International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 07 | July 2020 www.irjet.net p-ISSN: 2395-0072

AL/MI ALGORITHMS BASED EARLY STAGE CANCER DETECTION SYSTEM [AI/ML – ESCDS]

Pooja Hulmani[1], M Dakshayini[2]

[1] *Student M. Tech C.N.E, B M S College of Engineering, Bangalore, India*
[2] *Professor, Dept of I.S.E, B M S College of Engineering, Bangalore, India*

Abstract - *ML may be the part of Artificial Intelligence (AI) which is intensely utilized in the field of information science. It has solid potential in wellbeing related information examination for the prediction of automated diseases. The use of AI in the field of clinical finding is expanding steadily. This can be contributed basically to the improvement in the characterization and acknowledgment frameworks utilized in illness finding which can give information that guides clinical specialists in early identification of deadly maladies and consequently, increment the endurance pace of patient's altogether. Using EMR details the Lung Cancer is predicted in the early stage using machine learning algorithms Decision tree, K-Nearest Neighbor. This work is mainly focusing on the web interface which is used in the detection of the stages of lung cancer which can be treated later.*

Key Words: **MACHINE LEARNING, LUNG CANCER, EMR [ELECTRO MEDICAL RECORD], DETECTION, PATIENT, DOCTOR, DECISION TREE, K-NEAREST NEIGHBOR.**

1. INTRODUCTION

Cancer is considered the deadliest disease in the world. Cancer is usually unnoticed in the initial stages and is detected only in its later stages. The late detection of cancer is the main reason behind the highest number of deaths in the world. Lung cancer is caused when its most basic and important function of taking in of oxygen and eliminating the carbon-di-oxide out the body is not happening. When the tissues and cells of the lungs are uncontrollably proliferating,

Lung Cancer is bound to occur. When this process proliferation grows uncontrollably in its setting, there are more chances of it spreading to other tissues, causing damage.

The primary reason for death among men is the lung malignant growth and is also the second sort of disease among women. It is estimated that about 1.3 million individuals in the world die out of lung cancer every year. The statistics state that about around 30-40 thousand people in Turkey are detected with lung cancer every year. The growth of cancer in the lung does not provide any strong indications until the infection reaches the critical stage[1].

The very fact that the detection of its symptoms and its progression cannot be identified easily is the reason behind this cancer being extremely dangerous. It is a frightening fact that about a quarter of the patients are suffering from cancer without being asymptomatic. It is also known that a large number of people are not aware that lung cancer is brought upon by lung X-rays because of another disease. Early diagnosis of cancer is essential to cure lung disease as its growth can quickly spread to the other parts of the body such as bones, cerebrum, pancreas, etc. Nevertheless, the treatment for cancer has evolved and is becoming curable. Thanks to the advanced technology, many inventions are happening in the field of imaging techniques such as, Low Dose Spiral Compute Tomography, which will enable the diagnosis of cancer at the initial stage. Out of all the cancers, lung cancer is the most affected disease to mankind. It has caused 2.1 million deaths globally in the year 2018 alone.

The survival rate is five years for the patient suffering from this disease but can be improved if it is located during the nascent stage and best 15% of the patients are diagnosed early as the signs and indication will not appear until symptoms increase to a later or terminal state. According to the research results of the National Lung Screening Trial (NLST) and NELSON, states that to aid the application of the lung, most cancers detecting applications to recognize people with excessive chance of succumbing to this disease by utilizing the Low-Dose Computed Tomography (LDCT) imaging maximize the chance of its diagnose of cancer at the earliest.

1.1. OBEJCTIVES

The main objective of the project is:

- To develop a web interface for the prediction of the disease and it is user friendly.
- Helps to predict the early detection of the cancer.
- To identify various machine learning techniques in the case of analysis of Lung Cancer.
- Machine learning techniques are used in monitoring the stages of the cancer. Decision Tree and K-Nearest Neighbor algorithms are used to predict the cancer.
- Also helps in categorizing the disease into three stages such as low, medium and high.

Rest of this paper is organized as follows. Section 2 discusses the related work done in this direction. Section 3 explains how the existing system works and what its drawbacks are. System design of the proposed AI/ML-ESCDS is shown in section 4. Implementation and Results are discussed in section 5. Section 6 concludes the work.

2. RELATED WORK

Machine Learning makes AI programming a stride further as it empowers keen figuring out how to happen inside the part dependent on previous work is done or extrapolations produced using information. The product performs advanced dynamic procedures as it comes and gains from past activities. Some of the previous papers which are referred to during the detection of lung cancer using machine learning algorithms are discussed below:

Authors [1] discusses on what is cancer, how it is caused, how it can be prevented, how many deaths and cases are expected to occur in 2017, how the cancer is staged, what are the costs of cancer, types of cancers such as breast cancer, lung cancer, ovary cancer, pancreas cancer about their incidence trends, deaths, mortality trends, signs and symptoms, risk factors, early detection and treatment to that particular disease.

Authors in [2] manages the expectation of a post-usable future in lung malignancy patients utilizing prescient information mining algorithms to think about algorithms, most of the machine learning algorithms are decision tree, KNN, and CNN. A delineated 10-overlap cross-approval near examination to get the accurate result through these algorithms compared to the traditional data mining techniques.

According to Kwetishe Joro Danjuma, various outcomes are created for finding the stages of the disease called lung malignant using a large dataset. Most of the classifiers, for example, KNN, SVM, NN, and Logistic Regression were actualized, and comparing precision rates were acquired. [3]

[4] discusses different algorithms explained which include Hidden Markov Model, Naive Bayes, etc. Appropriate clarification is given about how and why different division algorithms are utilized in recognition of Lung Cancer.

[5] According to Elsevier B V, the list of the detection accuracy will identify the accurate result of the lung cancer and this system will give less accuracy and there is so much limitation in this system. It will predict whether they are having disease or not and it will not make the stages of the disease.

According to NLST research team, Screening alone is an unsuccessful effort to reduce mortality from Lung Cancer. The introduction of the Low-Dose Helical Computed Tomography has changed the Lung Cancer Screening scenario. The study indicates that Low-Dose CT can detect many tumors at initial stages of lung cancer. The NLST occurs to decide the conduction of the screening through Low-Dose Computed Tomography that can decrease the mortality from lung cancer. [6]

3. EXISTING SYSTEM

- It is basic to understand the exact finding of patients by clinical examination and evaluation. For convincing assurance choice, emotionally supportive networks that rely upon PC may accept a vital activity.
- The human services field makes gigantic data about clinical assessment, a report concerning understanding, fix, resulting meet-ups, medication, etc. It is unpredictable to coordinate suitably.
- The quality of the information affiliation has been affected because of the inappropriate administration of the data. Overhaul in the proportion of information needs some authentic method to focus and procedure data reasonably and productively.

3.1. DRAWBACKS

- It is difficult to predict the cancer in the early stage.
- There was no proper solution to detect the cancer in the early stage.

4. SYSTEM DESIGN

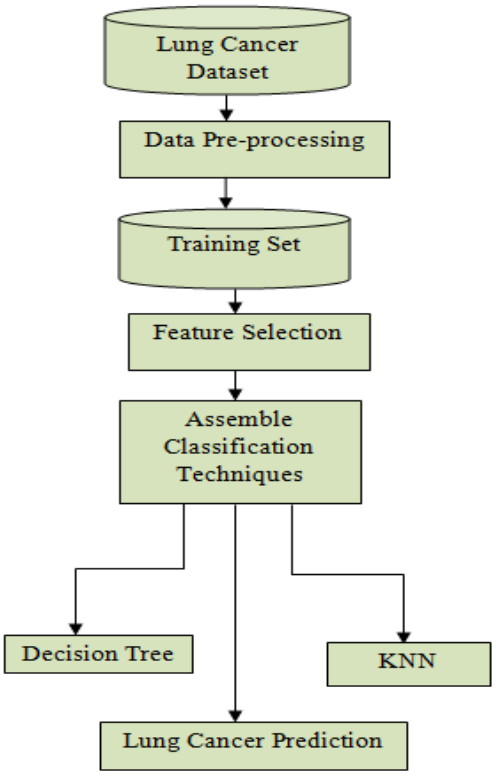

Fig 1: System Design

Step 1: Data assortment and data set planning. It contains the assortment of clinical data, artifacts from different sources like emergency clinics, release slips of patients, and from UCI repository.

Step 2: Once after collecting all the data from medical information, then pre-processing is enforced on the dataset which will expel all the superfluous information and concentrate significant highlights from the information.

Step 3: After the preprocessing the dataset we obtain the trained dataset. The lung cancer prediction system will use the trained dataset to get the accurate result.

Step 4: Applying the accurate model to predict the accurate result, implementing machine learning techniques KNN, and Decision Tree.

5. IMPLEMENTATION AND RESULTS

Using EMR details, the lung cancer is predicted in the early stage in which the data is taken from the kaggle website. Python language is used in coding. Anaconda Distribution tool is used. Web concepts are used to create user-friendly screens for the data entry. Machine Learning algorithms are used in categorizing the disease into three stages such as low, medium and high.

The project includes 4 modules:

1. User Module

The user can able to register their account and they can able to log in the application using their credentials.

2. Doctor Module

The doctor module creates the account to check the data with patient details and predicts the disease affected by or not.

3. Patient Module

The patient module is to give the symptoms for their lung cancer.

4. Email Module

The email module is to give the prediction for the final report. The report can be sent to the patient's emailed.

6. RESULTS

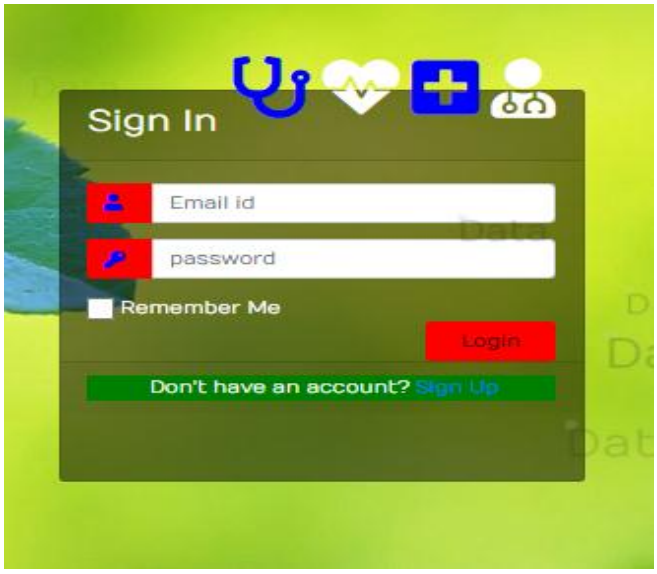

Fig 2: User Credentials

In the above fig 2, this is the login page for the users. Users can sign in as a patient to view the report and doctor to generate the report.

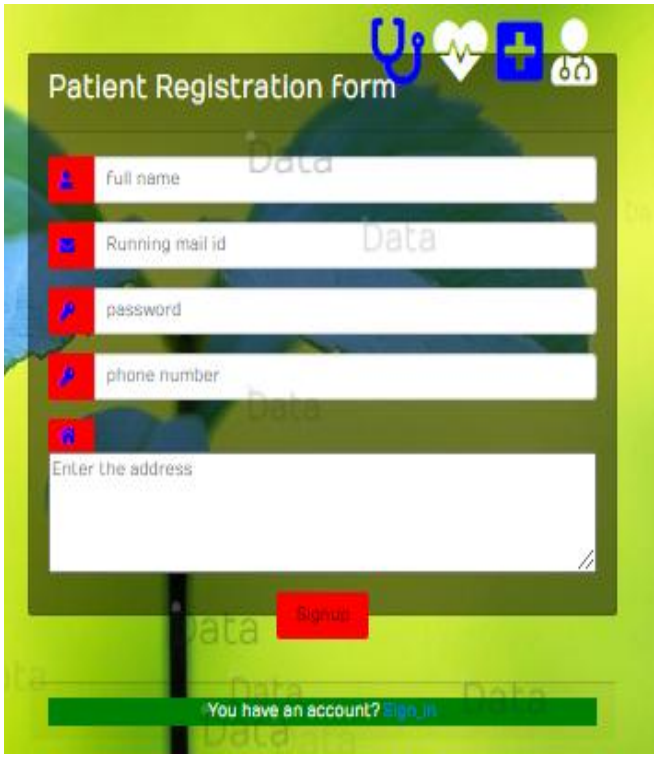

Fig 3: Registration form

In the above fig 3, is the registration form. User can register as a doctor or patient.

Fig 4: Patient Information Form

In fig 4, the information of the patient can be entered by the doctor based on the patient's symptoms and also predicts the disease is in which stage on clicking the submit button.

Lung Cancer detection is categorized into three stages such as low, medium and high.

Case 1: The detection of Lung cancer in low stage.

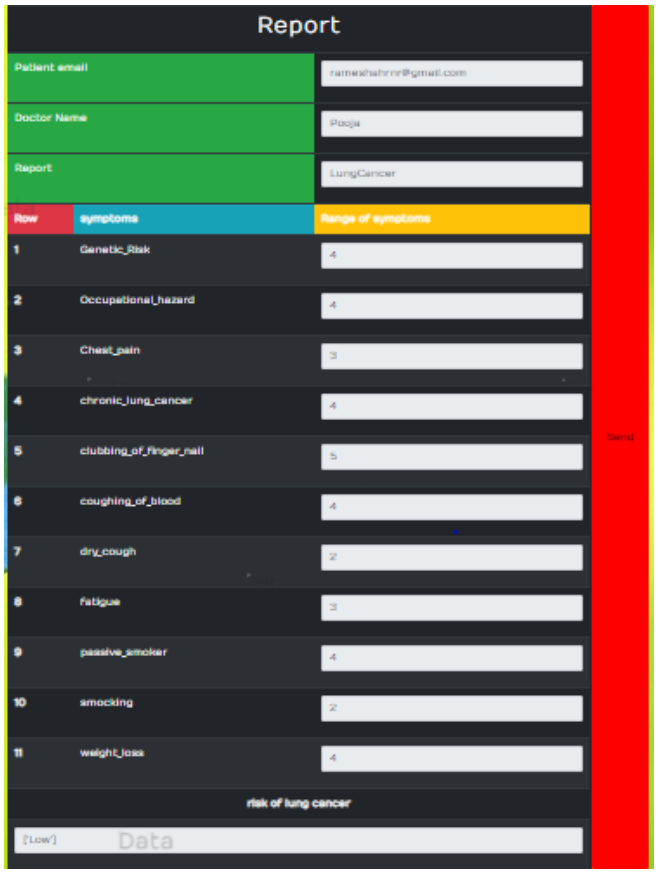

Fig 5: Detection of Cancer in Low stage

Case 2: The detection of Lung Cancer is in Medium stage.

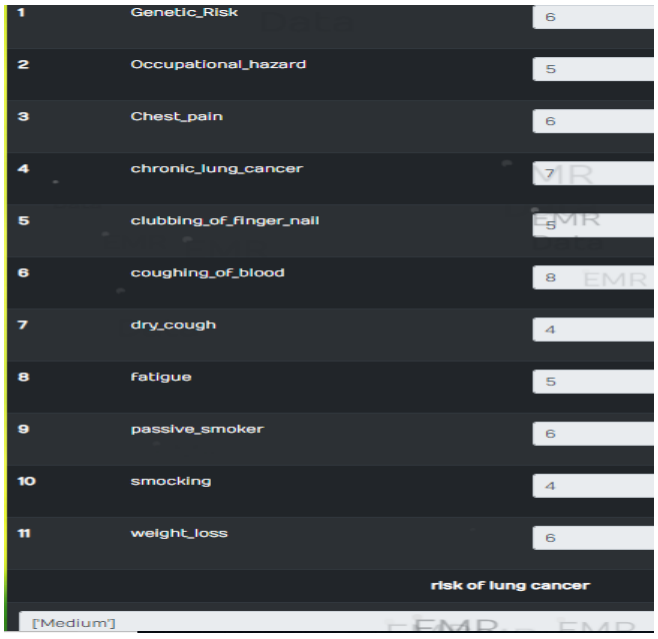

Fig 6: Detection of Cancer in medium stage

Case 3: The detection of Lung Cancer in high stage.

Fig 7: Detection of Cancer in high stage.

From the above fig 5, fig 6 and fig 7, doctor can generate a report based on the patient symptoms. And also the doctor can send the report to the patient through registered email.

7. CONCLUSION

Cancer is the most dangerous disease in the world where treatment is expensive and hence cannot be easily affordable. Among all the cancers, lung cancer is the most affected disease to mankind. It has caused 2.1 million deaths globally in the year 2018 alone. Cancer in the early stages is difficult to predict as it cannot be detected. And once it proceeds to the next stage, it is almost impossible to cure cancer and even if it is cured, the chances of relapse are always present.

This work has tried detecting cancer in the early stages due to which there are high chances of curing cancer. Thanks to the advancement of technology, detection of cancer is easy now due to machine learning techniques which help in predicting accurate values.

REFERENCES

[1] Cancer Facts & Figures, American Cancer Society, 2017.

[2] Arya K S, M N Kavitha, Devi Abirami, Vaishnavi D, "Lung cancer detection using machine learning", IJERT-2019.

[3] Kwetishe Joro Danjuma, "Performance Evaluation of Machine Learning algorithms in post-operative life expectancy in the Lung Cancer patients", Nigeria.

[4] Ranjeet Kaur, Ada, "A study of detection of Lung Cancer using Data Mining Classification techniques", IRJAET-March 2013.

[5] Elsevier B V, "Lung Cancer detection and classification with classification algorithms", Dec-2016.

[6] NLST research team, A M Adams, C D Berg, J D Clapp, R M Fagerstrorm, I F Green, J D Sicks, P M Marcus, W C Black, C Gatsonis and D R Aberle, "Reduced Lung Cancer Mortality with Low-Dose Computed Tomography Screening", Aug-2011.

[7] B Bunn, "Lung Cancer Screening with Low Dose Computed Tomography", 2018.

[8] A Dirksen, K S Bach, P F Clementson, M Dossing, K F Kofoed, J F Rasmussen, N Seersholm, B G Skov, J H Pedersen, "CT screening for Lung Cancer brings forward early disease", April-2012.

[9] C D Berg, T R Church, K L Clingan, D R Aberle, D S Gierada, J Sicks, I Mohan, S De Mello, "Results of the two incidence screenings in the National Lung Cancer Screening Trial",2013.

[10] L M Schwartz, J W Sullivan, J K Aronson, S Woloshin, J Brodersen and C Heneghan, "Overdiagnosis – What it is and What it isn't", 2012.

[11] Z Halpem, N Arber and M Leshno, "Cost-Effectiveness of Colorectal Cancer Screening n the average risk population", 2003.

[12] N K Stout, T Ayer, and O Alagoz, "A POMDP approach to personalize Mammography Screening decisions", 2012.