

Smart Health Prediction System Using Python

Manisha M S Pillai¹, Rahul Gopal², Roshitha Mariam Sunny³, Revathy Chandran⁴, Akhila Balachandran⁵

¹ Department of Computer Science, St. Thomas College of Engineering and Technology, Kerala, India

² Department of Computer Science, St. Thomas College of Engineering and Technology, Kerala, India

³ Department of Computer Science, St. Thomas College of Engineering and Technology, Kerala, India

⁴ Department of Computer Science, St. Thomas College of Engineering and Technology, Kerala, India

⁵ Department of Computer Science, St. Thomas College of Engineering and Technology, Kerala, India

*Corresponding Author: manishamspillai@gmail.com, Tel.: +91 95622 97088

DOI: <https://doi.org/10.26438/ijcse/v7i5.13721375> | Available online at: www.ijcseonline.org

Accepted: 24/May/2019, Published: 31/May/2019

Abstract— Breast cancer is increasing day by day due to life style, hereditary. So, health care need to be modernized it means that the health care data should be properly analyzed. Breast Cancer has become the common cause of death among women. Due to long hours invested in manual diagnosis and lesser diagnostic system available emphasize the development of automated diagnosis for early diagnosis of the disease. Each individual has different values for each attributes of breast cancer. Diagnosis is done by classifying the tumor. Tumors can be either benign or malignant but only latter is the cancer. Malignant are more cancerous than the benign. Unfortunately not all physicians are expert in distinguishing between the benign and malignant. So we need a proper and reliable diagnostic system that can detect the malignant. The frameworks use will provide multipurpose beneficial outputs which includes getting the healthcare data analysis into various forms. In this Smart Health Prediction Using Python, we are proposing a evaluate classification technique used for predicting the risk level of each person. The proposed system is using 13 attributes and 569 datasets to develop an accurate result. The patient risk level is classified using machine learning classification algorithm that is Logistic regression algorithm; the accuracy of the risk level is high when using a greater number of attributes. The proposed system will group together symptoms data and analyze it to provide cumulative information. After the analysis, algorithm could be applied to the resultant and grouping can be made to show a clear result. Our aim is to classify whether the breast cancer is benign or malignant for the analysis purpose of laboratories.

Keywords— Logistic regression algorithm, Malignant, Benign

I. INTRODUCTION

Breast Cancer is the most common type of cancer World Wide and is the leading cause of death among women. The most effective way to reduce breast cancer deaths is by detecting it earlier. This is possible by performing various tests like MRI, mammogram, ultrasound and biopsy. Breast Cancer refers to uncontrolled growth of cells in the breast tissue. If these cells are not stopped or controlled then it might cause an adverse effect on the whole body. Breast cancer can occur in men too having a higher mortality rate.

Several types of research have been done on early detection of breast cancer to start treatment and increase the chance of survival. Now a day's lots of woman's are suffering from breast cancer. It is vital to find alternative methods which are easier to implement and work with different data sets, cheaper and safer and provide accurate result. We have designed a system for hospital or any related authority to get diagnosed from patient's details. The

“SMART HEALTH PREDICTION SYSTEM USING PYTHON” application is an end user support project.

Here, we propose a web application that allows users to get instant result of their diseases through an intelligent system online. The application is fed with various details of the breast cancer associated with those details. The application allows user to share their symptoms. It then processes user specific details to check for cancer that could be associated with it. Here we use some intelligent machine learning algorithm including Logistic regression algorithm to guess the most accurate result that could be associated with patient's details. This study also discusses the datasets used for breast cancer detection and diagnosis. The proposed model can be used with different data types such as age, radius mean, texture mean, symmetry mean, radius session, texture session, smoothness session, fractional dimension session ,etc. After processing the data sets system predicts the result with the help of two terms benign and malignant which means benign stands for non-cancerous patient and malignant stands for cancerous patient.

The “SMART HEALTH PREDICTION SYSTEM USING PYTHON” is an end user support website and users to get diagnosed from the hospital. This system allows users to get analysis on the symptoms they give for predicting the chances of breast cancer. Here, we propose a web application that allows users to get instant result using an alert message . The main advantages include the no need for paper or other items, as the system is running in digital space, thus eco-friendly. The time required is reduced greatly by the use of our system. Based on result, system automatically shows the result specific accurate percentage.

Section I contain the introduction of Smart Health Prediction System using Python, Section II contain the related works, Section III contain the methodology used in proposed system, Section IV contains the result and discussion and Section V contains conclusion and future scope.

II. RELATED WORK

A number of related works exist using different data mining techniques and machine learning algorithms. The main aim is to achieve better accuracy for predict the chances of breast cancer. Factor considered for breast cancer prediction are radius mean, texture mean, symmetry mean, etc. In this section ,several papers are studied for better implementation of the project.

Abien Fred M. Agarap[1] proposed on breast cancer detection: An Application of Machine Learning Algorithms on the Wisconsin Diagnostic Dataset. In this paper a neural network architecture GRU-SVM combining the gated recurrent unit (GRU) variant of recurrent neural network (RNN) and the support vector machine (SVM), for the purpose of binary classification. System had high accuracy.

Hadi Lotfnezhad Afshar, Maryam Ahmadi, Masoud Roudbari and Farahnaz Sadoughi[2] proposed prediction of breast cancer survival through knowledge discovery in databases. In this paper there are three different types of methods: “Support Vector Machine (SVM), Bayes Net, and CHi-squared Automatic Interaction Detection (CHAID)” were used as prediction models. SVM is one of the supervised learning algorithms with well-built regularization properties. Bayes Network is capable of learning the probability density functions of individual pattern classes from a collection of learning samples. CHAID can be used for prediction as well as classification, and for detection of interaction between variables. . Accuracy is the percentage of testing data that are correctly predicted by the model.

S. Palaniappan and T. Pushparaj[3] proposed a novel prediction on breast cancer from the basis of association rules and neural network. In this system AR1 + AR2 + NN method are used in breast cancer diagnosis problem.

P. Saranya and B. Satheeskumar[4] proposed a survey on feature selection of cancer disease using data mining techniques. In this paper the k-Nearest Neighbors algorithm is a non-parametric method” used for classification and regression.

Chaurasia V, Pal S.[5] proposed data mining techniques: To predict and resolve breast cancer survivability. This paper provides easy RBF, Logistic and REP Tree for prognosis of breast cancer can be used to obtain fast automatic diagnostic systems for other diseases.

L. Breiman Random Forests[6] proposed J. Machine Learning making use of specific classification techniques for diagnosis of breast cancers.

L. Tolosi, and T. Lengauer[7] proposed classification with correlated features unreliability of feature ranking and solutions, bioinformatics Observations were made from a survey by applying the numerous techniques which can also be utilized by several investigators to diagnosis breast cancers.

III. METHODOLOGY

3.1. PROPOSED SYSTEM

The proposed system is “SMART HEALTH PREDICTION SYSTEM USING PYTHON”. In this system we propose a novel system for breast cancer using Logistic regression algorithm. Here we provide some data such as radius mean, texture mean, symmetry mean, etc., as inputs.

The system identifies the diseases associated with the given data. User will be asked to enter the data, then system will processes those attributes and dataset values for cancer that user could be alike with. It will be giving instant diagnosis on the user entered data. Logistic regression is another technique borrowed by machine learning from the field of statistics .It is the go to method for binary classification problems. Logistic function also called sigmoid function.

The algorithm used is logistic Regression algorithm. We are implementing this system by using 569 datasets. The data consist of 32 attributes for training purpose after analysis select 13 relevant attributes and removed the remaining irrelevant attributes. The output of our system have two classes, either 0 or 1. If the output is class 0 it denotes that no chance for cancer. Output class1 indicates cancer exists. The result of mean value of one patient after applying in a sigmoid function may result 0.90 it approximately considered as class1. In this system we assign a threshold value. Threshold value is 0.5. If the resultant values below the threshold is included as class 1. Value above 0.5 is included in class 0. Here we use Binary Logistic Regression algorithm, it is used for quite large samples.

Attributes are:

- id
- diagnosis: M = malignant, B = benign
- radius mean
- texture mean
- symmetry mean
- compactness mean
- fractional dimension mean
- radius session
- texture session
- smoothness session
- compactness session
- symmetry
- fractional dimension session

Required equations for Logistic regression algorithm are:

Sigmoid function:

$$p = 1 / (1 + e^{-y})$$

Linear Regression Equation:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$$

Where y-dependent variable; x1, x2- explanatory variables

Apply sigmoid function on Linear Regression.

$$p = 1 / (1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)})$$

IV. RESULTS AND DISCUSSION

The system predicts presence or absence of the breast cancer by evaluating input attributes by Logistic regression algorithm. Here only 569 datasets are given for testing and accuracy is determined.

The screenshot of adding clinical data from hospital authorities is shown in the figure 1

Fig 1:adding attribute data

Screenshot of final breast cancer analysis prediction is shown in the fig 2

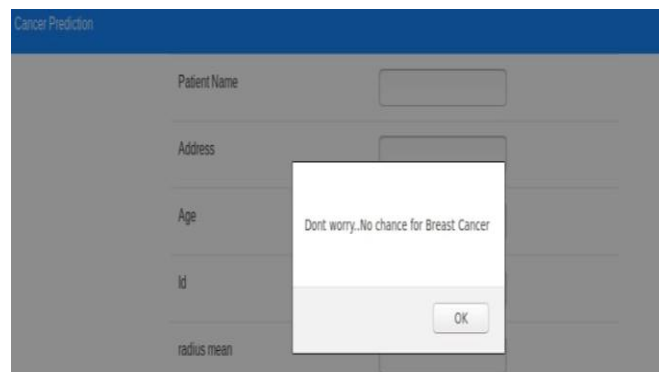


Fig 2:output of breast cancer prediction

V. CONCLUSION AND FUTURE SCOPE

In this project, we are proposing a prediction system for breast cancer using Logistic regression algorithm. It is more suited technique for prediction of breast cancer. Here we used two values 0 and 1, malignant is represented with a value of 1 and benign is represented with 0. It is more efficient. This technique is highly acceptable and can help the clinical staffs for early diagnosis so they can save the data for analysis purpose in future. User login into system using username and password, after login to the system, where the user can give clinical result data. Used dataset taken from Wisconsin repository. Comparing the obtained data with user given data, we can predict the presence of breast cancer, accuracy of the system is high. As a future work, it can be used as desktop applications, mobile app for multi users so through that each user can check the result from anywhere, it can also include SMS sending system for getting notifications to patients so they can identify the result if any problem in their result they can visit to a doctor. It can be implemented also using Cloud technology.

ACKNOWLEDGMENT

We gladly present this project titled "SMART HEALTH PREDICTION SYSTEM USING PYTHON" fulfillment for the award of the degree of Bachelor of Technology in Computer Science and Engineering.

Firstly we heartily thank GOD Almighty for his never-ending support for the successful completion of this seminar.

We express my hearty gratitude towards our staff adviser Mr. RAHUL GOPAL (Assistant Professor) for his timely advice and constructive guidance in carrying out the seminar into a success.

We express my sincere thanks to the Head of the Department of Computer Science and Engineering Mr. IMTHIYAS M P (Professor) for his support and advice.

We once again express my sincere thanks to all the staff in Department of Computer Science and Engineering who helped me directly or indirectly to bringing this seminar with stipulated time. We express grateful thanks to parents, friends and classmates for their help and support.

REFERENCES

- [1] Abien Fred M. Agarap, "On Breast Cancer Detection: An Application of Machine Learning Algorithms on the Wisconsin Diagnostic Dataset", ICMLSC, Phu Quoc Island, Viet Nam, February 2018.
- [2] Hadi Lotfnezhad Afshar, Maryam Ahmadi, Masoud Roudbari and Farahnaz Sadoughi "Prediction of Breast Cancer Survival Through Knowledge Discovery in Databases", Global Journal of Health Science; Vol. 7, Published by Canadian Center of Science and Education, 2015.

- [3] S. Palaniappan and T. Pushparaj, "A Novel Prediction on Breast Cancer from the Basis of Association rules and Neural Network", IJCSMC, Vol. 2, Issue. 4, 269 – 277, April 2013.
- [4] P. Saranya and B. Satheeskumar, "A Survey on Feature Selection of Cancer Disease Using Data Mining Techniques", IJCSMC, Vol. 5, Issue. 5, 713 – 719, May 2016.
- [5] Chaurasia V, Pal S., "Data Mining Techniques: To Predict and Resolve Breast Cancer Survivability". International Journal of Computer Science and Mobile Computing Vol 3, 10–22, 2014.
- [6] L. Breiman Random Forests, "J. Machine Learning", vol.45, pp. 5 -32, 2001.
- [7] L. Tolosi, and T. Lengauer, "Classification with correlated features: unreliability of feature ranking and solutions, Bioinformatics", vol. 27, no. 14, pp. 1986-1994, 2011.

Authors Profile

Ms. Manisha M S Pillai currently a final year computer science & engineering student of St.Thomas College of engineering & technology, Chengannur, Kerala. Her area of interest are Robotics, Machine learning.

Mr. Rahul Gopal currently working as Assistant Professor in Department of Computer science, St.Thomas College of engineering and technology, Chengannur, Kerala. Had a working experience in teaching of about 6 year. Completed B.Tech from Kerala University and M.Tech in Computer Science & Engineering from M.G University kottayam. Areas of interest include Computer networks and Computer system architecture.

Ms. Roshitha Mariam Sunny currently a final year computer science & engineering student of St.Thomas College of engineering & technology, Chengannur, Kerala. Her area of interest are Robotics, AI.

Ms. Revathy Chandran currently a final year computer science & engineering student of St.Thomas College of engineering & technology, Chengannur, Kerala. Her area of interest are AI, Machine learning.

Ms. Akhila Balachandran currently a final year computer science & engineering student of St.Thomas College of engineering & technology, Chengannur, Kerala. Her area of interest are Computer networks, Machine learning.