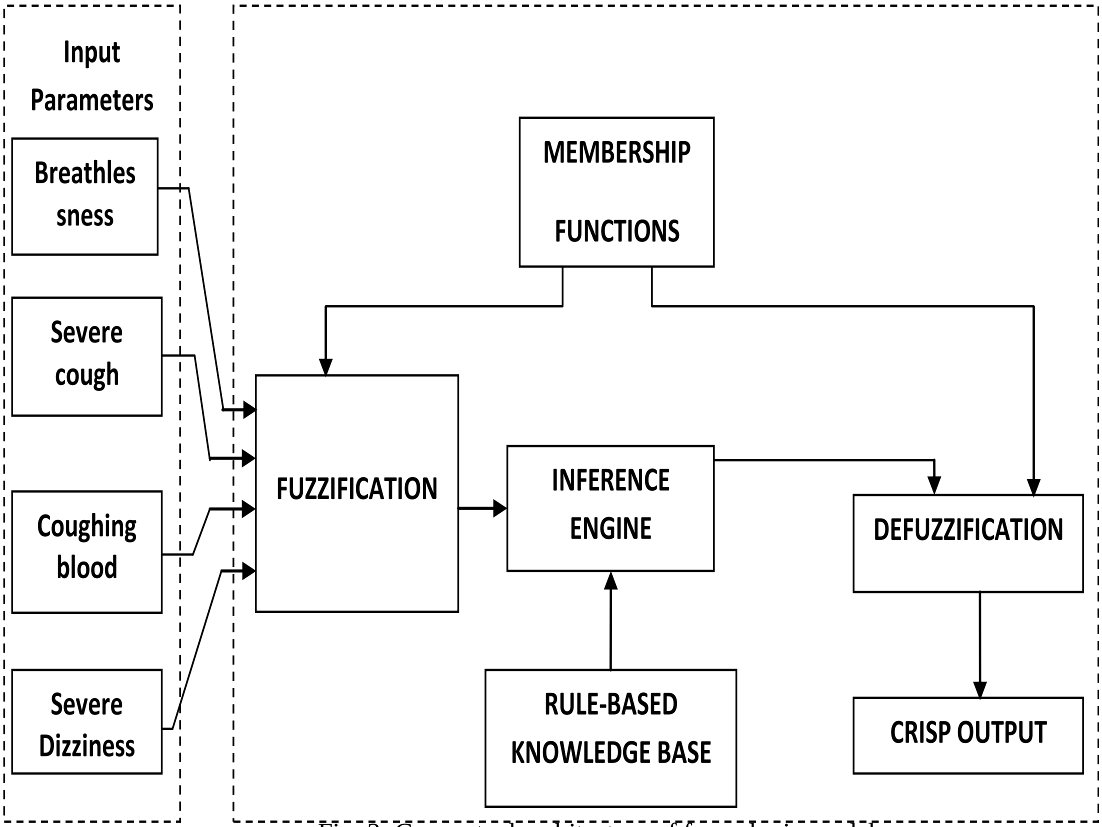

Fig. 3: Conceptual architecture of fuzzy logic model

X. COMPONENTS OF THE FUZZY LOGIC MODEL

The following components constitute the fuzzy logic model used in this work;

- **Fuzzification module:** this module maps the crisp input to a type-1 fuzzy set using the triangular membership functions defined for this work.
- **Inference engine:** this module evaluates the rules in a rule base against fuzzy set gotten from Fuzzification to produce another fuzzy set.
- **Defuzzification module:** it maps the fuzzy set from inference engine to a crisp output using center of gravity defuzzification method (or Centroid).
- **Knowledge base:** This is a database of rules (rules are generated from experts' knowledge) to be used by the inference engine.
- **Membership function:** This is a mathematical equation that maps a crisp input value to a degree of membership between 0 and 1, called the fuzzy set.

A. Fuzzy c–means algorithm fuzzification process

In this stage, for each input and output variable selected, we define four membership functions (MF), namely – Brethleness, Severe Cough, Coughing blood, and Severe Dizziness. A category is defined for each of the variable. We employ triangular membership function. For this reason, we need at least three points (a, b, c) to define one Membership Function (MF) of a variable.

$$Y_M = \sum_{i=1}^{N}\sum_{j=1}^{C} M_{ij}^{m} \left\| x_i - C_j \right\| \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (12)$$

Where,

m - Any real number greater than 1,
M_{ij}- degree of membership of x; in the cluster j,
x_i- data measured in d-dimensional,
R_j - d-dimension centre of the cluster (vector), The update of Fuzzy membership M ij and the cluster centres R are given by:

$$M_{ij} = \frac{1}{\sum_{k=1}^{C}\left(\frac{\|x_i - x_j\|}{\|x_i - C_j\|}\right)^{\frac{2}{m-1}}}\pi r^2 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (13)$$

$$R_i = \frac{\sum_{i=1}^{N} x_i . M_{ij}^{m}}{\sum_{i=1}^{N} M_{ij}^{m}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (14)$$

The above process ends when,

$$Max_{ij}\left|M_{ij}^{(k+1)} - M_{ij}^{(k)}\right| < \delta \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (15)$$

Where, = termination value or constant between 0 and 1, K= no of iteration steps.

The triangular membership function is defined as;

$$f(x; a, b, c) = \begin{cases} 0, & x \le a \\ \frac{x-a}{b-a}, & a \le x \le b \\ \frac{c-x}{c-b}, & b \le x \le c \\ 0, & c \le x \end{cases} \quad . \quad . \quad . \quad .(16)$$

Where:
a - the left leg of the membership function
b - the center of the function
c - the right leg of the function
x - the crisp input
f - a mapping function

XI. UNIVERSE OF DISCOURSE

The Universe of Discourse is the range of all possible values for the fuzzy linguistic variables. The following universe of discourse is defined for our linguistic variables;

Table 4: Fuzzy Universe of Discourse

Linguistic Variable	Lower Bound	Upper Bound	Variable Type
Breathlessness	0	1	Input
Severe Cough	0	1	Input
Coughing blood	0	1	Input
Severe Dizziness	0	1	Input
LCN Diagnose	0	1	Output

Table 5: Input variables simulation on digits

B	SC	CB	SD	LCNI
1.0	0	0	0	0
0.75	0.25	0.20	0.21	0.25
0	0.80	0.77	0.90	0.95

Table 6: Input variables simulation on variables

B	SC	CB	SD	LCNI	Interpretations
VN	NC	NS	NR	LI	Not Infected
N	C	SLS	NC	I	Infected
VP	SC	0.77	VSS	MI	Highly Infected

XII. DEFUZZIFICATION

We defuzzify the fuzzy set by using the center of gravity defuzzification method presented in the equation below;

$$COG = \frac{\sum_x^b \mu_A(x)x}{\sum_x^b \mu_A(x)}, \quad . \quad . \quad . \quad .(17)$$

Where $\mu_A(x)$is the degree of membership of x in a set A.

XIII. SYSTEM FLOW DIAGRAM

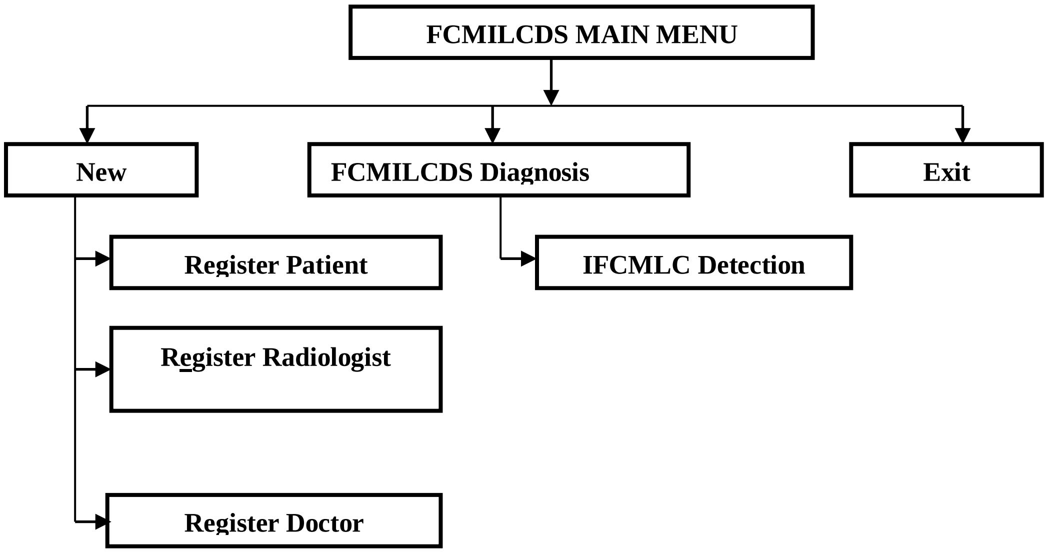

Fig. 4: System Flow Diagram for the FCMILCDS

XIV. CONCLUSION

Using FCM algorithm, it was realized that the reliability and credibility of the diagnosis and detection processes of lung cancer were guaranteed. This was accomplished by introducing FIS variables and values, implementing more fuzzy rules, and using the Triangular-shape membership function for the input variables difficulties in breathing, severe cough, coughing blood, and severe dizziness.

REFERENCES

[1.] American Cancer Society | Information and Resources for Cancer: Breast, Colon, Lung, Prostate, Skin." [Online]. Available: http://www.cancer.org/. [Accessed: 04-Oct-2021].

[2.] Cancer of the Lung and Bronchus - SEER Stat Fact Sheets." [Online]. Available: http://seer.cancer.gov/stat facts/html/lungb.html. [accessed: 04-Oct.-2021]

[3.] Liu D, Peng H, Sun Q, et al. The Indirect Efficacy Comparison of DNA Methylation in Sputum for Early Screening and Auxiliary Detection of Lung Cancer: A Meta-Analysis. Int J Environ Res Public Health. 2017;14(7):679.

[4.] Chiang A, Detterbeck FC, Stewart T, Decker RH, Tanoue L. Chapter 48: Non-small cell lung cancer. In: DeVita VT, Lawrence TS, Rosenberg SA, eds. DeVita, Hellman, and Rosenberg's Cancer: Principles and Practice of Oncology. 11th ed. Philadelphia, Pa: Lippincott Williams & Wilkins; 2019.

[5.] Hann CL, Wu A, Rekhtman N, Rudin CM. Chapter 49: Small cell and Neuroendocrine Tumors of the Lung. In: DeVita VT, Lawrence TS, Rosenberg SA, eds. DeVita, Hellman, and Rosenberg's Cancer: Principles and Practice of Oncology. 11th ed. Philadelphia, Pa: Lippincott Williams & Wilkins; 2019.

[6.] Kelly M. Latimer and Timothy F. Mott (2015). Lung Cancer: Diagnosis, Treatment Principles, and Screening, Patient information: A handout on this topic, written by the authors of this article, is available at http://www.aafp.org/ afp/2015/0215/p250-s1. html.

[7.] Rivera MP, Mehta AC, Wahidi MM. Establishing the diagnosis of lung cancer: diagnosis and management of lung cancer, 3rd ed.: American College of Chest Physicians evidence-based clinical practice guidelines. Chest. 2013;143(5 suppl):e142-165S.

[8.] Lamprecht B, Porsch P, Pirich C, Studnicka M. Electromagnetic navigation bronchoscopy in combination with PET-CT and rapid on-site cytopathologic examination for diagnosis of peripheral lung lesions. Lung. 2009;187(1):55-59. 24.

[9.] Herth FJ, Eberhardt R, Vilmann P, Krasnik M, Ernst A. Real-time endobronchial Ultrasound-guidedtransbronchial needle aspiration for sampling mediastinal lymph nodes. Thorax. 2006;61(9):795-798.

[10.] Detterbeck FC, Postmus PE, Tanoue LT (2019). The stage classification of lung.