



Cancer Recognition Through Machine Learning

¹J. Navya Mani, ²G. Bhanu Prakash,

³D. Bharath Simha Reddy, ⁴B. Venkata Koteswararao

^{1,2,3,4}ECE Department, Tirumala Engineering College

¹19NE1A0452, (navyamanijupudi@gmail.com)

²19NE1A0442, (prakashgatta@gmail.com)

³19NE1A0431, (devarapalli@gmail.com)

⁴19NE1A0421, (bvenkat0725@gmail.com)

Under the guidance of (Mr. T. Jagadeesh), Associate Professor, Head of the Department,
Department of Electronics and Communications, Tirumala Engineering College

ABSTRACT

The combination of Machine Learning techniques with any field is one of the easiest ways for analyzation. There are various kinds of cancer diseases in our day to day life. These are very harmful to health, are not curable with medicine till now. It is very important to know about the classification whether it is curable or not. "Cancer recognition through machine learning" is proposed in this project. The main objective is to make every person, who is suffering with cancer, alert about their health without any expensive tests. For analysis of cancer with symptoms, machine learning techniques such as Logistic Regression, KNN, SVM, Random Forest, Decision Tree are used. Therefore by analysing, one can know whether the person has cancer or not.

Keywords— Cancer, recognition, Logistic Regression, KNN, SVM, Random Forest, Decision Tree

I. INTRODUCTION

Cancer is the most threatening disease. It forms in the entire body, starts from a small portion in the body and grows uncontrollably i.e. changes from benign to malignant. There are many reasons for why cancer grows in a body. It is prior to know that cancer is a non-communicable disease. It means that it is a biological disease and does not relate to be speeded through human interactions. There are various reasons for occurring of cancer in a human body.

A. CANCER: Cancer is classified into two categories; benign and malignant. Benign tumors are those that stay in their primary location without spreading to other sites of the body. They do not spread to local structures or to distant parts of the body. Benign tumors tend to grow slowly and have distinct borders in shape. Benign tumors are not generally problematic. But, they can become large and compress structures nearby, causing pain or other medical complications.



II. TYPES OF CANCERS

A. Benign Cancer

Benign cancers can be curable if they are identified in its initial stage. These will be grown and transfer into malignant. A benign tumor is an abnormal but noncancerous collection of cells. It can form anywhere on or in your body when cells multiply more than they should or don't die when they should. A benign tumor is not malignant.

B. Malignant Cancer

Malignant tumors have cells that grow uncontrollably and spread locally to distant sites. Malignant tumors are cancerous. They spread to distant sites through the bloodstream or the lymphatic system. This spreading process is called *metastasis*. Metastasis can occur anywhere in the body and most commonly it is found in the liver, lungs, brain, and bone. Malignant tumors can spread rapidly and require treatment to avoid spread. By using modern technology, the prediction becomes easier. The machine learning makes it easier through its wide range of algorithms. The algorithms are taken here in order to choose the high precision output for the user.

There are two cancers taken for the prediction. They are namely Lung cancer and Cervical cancer.

Lung Cancer: Lung cancer is the highly damaging cancer that transfers into malignant in a very limited time. Then it is important to note that there is an urgency of finding the cancer and take preventive steps. The lung cancer occurs at the lungs. For lung cancer, we are considering smoking, yellow fingers, anxiety, peer pressure etc. [1].

Cervical Cancer: Cervical cancer is the 4th most hazardous cancer faced by women across the world. It is clear that the cancer should be treated quickly. It is the cancer that grows at the cervical area, the lower part of womb. In order to detect the cancer, we have the symptoms of the cancer. The symptoms differ in each cancer. In lung cancer, the symptoms taken are smoking, alcohol consumption, yellow fingers, swallowing difficulty and anxiety etc. Cervical cancer has the symptoms like age, number of partners, hinselmann, schiller, citology etc.

III. LITERATURE SURVEY

The existing system consists the data set and is trained. It takes input from user. Then sends data to server. The server takes the data from the input, and then through http requests, handles the data based on the route opted. After evaluation, it returns the value as a response on the screen. Then by http response simulation, the output is displayed as the response received by the server. It is the end of the problem. If the result produced is satisfied by the user, it is the end. If the user is not satisfied, then again the analysis begins from dataset. The literature base paper consists of three cancers. These are analysed based on the cell parameters like cell size, the area, width, colour etc. These are very peculiar and not exactly known to user unless by performing some calculations.

IV. PROPOSED SYSTEM

The proposed system is as shown in above figure. At first, the database is collected from cancer organizations. The dataset may consists some alpha data values. This is handled by handling the dataset. Then the process

begins with the feature selection using RFE method. This makes the dataset precise and then becomes as a finite dataset. The dataset is given for testing and training in the ratio 80: 20 for the machine learning algorithms. The data is then perfectly trained and ready to analyse the data given by user. In order to analyse, the best algorithm performance is selected from multiple algorithms in machine learning.

The main objective of the project is to develop a facility that can easily and accurately predict cancer. The machine learning makes it easier with the algorithms. The proposed system is developed using the symptoms that a cancer victim should have. These can be easily understandable to the user. The existing system is developed using the cell radius, colour etc. which are the values of the tumor. It gives accurate results but the victim is not able to identify the cell parameters. Hence the proposed system will be advantageous and friendly to predict the presence of cancer.

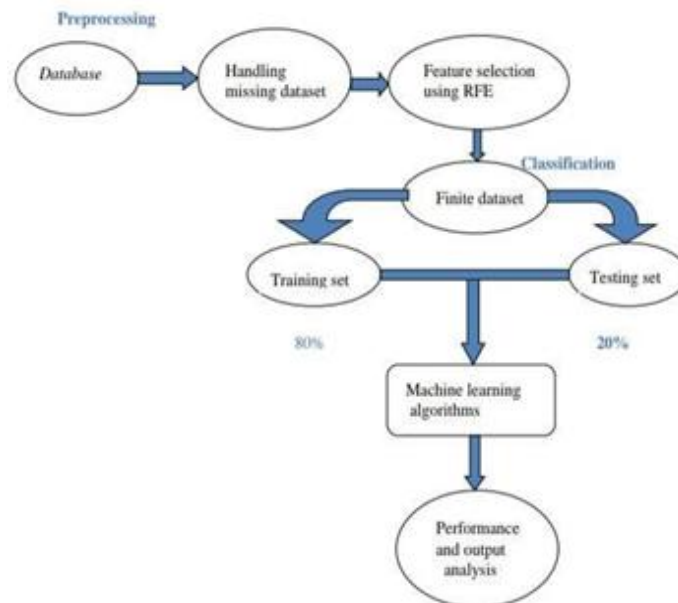


Fig: Proposed System

V. TRAINING AND TESTING THE DATASET

The dataset is given to the machine learning algorithms to analyse their performance in prediction. The best algorithm is selected for the creation of webpage. Before training, the data should be finite in order to get accurate output. To get the dataset, we need to develop RFE correlation. The machine learning algorithms are given with datasets and their performance is measured. The datasets are trained with their respective parameters and are tested. The best algorithm is furtherly decided for précised output. The algorithms are KNN, Xgboost, Logistic Regression, SVM, Random Forest and Decision Tree. For lung cancer, Xgboost has high accuracy and hence selected.

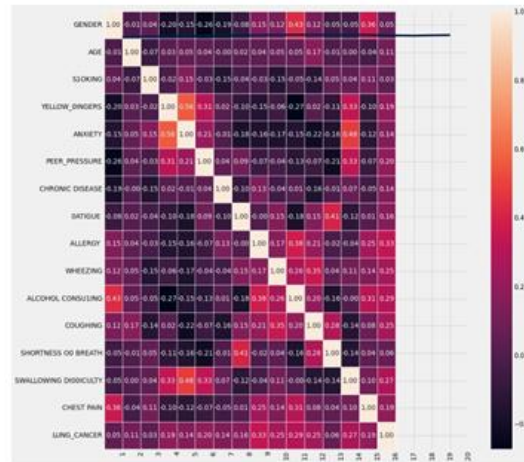


Fig:2.1. Feature correlation of lung cancer

VI. RESULTS

Among all those algorithms, the best performance is selected. For Cervical Cancer Logistic Regression is selected since it has high testing accuracy. It is because both cancers have different symptoms as parameters. The web page for userinterface is created using selected best model.

The lung cancer and cervical cancer are subjected to the training and they got the result as:

Here the lung cancer has 280 – lung cancer victims and 39 non lung cancer victims. For the lung cancer, the best algorithm is Xgboost algorithm.

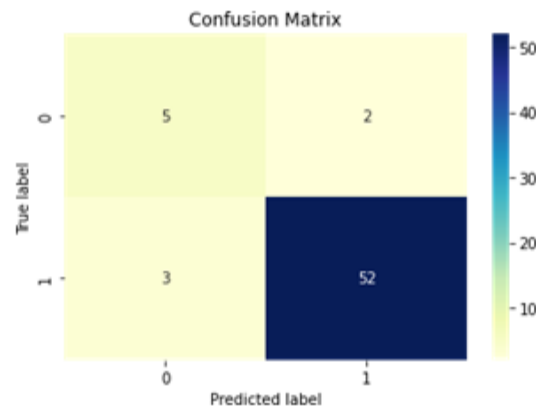


Fig: Confusion Matrix of Lung Cancer dataset

Training Results of Lung Cancer:

Xgboost:

Precision : 0.91

Accuracy : 0.91

Recall : 0.95F1 – Score: 0.93

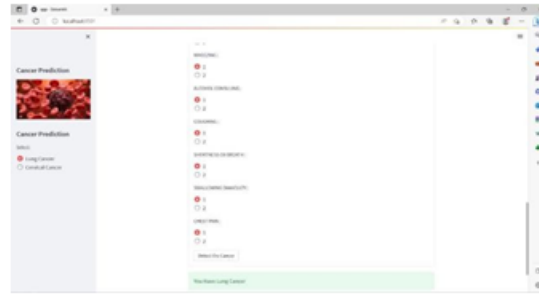


Fig: Lung Cancer webpage

The web page is created. Then opting for the cancers is present on the left side. By choosing the symptoms, the lung cancer can be predicted.

And for the cervical cancer, the best algorithm is Logistic Regression.

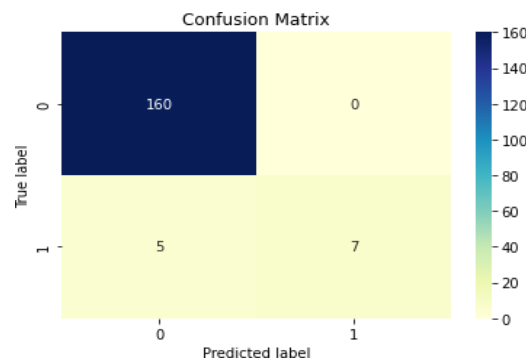


Fig: Confusion Matrix of Cervical Cancer

Training Results of Cervical Cancer:

Logistic Regression:

Precision : 0.95

Accuracy : 0.97

Recall : 0.98F1 – Score : 0.97

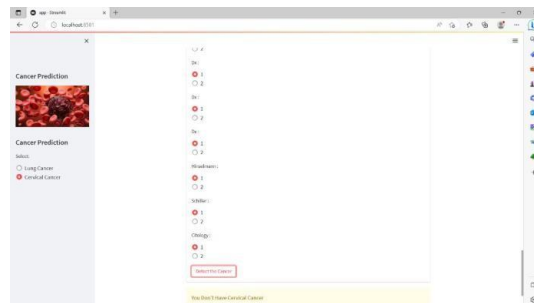


Fig: Cervical Cancer web page

The cervical cancer is predicted by opting the symptoms.

Then result is given.

The website link for Cancer prediction is: Local URL: <http://localhost:8501> Network URL:

<http://192.168.156.198:8501>



CONCLUSION

This project is designed to make the users understand about the situation of the victim health regarding cancer disease. The project ensures to process data. When the characteristics are identified, the clusters are formed. These clusters are processed with the machine language algorithms. The algorithms are further classify the data and the acquire information by processing, then the result will be displayed onthe screen.

REFERENCES

- [1] G. Sruthi, Ch. Likitha Ram, M. Koushik Sai, BhanuPratap Singh, Nikhil Majhotra, Neha Sharma; “Cancer prediction using machine learning”; 2022 2nd International Conference on Innovative Practices in Technology andManagement (ICIPTM)
- [2] Wunsch, Georgios, RuiXu,”Multiclass Cancer Classification Using Semi supervised Ellipsoid ARTMAP and Particle Swarm Optimization with Gene Expression Data”,Jan- March, 2007, IEEE/ACM Transactions on ComputationalBiology and Bioinformatics (Volume:4, Issue: 1)
- [10] K Kourou, Konstantinos P.Exarchos, Michalis V.Karamouzis, “Machine learning applications in cancer prognosis and prediction”,2015 Syed Saba Raof, M A. Jabbar, Syed Aley Fathima, “Lung Cancer Prediction using Machine Learning: A Comprehensive Approach”, 23 April 2020, 2020 2nd International Conference on Innovative Mechanisms forIndustry Applications (ICIMIA).
- [11] C. Suguna, S. P. Balamurugan; “ Diagnosing Cervical cancer using machine learning methods”; 2022 International Congress On Human – computer interaction”, Optimization and Robotic Applications (HORA)
- [4] Aman Sharma & Rinkle Rani, “A Systematic Review of Applications of Machine Learning in Cancer Prediction and Diagnosis”, Survey Articla, Published: 25th January, 2022
- [5] Nimmagadda Satyanarayana Murthy & Chaitanya Bethala, “Review paper on research direction towards cancer prediction and prognosis using machine learning and deep learning model”, Original Research, Published: 29 March 2021
- [6] Dakhaz Mustafa Abdullah, Adnan Mohsin , Amira Bibo Sallow, “Lung cancer Prediction and Classification based on Correlation Selection method Using Machine Learning Techniques”.
- [7] YawenXiao, JunWu, ZongliLin, XiaodongZhao , “A deep learning-based multi-model ensemble method for cancer prediction”, 3 October 2017
- [8] Aman Sharma & Rinkle Rani, “A Systematic Review of Applications of Machine Learning in Cancer Prediction and Diagnosis”, Survey article, Published: 25 January 2021.
- [9] Akshitha Shetty; Vrushika Shah, “Survey of Cervical Cancer Prediction Using Machine Learning: A Comparative Approach”, 18 October 2018, 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT)