Analysis of the Lung Cancer patient's for Data Mining Tool

Soobia Saeed[1†], Afnizanfaizal Abdullah[2††], NZ Jhanjhi[3†††]

Department of Software Engineering, Universiti Teknologi Malaysia[1], [2]
School of Computing & IT (SoCIT) Taylor's University, Malaysia[3]

Abstract
Data mining technology recently focuses on the methods of classification of the decision tree in data mining and propose a new algorithm for the classification of the decision tree with variable accuracy. The researcher uses the data analysis tool Rattle R and Weka. The researcher use data sets for different age groups are divided into gender-related treatment for lung cancer using various modes of treatment in this research. The age group is in between (30- 60 years) with categories in males and females. The decision tree is a suitable and sufficient algorithm for analyzing the results of treatment with radiation and chemotherapy for a specific age group. The Rattle R and Weka tools predict each group for best treatment method by which the appropriate treatment method can be analyzed. The predictions are also compared using graph plots with related tables also. These graphs are correlated with the forecasts. The researcher introduces the most efficient and widely used classification methods for data mining techniques and the main concepts of the decision tree method. In addition, the two data mining software rattle R and Weka are briefly described. To illustrate the procedure of this research, 200 real data sets were then compared in terms of the accuracy of the classification between the two different algorithms of the decision tree.

Keywords
Rattle R , Weka, tool, lung, cancer, treatment

1. Introduction

Tumor growth takes more space in the body. Cancer can cause stress in the surrounding structures. It can also grow in nearby structures of the body. That's called the area's invasion. It is not fully understood how cancer really grows in surrounding tissues. Cancer can grow from where it started in a random direction. Researchers know, however, that tumors can spread more easily in some tissues than others. For example, large blood vessels with strong walls and dense tissues such as tumor cartilage are difficult to develop.

Tumors usually grow along the "least resistance path." This means that the easier route is likely to be taken. Research has found three different ways that tumors can grow in the surrounding tissues. The tumor can be propagated in three forms. The most common method depends on the type of tumor and where the body is growing. Below are three ways in which tumors can grow into surrounding tissues [1].

When the tumor grows and takes up more space, it begins to press the nearby body's normal tissues. The growth of the tumor is pushed through normal tissue. The emergence of growth occurs, like a finger, because some pathways are easier to penetrate than others to grow cancer. Cancer, for example, can grow directly between muscle tissue sheets rather than between muscles. As cancer grows, the small blood vessels in the area are squeezed and closed. The Part of normal tissue starts to die due to low blood and oxygen levels. This makes it easy for cancer to keep working [2]. It focuses on tree classification methods in the field of data extraction in data mining technology and proposes a new classification algorithm for variable - precision tree resolution. A collection of lung cancer data for comparative analysis using a data extraction technique in this study leads to predictive treatment that contributes more to lung cancer perfection [1][2].

To analyze the data, the researcher uses Rattle R. Data sets are divided into a genus related to the treatment of lung cancer using different methods studied. The age group ranges from men to women (30-60 years). The decision tree is an appropriate and sufficient algorithm for the analysis of radiotherapy and chemotherapy results for a particular age group. Rattle R predicts each group's best treatment to analyze the appropriate method of treatment. In health sciences, information technology is becoming increasingly involved. Artificial intelligence is the computer industry that participates more effectively and efficiently in medical science. Using the data extraction methodology, many health care support systems have been built. In hospitals and clinics, these systems are now widely used.

It has proven to be very useful in decision - making for the patient and medical experts. To develop these systems, different methodologies are used. In different methodologies, the method of collecting input data and providing output information differs. Any computer program that helps experts make medical care decisions is subject to the control of the support system for health care decisions. An important feature of data mining is that creativity and the use of health knowledge can be supported. The main objective of this document is to provide the best treatment for lung cancer with modern trends in health care support systems [3].

- ✓ Discuss methodologies used in health care.
- ✓ Use of the electronic record used in health.

The key step in finding predictive and hidden information from large databases is data mining. The following is a formal definition of knowledge discovery in databases:

Manuscript received July 5, 2019
Manuscript revised July 20, 2019

"Data mining is a small extract of implicit information from previously unknown data. “Data extraction technology offers a new and hidden data styles user-oriented approach. Knowledge discovered by health officials can be used to improve service quality. Physicians can likewise utilize learning to decrease the number of antagonistic medication impacts to propose more affordable treatment options. The following are some significant zones where information mining procedures can be most valuable in the administration of restorative consideration and information demonstrating for well-being applications and executive information system for health care. Physicians can likewise utilize learning to decrease the number of antagonistic medication impacts to propose more affordable treatment options. The following are some significant zones where information mining procedures can be most valuable in the administration of restorative consideration and information demonstrating for well-being applications and executive information system for health care [4].

- ✓ Expectations of treatment costs and resource demand.
- ✓ Predict future behavior of the patient in view of its history
- ✓ Public health information
- ✓ E-governance structures.
- ✓ Health insurance

1.1 Data Mining Techniques

The most important knowledge discovery technology in the database (KDD) is data extraction technology. KDD has various procedures, such as data cleaning, data integration, data selection, data conversion, data mining, model assessment, cognitive visualization, etc. In the data extraction project, various types of techniques are used. These include the decision tree, networks of Bayes, naive berry, neural network sets [5].

1.2 Decision Tree

The most widely used technique for data analysis. It is used to classify records in a suitable category and applies to regression and correlation tasks. Decision trees determine the sequence of personality in the medical field. A happy tree consists of nodes specifying conditional attributes: Symptoms [3]

X = {x1, x2,.....xk}, branches showing the S ih values, i.e. the h-th range for i-th symptoms and leaves showing the decision.

Y = {y1, y2.....yk} and their binary values Zdk = {0,1}. A decision tree for the sample is shown in fig.1 [4].

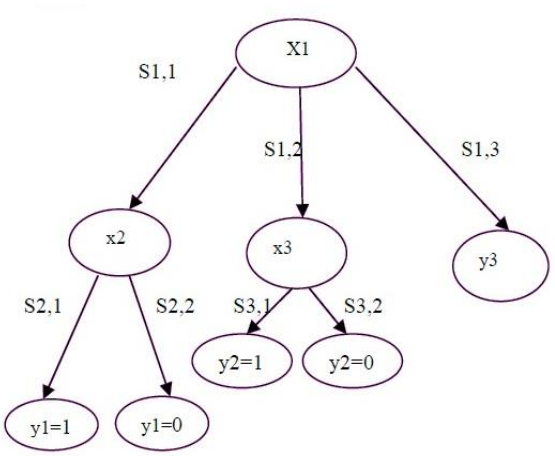


Fig.1 Decision tree applicable in health care

1.3 Causes of Lung Cancer

There are a wide variety of causes and cancer groups. In terms of their biological and path physiological nature, each of them is different. Even all plants and animals can be cancer susceptible [5].

1.4 Cancer at the molecular level

The living body has millions and billions of living cells growing and dying under the control of the cell's DNA mechanism. This growth process is faster in the body of the child, but the dead cells are replaced by the new dose only to repair the injuries once they become adult. When his cells become ill and out of control, the body can become an easy target for Kanner. Compared to normal cell growth, cancer cells have different growth types. Keep growing and form new but abnormal cells. These abnormal cells also attack other tissues, making them different from normal cells. This causes cancer to develop when the planner's DNA in the natural cell begins to damage. For cell growth, death and synthesis, the cell's DNA are responsible. The cell dies if the cell DNA is damaged. DNA does not die in the case of cancer, which is damage, but it produces more abnormal cells with the same damage as abnormal DNA. This phenomenon can be inherited from parents or caused by a sudden problem. This phenomenon is referred to as a mutation. Exposure to external toxins such as cigarette smoke can cause damage to DNA. Some other factors also cause cancer, which is hard to track [6].

1.5 Risk factors for cancer

Some of the common risk factors for cancer are Dr Ananya Mandal, according to the World Health Organization (WHO).

- ✓ Tobacco
- ✓ Alcohol
- ✓ Chronic hepatitis B virus infections, hepatitis B virus (HBV), Hepatitis C virus and some types of HPV
- ✓ Increase weight and obesity.
- ✓ In physical inactivity
- ✓ Factors dietary factors, including inadequate intake of fruits and vegetables
- ✓ Environmental risks and occupational hazards, including ionizing and non-ionizing radiation [7].

1.6 Cancer-causing agents of Chemical Carcinogens

The destruction of DNA is responsible for many environmental and chemical toxins. These substances are called carcinogenic substances, which are responsible for DNA damage or mutagenic mutations and mutations leading to cancer.
Some substances are directly linked to a particular type of cancer, such as smoking, which is usually associated with lung cancer and causes 90 %. Tobacco smoking is also responsible for the head, neck, and bladder and kidney cancer [7].

1.7 Ionizing radiations

Melanoma and other malignancies in the skin may be caused by long exposure to the sun's ultraviolet (UV) radiation or radon. Treating the cancer patient with this radiation can cause another type of cancer. Radiotherapy for a chest patient, for example, can lead to breast cancer [8].

1.7 Viral and bacterial infections

Virus or bacterial infections can also cause various types of cancer, such as the liver. Hepatitis B and C infections can cause cancer, and Barr virus can cause stomach or stomach cancer, and HPV infection can cause cervical cancer [8].

1.8 Genetic or inherited cancers

Ovarian cancer and breast cancer can be an example of BRCA 1 and 2 genetic cancers. Malignant tumors, such as testicular cancer and leukaemia, are known to develop Down syndrome [9].

1.9 Hormonal changes

Changes in the level of female hormones can become a cause of cancer, as excess estrogen can cause uterine cancer [9].

1.10 Immune system dysfunction

HIV infection can cause Kaposi sarcoma, non-Hodgkin lymphoma and malignant HPV-related tumors that lead to anal and cervical cancer [9].

2. Application Concept

The concept of application for data mining tools are given below:

2.1 Data Mining

Data extraction is often referred to as knowledge discovery as it is useful to find hidden patterns and raw data basic information. Domain experts are often data mining users, so it is currently easy to collect and manipulate data. Users of data mining often have expertise in areas where data is not only available but also collected. Human analysts can identify and distinguish important patterns in small data sets and can then be applied to large sets of data using specialized data extraction tools. You can also simultaneously analyze relationships between various dimensions [10].
The extraction of data is also known as the discovery of knowledge. Knowledge Discovery Process (KDP) is a non-trivial process in which desirable, useful and valid data patterns are identified. The process of knowledge discovery includes how to manage and retrieve the data? How can efficient and related algorithms be implemented? How to translate and use the results and how to make it easy for people and machines to understand [11].

2.2 Decision Trees Algorithm

The algorithm of the decision tree is one of the most widely used algorithms to overcome accuracy and predictability. Explore and identify the interface between similar properties in a very simple way and apply the rule "if...then. Else." These decision rules are easy and comprehensible and apply on the basis of the sum of the calculated points of the various components. For example, say: "If it was more than two months, you received an asset without any other questions, but now they are asking me for additional values. Such a scenario could force the applicant to make a false statement or find another financial services option [12].

2.3 Data Mining with Rattle and R

The rattle is an abbreviation for the ‘R’, R tool for easy-to-learn analysis. Statistical language R is the famous graphical data mining application. RATEL is not mandatory. R provides an easy review of data extraction projects using powerful and powerful data mining

language; facilitates migration so that code can be written with Rattle commands and can be easily debugged and implemented in the R console. Creating the Gnome database with support from many operating systems such as MS / Windows, GNU / Linux, and Macintosh OS / X.
The instinctive UI of RATL enables you to explore the fundamental information extraction steps. The code can be high-goals spared in R and utilized as content. This content can be downloaded from the R console. Rattle can address the issues of every client and give a mind-boggling condition to preparing and displaying. There are boundless thoughts on how things should be possible and a progressively proficient client can communicate legitimately with this incredible language [13].

2.4 Data Mining with Weka

Weka is a free open source Java program that was used by the researcher under the GNU General Public License. The license was the third program for the researcher. The environment of Waikato has grown to analyze Weka's knowledge of the remarkable need for a unified office that allows researchers to easily access the most advanced automated learning techniques [14].
Weika aims to provide a complete set of automated learning algorithms and data processing tools to researchers and professionals. Regression algorithms, classification, compilation, extraction of code and attribute selection are included in the work environment. Preliminary data are well examined by the display of data and many pre-treatment tools. Weka contains many graphical user interfaces that make basic functions easy to access. The main user interfaces in Explorer. Contains a panel-based interface compatible with different data mining tasks for different panels. Data can be loaded and converted using pre-processing tools for so-called WIKA data filters in the first panel, called the pre-processing panel. Data from several sources, including files, URLs and databases, can be downloaded. The ARFF format in Wefa and others such as CSV, LibSVM and C4.5 have supported file formats. Data can also be created using an artificial data source and manually modified using a data set editor [15].

3. Literature Review

This section identifies the best treatments for different cancers available. Patients with cancer are encouraged to conduct clinical trials before developing a treatment plan. The clinical trial enables a patient to evaluate the best and safest approach to treatment. These methods are further tested to identify new drugs and different combinations of new treatments or doses [16].
Lung cancer can be treated in four ways: Surgery, radiation treatment, chemotherapy, and guided treatment. The choice of treatment depends on the type, stage, potential side effects and condition of the patient [17]. The cancer treatment plan may include symptoms and treatment for side effects. The patient should be aware of the choice of treatment available and the doctor's procedure. It also raises unclear options questions [18].
Four treatment methods were described in detail by the researcher in the next step. The doctor who uses surgery is defined as a surgical option for surgical oncology for cancer patients. For example, the thoracic surgeon performs lung cancer surgery in the case of lung cancer. This surgeon removes the specialist in lung tumors and all lymph nodes in the chest. A peripheral margin of normal lung tissue must remove this tumor. This means that cancer should not occur in the healthy part or tissue around the tumor after the surgery. For lung cancer, various types of surgery can be used, some of which are [18]:
This is the most effective process in which effective lobes are removed from the lung, even if the tumor is too small in the current five lobes. In this process, the tumor is removed only with its surrounding lung margin. Only the affected part of the lobe is eliminated in this surgery. The whole lung is removed in this type of surgery. This usually happens when the tumor is near the chest [19].
By inserting the needle into the tumor and saving the voltage, the tumor is destroyed in this process. This process is used when tumor removal is difficult. After surgery, recovery time depends on the type of surgery and the patient's fitness level prior to surgery. Co-treatment is a treatment given to an operating patient to avoid the risk of lung cancer recurrence. Radiation, chemotherapy and guided therapy, which are also discussed in the next section, may be included. This treatment allows any cancer cells in the body to be eliminated in order to reduce the possibility of recurrence. Many tools are available to help your doctor determine the patient's best treatment technique. The patient can discuss the physical condition and benefits of postoperative treatment with the phase with the doctor as the Phase I, Phase II and Phase III.
To destroy cancer cells using high - energy x - rays or other radiation therapy practices. An oncologist will better inform the patient about the treatment's effectiveness or damage. The most common radiation therapy is known to be given by an external radiation therapy device. The length of such treatment can increase from days to weeks. Another type of treatment is the treatment with implants for internal radiation. Radiotherapy for cancer that affects a large area of the body is not recommended since radiation therapy alone has destroyed radiation cells. This radiation will kill your healthy cell. CT scans are used to detect the exact location of the affected part of the body and to reduce the deformation of healthy cells in the body by using radiation to manage the small tumor. Radiation therapy (IMRT) is called this method. For all patients, such treatment is not recommended [20].

Radiotherapy has many side effects. Lung cancer patients often experience fatigue and loss of appetite. The patient also causes irritation to the skin, such as sunburn. Radiation therapy can cause a problem in cough, fever or asthma after radiation for some time, so radiation tumors accurately determine the locations of normal lung cells exposed to radiation through CT scan [20].

Drugs are thus used to destroy cancer cells. These drug experiences stop these affected cells from growing and proliferating. A doctor specializing in the treatment of cancer with drugs manages this treatment. Chemotherapy at any stage improves the life of a lung cancer patient. Doses are administered through the bloodstream in systemic chemotherapy. Another way to reach the cancer cell is through intravenous injection with an intravenous needle. This intravenous tube injects pills or capsules into the body. Up to a specified number of cycles and times, one or more drugs can be injected into the body of the patient. Advanced chemotherapy methods are more effective and have fewer side effects. Some of the side effects that can be attacked after chemotherapy are fatigue, nausea, vomiting, and risk of infection, hair loss and anorexia. Chemotherapy can also affect blood cells, skin cells, and nerve cells. To help relieve many of these side effects, full medical care is required. Drugs used to treat cancer are always studied and before taking any medication, the patient should consult his doctor [21].

This type of treatment targets specific genes, tissues or bacteria that contribute to cancer and prevent healthy cells from growing. A cancer patient must carry out several tests before treatment to detect microorganisms, proteins and other tumor organs. Not all tumors are of the same kind. In the performance test, most abnormal proteins are found and used to detect these proteins so that each patient can be compared to the best possible treatment. Many types of research continue to discover more specific molecular objectives and new therapies. For lung cancer, especially in clinical studies, the following types of treatment are recommended.

This technique focuses on creating new blood vessels and stopping angiogenesis in the body. Bevacizumab is a dose of anti-angiogenesis administered to patients with lung cancer to starve the tumor. For patients with squamous cell carcinoma, this technique is most common [21].

Medicines that block EGFR are effective in slowing lung cancer growth. EGFR is one of the most common medicines. This is more effective than chemotherapy. This drug is a pill and can be taken from the mouth. Iressa is another drug that blocks EGFR effectively.

There may be several types of treatment for patients with lung cancer. This is called a multidisciplinary approach. For example, before or after radiation therapy, chemotherapy may be given. The patient-doctor understanding is very important for successful treatment. The patient must know a doctor's coordinated care and treatment plan. The patient should talk to the doctor and also ask for an opinion on the other options.

Treatment with cancer often has side effects. The focus of caring for cancer patients is to alleviate the symptoms of the person and palliative or supportive care is called the side effect. These include a person's physical, social and emotional care [22].

This palliative care method can benefit the patient at any stage. In order to avoid parallel side effects, cancer treatments and treatments can be avoided. Can patients receiving treatment improve their life and quality? There are often dietary changes in painkillers, In addition to other treatments, medicines, and relaxation. The treatment of chemotherapy, surgery, and radiation was also milder.

Some common treatments for lung cancer symptoms are:

- ✓ The tumor can be damaged near the chest, which causes bleeding, or obstruction of the lungs through radiation therapy.
- ✓ During bronchoscopy, breathing can be enhanced by opening the cancer block.
- ✓ By laser, the surgeon can burn up the tumor.
- ✓ To relieve cancer pain, many medications can be used. Hospitals have a specialist doctor who deals only with pain control through the use of morphine and other medications.
- ✓ Drugs can be used to contain coughs and reduce bronchial discharges.
- ✓ Artificial oxygen can be provided to oxygen-free lungs from the air.
- ✓ To prevent osteoporosis in the future, many medications can be taken [22].

4. Treatment by Stage

The treatment of lung cancer and their stages to analysis are given below:

4. 1 Stage I and II

Patients can be treated in the first and second stages of surgery. Before and after the procedure, the patient should consult the oncologist. A patient with a large tumor can be treated before surgery with chemotherapy, usually using inductive chemotherapy to reduce the chance of cancer returning. In cases where chemotherapy is not recommended, radiation therapy can be used as a treatment method [23].

4.2 Stage III

A serious patient is at great risk in the third stage. To treat this patient, surgery and radiotherapy are not enough. Patients in the third stage are also more likely to develop

cancer, so the doctor usually recommends chemotherapy before surgery [24].
Chemotherapy reduces cancer size and ensures that surgery continues to be effective. In the third stage, some patients must combine chemotherapy and radiotherapy instead of surgery at the same time. This improves cancer reduction and reduces the likelihood of recurrence.
Chemotherapy has been shown to be more effective than treatment at different times at the same time, but it causes many side effects. All changes can also be performed by a patient who has received chemotherapy and radiation.
Doctors are still confused about whether a patient needs surgery after successful radiation therapy and whether radiation therapy after successful chemotherapy is useful. The cancer of the patient's third stage cannot, for the most part, be surgically removed because the surgeon usually feels that working is a major risk. A combination of chemotherapy and radiation therapy may still be used for such a patient [24].

4.3 Stage IV

In general, the patient does not receive surgery or radiation therapy in stage IV. Doctors are operated at this stage from time to time if cancer does not spread to several parts of the body. Since patients at this stage have a very high risk of cancer spread, the doctor only recommends chemotherapy for these patients [25].
Chemotherapy treats prevent the spread of cancer and extend the patient's life. Chemotherapy will cause cancer to disappear, but doctors know from experience that cancer will return. If cancer causes many side effects, treatment can be stopped [25].

4.4 Treatment of small cell lung cancer

Because it spreads rapidly and depends on the stage, chemotherapy is a major treatment for small cell lung cancer. Ethoposide (Toposar, VePesid, Etopophos), cisplatin (Platinol) or carboplatin (Paplatin) are the most common chemotherapy. Twice a day, patients with small-stage lung cancer, radiation, and chest chemotherapy are recommended. Occasionally, surgery is recommended for patients with very early disease [25].

4.5 Metastatic lung cancer

Cancer has spread to various parts of the body called metastatic cancer. There may be a difference of opinion on the best treatment plan. If cancer is spread to the brain, the best choice is not radiation or surgery. Side effects such as hair loss, fatigue, and scalp redness can occur. The diagnosis of metastatic cancer has been very stressful and painful for some time. Patients must talk to doctors and their families [26].

5. Cancer Cell Migrate from Breast to Lungs

Here are some condition of cancer cell migrate from breast to lungs and their risk factors are also given below:

5.1 Causes and risk factors

It's very hard to know why one woman had breast cancer and another didn't? Some risk factors may affect breast cancer probability:

5.2 Age

Increasing age has increased breast cancer change. Most cases are found in women over the age of 50 because menopause has been the most common condition in age 50.

5.3 Family history

If a close relative with breast cancer or ovarian cancer is present, the high risk of breast cancer is taken into account once. Because breast cancer is most common in women and there can be more than one family member [27].

5.4 Previous diagnosis of breast cancer

If you have breast cancer, it is likely that you will develop it again. It can be in the same breast or in the same breast [27].

5.5 Previous benign breast lump

Breast cancer does not always mean that you have breast cancer, but it can be before you. It can cause changes in breast tissue and increase breast cancer risk [28].

5.6 Breast density

Women with high breast density have greater breast cancer changes because more cancer tissue is available [29-31].

5.7 Exposure to oestrogen

Oestrogen may increase the chance of breast cancer in some women's hormones. The amount of oestrogen exposed to the body may increase the likelihood of breast cancer. Chances of breast cancer may also be increased by not having children or pregnancy later in life [29].

6. Treatment of Breast Cancer

The treatment of breast cancer and their conditions are given below:

6.1 Radiation

It is commonly assumed that radiation use, such as Xrays and CT scans, increases the chance of breast cancer slightly [30].

6.2 Hormone replacement therapy (HRT)

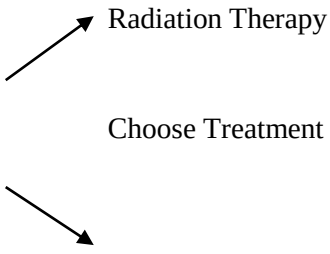
Hormone replacement therapy (HRT) known as high-risk therapy often used in breast cancer. If the patient takes HRT combination, this risk will be much higher. In every 1,000 women who take combined HRT for 10 years, 19 additional cases of breast cancer will be recorded [30]. Among the highest morbidity and mortality rates among women, breast cancer and lung cancer are among the most common cancers. Long-term survival means that patients with BC may develop second primary cancers after breast cancer, Lung cancer is the leading cause of cancer death. Therefore, understanding the risk of second primary lung cancer in BC patients is very important. One of the things that make cancer cells different from normal cells is that they can move about more easily. It seems likely that one of the ways that cancers spread through nearby tissues is by the cells directly moving. Scientists have discovered a substance made by breast cancer cells that stimulate them to move. It seems likely that this substance plays a big part in the local spread of cancers from breast to lung cancer.

7. Challenges &Solution

It has become very difficult to draw any informative knowledge from the available data for patients with lung cancer with the growing number of patients with lung cancer. As result, people cannot make an informative decision on the primary causes and relationships of the different factors of lung cancer. Some informative knowledge is very important to help doctors and patients carry out their research, diagnosis and appropriate treatment much more easily so that data mining can help in this regard. For the above challenges, the author predicts an appropriate treatment method using data mining techniques to survive longer for patients with lung cancer.

7.1 Treatment Process

Lung Cancer Patient →Diagnosis of Cancer Stages



Predict the best treatment of Survival longer period of time for a lung cancer patient

7.2 Variable

1. Independent Variable

- ✓ Age
- ✓ Gender
- ✓ Cholesterol
- ✓ Weight
- ✓ Smoke habit
- ✓ Previous Radiation Therapy
- ✓ Blood Group
- ✓ Family Background
- ✓ HIV

2. Dependent Variable

- ✓ Treatment (Radiation and Chemo Therapy)

7.3 Data Collection

Data from government and private hospitals in Karachi, Pakistan were collected from this study. The researcher used data from 500 patients with lung cancer (men and women) in this study. The oncologist will check the data in the respective hospitals.

8. Methodology

The framework of data mining with rattle R are given below:

8.1 Data Mining with Rattle and R

The rattle is an abbreviation for "R Tool for Easy Analysis" and uses the famous graphical data mining application for statistical language R. The R experience is not obligatory to use Rattle. R provides an easy overhaul of data mining projects with a powerful and powerful data mining language; it also makes it easy to write code using Rattle commands and easy to debug and deploy in the R console. Set the graphical user interface rule (Gnome) with support from different operating systems such as Windows.
The intuitive user interface of Rattle enables you to naviga te the basic data mining steps. The R code can be saved an d used as a high resolution script.This script can be loaded into the R consoleRattle to meet all user needs and provide a sophisticated processing and modelling environment individually. There are unlimited ideas on how to do things

and a more professional user can directly interact with this powerful language.

8.2 Rattle Interface and Process

Rattle's layout is designed to flow the data mining process from side to side, progressing from left to right from side to side.
These are the following processes in Rattle:

Table 1: Description of Rattle R

STEP	DESCRIPTION	UTILITY	ACTION
1	Load a Dataset	Data	CSV file
2	Select variables and Explore the data	Explore	Sum & Distribution
3	Transform the data into training & test datasets	Transform	Re-Scale
4	Build your Models	Model	Tree/Forest/Linear
5	Evaluate the models	Evolution	Tree/Forest/Linear
6	Review the Log of the data mining process	Log	Log (Export Comment)

9. Analysis and Results

The researcher takes the following steps to describe the use of the rattle package: Load the data file; divide the instances into learning samples and tests; specify the variables types (target or input); Computation of descriptive statistics; learning from the learning sample predictive models; evaluation of test sample models (confusion matrix, error rate, certain curves).

STEP 1: Dataset
The researcher uses the data file "Lung Cancer Dataset." The researcher wishes to explain the characteristics of DISEASE in patients. The researcher shows the first instances of the dataset.

Algorithm.1
Loading the rattle package:
Loads the rattle package [library ()]

APPLY GUI with the command rattle ()

> #loading the package

>Maintain library (rattle)

> #launching GUI

>Rattle ()

From now on, the researcher will perform all operations by clicking on the relevant menu or button. All these operations are recorded as R commands by rattle. It shows the rattle of the GUI. Rattle usage is always the same. By working on the appropriate tab (data: loading the data set; exploring: some Meta statistics; testing: some statistical tests, etc.), the researcher determines the command. Importing the data file: The author clicks on the filename button in the "Data" tab. The investigator selects the data file "lung for rattle.txt."
The researcher specifies the split column: "Separator = \t." Then click on the author to execute. It loads the dataset. The variable type is detected automatically in each column (discrete or continuous) from the different values. The target attribute and the input can be defined by the researcher. Finally, the researcher specifies the size of the training samples (70% of cases, randomly drawn) and the test (30%).

STEP 2: Dataset description
Into the Explore tab: The researcher obtains some descriptive statistical indicators for the variables (option Summary / Summary). Rattle lists the values (levels) for the discrete variables. The researcher has the quartiles of min, max, mean. The learning sample calculates all indicators. The researcher will obtain a more detailed description of the Summary / describe option. Among other things, indications are useful for detecting unusual values (outliers) for continuous variables. The researcher still gets some graphical representations of the distributions in the Explore tab with the distributions option. For example, the researcher has age and Cholesterol conditional box plots according to disease values.

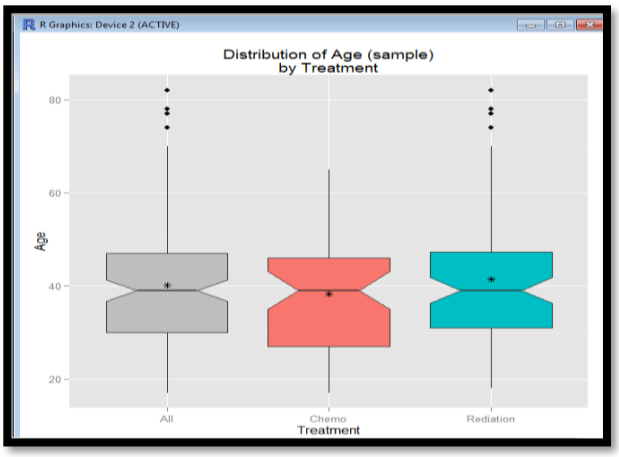


Fig. 2 Distribution of age by the treatment of lung cancer

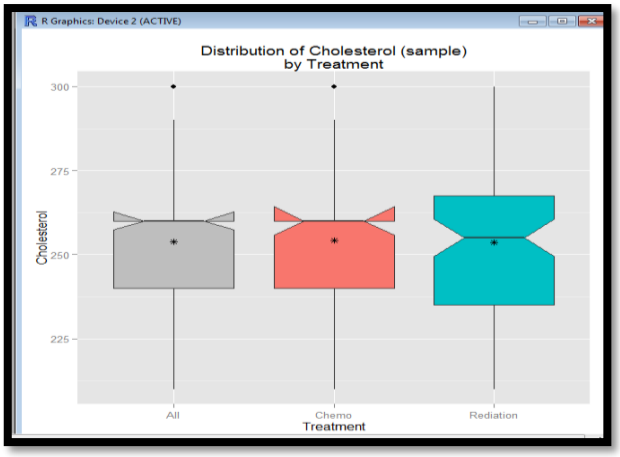


Fig. 3 Distribution of cholesterol by the treatment of lung cancer

The researcher can also have functions of the conditional distribution. Depending on the values of the destination attribute, the researcher can obtain "Mosaic" variables on the discrete variables. For example, men (males) are more numerous in a sample about sex than women (females). Men have a higher proportion of the disease. The researcher can also obtain links to persistent input attributes. Links in a hierarchical structure described. For example, it is useful to detect iterative variables.

STEP 3: Data transformation
The Conversion tab has been customized to convert the variable. Some common triggers (eg., logarithm, rank, etc.) are available.

STEP 4: Supervised learning
Supervised machine learning (SML) is the search for algorithms to generate general hypotheses based on externally presented cases, which then predicts future situations. One of the tasks that smart systems frequently perform is supervised classification. This author describes several techniques used by the Waikato Knowledge Analysis (WEKA) machine-learning tool for machine learning modified (ML) techniques. A lung cancer data collection kit was used for the implementation of the algorithms with two attributes as an independent variable and one as an analytical dependent variable. Research shows that a key factor in this investigation is the time needed to build a model and precision. ML algorithms, therefore, require accuracy, accuracy and minimum error to monitor predictive machine learning. This step is at the heart of our analysis. The Finder selects the Model tab. The researcher wants to evaluate three.
Methods: Demonstration trees demonstration, Informal forests, and Logistic reducers. The decision tree, rattle, uses the r-part command of the r-part package. The researcher takes note of the default parameters used. Press the port button. We get the rules associated with the tree by clicking the rules button.

STEP 5: Measuring the generalization performance
The researcher wants to evaluate the classification performance of the test sample (30 percent of the entire dataset). First, we want the confusion matrix and the error rate associated with it. We choose the option "Error Matrix." For the "Data" item, the "Test" option must be selected. Only models learned in the tab "Model" can be found here.

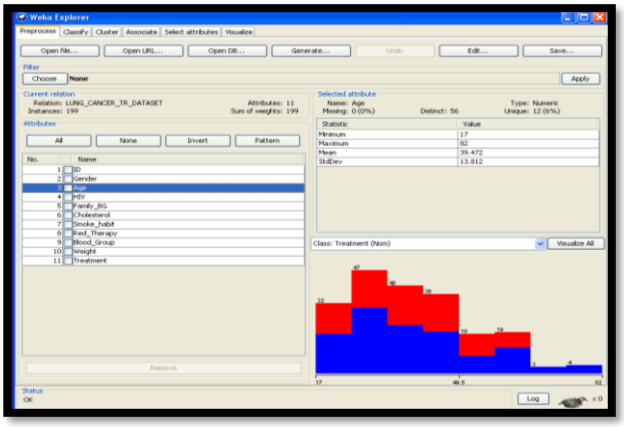


Fig. 4 WEKA Explorer

In fact, the error rate is not a good standard here. Note that the differences between the methods are based only on the bad classification status, and the decision tree is definitely worse compared to the other two compilers, which are similar in terms of performance. Not surprising. The researcher knows that the decision tree is not well prepared for the registration process.

9.1 Data Mining with Weka

WEKA implies information learning in the Waikato condition created by Waikato University, New Zealand. WEKA bolsters numerous information mining errands, including information re-preparing, arranging, and total, relapse and occupation choice, to give some examples. The workflow at WEKA will be as follows: →Data Pre-processing →Data Mining →Knowledge

9.2 Weka Interface and Process

STEP 1: Dataset
The first interface that appears looks like the one given below.

1. *Explorer:* An environment for data exploration, Supports raw data processing, character selection, learning and visualization
2. *Experiment:* An experimental and statistical testing environment between machine learning algorithms.
3 *Knowledge Flow*: Explorer is similar but has a drag-and-drop interface and gives the KDD process a visual design.
4. *Simple CLI:* Provides a simple interface to execute WEKA commands on the command line.
5. Explorer on the first screens of the interface and load a library dataset. The data set ' lung cancer 'tr' dataset.csv ' is mention in this research

STEP 2: Dataset description
Click on each attribute to view the distribution of each sa mple. You can also view all of them simultaneously by cli cking on the right pane on the Visualize All.

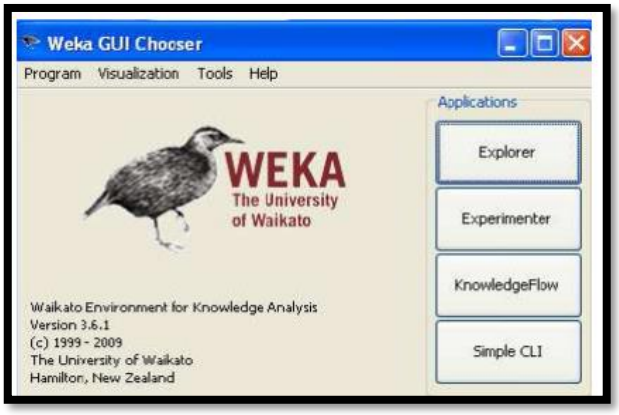


Fig. 5 WEKA Explorer by tools

The right pane shows the results for training and testing. It also indicates the number of correctly classified and misclassified samples.

STEP 3: Measuring the generalization performance
The right pane demonstrates the outcomes for preparing and testing. It likewise shows the quantity of effectively grouped and misclassified samples. Once the model has been prepared and tested, the specialist needs to quantify the exhibition of the model. For this reason, the scientist utilized three measures to be specific: precision, recall and accuracy in this research.
Algorithm.2

Precision (P) = tp/(tp+fp)

RECALL (R) = tp/(tp+fn)

Accuracy (A) = (tp+tn)/Total # samples

Where tp, fp, tn and fn are true positive, false positive, true negative and false negative respectively.

Table 2: Comparison between the accuracy of Weka and Rattle R

	WEKA	**RATTLE- R**
Efficiency (Sec)	0.5	0.2
Accuracy (%)	60%	56%

Table 3: Data sets record of lung cancer patient surgery

Whole Population			
Total STAGE 1	**3839**		
Groups	**Number (n)**	**Percentage (%)**	**Definition**
Group 1: Surgery			
Surgery	301	8.00	Surgery alone
Surgery and Adjuvant Chemotherapy	349	9.00	Surgery then chemo.
Neoadjuvant chemotherapy and surgery	34	0.90	Chemotherapy then surgery
Surgery and Radiotherapy	38	0.98	Surgery and any radiotherapy.
Triple therapy	48	1.20	Chemotherapy/radiotherapy and surgery (any order)
Group 2: Radical Radiotherapy			
Radical Radiotherapy	295	7.60	(radRT and no chemo)
Radical Radiotherapy and Chemotherapy	429	11.00	(Chemo any gap then rad RT).
Group 3: Palliative intent treatment			
Palliative radiotherapy and chemotherapy	190	4.90	
Palliative radiotherapy	397	10.30	
Chemotherapy alone	428	11.10	
Group 4: Best Supportive Care (No treatment)	1330	34.60	
Total STAGE 2	2449		
Groups	**Number (n)**	**Percentage (%)**	**Definition**
Group 1: Surgery			
Surgery	19	0.77	Surgery alone
Surgery and Adjuvant Chemotherapy	10	0.40	Surgery then chemo.
Neoadjuvant chemotherapy and surgery	2	0.08	Chemotherapy then surgery
Surgery and Radiotherapy	8	0.32	Surgery and any radiotherapy.

Triple therapy	4	0.16	Chemotherapy/radiotherapy and surgery (any order)
Group 2: Radical Radiotherapy			
Radical Radiotherapy	81	3.30	(radRT and no chemo)
Radical Radiotherapy and Chemotherapy	265	10.80	(Chemo any gap then rad RT).
Group 3: Palliative intent treatment			
Palliative radiotherapy and chemotherapy	242	9.80	
Palliative radiotherapy	321	13.10	
Chemotherapy alone	551	22.40	
Group 4: Best Supportive Care (No treatment)	946	38.60	
Total STAGE 3	**2610**		
Groups	**Number (n)**	**Percentage (%)**	**Definition**
Group 1: Surgery			
Surgery	260	9.96%	Surgery alone
Surgery and Adjuvant Chemotherapy	314	12.03%	Surgery then chemo.
Neoadjuvant chemotherapy and surgery	34	1.30%	Chemotherapy then surgery
Surgery and Radiotherapy	34	1.30%	Surgery and any radiotherapy.
Triple therapy	42	1.61%	Chemotherapy/radiotherapy and surgery (any order)
Group 2: Radical Radiotherapy			
Radical Radiotherapy	254	9.73%	(radRT and no chemo)
Radical Radiotherapy and Chemotherapy	384	14.71%	(Chemo any gap then rad RT).
Group 3: Palliative intent treatment			
Palliative radiotherapy and chemotherapy	169	6.48%	
Palliative radiotherapy	274	10.50%	
Chemotherapy alone	379	14.52%	
Group 4: Best Supportive Care (No treatment)	466	17.85%	
Total STAGE 4	**1684**		
Groups	**Number (n)**	**Percentage (%)**	**Definition**
Group 1: Surgery			
Surgery	15	0.89%	Surgery alone
Surgery and Adjuvant Chemotherapy	9	0.53%	Surgery then chemo.
Neoadjuvant chemotherapy and surgery	2	0.12%	Chemotherapy then surgery
Surgery and Radiotherapy	8	0.48%	Surgery and any radiotherapy.
Triple therapy	4	0.24%	Chemotherapy/radiotherapy and surgery (any order)
Group 2: Radical Radiotherapy			
Radical Radiotherapy	73	4.33%	(radRT and no chemo)
Radical Radiotherapy and Chemotherapy	238	14.13%	(Chemo any gap then rad RT).
Group 3: Palliative intent treatment			
Palliative radiotherapy and chemotherapy	219	13.00%	
Palliative radiotherapy	238	14.13%	
Chemotherapy alone	497	29.51%	
Group 4: Best Supportive Care (No treatment)	381	22.62%	

The above table.3 shows the overall record of lung cancer for surgery patients. These records show the all history of LC stages from 1 to 4 and the ratio of patient survival for the duration of months from one month to an around a complete year. The above table shows that the four groups for each stage of LC and also mention the performance ratio of each group.

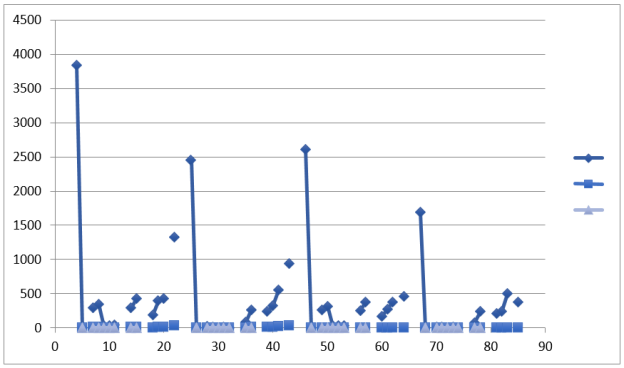


Fig. 6 Graph of data sets record of lung cancer patient surgery

As the above figure.6 shows that the ratio of lung cancer for surgery patients, These ratio is based on all stages of lung cancer from I to IV for the patient survival for the duration of months from one month to an around a complete year. The above graph shows that the four groups for each stage of LC and also mention their performance of each group

Table 4: Accuracy results of lung cancer survival patients for stage 1-4

Whole Population			
SNO.	**Number of LC patients**	**One year survival**	%
Overall	26901	10042	37.3 %
Stage 1	5144	4205	81.7 %
Stage 2	2028	1300	64.1 %
Stage 3	5284	2248	42.5 %
Stage 4	13193	2043	15.5 %
Stage Missing	1252	246	19.6 %

The above table.4 shows the overall calculation of lung cancer patient for the duration of survival from one month to year as well. This study shows the all stages condition of lung cancer with the accuracy of treatment criteria, which is shown in table.3.

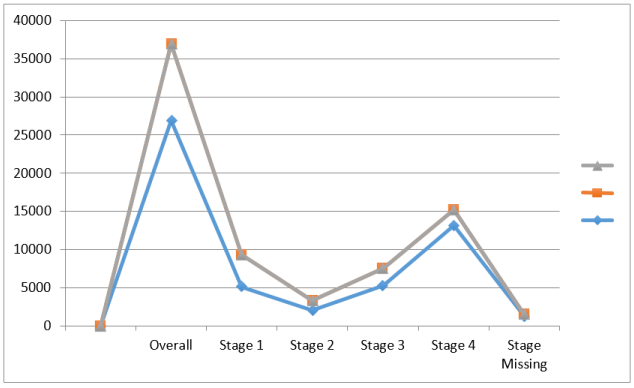


Fig. 7 Graph of Accuracy results of lung cancer survival patients for stage 1-4

As the above figure.7 shows that the duration of the survival ratio of lung cancer from stage 1 to stage 4 (month to a year). This study shows the all stages and condition of lung cancer with the accuracy of criteria are:

Stage 0: At this stage, cancer is also known as in situ cancer. Cancer is small in size and has not spread to, or outside, deeper lung tissue.

Phase I (Stage 1): There may be cancer in the underlying lung tissue, but the lymph nodes are not affected.

Phase II (Stage 2): Cancer can spread to nearby lymph nodes or walls of the chest

Phase III (Stage 3): Cancer continues to spread from the lungs to the lymph nodes or nearby structures and organs such as the heart, trachea and oesophagus.

The most developed type of disease is the fourth phase of non-small cell lung cancer (NSCLC). In the fourth stage, malignant growth spreads or spreads to different pieces of the body behind the lungs. Roughly 40% of NSCLC patients are determined to have lung malignant growth in stage IV. For individuals with stage IV lung malignancy, the five-year survival rate is under 10%.

Table 5: Experimental results of lung cancer surgery patients

SNO.	No. of Features	Model Type	Accuracy	ROC	Sensitivity	Specificity
1	Decision Tree	Preset: Fine Max number of splits: 100 Splits criterion: Diversity Index Surrogate decision splits: off	100%	1.00	0.01	0.01
2	Decision Fine-Tree	Preset: Fine Tee Max number of splits: 100 Splits criterion: Diversity Index Surrogate decision splits: off	95%	0.95	0.095	0.009
3	Decision Medium-Tree	Preset: Medium Tree Max number of splits: 20 Splits criterion: Diversity Index Surrogate decision splits: Off	83%	0.83	0.085	0.006
4	Coarse Tree	Preset: Coarse Tree	83%	0.83	0.093	0.006

		Max number of splits: 4 Splits criterion: Diversity Index Surrogate decision splits: Off				
5	SVM Linear	Preset: Linear SVM Kernel Function: Linear Kernel Sale: Automatic Box Constraint Level: 01 Multi-Class Method: One-to-One Standard Data: True	50%	0.50	0.005	0%
6	SVM Coarse	Preset: Coarse SVM Kernel Function: Coarse Kernel Sale: Automatic Box Constraint Level: 01 Multi-Class Method: One-to-One Standard Data: True	100%	1.00	0.01	0.01
7	SVM Quadratic	Preset: Quadratic SVM Kernel Function: Quadratic Kernel Sale: Automatic Box Constraint Level: 01 Multi-Class Method: One-to-One Standard Data: True	50%	0.50	0.005	0%
8	SVM Cubic	Preset: Cubic SVM Kernel Function: Cubic Kernel Sale: Automatic Box Constraint Level: 01 Multi-Class Method: One-to-One Standard Data: True	50%	0.50	0.005	0%
9	SVM Fine Gaussian	Preset: Gaussian SVM Kernel Function: Gaussian Kernel Sale: 0.79 Box Constraint Level: 01 Multi-Class Method: One-to-One Standard Data: True	100%	1.00	92%	60%
10	SVM Medium Gaussian	Preset: Medium Gaussian SVM Kernel Function: Gaussian Kernel Sale: 3.2 Box Constraint Level: 01 Multi-Class Method: One-to-One Standard Data: True	100%	1.00	0.01	0.01
11	SVM Coarse Gaussian	Preset: Coarse Gaussian SVM Kernel Function: Gaussian Kernel Sale: 13 Box Constraint Level: 01 Multi-Class Method: One-to-One Standard Data: True	100%	1.00	0.01	0.01

Fig. 8 Graph of Experimental results of lung cancer surgery patients

The above figure.8 shows the number of features of Experimental results of lung cancer surgery patients and their performance as well.

10. Discussion

These experimental comparisons depend on supervised learning of the machine as the analyst works through Weka on lung cancer. The aftereffects of the investigations are led by a dataset "Malignant lung cancer with all stages" with 2 factors with 2 attributes. At that point, unique dataset with 324x34 dimensional information is arbitrarily separated into 75-80 percent proportion preparing and test sets as introduced in this research. The author has trained several supervised machine-learning models using the training dataset, namely Decision Tree SVM (Coarse), SVM (Coarse), SVM (Quadratic), SVM (Cubic), SVM (Fine Gaussian), SVM (Medium Gaussian) and SVM (Coarse Gaussian). The new data from the testing dataset is predicted once the models are trained. The performance of the studied methods will be evaluated by measuring four Evaluation Metrics (EM), namely sensitivity (sen), specificity (spe) ROC, and accuracy (Acc). Table 5 mentions the experimental results. Five models often machine-learning methods: Decision Tree, SVM (Coarse), SVM (Fine Gaussian), SVM (Medium Gaussian), SVM (Coarse Gaussian). The computational time of each classifier is also calculated to evaluate their complexity. The remaining models: SVM (Linear), SVM (Quadratic)

and SVM (Cubic), 50 percent accuracy are achieved. The remaining two models: Medium-Tree and Coarse Tree decision reached 83%. We also compared our results achieved with the results obtained in the literature through related works. The evaluation shows that the Decision Tree and SVM models are 100% more accurate than previous works in our study. This model has the biggest advantage of taking a shorter time and taking the value too far away. For previous work, our proposed system is better. These models are improved because of a shorter period of time; they deliver results quickly and take value for a long time. However, the researcher does not meet the 100% accuracy for all models but give better results for lung cancer DT and SVMM.

11. Conclusion

The algorithm of the decision tree is linked to lung cancer data in this research. By integrating patient data, the researcher can detect potential treatment for lung cancer. This study carried out a comparative study of a set of data between two Weka and Rattle R data extraction tools for classification using the algorithm of the decision tree. The researcher finds that both can generate a model for the tree in much less time after analyzing the results of both methods. However, the rattle is somehow faster than Weka, which can be due to Rattle R's internal structure, which is organized into memory columns. The data set used in this research is the same in both methods and then compare the results of both methods as well. The researcher can clearly see that Weka is better in terms of accuracy than Rattle R. In the future, the researcher can use this model to predict appropriate treatment methods in a larger set of patient data. The focus of this research is to find the target group of people requiring additional lung cancer screening tests to reduce prevalence and mortality. It is possible to improve the system for predicting and extending lung cancer. Other techniques to extract data, such as time series rules, division and pairing, can also be included. Instead of just class data, continuous data can also be used. Another area is the use of mining text to extract large quantities of unstructured data in databases on health care. The integration of data extraction and text extraction would be another challenge.

Acknowledgement

The authors would like to thank Universiti Teknologi Malaysia, Malaysia for Research in journal citation report in terms of ISI-Impact factor journal citation for providing support in the completion of this research work.

Reference

[1] Omar, Y., Tasleem, A., Pasquier, M. and Sagahyroon,A., 2018," Lung Cancer Prognosis System using Data Mining Techniques. In Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2018) - Volume 5, HEALTHINF, pages 361-368 ISBN: 978-989-758-281-3

[2] PeymanRezaei. TahaSamad-Soltani. 2017. Analyzing a Lung Cancer Patient Dataset with the Focus on Predicting Survival Rate One Year after Thoracic Surgery. Asian Pac Journal of Cancer Prevention, Vol 18(6):1531-1536.

[3] Jennifer S. Temel, M.D., Joseph A. Greer, Ph.D., AlonaMuzikansky, M.A., Emily R. Gallagher, R.N., SonalAdmane, M.B., B.S., M.P.H., Vicki A. Jackson, M.D., M.P.H., Constance M. Dahlin, A.P.N., Craig D. Blinderman, M.D., Juliet Jacobsen, M.D., William F. Pirl, M.D., M.P.H., J. Andrew Billings, M.D., and Thomas J. Lynch, M.D. 2010. Early Palliative Care for Patients with Metastatic Non–Small-Cell Lung Cancer. The New England journal of medicine, pp. 733-742

[4] American Cancer Society, 2017."Cancer Facts and Figures 2017". Atlanta, Ga: American Cancer Society. American Cancer Society, 2016. "Lung Cancer Prevention and Early Detection", Atlanta, Ga: American Cancer Society.

[5] Khaleej Times, 2016. "Lung Cancer Among Top 5 Fatal Diseases in UAE", Khaleej Times, UAE.

[6] [6] WAM, 2015. "Tawam Hospital Now Offers Lung Cancer Screening, a First for the UAE", WAM Emirates News.HAAD, 2016. "What is Lung Cancer?", Health Authority Abu Dhabi, UAE.

[7] HMW, 2013. "GenieMD Announces First Place Finish in ONC Blue Button Co-Design Challenge," Health & Medicine Week, p. 476.

[8] WebMD, 2017. "Datamonitor", WebMD Corporation.

[9] Mesko, B., 2015. "The Lungscreen App: Find out your risk," ScienceRoll. OFWW, 2011. "Lung Cancer; New Lung Cancer Findings from University of Toronto Described," Obesity, Fitness & Wellness Week, p. 846. Park, A., 2001. "Cancer spotter," Time, vol. 158, no. 7.

[10] Witten, I., Frank, E., 2016. Data Mining: Practical Machine Learning Tool and Techniques, Amsterdam: Morgan Kaufman.

[11] Rangra, K., Bansal, D., 2014. "Comparative Study of Data Mining Tools", International Journal of Advanced Research in Computer Science and Software Engineering, vol. 4, no. 6.

[12] Zhu, Z., Ong, Y., 2007. "Markov Blanket-Embedded Genetic Algorithm for Gene Selection", Pattern Recognition, vol. 49, no. 11, pp. 3236-3248.

[13] Venables, W., Ripley, B., 2012. Modern Applied Statistics with S, 4th Ed. New York: Springer.

[14] Cerfolio RJ, Bryant AS, Skylizard L, et al (2011). Initial consecutive experience of completely portal robotic pulmonary resection with 4 arms. J ThoracCardiovascSurg, 142, 740-6

[15] Guyer C (2016). Data Mining (SSAS) [Online]. Microsoft documentation: Microsoft. Available: https://docs.microsoft. Com/en-us/SQL/analysis-services/data-mining/data-miningssas 2017].

[16] Jaklitsch MT, Jacobson FL, Austin JH, et al (2012). The American Association for thoracic surgery guidelines for lung cancer screening using low-dose computed tomography scans for lung cancer survivors and other high-risk groups. J ThoracCardiovascSurg, 144, 33-8.

[17] Jin X, Han J (2011). Expectation maximization clustering.In 'Encyclopedia of Machine Learning', Eds Springer, pp 382-3.Koh HC, Tan G (2011). Data mining applications in healthcare. J HealthcManag, 19, 65.
[18] Mongan JJ, Ferris TG, Lee TH. Options for slowing the growth of health care costs. N Engl J Med 2008;358:1509-14.
[19] Weinstein MC, Skinner JA. Comparative effectiveness and health care spending — implications for reform. N Engl J Med 2010;362:460-5.
[20] Ferris FD, Bruera E, Cherny N, et al. Palliative cancer care a decade later: accomplishments, the need, next steps from the American Society of Clinical Oncology. J ClinOncol 2009;27:3052-8. 4. Levy MH, Back A, Benedetti C, et al. NCCN clinical practice guidelines in oncology: palliative care. J NatlComprCancNetw 2009;7:436-73
[21] Follwell M, Burman D, Le LW, et al. Phase II study of an outpatient palliative care intervention in patients with metastatic cancer. J ClinOncol 2009;27:206- 13.
[22] Ridge CA, McErleanAM, Ginsberg MS (2013). Epidemiology of Lung Cancer.SeminInterventRadiol, 30, 93-8.
[23] Samad-Soltani T, Ghanei M, Langarizadeh M (2015). Development of a fuzzy decision support system to determine the severity of obstructive pulmonary in chemically injured victims. Acta Inform Med, 23, 138.
[24] SamadSoltani T, Langarizadeh M, Zolnoori M (2015). Data mining and analysis: Reporting results for patients with asthma. PayavardSalamat, 9, 224-34.
[25] Satar R, Ali A, Abraha W, et al (2016). Estimating the economic burden of lung cancer in Iran. Asian Pac J Cancer Prev, 17, 4729.
[26] Shin B, Cole S, Park S-J, et al (2010). A new symptom-based questionnaire for predicting the presence of asthma. J InvestigAllergolClinImmunol, 20, 27-34.
[27] Shouman M, Turner T, Stocker R (2012). Applying k-nearest neighbour in diagnosing heart disease patients. Int J InfEducTechnol, 2, 220.
[28] Temel JS, Jackson VA, and Billings JA, et al. Phase II study: integrated palliative care in newly diagnosed advanced non-smallcell lung cancer patients. J ClinOncol 2007; 25:2377-82.
[29] Jemal A, Siegel R, Ward E, Hao Y, Xu J, Thun MJ. Cancer statistics, 2009. CA Cancer J Clin 2009; 59:225-49.
[30] Noor Zaman, Mudassir Ilyas, Muneer Ahmad, Fazdil Mohammad, Azween Abdullah. "An Experimental Research in Health Informatics for Designing an Enhanced Intelligent Could-Based Collaborative Multi-Modal Framework for Medical Imaging Diagnostics", in Journal of Medical Imaging and Health Informatics, 2017. Vol.7, issue, 6, pp. 1358-1364.
[31] Muneer Ahmad, Noor Zaman, Muhammad Al-Amin. "An experimental research in health informatics for enhancing ovarian cancer identification in ovarian imaging analysis using fuzzy histogram equalization", in Journal of Medical Imaging and Health Informatics, 2017. Vol.7, issue, 6, pp. 1385-1390.

Soobia Saeed is working as an Assistant Professor, Head of publication Department, and Coordinator of Seminars and Training at Institute of Business & Technology-IBT, Karachi, Pakistan. Currently, she is a Ph.D. Scholar in software engineering, from University Teknologi Malaysia-UTM, Malaysia She did MS in Software Engineering from Institute of Business & Technology- IBT, Karachi, Pakistan, and Masters in Computer Science fromInstitute of Business & Technology-IBT, Karachi, Pakistan and Bachelors in Mathematical Science from Federal Urdu University of Art, Science & Technology (FUUAST), and Karachi, Pakistan. She is a farmer research Analytic from University Teknologi Malaysia and supervises ICT & R and D funded Final Year Project (FYP).

Afnizanfaizal Abdullah is a senior lecturer at the School of Computing, with a PhD. in Computer Science, specializing in artificial intelligence techniques for analyzing biological data. My research interests are in the designing of machine learning algorithms for healthcare applications in the cloud environments. In 2015, I have co-founded Synthetic Biology Research Group to drive innovation in research and development of healthcare, biotechnology, and environment areas through computing and engineering. I am also active in engaging with industrial partners and professional communities to contribute the knowledge and skills for the public.

Noor Zaman has completed his PhD. in IT from University Technology Petronas (UTP) Malaysia. He has 18 years of teaching and administrative experience internationally. He has an intensive background of academic quality accreditation in higher education besides scientific research activities, he had worked for academic accreditation for more than a decade and earned ABET accreditation twice for three programs at College of computer sciences and IT, King Faisal University Saudi Arabia. He also worked for National Commission for Academic Accreditation and Assessment (NCAAA), Education Evaluation Commission Higher Education Sector (EECHES) formerly NCAAA Saudi Arabia, for institutional level accreditation. He also worked for National Computing Education Accreditation Council (NCEAC) Pakistan. He has experienced in teaching advanced era technological courses including, Mobile Programming (Android), Mobile Computing and .Net Framework programming besides other undergraduate and postgraduate courses, graduation projects and thesis supervision.

Noor Zaman has authored several research papers in ISI indexed and impact factor research journals\international conferences, edited 10 international reputed Computer Science area books, focused on research students, has many journal, IEEE conferences and book chapter publications to his credit. He has successfully completed more than 18 international funded research grants. He is Associate Editor, Regional Editor, Editorial board member, PC

member, reviewer, Keynote speaker for several reputed international journals and conferences around the globe. He also chaired international conference sessions and presented session talks internationally. He has strong analytical, problem solving, interpersonal and communication skills. His areas of interest include Wireless Sensor Network (WSN), Internet of Things IoT, Security, Mobile Application Development, Ad hoc Networks, Cloud Computing, Big Data, Mobile Computing, and Software Engineering.