# Sudan University of Science and Technology

## Engineering College

## College of Graduate Studies and Scientific Research

**Thesis Title:**

# Early Diagnosis of Lung Cancer Using Adaptive Neuro Fuzzy Inference System

# التشخيص المبكر لسرطان الرئه باستخدام نظام الاستدلال عصبي غامض

Thesis submitted in partial fulfillment of the requirements for the award of degree of Master of Science in Biomedical Engineering


**Submitted by:**

Waha Alshareef Shazali ALshareef

**Supervised by:**

Dr. Eltaher Mohammed Hussein


October 2022

**الآية**

{إِنَّمَا إِلَـهُكُمُ اللَّهُ الَّذِي لا إِلَهَ إِلاَّ هُوَ وَسِعَ كُلَّ شَيْءٍ عِلْمًا}

صدق الله العظيم

سورة طه(98 )

# DEDICATION

To be or not to be ‘that is the secret we learned since childhood…

This work is dedicated to

(My parent)

To … those they give me ideas and support …..

(My family)

(My friends)

Special thanks to …….

(Eithar seif eldeen)

To everybody who have made this work possible…

# ACKNOWLEDGMENT

Thanking and appreciation to University of Sudan for Science and Technology, which including more efficient teachers and whom are doing hard under the banner of science and we appreciate special thanks to the biomedical engineering department.

Acknowledgment to the greatest and venerable teachers

**Dr. Eltahir Mohammed Hussein** who supervised this research and gave us counselling and advising with her brilliant idea that which helped us with her thoughts and efforts and her experience that light emitting our road until this research marched-out to the light.

# TABLE OF CONTENTS

الآية ..... I
DEDICATION ..... II
ACKNOWLEDGMENT ..... III
TABLE OF CONTENTS ..... IV
LIST OF FIGURES ..... **Error! Bookmark not defined.**
LIST OF TABLES ..... VIII
ABBREVIATION ..... IX
ABSTRACT ..... X
المستخلص ..... XI

**CHAPTER ONE**

**INTRODUCTION**

1.1 General View ..... 1
1.2 Problem Statement ..... 2
1.3 The Objective ..... 2
1.3.1 Specific objectives are to: ..... 2
1.4 Methodology ..... 2
1.5 Thesis layout ..... 3

**CHAPTER TWO**

**LITERATURE REVIEWS**

2.1 Literature review ..... 4
2.2 Summary ..... 8

**CHAPTER THREE**

**THEORETICAL BACKGROUND**

3.1 Respiratory system ..... 9
3.1.1 Anatomy of the lung: ..... 10
3.2 Cancer description ..... 11
3.3 Lung cancer: ..... 11

3.4 Risk factors .....12
3.5 Signs and Symptoms of Lung Cancer.....15
3.6 Types of lung cancer.....16
3.7 Stages of lung cancer .....17
3.8 Diagnostic of lung cancer .....18
3.9 Treatment .....20
3.10 Neural Network (NN) .....21
3.11 Fuzzy inference system.....22
3.12 Adaptive Neuro-Fuzzy Inference Systems (ANFIS).....23
3.12.1 ANFIS Architecture.....24
3.13 MATLAB.....27
3.14 Metric for evaluation of classifier performance .....27
3.14.1 The sensitivity.....28
3.14.2 The specificity.....28
3.14.3 The accuracy .....28

## CHAPTER FOUR

## METHODOLOGY

4.1 Introduction.....29
4.2 Data description .....30
4.3 Data per processing.....31
4.4 designing of ANFIS .....31

## CHAPTER FIVE

## RESULT AND DISCUSSION

5.1 Introduction:.....37
5.2 Results:.....37

## CHAPTER SIX

## CONCLUSION AND RECOMMENDATION

6.1 Conclusion .....41

6.2 Recommendation .....41
**REFERENCES**.....42

# LIST OF FIGURE

Figure 3. 1 human respiratory airway .....10
Figure 3. 2 Normal and abnormal cells.....11
Figure 3. 3 artificial neural network architecture .....21
Figure 3. 4 Fuzzy inference system .....23
Figure 3. 5 Basic structure of ANFIS .....24
Figure 4. 1the proposed system .....29
Figure 4. 2 Sugeno model 15 input and one output .....32
Figure 4. 3 input value age.....33
Figure 4. 4 rule of Sugeno model.....35
Figure 4. 5 ANFIS structure.....36
Figure 4. 6 structure of NN .....36
Figure 5. 1 increasing the age value.....37
Figure 5. 2 decreasing the age value .....38
Figure 5. 3 the performance evaluate FIS .....38
Figure 5. 4 confusion matrix of ANFIS classifier .....39

# LIST OF TABLES

Table (4. 1) Original Dataset attributes description.....30
Table (4. 2) example of survey lung cancer dataset .....31
Table (4. 3) example of data after pre possessing.....31

Table (5. 1) Comparison results for proposed method and (NN).....39
Table (5. 2) Comparison between the proposal with literature reviews.....40

# ABBREVIATION

| | |
|---|---|
| ANFIS | Adaptive Neuro Fuzzy Inference System |
| NN | Neural Network |
| MF | Membership Function |
| FIS | Fuzzy Inference System |

# ABSTRACT

Lung cancer is one of the most serious cancer worldwide. Diagnostic of lung cancer very weak because doctor will able to know the disease only at the advanced stage .the main objective of this study is to predict and early detection of lung cancer by using an adaptive neuro fuzzy inference system (ANFIS) and artificial neural network (ANN). Symptoms and other information about the person were used to diagnose the lung cancer, as input variables for (ANFIS) and (ANN).The data set obtained from data world which content 309 records and two classes. Data set fed into the system as input after pre-processed. According to this study the accuracy of (ANN) and ANFIS are 98.913 97.087 respectively. Model evaluation showed that the (ANN) and (ANFIS) have been able to detect the absence or presence of lung cancer.

# المستخلص

يعد سرطان الرئة من اخطر السرطانات في العالم .تشخيص سرطان الرئة عملية ضعيفة للغاية لأن الطبيب قادر علي معرفة المرض فقط في المرحلة المتقدمة .والهدف الرئيسي من هذه الدراسة هو التنبؤ المبكر عن سرطان الرئة باستخدام نظام الاستدلال العصبي الغامض التكيفي والشبكة العصبية الاصطناعية . تم استخدام الأعراض ومعلومات أخرى عن الشخص لتشخيص سرطان الرئة كمتغيرات إدخال لنظام الاستدلال العصبي الغامض التكيفي والشبكة العصبية الاصطناعية. مجموعة البيانات تم الحصول عليها من عالم البيانات والتي تحتوي على 309 سجل وفئتين. يتم تغذية مجموعة البيانات في النظام كمدخلات بعد المعالجة المسبقة .وفقا لهذه الدراسة فإن دقة الشبكات العصبية الاصطناعية ونظام الاستدلال العصبي الغامض التكيفي هي 98.913 97.087 على التوالي. أظهر تقييم النموذج أن الشبكات العصبية الاصطناعية ونظام الاستدلال العصبي الغامض التكيفي كانا قادرين على اكتشاف غياب أو وجود سرطان الرئة.

# CHAPTER ONE

# INTRODUCTION

## 1.1 General View

Lung cancer is a malignant lung tumour characterized by the uncontrolled growth of abnormal cells in one or both of the lungs. These abnormal cells do not carry out the functions of normal lung cells and do not develop into healthy lung tissue. With time, the abnormal cells start to form tumours that interfere with the functioning of the lung. If left untreated, this growth can spread beyond the lung into nearby tissue or other parts of the body. According to the world health organization (WHO) lung cancer is the leading cause of death from cancer (18.4%) of all death followed by colorectal (9.2%) and stomach cancers (8.2%) this makes it most common cause of death [1, 2].

In disease, attributes are the real information container of the disease. Lung cancer symptoms usually do not appear until the disease has progressed. Therefore early prediction is highly important so that the mortality rate can be easily prevented with effective control. So, the cure rate and prognosis depend on the early detection and diagnosis of the disease. But this early diagnosis is not an easy task, Survival rate of lung cancer differs from person to person. It depends on age, sex and race as well as health condition [3, 4].

Machine learning now plays a crucial role in the detection and prediction of medical diseases at early stages and enhances the diagnosis process. An adaptive neuro-fuzzy inference system was proposed in this study to help the physician in the diagnosis of lung cancer diseases based on attributes, which are a set of human symptoms, and information. These systems are widely accepted in medical institutions operating at all levels of healthcare. The accuracy of some up-to-date neuro-fuzzy systems today matches or even surpasses the

diagnostic abilities of physicians, thus holding an important role in risk assessment and diagnostics in medicine [4, 5].

## 1.2 Problem Statement

The process of diagnosing lung cancer is very weak because doctors will able to know the existence of the disease at the advanced stages usually the symptoms of lung cancer do not appear at an early stage in some cases even at an advanced level patients with lung do not have symptoms associated with the Lung cancer.

## 1.3 The Objective

The main objective of this research is to design an early diagnosis system to predict lung cancer based on an attribute which is a set of symptoms and information about person. An adaptive neuro-fuzzy inference system was suggested.

### 1.3.1 Specific objectives are to:

1. Design system that predicts lung cancer in early stage.
2. Evaluate the accuracy, sensitivity and specificity of the classifier.

## 1.4 Methodology

This research consists of five phases as shown in the block diagram below. Selected data were obtained from the data world website[6] this data was fed into the system as input after being pre-processed and then passed to the adaptive neuro-fuzzy model for classification and evaluation.

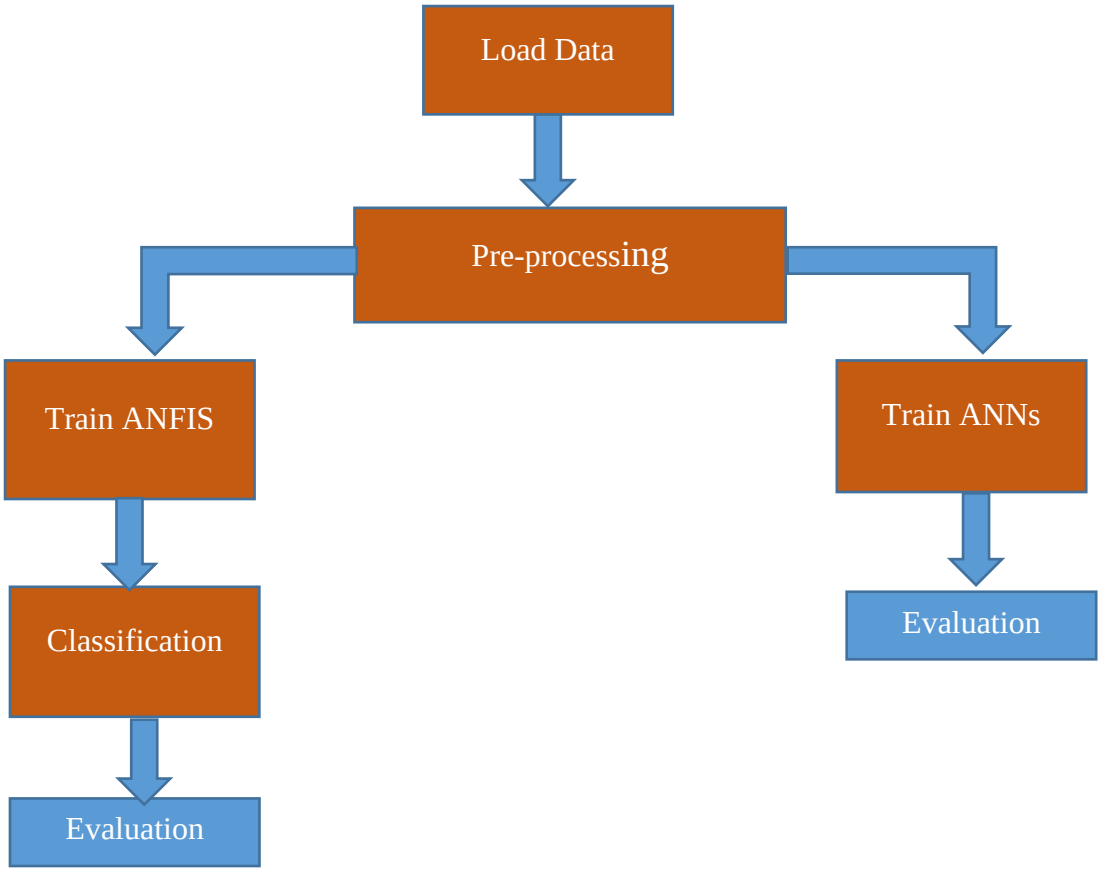


Figure (1. 1) The Proposed System Block Diagram

## 1.5 Thesis layout

This research consists of six chapters. Chapter one is an introduction, chapter two discusses the related literature review and chapter three deals with the theoretical background. The design and implementation of the classification system ware explained in chapter four. The result and discussion were illustrated in chapter five, and finally, the conclusion and recommendations were presented in chapter five.

# CHAPTER TWO

# LITERATURE REVIEWS

## 2.1 Literature review

Overview of the most relevant systems and methods in the detection of lung cancer.

In 2016 Aqeel Hamad proposed a diagnosis system to detect lung cancer based on fuzzy logic and neural network, he has used different methods to processing the input image by enhancement the image and separate the different properties of the image, then extract the features of the image by using different approaches , this features will be used as input to neural network that will be used as classifier tool to detect the normal and abnormal image that may contain lung nodule which is small growths in the lung, then we used other lung cancer symptoms such as coughing, blood in sputum, shortness of breath, pain in chest etc. as input to expert system to determine the condition of the patient [7].

In 2020 Radhanath Patra used weka tool to Prediction of Lung Cancer Using Machine Learning Classifier. Weka is an open source tool used for classification, clustering, regression and data visualization. Weka generally supports input file either in .csv or .arff extension format. In this paper author has analysed various machine learning classifiers techniques such as Neural Network, Radial Basis Function Network, J48 Classifier and Naïve Bayes Classifier to classify available lung cancer data in UCI machine learning repository in to benign and malignant. The input data is prepossessed and converted in to binary form followed by use of some well-known classifier technique in Weka tool to classify the data set in to cancerous and non-cancerous. The comparison technique reveals that the proposed RBF classifier

has resulted with a great accuracy of 81.25% and considered as the effective classifier technique for Lung cancer data prediction [4].

Ibrahim M.et al in 2019 they developed an Artificial Neural Network (ANN) for detect the absence or presence of lung cancer in human body. Symptoms were used to diagnose lung cancer. These symptom such as yellow finger anxiety , Chronic Disease, Fatigue, Allergy, Wheezing, Coughing, Shortness of Breath, Swallowing difficulty and Chest pain . They were used and other information about the person as input variables for the ANN. The dataset used in this study created by the user *sta427ceyin* on data world website .The Model evaluation showed that the ANN model is able to detect the absence or presence of lung cancer with 96.67 % accuracy[8].

Saadaldeen Rashid .et al in 2019 develop algorithms to determine whether a patient has or is likely to develop lung cancer using dataset images using data mining and machine learning for the classification and examination. The dataset obtain from UCI machine learning. The proposed system use combination of machine learning which includes support vector machine, k-nearest neighbour and decision trees. These approach use as classification for lung cancer. To detect [9][9][9][9]lung cancer in this approach start by detecting nodules in each patient's CT scans and extracting patches around them. Then we encode these patches using one of three encodings and use the average encoding of all the patches returned for a patient as the feature vector representing them. Finally, we train an SVM classifier on these feature vectors. 97.5% accuracy is achieved throughout the process [9].

Manikandan T *.et al* in 2017 designed a Hybrid Neuro-Fuzzy System (HNFS) for the prediction of lung diseases such as asthma, TB and lung cancer, based on observed symptom values, which could be used in Computer-Aided Diagnosis (CAD) systems to save the life of the patient. To predict lung diseases, Cancer Assessment Questionnaire for Lungs (CAQ-L) was prepared, which consists of

questions of symptoms on lung diseases .This study was carried out for 221 abnormal (asthma, tuberculosis and cancer) subjects and 50 normal subjects, aged between 37-81 years, asked to respond to the CAQ-L. The significant symptoms were identified based on Pearson's correlations performed on all the observed data. The proposed HNFS has achieved accuracy 75% [10].

S. Senthil in 2018 introduced an integrated framework for predicting lung cancer using Neural Network with Particle Swarm Optimization. Neural networks use to training and testing the input data sample .dataset used in this study obtain from UCI machine learning repository. Particle Swarm Optimization (PSO) is applied to extract the features of the given input images and further process is proceeded to detect the lung cancer[11]

Mustain B *.et al* in 2016 were developed an Adaptive Neuro Fuzzy Inference System (ANFIS) and Linear Discriminant Analysis (LDA) to diagnoses lung cancer .this system has mainly two steps they are Feature extraction reduction and classification. Lung cancer historical datasets are collected from different hospitals and they are the pre-processed dataset .To reduce the lung cancer features dimensionality, Linear Discriminant Analysis (LDA) is applied. Reduced features are then fed into AN-FIS classifier System. The proposed system obtains accuracy of about 95.4%[12].

AMJED KHAN *.et al* in 2021 proposed system to classify lung cancer in early stage used deep Convolutional neural networks (DCNN). In this research work, there are 900 CT images were used for training and testing purposes the dataset obtain from lung image database consortium and image database recourse (LIDC-IDR) consists of CT images in DICOM format of 1018 cases. The size of the original images are 512 x 512 but it is difficult to train large size images in DCNN so pre-processed the images to reduce size suitable for the network. Hence training and testing images are categorized for evaluating the network for efficient classification of images into cancerous and non-cancerous images and

helps for diagnosing the patient in the early stages. The complete process of DCNN gives 100% of accuracy with computation time of 45,141 seconds in single CPU workstation, which is the best level of accuracy obtained [13]

In 2020 Ahmad S. Ahmad proposed a new tool for the early prediction of lung cancer consist of three stages firstly analysis of an international cancer database that obtain from data world website. The database consists of 1000 records and 23 attributes that represent the symptoms, risk factors of lung cancer and three categories representing the risk levels of lung cancer: Low, Medium and High. Second stage, several medical questionnaires are distributed among a number of doctors and specialists in the fields of internal and thoracic tumours, in order to determine the most effective symptoms of lung cancer from a local medical point of view. In the third stage, medical knowledge from global research projects and reports is extracted, and the most appropriate pathway is defined in order to determine the risk of lung cancer by analysing the values of available diagnostic factors. Decision trees and random forest algorithms will help us find the most important factors that could affect the final decision (risk degree), and this will confirm the validity of the lung cancer prediction tool( LCPT) results .the system obtain 90.47%, 100% and 93.33% for specificity, sensitivity and accuracy respectively[14].

Table (2. 1) Some of previse studies

| Author/year | Classification | Data | Result % |
|---|---|---|---|
| Ibrahim M.et al in 2019 [6] | Artificial Neural Network (ANN) | Sta427ceyin on data world | 96.67 |
| Manikandan T .et al in 2017 [8] | Hybrid Neuro-Fuzzy System (HNFS) | Cancer Assessment Questionnaire for Lungs (CAQ-L) | 75 |
| S. Senthil in 2018 [9] | Neural network with particle swarm optimization | UCI machine learning | 97.8 |
| Mustain B .et al in 2016 [5] | Adaptive Neuro Fuzzy Inference System (ANFIS) and Linear Discriminant Analysis (LDA) | Different hospitals | 95.4 |
| Saadaldeen Rashid .et al in 2019 [7] | Data mining and machine learning | UCI machine learning | 97.5 |

## 2.2 Summary

Lung cancer has been a major concern to researchers in both oncology and the field of medical aid that is based on artificial intelligence. Some studies focus on feature extraction from the lung cancer patient image and other studies based on observed symptoms .Studies about the diagnosis of lung cancer have been based on techniques such as fuzzy logic and neural networks, deep convolution neural network other studies have used hybrid neuro-fuzzy techniques.

# CHAPTER THREE

# THEORETICAL BACKGROUND

## 3.1 Respiratory system

The respiratory system is the group of organs in our body that enable us to breathe. When we breathe we take in oxygen and give out carbon dioxide[15].

The respiratory system, functionally, can be separated in two zones; conducting zones (nose to bronchioles) form a path for conduction of the inhaled gases and respiratory zone (alveolar duct to alveoli) where the gas exchange takes place. Anatomically, respiratory system is divided into upper and lower respiratory tracts. Together, the two tracts are responsible for ventilation. The upper respiratory tract, known as the upper airway, warms and filters inspired air. The lower respiratory tract (the lungs) can accomplish gas exchange. Upper airway structures consist of the nose, sinuses and nasal passages, pharynx, tonsils and adenoids, larynx, and trachea. The lower respiratory tract consists of the lungs, which contain the bronchial and alveolar structures needed for gas exchange. Gas exchange involves delivering oxygen to the tissues through the bloodstream and expelling waste gases, such as carbon dioxide, during expiration [16, 17].

The human respiratory tract can be divided into 24, generations Figure 3.1 is a schematic representation of the respiratory tract. Regarding physiological functions, the contiguous airway from the trachea to the terminal bronchioles (generation index 0–16) is called the conducting zone, and the area from the respiratory bronchioles to the alveolar sacs (generation index 17–23) is called the transitional and respiratory zone.

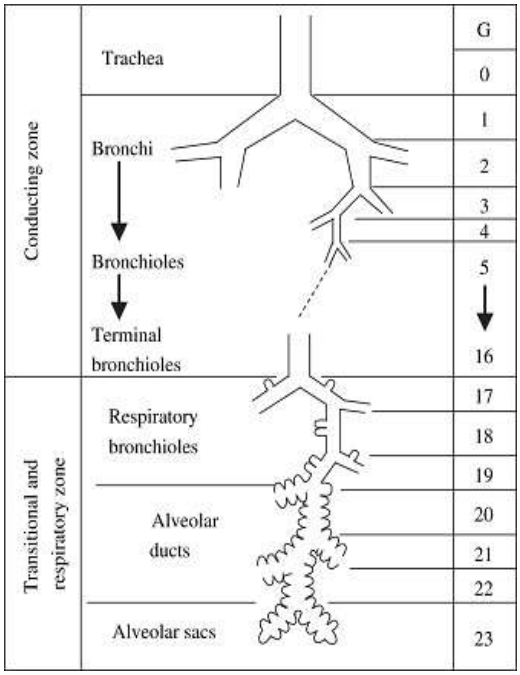


Figure 3. 1 human respiratory airway

### 3.1.1 Anatomy of the lung:

The lungs are the major organs of the respiratory system, and are a pair of spongy, air-filled organs located on either side of the chest [18]. Each lung is divided into lobes by fissures. Both lungs have oblique fissure and the right is further divided by a horizontal fissure. The oblique and horizontal fissure divides the lungs into superior, middle and inferior lobes. Thus the right lung has three lobes while the left has two. Each lobe is supplied by a lobar bronchus. The lobes are subdivided by Broncho pulmonary segments which are supplied by the segmental bronchi. [19] . The lungs are separated by the mediastinum. This area contains the heart, trachea, oesophagus, and many lymph nodes. The lungs are covered by a protective membrane known as the pleura and are separated from the abdominal cavity by the muscular diaphragm[20].

With each inhalation, air is pulled through the windpipe (trachea) and the branching passageways of the lungs (the bronchi), filling thousands of tiny air sacs (alveoli) at the ends of the bronchi .These sacs, are surrounded by small

blood vessels (capillaries). Oxygen passes through the thin membranes of the alveoli and into the bloodstream. The red blood cells pick up the oxygen and carry it to the body's organs and tissues. As the blood cells release the oxygen they pick up carbon dioxide, a waste product of metabolism. The carbon dioxide is then carried back to the lungs and released into the alveoli. With each exhalation, carbon dioxide is expelled from the bronchi out through the trachea [19].

## 3.2 Cancer description

Cancer is a disease in which some of the body’s cells grow uncontrollably and spread to other parts of the body. There are more than 100 types of cancer. Types of cancer are usually named for the organs or tissues where the cancers form. For example, lung cancer starts in the lung, and brain cancer starts in the brain [21].

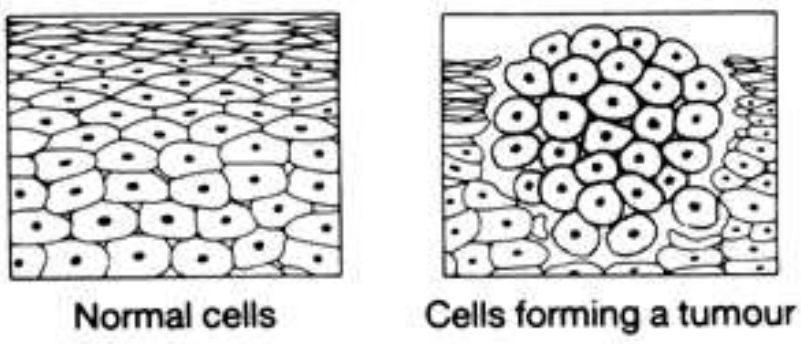


Figure 3. 2 Normal and abnormal cells

## 3.3 Lung cancer:

Lung cancer occurs when abnormal cells develop in one or both lungs and then grow so quickly that the body's immune system cannot keep up. The abnormal cells can form tumours that keep your lungs from working. Left untreated, the cancer can spread to nearby lymph nodes and other parts of your body. Worldwide lung cancer is the most common cancer among men in terms of both incidence and mortality and among women has the third highest

incidence, and is second after breast cancer mortality. In 2012, there were 1.82 million new cases and 1.56 million deaths due to lung cancer, representing 1904% of all deaths from cancer [22].

## 3.4 Risk factors

A risk factor is anything that increases a person's chance of getting a disease such as cancer. Different cancers have different risk factors. Several risk factors can make you more likely to develop lung cancer. But having a risk factor, or even several, does not mean that you will get the disease[23].

**Smoking**

Smoking, particularly of cigarettes, is far the main contributor to lung cancer. Across the developed world 90% of lung cancer deaths in men during the year 2000 were attributed to smoking 7 % for women. Smoking addiction is due to the presence of nicotine in tobacco. Nicotine itself is not responsible for the development of lung cancer but the exposure to tar leads to carcinoma of the lung. The total particulate matter of cigarette smoke after removing nicotine and water is called tar. Tar consists of about 3500 different compounds and most of them are carcinogenic. Smoking brings these substances into the body that damage cells lining the lungs[24, 25].

**Second-hand smoke (SHS)**

When people who don't smoke are exposed to second hand smoking it's called passive smoking. When you breathe in SHS, you take in nicotine and toxic chemicals the same way people who smoke do. SHS is thought to cause more than 7,000 deaths from lung cancer each year[26].

**Exposure to radon**

Radon gas you can't see, taste, or smell it generated by the breakdown of radioactive radium which in turn is the decay product of uranium, found in the Earth's crust. The radiation products ionize genetic material causing mutations

that sometimes turn cancerous. Radon is the second-most common cause of lung cancer in the USA, causing about 21,000 deaths each year[24].

**Exposure to asbestos**

Asbestos is a group of minerals that occur naturally as bundles of fibres. These fibres are found in soil and rocks in many parts of the world. They are made mainly of silicon and oxygen, but they also contain other elements.

Asbestos is a natural mineral product that's resistant to heat and corrosion[24]. It was used extensively in the past in products such as insulation, cement and some floor tiles.

The federal government began regulating the use of asbestos and asbestos products in the 1970s. Today, its handling is strictly regulated. Getting asbestosis is extremely unlikely if you follow your employer's safety procedures. Only trained and accredited asbestos professionals should manage asbestos products.

It’s not clear how much low-level or short-term exposure to asbestos might raise lung cancer risk[26].

**Genetic factors**

The multiple genetic factors play a role in lung cancer risk. Lung cancer genetic studies are focused on identification of mutations and single nucleotide polymorphisms (SNP) which increase the risk of lung cancer. Development of new technologies such as genomic profiling and genome-wide association studies has been helpful in the detection of new genetic variants likely involved in lung cancer risk [27, 28].

**Previous Respiratory Disease**

The chronic pulmonary diseases (COPD) and other chronic pulmonary diseases such as emphysema, chronic bronchitis and asthma could be caused by inflammation in lung tissues, these conditions may act as intermediates or

catalysts in the development of lung neoplasms or be related with lung cancer[29].

**Viral Infections**

The scientist that Retroviral DNA and Human Papillomavirus DNA were found in most of the cases of squamous cell carcinoma (85% and 69% respectively). Retroviral DNA was also found in 47% cases of adenocarcinoma. According to an international pooled analysis human papillomavirus (HPV) were found to more prevalent in lung cancer than the normal lung cells. Various studies have shown the presence of HPV type 18 and 16 DNA in lung tumour tissues. Some studies have also detected the presence of HPV E6-E7 mRNA in lung cancer cells[28].

**Diet and obesity**

Various reports suggest that obesity and dietary factors are associated with increased risk of lung cancers. Low serum concentrations of beta-carotene, vitamin C, and alpha-tocopherol have been associated with the development of lung cancer. In a case-control study of 242 lung cancer patients and their 484 matched controls on age, sex, and place of residence have shown that consuming bread, rice, beef, liver, dairy products, vegetable ghee, and animal ghee were possible risk factors for the development of lung cancer in Iran [27].

**Alcohol**

Previous studies have indicated that alcohol consumption, particularly consumption of beer, had been associated with the increased risk of lung cancer if tobacco smoking is controlled. The studies showed that there was 15% increase in the risk of lung cancer in heavy drinkers relative to non-drinkers or occasional drinkers. other studies showed that individuals who consumed 40+ oz./month of alcohol were at 1.9 times higher risk of developing lung cancer than non- drinkers. Alcohol consumption is linked with p53 mutation, suggesting alcohol may increase the mutagenic effects of cigarette smoke in the

lung. Studies have also shown an association between lung cancer and marijuana smoking. Marijuana smoke contains polycyclic aromatic compounds like benzopyrene, tetrahydrocannibol which cause mutations linked to lung cancer [27].

**Environmental Air Pollution**

Outdoor air pollution has a small effect on increasing the risk of lung cancer. Fine particulates and sulfate aerosols, which may be released in traffic exhaust fumes, are associated with highly increased risk. For nitrogen dioxide, an incremental increase of 10 parts per billion increases the risk of lung cancer by 14%. Outdoor air pollution is estimated to account for 1-2% of lung cancers. Tentative evidence supports an increased risk of lung cancer from indoor air pollution related to the burning of wood, charcoal, dung or crop residual for cooking and heating. Women are exposed to indoor coal smoke have twice the risk and number of the by-products of burning biomass are known or suspected carcinogens. The risk affects about 2.4 billion people globally, and is believed to account for 1.5% of lung cancer deaths [24].

**Family History of lung Cancer**

Familial history of lung cancer is also a major risk factor. A study by International Lung Cancer Consortium showed that there is 1.5 times more risk of lung cancer in individuals with a first-degree relative (mother, father, sibling) suffering from lung cancer. This study also showed that the risk of lung cancer is higher when a brother or sister is suffering from lung cancer [27].

## 3.5 Signs and Symptoms of Lung Cancer

Lung cancer usually doesn't cause signs and symptoms in its earliest stages. Signs and symptoms of lung cancer typically occur when the disease is advanced but some people with early lung cancer do have symptoms. Most of these symptoms are more likely to be caused by something other than lung

cancer. Still, if you have any of these problems, it’s important to see your doctor right away so the cause can be found and treated, if needed[30, 31].

The most common Signs and symptoms of lung cancer may include:

- A cough that does not go away or gets worse.
- Coughing up blood or rust-colored sputum (spit or phlegm).
- Chest pain that is often worse with deep breathing, coughing, or laughing.
- Hoarseness.
- Loss of appetite.
- Unexplained weight loss.
- Shortness of breath.
- Feeling tired or weak.
- Infections such as bronchitis and pneumonia that don’t go away or keep coming back.
- New onset of wheezing.

If lung cancer spreads to other parts of the body, it may cause:

- Bone pain (like pain in the back or hips).
- Nervous system changes (such as headache, weakness or numbness of an arm or leg, dizziness, balance problems, or seizures), from cancer spread to the brain.
- Yellowing of the skin and eyes (jaundice), from cancer spread to the liver.
- Swelling of lymph nodes (collection of immune system cells) such as those in the neck or above the collarbone.

## 3.6 Types of lung cancer

Cancer that starts in the lung is called primary lung cancer. If cancer spreads to your lungs from somewhere else in your body, this is secondary lung

cancer. There are different types of primary lung cancer and they are divided into two main group's small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC). The most common type is non-small cell lung cancer. About 80% to 85% of lung cancers are NSCLC. The main subtypes of NSCLC are adenocarcinoma, squamous cell carcinoma, and large cell carcinoma. These subtypes, which start from different types of lung cells are grouped together as NSCLC because their treatment and prognoses (outlook) are often similar. About 10% to 15% of all lung cancers are SCLC and it is sometimes called oat cell cancer. This type of lung cancer tends to grow and spread faster than NSCLC. About 70% of people with SCLC will have cancer that has already spread at the time they are diagnosed. Since this cancer grows quickly, it tends to respond well to chemotherapy and radiation therapy. Unfortunately, for most people, the cancer will return at some point [30].

**Other types of lung tumours**

Among with the main types of lung cancer, other tumours can occur in the lungs.

- Lung carcinoid tumours: Carcinoid tumours of the lung account for fewer than 5% of lung tumours. Most of these grow slowly.
- Adenoid cystic carcinomas, lymphomas, and sarcomas, as well as benign lung tumours such as hamartomas are rare. These are treated differently from the more common lung cancers.
- Cancers that spread to the lungs: Cancers that start in other organs can sometimes spread (metastasize) to the lungs, but these are not lung cancers.

## 3.7 Stages of lung cancer

Staging is an important step in choosing the best treatment for you. The stage of your cancer also helps doctors predict the outcome of your

treatment. One thing your lung cancer stage can't tell you is how long you'll live. Doctors use the same staging system for both non-small cell and small cell lung cancer. Stages of Lung Cancer based on The size and location of the original tumour (also called the primary tumour) secondly based on Whether the cancer has metastasized (spread) to nearby lymph nodes lastly based on Whether the cancer has metastasized (spread) to other organs[32]. There are four Stages of Lung Cancer are:

- **Stage 1**

In this stage the tumour is only in the lung. It is bit small (4 centimetres or less). It has not spread to nearby lymph nodes or outside the chest.

- **Stage 2**

In this stage lung cancer, there are larger tumours (more than 4 centimetres). Or, there are signs the cancer has spread to nearby lymph nodes, but not outside the lung.

- **Stage 3**

In stage 3 lung cancer, there is cancer in lymph nodes of the chest further away from the lung. Or, there may be large tumours that spread to nearby lymph nodes.

- **Stage 4**

In stage 4 lung cancer, there is cancer outside the chest cavity where it started. The most common areas are the other lung, bones, brain, and the adrenal gland (a gland on top of the kidneys).

## 3.8 Diagnostic of lung cancer

The diagnostic process usually begins with a simple physical examination of the person seeking medical care and then accordingly diagnostic tests. There are a variety of test to diagnoses lung cancer. Imaging tests suggest to look for

lung cancer. Imaging tests such as x-rays, magnetic fields, sound waves, or radioactive substances used to create pictures of the inside of your body. Imaging tests might be done for a number of reasons both before and after a diagnosis of lung cancer [26].

A chest X-ray is usually the first test used to diagnose lung cancer. Most lung tumours appear on X-rays as a white-grey mass. However, chest X-rays cannot give a definitive diagnosis because they often cannot distinguish between cancer and other conditions, such as a lung abscess (a collection of pus that forms in the lungs) [33].

A CT scan is usually the next test you'll have after a chest X-ray. A CT scan uses X-rays and a computer to create detailed images of the inside of your body. Before having a CT scan, you'll be given an injection containing a special dye called a contrast medium, which helps to improve the quality of the images [34].

MRI scans, which use radio waves and strong magnets to create detailed images of soft tissue. Like CT scans, they can produce detailed images of the tissue in the chest cavity. They are most often used to see if lung cancer has spread beyond its initial site [33].

The PET-CT scan (which stands for positron emission tomography-computerized tomography) may be done if the results of a CT scan show you have cancer at an early stage .also PET-CT scan  can show where there are active cancer cells and help with diagnosis and choosing the best treatment[33].

Once a doctor determines that there is reason to suspect that there may be cancer (or some other condition), further testing, which may include one or more of the following procedures.

Biopsies are the most common tool to obtain tissue for diagnosing lung cancer. Depending on where the nodule is located and the patient’s physical condition,

the doctor will do either a needle biopsy or a bronchoscopy. During a needle biopsy, the surgeon uses a syringe to remove tissue from the nodule. A CT scan guides the surgeon to the nodule [35].

Bronchoscopy is a biopsy done by passing a tube with a camera at the end, called a bronchoscope through the patient's mouth or nose, down into the trachea (windpipe) and then into the lungs where the suspicious nodule is located. This procedure that allows a doctor to see the inside of your airways and remove a small sample of cells (biopsy)[35].

End bronchial ultrasound scan (EBUS) is a newer procedure which combines a bronchoscopy with an ultrasound scan. Like a bronchoscopy, an EBUS allows a doctor to see the inside of your airways. However, the ultrasound probe on the end of the camera also allows the doctor to locate the lymph nodes in the centre of the chest so they can take a biopsy from them [35].

## 3.9 Treatment

The treatment of lung cancer depends on the type of lung cancer and its stage. People with lung cancer can be treated with surgery, chemotherapy, radiation therapy, targeted therapy, or a combination of these treatments [36].

**Surgery:** An operation where doctors cut out cancer tissue. The surgeon can remove part of the lung or the entire lung and surgeon also removes nearby lymph nodes.

**Chemotherapy:** using special medicines to shrink or kill the cancer. The drugs can be pills you take or medicines given in your veins, or sometimes both.

**Radiation therapy:** using high-energy rays (similar to X-rays) to kill the cancer.

**Targeted therapy:** using drugs to block the growth and spread of cancer cells. The drugs can be pills you take or medicines given in your veins. You will get

tests to see if targeted therapy is right for your cancer type before this treatment is used [35].

## 3.10 Neural Network (NN)

A neural network is a type of information processing unit, whose architecture is inspired by biological neural systems. It consists of three layers, namely, the input layer, the hidden layer and the output layer. The input is applied to the first layer (input), and the signals propagate through the middle (hidden) layer(s) to the output layer, which produces the result of the process. NN are primarily used to learn how tasks are done on the basis of data provided for training, and to generalize reasonable outputs for inputs that it has not handled before[10, 37]. The key element of this idea is to create new structures for the information processing system. The Artificial neural network architecture is shown in the figure 3.4.

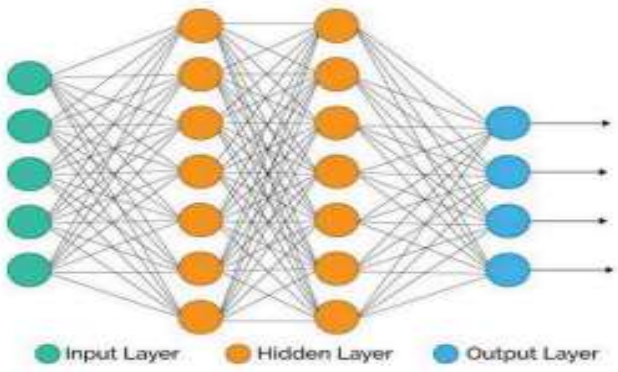


Figure 3. 3 artificial neural network architecture

The system is made up of a large number of highly interconnected processing elements called neurons that work together to solve a problem and transmit information through synapses (electromagnetic connections). The neurons are interconnected closely and organized into layers. The input layer receives the data, while the output layer generates the final result. Between the two, one or

more secret layers are typically sandwiched. This arrangement makes predicting or knowing the exact flow of data difficult [37].

Each connection has a connection weight, and each neuron has a threshold value and an activation function. It is calculated if each input has a positive or negative weight based on the sign of the input's weight. The weight affects the signal intensity at a connection .Neurons which have a threshold above which a signal is only transmitted if the aggregate signal exceeds it. The Activation Value is the weighted sum of the summing unit, and the output is generated based on the signal from this activation value. In these networks, if one cell is damaged, other cells can make up for its absence and contribute to its regeneration. These networks are capable of learning. Basically, the ability to learn is the most important feature of an intelligent system[38].

## 3.11 Fuzzy inference system

Fuzzy system was first developed by Zadeh in the mid-1960s for representing uncertain and imprecise knowledge. It provides an approximate but effective means of describing the behaviour of systems that are too complex, ill-defined, or not easily analysed mathematically. Fuzzy if-then rules or fuzzy conditional statements are expressions of the form IF A THEN B, where A and B are labels of fuzzy sets characterized by appropriate membership functions. Due to their concise form, fuzzy if-then rules are often employed to capture the imprecise modes of reasoning that play an essential role in the human ability to make decisions in an environment of uncertainty and imprecision. Fuzzy variables are processed using a system called a fuzzy inference system.it composed of five functional blocks show in Figure 3.4[39, 40].

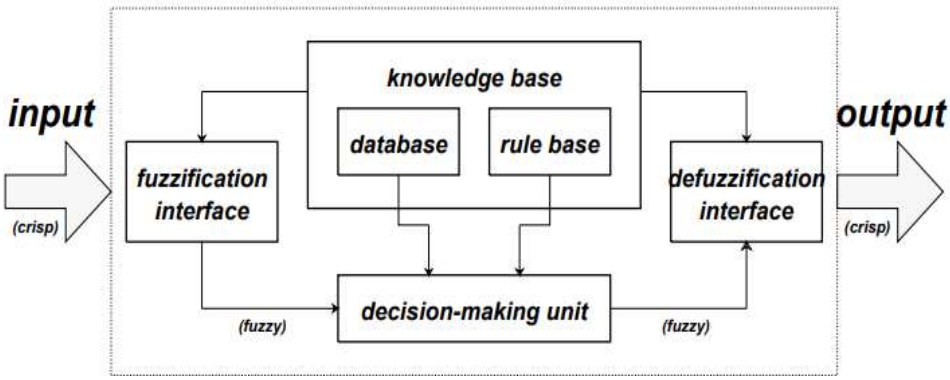


Figure 3. 4 Fuzzy inference system

- Rule base containing a number of fuzzy if-then rules.
- Database which defines the membership functions of the fuzzy sets used in the fuzzy rules.
- Decision-making unit which performs the inference operations on the rules.
- Fuzzification interface which transforms the crisp inputs into degrees of match with linguistic values.
- Defuzzification interface which transform the fuzzy results of the inference into a crisp output.

## 3.12 Adaptive Neuro-Fuzzy Inference Systems (ANFIS)

A neuro-fuzzy system is one of the most successful schemes which combine the benefits of these two powerful paradigms into a single capsule it works in Takagi-Sugeno type fuzzy inference system. ANFIS it is a neural network that is functionally equivalent to a fuzzy inference model trained by neural network learning algorithm. In a neuro-fuzzy system, neural networks extract fuzzy rules automatically from numerical data and, the membership functions are adjusted adaptively through the learning process. Training helps the system to develop fuzzy IF-THEN rules and determine membership functions for input and output variables of the system [41].

There are several features of the ANFIS which enable it to achieve great success in a wide range of scientific applications. The attractive features of an ANFIS

include: easy to implement, fast and accurate learning, strong generalization abilities, excellent explanation facilities through fuzzy rules, and easy to incorporate both linguistic and numeric knowledge for problem solving [42].

### 3.12.1 ANFIS Architecture

ANFIS has a similar structure to a multilayer feed forward neural network. According to the neuro-fuzzy approach, a neural network is proposed to implement the fuzzy system, so that structure and parameter identification of the fuzzy rule base are accomplished by defining, adapting and optimizing the topology and the parameters of the corresponding neuro-fuzzy network, based only on the available data. The network can be regarded both as an adaptive fuzzy inference system with the capability of learning fuzzy rules from data, and as a connectionist architecture provided with linguistic meaning. A typical architecture of an ANFIS, in which a circle indicates a fixed node, whereas a square indicates an adaptive node, is shown in Figure 3.5[42].

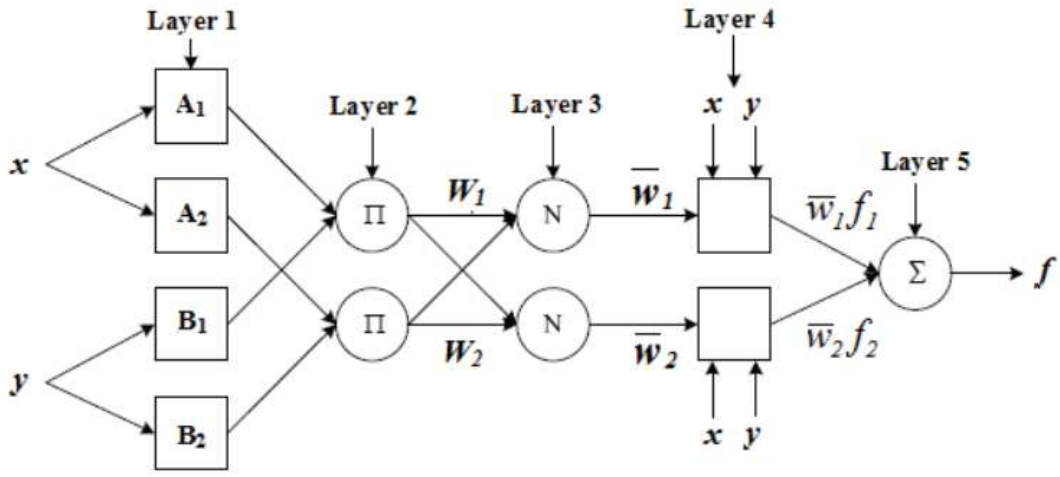


Figure 3. 5 Basic structure of ANFIS

In this connectionist structure, there are input and output nodes, and in the hidden layers, there are nodes functioning as membership functions (MFs) and rules. This eliminates the disadvantage of a normal feed forward multilayer network, which is difficult for an observer to understand or to modify [42]. For

simplicity, we assume that the examined FIS has two inputs and one output. For a first-order Sugeno fuzzy model, a typical rule set with two fuzzy "if -then" rules can be expressed as follows:

$$Rule\ 1: If x\ is\ A_1 and\ y\ is\ B_1, then\ f_1 = p_1x + q_1y + r_1 \quad \text{Equation( 3. 1 )}$$

$$Rule\ 2: If x\ is\ A_2 and\ y\ is\ B_2, then\ f_2 = p_1x + q_2y + r_2 \quad \text{Equation (3. 2)}$$

Where x and y are the two crisp inputs, and Ai and Bi are the linguistic labels associated with the node function.

As indicated in Fig. 3, the system has a total of five layers. The functioning of each layer is described as follows:

Input node (Layer 1):

Nodes in this layer contains membership functions. Parameters in this layer are referred to as premise parameters. Every node i in this layer is a square and adaptive node with a node function:

$$O_i^1 = \mu_{Ai}(x) \qquad For\ i = 1,2 \quad \text{Equation (3. 3 )}$$

Where x is the input to node i, and Ai is the linguistic label (small, large, etc.) associated with this node function. In other words, 1 $o^i$ is the membership function of Ai and it specifies the degree to which the given x satisfies the quantifier $A_i$ [42].

Rule nodes (Layer 2):

Every node in this layer is a circle node labelled II, whose output represents a firing strength of a rule. This layer chooses the minimum value of two input weights. In this layer, the AND/OR operator is applied to get one output that represents the results of the antecedent for a fuzzy rule, that is, firing strength. It means the degrees by which the antecedent part of the rule is satisfied and it

indicates the shape of the output function for that rule. The node generates the output (firing strength) by cross multiplying all the incoming signals [42].

$$O_i^2 = \omega_i \mu_{Ai}(x) * \mu_{Bi}(y), \quad i = 1,2 \quad \text{Equation (3. 4 )}$$

Average nodes (Layer 3):

Every node in this layer is a circle node labelled N. The $i^{th}$node calculates the ratio between the $i^{th}$rule's firing strength to the sum of all rules' firing strengths. Every node of these layers calculates the weight, which is normalized. For convenience, outputs of this layer are called normalized firing strengths [42].

$$\overline{\omega_i} = \frac{\omega_i}{\omega_1 + \omega_2}, \quad i = 1,2 \quad \text{Equation (3. 5)}$$

Consequent nodes (Layer 4):

This layer includes linear functions, which are functions of the input signals. This means that the contribution of ith rule's towards the total output or the model output and/or the function defined is calculated. Every node i in this layer is a square node with a node function [42]

$$O_i^4 = \overline{w}_i f_i = \overline{w}_i (p_i x + q_i y + r_i) \quad \text{Equation( 3. 6)}$$

Where$\overline{\omega_i}$ is the output of layer 3, and [7] is the parameter set of this node. These parameters are referred to as consequent parameters.

Output node (Layer 5):

The single node in this layer is a fixed node labelled ∑, which computes the overall output by summing all incoming signals [42].

$$O_i^5 = Overalloutput = \sum_i \overline{w_i} f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \quad \text{Equation (3. 7)}$$

## 3.13 MATLAB

Matlab is a programming language similar to other well-known languages such as Java, C#, etc., which comes with its own IDE (that is Integrated Development Environment) and set of libraries. Matlab is an abbreviation of the term “Matrix Laboratory” since it was initially referred to as the matrix programming language. It was first discovered by Cleve Moler.

Using Matlab you can implement and design different algorithms. You can load data from different sources such as files, databases or the web to analyse your data and visualize it using Matlab visualization application which gives you a wide range of graph plots to choose from. It also makes it easier to work with larger data sets. It as a math product contains a mathematical function library that allows you to perform linear algebra and computation of matrices. This also helps to facilitate data analysis. Creating data models, prototypes, and simulations of data can be achieved. You can also design interfaces for both users as well as other programming applications to make working with Matlab easier [43].

## 3.14 Metric for evaluation of classifier performance

Model evaluation for machine learning algorithms should ensure that data were transformed to the model properly and the model represents the system with an acceptable accuracy .in this study confusion matrix was, used to calculate TP, FN,TN and FN ,these value will be used to compute sensitivity, specificity and accuracy[44]

**Definitions**

- **Patient**: positive for disease
- **Healthy**: negative for disease
- **True positive(TP)**: the number of cases correctly identified as patient

- **false positive(FP)**: the number of cases incorrectly identified as patient
- **True negative(TN)**: the number of cases correctly identified as healthy
- **false negative(FN)**: the number of cases incorrectly identified as healthy

### 3.14.1 The sensitivity

The sensitivity of the test (or symptom) is the probability of a positive test results (or presence of the symptom) given the presence of the disease. Compute the conditional probability estimate [45].

$$\text{Sensitivity} = \frac{TP}{TP+FN} \qquad \text{Equation (3. 8)}$$

This ratio is an estimate of the sensitivity of screening test.

### 3.14.2 The specificity

The specificity of the test (or symptom) is the probability of a negative test results (or presence of the symptom) given the presence of the disease. Compute the conditional probability estimate[45].

$$\text{specificity} = \frac{TN}{TN+FP} \qquad \text{Equation ( 3. 9)}$$

This ratio is an estimate of the specificity of screening test.

### 3.14.3 The accuracy

The accuracy of a test is its ability to differentiate the patient and healthy cases correctly. To estimate the accuracy of a test, we should calculate the proportion of true positive and true negative in all evaluate cases[45]. Mathematical this can be stated as:

$$\text{Accurancy} = \frac{TP+TN}{TP+TN+FP+FN} \qquad \text{Equation( 3. 10)}$$

# CHAPTER FOUR

# METHODOLOGY

## 4.1 Introduction

This chapter discusses a  system was designed to diagnosis lung cancer data set using adaptive neuro fuzzy inference system and neural network and data classify into lung cancer  or not cancer , figure 4.1 shows the steps for the proposed system.

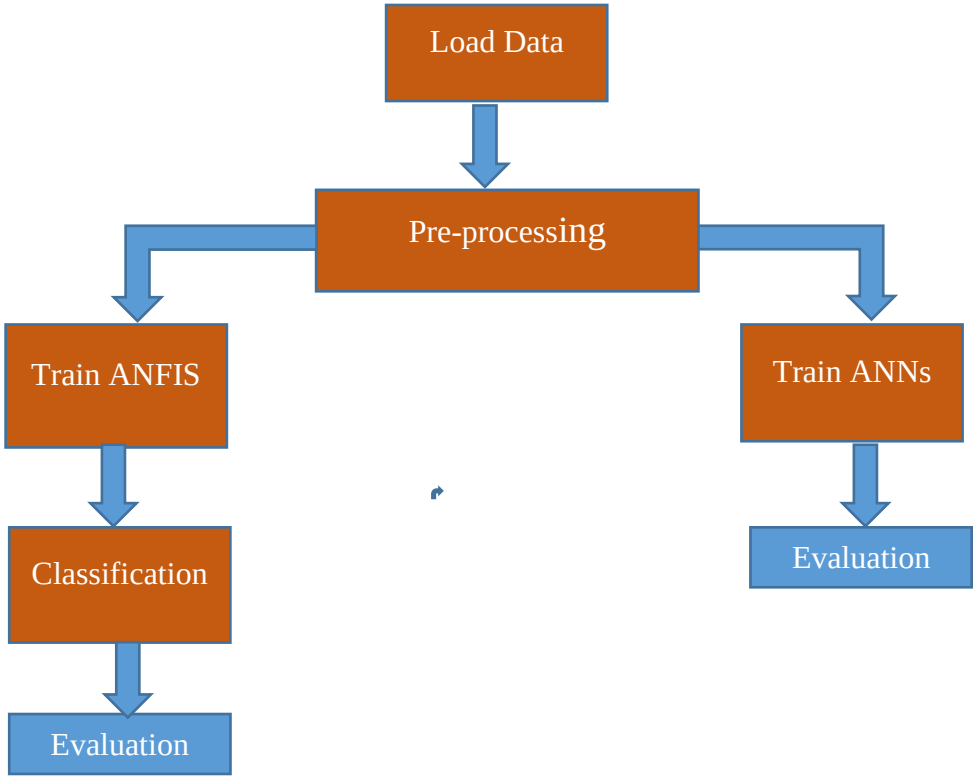


Figure 4. 1the proposed system

## 4.2 Data description

The data set obtain from website (data world) the data contain 309 records and 16 attributes. The first 15 attributes that represent symptoms and risk factors of lung cancer and lung cancer attribute represent persons that have lung cancer and that have not. Some pre-processing and transformation should done so the data is more suitable for predictive analysis. Table 4.1 show the original data before per processing and Table 4.2 shows the original dataset attributes description before pre-processing.

Table (4. 1) Original Dataset attributes description

| | Attribute | Scope |
|---|---|---|
| 1. | Gender | M(male), F(female) |
| 2. | Age | Age of the patient |
| 3. | Smoking | YES=2, NO=1. |
| 4. | Yellow fingers | YES=2, NO=1. |
| 5. | Anxiety | YES=2, NO=1. |
| 6. | Peer pressure | YES=2, NO=1. |
| 7. | Chronic Disease | YES=2, NO=1. |
| 8. | Fatigue | YES=2, NO=1. |
| 9. | Allergy | YES=2, NO=1. |
| 10. | Wheezing | YES=2, NO=1. |
| 11. | Alcohol | YES=2, NO=1. |
| 12. | Coughing | YES=2, NO=1. |
| 13. | Shortness of Breath | YES=2, NO=1. |
| 14. | Swallowing Difficulty | YES=2, NO=1. |
| 15. | Chest pain | YES=2, NO=1. |
| 16. | Lung Cancer | YES, NO. |

Table (4. 2) example of survey lung cancer dataset

| gender | age | smoking | yellow_fi | anxiety | peer_pres | chronic_d | fatigue | allergy | wheezing | alcohol_c | coughing | shortness | swallowir | chest_pai | lung_cancer |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| M | 69 | 1 | 2 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | TRUE |
| M | 74 | 2 | 1 | 1 | 1 | 2 | 2 | 2 | 1 | 1 | 1 | 2 | 2 | 2 | TRUE |
| F | 59 | 1 | 1 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | FALSE |
| M | 63 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 2 | 2 | FALSE |
| F | 63 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 1 | FALSE |
| F | 75 | 1 | 2 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 2 | 2 | 1 | 1 | TRUE |
| M | 52 | 2 | 1 | 1 | 1 | 1 | 2 | 1 | 2 | 2 | 2 | 2 | 1 | 2 | TRUE |
| F | 51 | 2 | 2 | 2 | 2 | 1 | 2 | 2 | 1 | 1 | 1 | 2 | 2 | 1 | TRUE |
| F | 68 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | FALSE |
| M | 53 | 2 | 2 | 2 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | TRUE |
| F | 61 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 2 | 1 | 2 | 2 | 2 | 1 | TRUE |
| M | 72 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 2 | TRUE |
| F | 60 | 2 | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 2 | 1 | 1 | FALSE |
| M | 58 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 2 | TRUE |
| M | 69 | 2 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 2 | FALSE |
| F | 48 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 2 | 2 | 2 | 1 | TRUE |
| M | 75 | 2 | 1 | 1 | 1 | 2 | 1 | 2 | 2 | 2 | 2 | 2 | 1 | 2 | TRUE |
| M | 57 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 2 | 1 | 1 | 2 | 2 | TRUE |
| F | 68 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | TRUE |
| F | 61 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 1 | 1 | FALSE |
| F | 44 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | TRUE |

## 4.3 Data per processing

The’ gender’ attribute value change from text into numerical value ‘,M ‘that is refer to male become 1 and’ F’ that is refer to female become 2 also the’ lung cancer ’classifier value changes to numerical value ‘true’ become 1 and’ false ‘become 0. Table 4.3 shows the example of data after pre possessing.

Table (4. 3) example of data after pre possessing

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 69 | 1 | 2 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 1 |
| 2 | 1 | 74 | 2 | 1 | 1 | 1 | 2 | 2 | 2 | 1 | 1 | 1 | 2 | 2 | 2 | 1 |
| 3 | 2 | 59 | 1 | 1 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 0 |
| 4 | 1 | 63 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 2 | 2 | 0 |
| 5 | 2 | 63 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 1 | 0 |
| 6 | 2 | 75 | 1 | 2 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 2 | 2 | 1 | 1 | 1 |
| 7 | 1 | 52 | 2 | 1 | 1 | 1 | 1 | 2 | 1 | 2 | 2 | 2 | 2 | 1 | 2 | 1 |
| 8 | 2 | 51 | 2 | 2 | 2 | 2 | 1 | 2 | 2 | 1 | 1 | 1 | 2 | 2 | 1 | 1 |
| 9 | 2 | 68 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| 10 | 1 | 53 | 2 | 2 | 2 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 1 |
| 11 | 2 | 61 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 2 | 1 | 2 | 2 | 2 | 1 | 1 |
| 12 | 1 | 72 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 2 | 1 |
| 13 | 2 | 60 | 2 | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 0 |
| 14 | 1 | 58 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 2 | 1 |
| 15 | 1 | 69 | 2 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 2 | 0 |
| 16 | 2 | 48 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 2 | 2 | 2 | 1 | 1 |
| 17 | 1 | 75 | 2 | 1 | 1 | 1 | 2 | 1 | 2 | 2 | 2 | 2 | 2 | 1 | 2 | 1 |
| 18 | 1 | 57 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 2 | 1 | 1 | 2 | 2 | 1 |
| 19 | 2 | 66 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 1 |
| 20 | 2 | 61 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 0 |
| 21 | 2 | 44 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 1 |
| 22 | 2 | 64 | 1 | 2 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 1 |



## 4.4 designing of ANFIS

For designing of ANFIS model procedure is divided into four steps:

- load data

- generate FIS
- train FIS
- test FIS

**Load data**

ANFIS modelling process starts by obtaining a data set and it was divided into training and testing data sets. The data set fed into the system after pre-processing.

**Generating ANFIS**

Sugeno model, selected to implement the ANFIS. In Sugeno model membership parameters select automatically. The ANFIS model's structure consists of input parameters (patient's symptoms), membership functions (MF) of input, fuzzy rules and output (diagnosis the lung cancer). This parameters are adapted by Scaled conjugate gradient method (training algorithm). The training algorithm use combination of back propagation and least squares method. To generating FIS Sugeno model there are 15 inputs includes gender, age, smoking, yellow fingers, anxiety, peer pressure, chronic disease, fatigue, allergy, wheezing, alcohol consuming, coughing, shortness of breath, swallowing difficulty and chest pain, and one output represent risk prediction .figure 4.3 showing Sugeno model input and output.

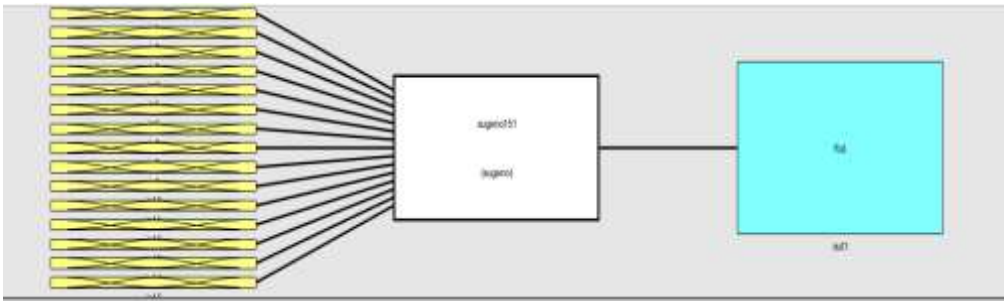

Figure 4. 2 Sugeno model 15 input and one output

In this system to generation FIS hybrid method is used. The MF of input data set adjusted by training algorithm. There are three MF are estimated for the suitable range of inputs value, all inputs have range (1 2) except age has range (21 87) figure 4.3 shown the MF of age.

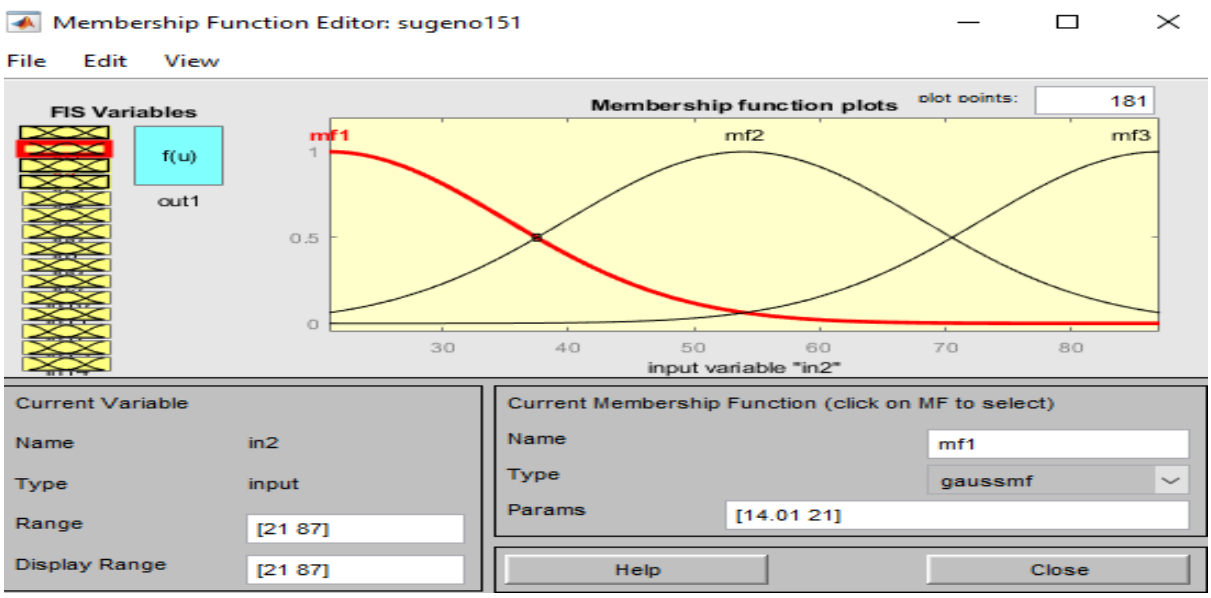


Figure 4. 3 input value age

The rule are made of for each of 15 input also the rule viewer predict the shape of MF as shown in figure 4.4. The rules are randomly generated based on the symptoms that detect the disease. Function rules are defined based on the if-then rule. In this system there is six rule sets are illustrated below:

- 1. If (in1 is in1cluster1) and (in2 is in2cluster1) and (in3 is in3cluster1) and (in4 is in4cluster1) and (in5 is in5cluster1) and (in6 is in6cluster1) and (in7 is in7cluster1) and (in8 is in8cluster1) and (in9 is in9cluster1) and (in10 is in10cluster1) and (in11 is in11cluster1) and (in12 is in12cluster1) and (in13 is in13cluster1) and (in14 is in14cluster1) and (in15 is in15cluster1) then (out1 is out1cluster1) (1)
- 2. If (in1 is in1cluster2) and (in2 is in2cluster2) and (in3 is in3cluster2) and (in4 is in4cluster2) and (in5 is in5cluster2) and (in6 is in6cluster2) and (in7 is in7cluster2) and (in8 is in8cluster2) and (in9 is in9cluster2)

and (in10 is in10cluster2) and (in11 is in11cluster2) and (in12 is in12cluster2) and (in13 is in13cluster2) and (in14 is in14cluster2) and (in15 is in15cluster2) then (out1 is out1cluster2) (1)

- 3. If (in1 is in1cluster3) and (in2 is in2cluster3) and (in3 is in3cluster3) and (in4 is in4cluster3) and (in5 is in5cluster3) and (in6 is in6cluster3) and (in7 is in7cluster3) and (in8 is in8cluster3) and (in9 is in9cluster3) and (in10 is in10cluster3) and (in11 is in11cluster3) and (in12 is in12cluster3) and (in13 is in13cluster3) and (in14 is in14cluster3) and (in15 is in15cluster3) then (out1 is out1cluster3) (1)
- 4. If (in1 is in1cluster4) and (in2 is in2cluster4) and (in3 is in3cluster4) and (in4 is in4cluster4) and (in5 is in5cluster4) and (in6 is in6cluster4) and (in7 is in7cluster4) and (in8 is in8cluster4) and (in9 is in9cluster4) and (in10 is in10cluster4) and (in11 is in11cluster4) and (in12 is in12cluster4) and (in13 is in13cluster4) and (in14 is in14cluster4) and (in15 is in15cluster4) then (out1 is out2cluster1) (1)
- 5. If (in1 is in1cluster5) and (in2 is in2cluster5) and (in3 is in3cluster5) and (in4 is in4cluster5) and (in5 is in5cluster5) and (in6 is in6cluster5) and (in7 is in7cluster5) and (in8 is in8cluster5) and (in9 is in9cluster5) and (in10 is in10cluster5) and (in11 is in11cluster5) and (in12 is in12cluster5) and (in13 is in13cluster5) and (in14 is in14cluster5) and (in15 is in15cluster5) then (out1 is out2cluster2) (1)
- 6. If (in1 is in1cluster6) and (in2 is in2cluster6) and (in3 is in3cluster6) and (in4 is in4cluster6) and (in5 is in5cluster6) and (in6 is in6cluster6) and (in7 is in7cluster6) and (in8 is in8cluster6) and (in9 is in9cluster6) and (in10 is in10cluster6) and (in11 is in11cluster6) and (in12 is in12cluster6) and (in13 is in13cluster6) and (in14 is in14cluster6) and (in15 is in15cluster6) then (out1 is out2cluster3) (1)

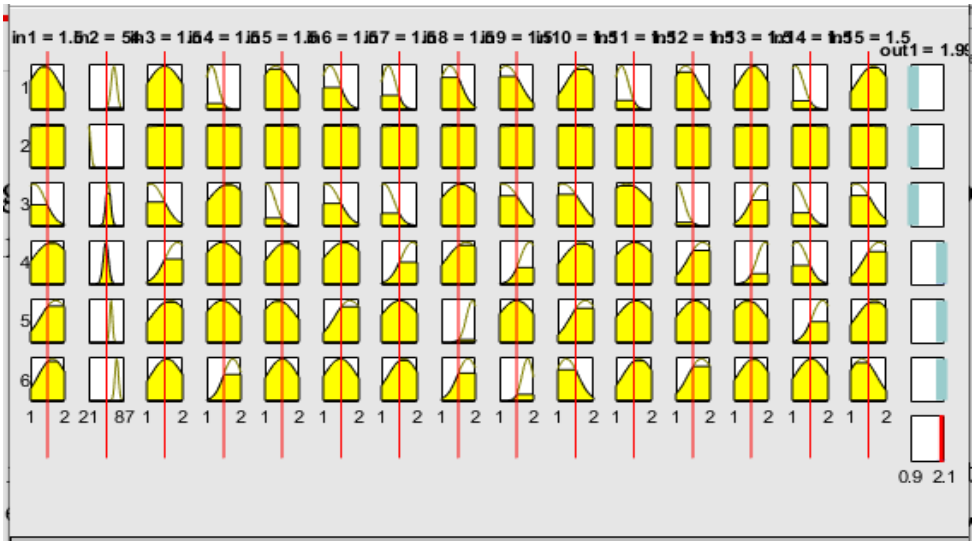

Figure 4. 4 rule of Sugeno model

- **Train FIS**

Training process is done by selecting the number of iterations. Different number of epochs in system performed in order to achieve the lower training error. This system, training is started with 25 epochs then epochs increases until 600. Training error occur during the training of ANFIS is 0.045688 at 600 epochs. The training process is stop when the error goal is achieved with maximum epoch number is reached show in result.

- **Test FIS**

Testing the model using a testing data set that is not the part of training data set to ensure the accuracy of FIS system. The testing data set known as validation set used to test the output of trained FIS to ensure the e the accuracy and to check the total number of cases provided true results and having false results. Figure 4.5 represents the ANFIS structure after training and testing the data.

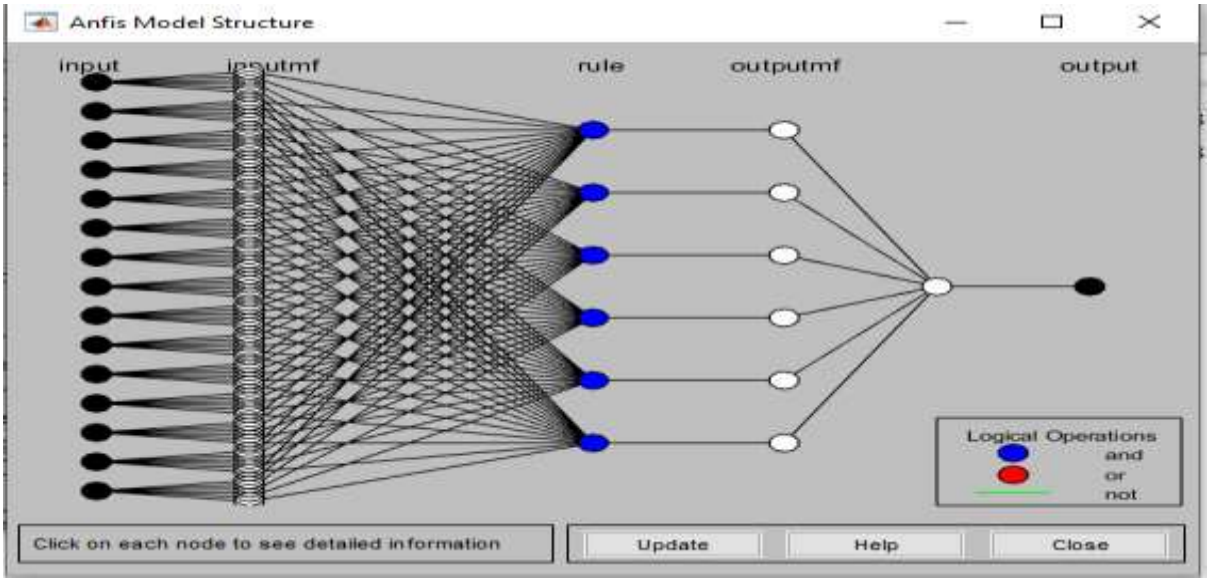


Figure 4. 5 ANFIS structure

To evaluate the performance of the system Feed forward neural network classifier is used .it has input layer, hidden layer and output layer .the data set with 15 attributes fed into the FNN as input passed through three layer .70%from the dataset used for training and 30 %used to test the model .figure 4.6 shown the structure of the FNN.

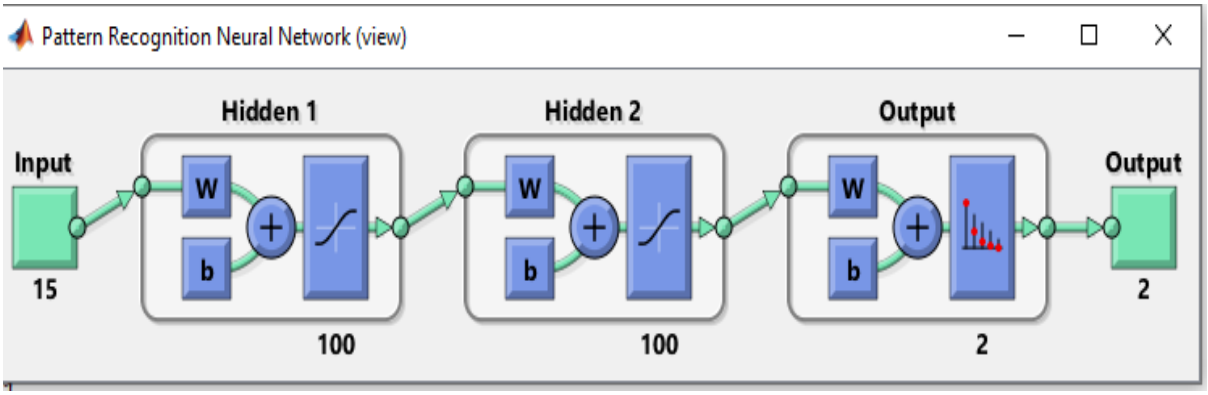


Figure 4. 6 structure of NN

# CHAPTER FIVE

# RESULT AND DISCUSSION

## 5.1 Introduction:

This section describes and discusses the performance evaluate FIS results that obtained from the applying of all the steps. And the result of proposed system will compare with the result of neural network (NN) with the help of performance parameters: accuracy and Sensitivity and specificity.

## 5.2 Results:

From rule viewer all inputs have average value, when we change the value of one variable and all the other inputs fixed in the middle, we notice that the most effective attribute of lung cancer from rule viewer is an input that represent age. Increasing or decreasing the age value directly affects the value of the output .figure 5.1 shown increasing the age value and the output became more than 1.5 that represent abnormal person. Decreasing the value of age the output became less than 1.5 that means normal person, that shown in figure 5.2.


Figure 5. 1 increasing the age value


Figure 5. 2 decreasing the age value

Figure 5.3 shows the performance of the ANFIS that occur when the epochs reach 600 and root mean square error (RMSE) is 0.045688.

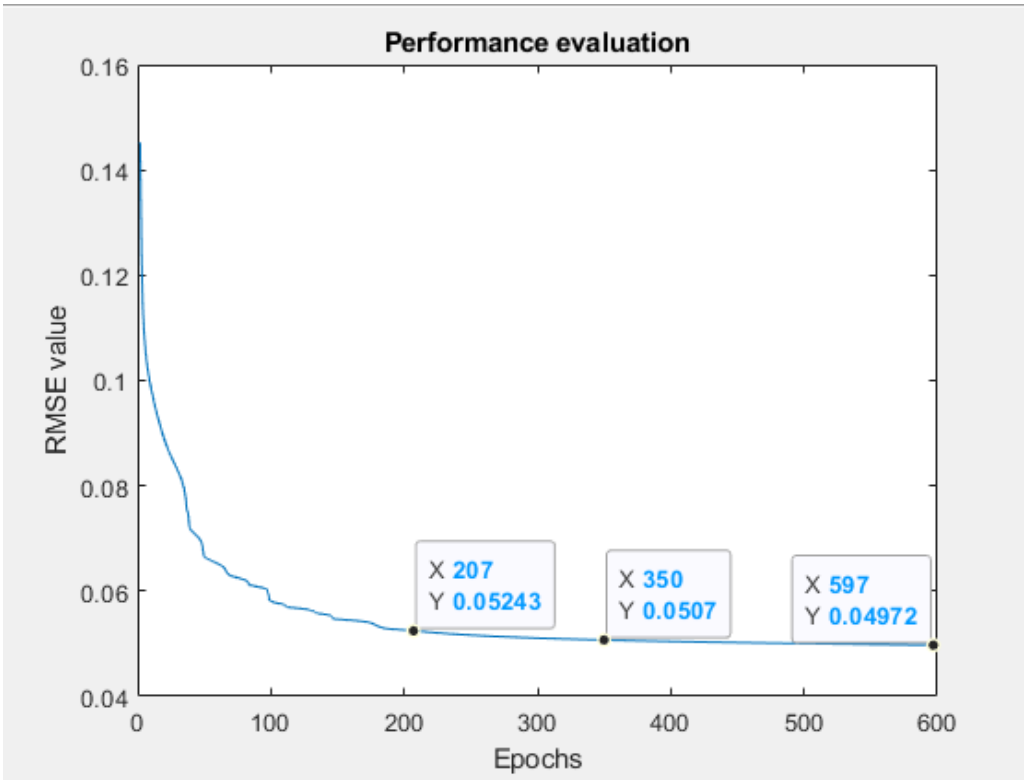


Figure 5. 3 the performance evaluate FIS

ANFIS algorithm was used to classification of 270malgnant (87%) and 39 benign (13%) case from 309 record with 15 attribute. This study lung cancer

risk has two class (1 considered as abnormal patient), (0 considered as normal patient) which classification given from dataset .For estimate the performance of the model confusion matrix was used as shown in figure 5.1. From the confusion matrix 309 patient's cases tested on the system, 265 true positive (TP) the extracted dataset containing cancer nodule is classified as cancerous, 5 False Negative (FN) the extracted dataset containing cancer nodule is classified as non-cancer, 35 True Negative (TN) the extracted dataset without cancer nodule is classified as non-cancerous, 4 False Positive (FP) the extracted dataset without cancer nodule is classified as cancerous.

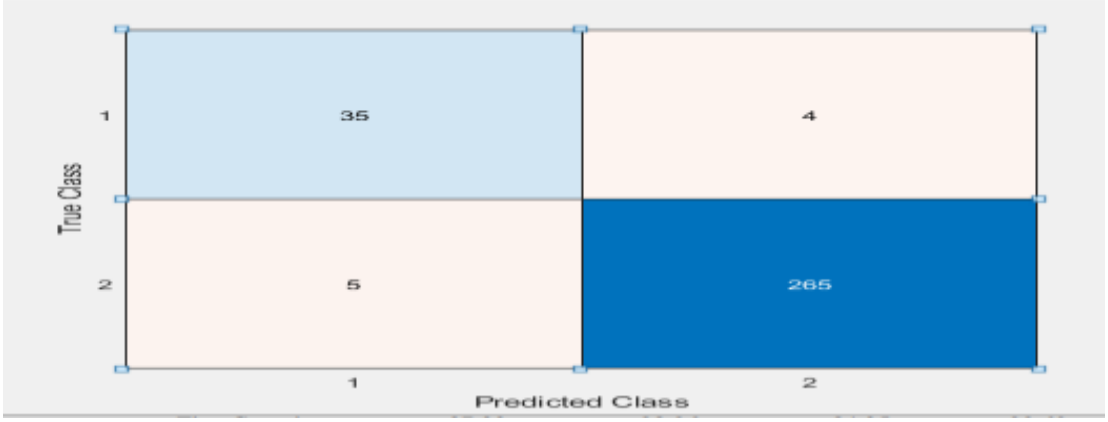


Figure 5. 4 confusion matrix of ANFIS classifier

Table 5.1show the comparison table of performance parameters for proposed Adaptive Neuro-Fuzzy Inference System (ANFIS) model and neural network (NN) for diagnosis lung cancer .the performance of the two systems is estimate by using sensitivity, specificity and accuracy.

Table (5. 1) Comparison results for proposed method and (NN)

| classifier | Accuracy (%) | Sensitivity (%) | Specificity (%) |
|---|---|---|---|
| ANFIS | 97.087 | 89.744 | 98.148 |
| NN | 98.913 | 99.98 | 98.795 |

From the table 5.1 above the proposed model shows the value of accuracy and sensitivity and specificity is lower than neural network (NN), so the outcome

achieved from NN is more efficient as compared to proposed model. Also the proposed model comparison with other models from the literature reviews. Table 5.2 shows Comparison between the proposed models with literature reviews.

Table (5. 2) Comparison between the proposal model with literature reviews

| Author/year | Classification | Data | Result % |
|---|---|---|---|
| The Proposal model | Adaptive neuro fuzzy inference system<br>Artificial Neural Network (ANN) | Sta427ceyin on data world | 97.087<br>98.913 |
| Ibrahim M.et al in 2019 [6] | Artificial Neural Network (ANN) | Sta427ceyin on data world | 96.67 |
| Manikandan T .et al in 2017 [8] | Hybrid Neuro-Fuzzy System (HNFS) | Cancer Assessment Questionnaire for Lungs (CAQ-L) | 75 |
| S. Senthil in 2018 [9] | Neural network with particle swarm optimization | UCI machine learning | 97.8 |
| Mustain B .et al in 2016 [5] | Adaptive Neuro Fuzzy Inference System (ANFIS) and Linear Discriminant Analysis (LDA) | Different hospitals | 95.4 |
| Saadaldeen Rashid .et al in 2019 [7] | Data mining and machine learning | UCI machine learning | 97.5 |

# CHAPTER SIX

# CONCLUSION AND RECOMMENDATION

## 6.1 Conclusion

Lung cancer is one of the most deadly diseases .early diagnosis may contribute to survival. Now day's expert system and machine learning play important role in diagnoses diseases.in this study ANFIS method and NN were used to classify lung cancer based on symptoms. The dataset obtain from data world web site to diagnose lung cancer are presented. This data fed into the system as input .The performance evaluation of ANFIS method and NN shows acceptable results and it indicates that neural network can be more effectively than ANFIS. The prediction could help doctor to plan for a better medication and provide the patient with early diagnosis.to evaluation the ANFIS the accuracy is 97.087 and Neural network is 98.913.

## 6.2 Recommendation

- The future work may involve feature extraction to achieve more optimal results.
- Link the model with a website and create a user interaction interface.

# REFERENCES


[1] W. H. Organization, "WHO report on cancer: setting priorities, investing wisely and providing care for all," 2020.

[2] D. World. "Lung Cancer: Types, Symptoms, Causes, Treatments." https://www.disabled-world.com/health/cancer/lung/ (accessed 29.sep.2022).

[3] M. Billah and N. Islam, "An Early Diagnosis System for Lung Cancer Risk Using Neuro Fuzzy Inference System and Linear Discriminant Analysis," *Journal of MPE Molecular Pathological Epidemiology,* vol. 1, no. 1, 2016.

[4] R. Patra, "Prediction of lung cancer using machine learning classifier," in *International Conference on Computing Science, Communication and Security*, 2020: Springer, pp. 132-142.

[5] H.-N. L. Teodorescu, A. Kandel, and L. C. Jain, *Fuzzy and neuro-fuzzy systems in medicine*. CRC Press, 1998.

[6] staceyinrobert. https://data.world/sta427ceyin (accessed 25 sep, 2022).

[7] A. M. Hamad, "Lung cancer diagnosis by using fuzzy logic," *IJCSMC*, vol. 5, no. 3, pp. 32-41, 2016.

[8] I. M. Nasser and S. S. Abu-Naser, "Lung cancer detection using artificial neural network," *International Journal of Engineering and Information Systems (IJEAIS)*, vol. 3, no. 3, pp. 17-23, 2019.

[9] S. R. A. Ahmed, I. Al Barazanchi, A. Mhana, and H. R. Abdulshaheed, "Lung cancer classification using data mining and supervised learning algorithms on multi-dimensional data set," *Periodicals of Engineering and Natural Sciences (PEN)*, vol. 7, no. 2, pp. 438-447, 2019.

[10] T. Manikandan, N. Bharathi, M. Sathish, and V. Asokan, "Hybrid neuro-fuzzy system for prediction of lung diseases based on the observed symptom values," *J Chem Pharm Sci ISSN*, vol. 974, p. 2115, 2017.

[11] S. Senthil and B. Ayshwarya, "Lung cancer prediction using feed forward back propagation neural networks with optimal features," *International Journal of Applied Engineering Research*, vol. 13, no. 1, pp. 318-325, 2018.

[12] M. Billah and N. Islam, "An early diagnosis system for predicting lung cancer risk using adaptive neuro fuzzy inference system and linear discriminant analysis," *Journal of MPE Molecular Pathological Epidemiology*, vol. 1, no. 1: 3, pp. 1-4, 2016.

[13] A. Khan, "Identification of Lung Cancer Using Convolutional Neural Networks Based Classification," *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, vol. 12, no. 10, pp. 192-203, 2021.

[14] A. S. Ahmad and A. M. Mayya, "A new tool to predict lung cancer based on risk factors," *Heliyon*, vol. 6, no. 2, p. e03402, 2020.
[15] D. M. Ornitz and Y. Yin, "Signaling networks regulating development of the lower respiratory tract," *Cold Spring Harbor Perspectives in Biology*, vol. 4, no. 5, p. a008318, 2012.
[16] A. Patwa and A. Shah, "Anatomy and physiology of respiratory system relevant to anaesthesia," *Indian journal of anaesthesia*, vol. 59, no. 9, pp. 533-541, 2015.
[17] G. Hedenstierna and H. U. Rothen, "Respiratory function during anesthesia: effects on gas exchange," *Comprehensive Physiology*, vol. 2, no. 1, pp. 69-96, 2011.
[18] H. M. I. T. Much, "Medically Reviewed by Sabrina Felson, MD on July 30, 2020."
[19] physiopedia. https://www.physio-pedia.com/Lung_Anatomy (accessed 21oct, 2022).
[20] N. C. I. . National Institutes of Health. https://training.seer.cancer.gov/citation.html (accessed 23oct 2021).
[21] N. C. Institute. https://www.cancer.gov/syndication (accessed 23sep, 2021).
[22] N. C. Institute. https://www.cancer.gov/ (accessed 5sep, 2022).
[23] A. C. Society. https://www.cancer.org/cancer.html (accessed.
[24] M. Mustafa, A. Azizi, E. IIIzam, A. Nazirah, S. Sharifa, and S. Abbas, "Lung cancer: risk factors, management, and prognosis," *IOSR Journal of Dental and Medical Sciences*, vol. 15, no. 10, pp. 94-101, 2016.
[25] D. Gupta, P. Boffetta, V. Gaborieau, and S. Jindal, "Risk factors of lung cancer in Chandigarh, India," *Indian Journal of Medical Research*, vol. 113, p. 142, 2001.
[26] A. C. Society. https://www.cancer.org/ (accessed 26sep, 2022).
[27] A. L. Marshall and D. C. Christiani, "Genetic susceptibility to lung cancer—light at the end of the tunnel?," *Carcinogenesis*, vol. 34, no. 3, pp. 487-502, 2013.
[28] A. Izzotti *et al.*, "Identification by MicroRNA Analysis of Environmental Risk Factors Bearing Pathogenic Relevance in Non-Smoker Lung Cancer," *Journal of personalized medicine*, vol. 11, no. 7, p. 666, 2021.
[29] H. Wang *et al.*, "Association between chronic obstructive pulmonary disease and lung cancer: a case-control study in Southern Chinese and a meta-analysis," 2012.
[30] R. L. Krech, J. Davis, D. Walsh, and E. B. Curtis, "Symptoms of lung cancer," *Palliative medicine*, vol. 6, no. 4, pp. 309-315, 1992.
[31] T. Tarver, "Cancer facts & figures 2012. American cancer society (ACS) Atlanta, GA: American Cancer Society, 2012. 66 p., pdf. Available from," ed: Taylor & Francis, 2012.

[32] J. C. Nesbitt, J. B. Putnam Jr, G. L. Walsh, J. A. Roth, and C. F. Mountain, "Survival in early-stage non-small cell lung cancer," *The Annals of thoracic surgery*, vol. 60, no. 2, pp. 466-472, 1995.
[33] Nomensa. (accessed 25 May, 2022).
[34] N. services. https://www.nhs.uk/conditions/lung-cancer/diagnosis (accessed 15 aug, 2022).
[35] M. P. Rivera, F. Detterbeck, and A. C. Mehta, "Diagnosis of lung cancer: the guidelines," *Chest*, vol. 123, no. 1, pp. 129S-136S, 2003.
[36] C. f. D. C. a. P. Division of Cancer Prevention and Control. https://www.cdc.gov/cancer/lung/basic_info/what-is-lung-cancer.htm (accessed Sep 30, 2022).
[37] H. A. Abdulbaqi, A. S. A. Jabar, and Z. S. A. Jabar, "INTEGRATED SOFTWARE PROJECT RISKS METHOD BASED ON PDF-ANN TECHNIQUES."
[38] R. Dastres and M. Soori, "Artificial Neural Network Systems," *International Journal of Imaging and Robotics (IJIR)*, vol. 21, no. 2, pp. 13-25, 2021.
[39] J.-S. Jang, "ANFIS: adaptive-network-based fuzzy inference system," *IEEE transactions on systems, man, and cybernetics*, vol. 23, no. 3, pp. 665-685, 1993.
[40] R. Hndoosh, M. Saroa, and S. Kumar, "Fuzzy and adaptive neuro-fuzzy inference system of washing machine," *European Journal of Scientific Research*, vol. 86, no. 3, pp. 443-459, 2012.
[41] T. Mitiku and M. S. Manshahia, "Neuro fuzzy inference approach: a survey," *Int. J. Sci. Res. Sci. Eng. Tech*, vol. 4, pp. 505-519, 2018.
[42] A. T. Azar, "Adaptive neuro-fuzzy systems," *Fuzzy systems*, vol. 42, no. 11, pp. 85-110, 2010.
[43] -. EDUCBA. https://www.educba.com/what-is-matlab/ (accessed sep 23, 2022).
[44] G. Varoquaux and O. Colliot, "Evaluating machine learning models and their diagnostic value," ed, 2022.
[45] B. J. Erickson and F. Kitamura, "Magician's corner: 9. Performance metrics for machine learning models," *Radiology: Artificial Intelligence*, vol. 3, no. 3, 2021.