

Prediction of Cancer Risk in Perspective of Symptoms using Naïve Bayes Classifier

^[1] Pallavi Mirajkar ^[2] Dr. G. Prasanna Lakshmi

^[1] Research Scholar, Faculty of Computer Science ^[2] Guide, Associate Professor

^[1] Pacific Academy of Higher Education and Research University, Udaipur.

^[2] Andhra University

Abstract— Health Care professionals face the complex task of predicting the type of cancer in patient. This earlier prediction of cancer helps to the practitioner to recommend cancer treatment. Various studies have been carried out that examine patient to improve help practitioner's prognostic accuracy. Here we worked on major factor that is symptoms. In this thesis paper we have used Naïve Bayes Classification algorithm of data mining to predict the type of cancer. Proposed method of risk prediction aims to predict probability of cancer. Based on the classification algorithm, symptoms of the cancer are classified using Navie Bayes algorithm to recognize the risk of cancer such as lung, breast, ovarian, stomach and oral. Therefore accurate prediction of cancer in patient is important for good clinical decision making in health care strategies.

Index Terms— Data Mining, Naïve Bayes Classifier, Probability, Cancer, Symptoms.

INTRODUCTION

Cancer is a collection of diseases concerning anomalous cell growth with the potential to attack or spread to other parts of the body. Cancer becomes the most life threatening disease now days. Accurate prediction of cancer in patient is important for good clinical decision making and for recommending the treatment, at present, accurate detection remains challenge even for experienced clinicians. The most effective way to decrease cancer deaths is its early recognition. Early diagnosis requires an accurate and reliable diagnosis procedure that allows physicians to diagnose symptoms of cancer.

Data mining plays a crucial role in the discovery of knowledge from huge databases. Data mining has created its important hold in every field including healthcare. Data mining has its most important role in extracting the concealed information. Data mining process is more than the data analysis which includes classification, clustering, association rule mining and prediction. A patient affected with any Cancer may suffer symptoms in the body. Cancer symptoms are used to forecast the risk level of the cancer disease. The main aim of this study is to predict the risk of type of cancer using Naïve Bayes classification algorithm. Naive Bayes is one of the most effective statistical and probabilistic classification algorithms. Naïve Bayes algorithm is called "Naïve" because it makes the assumption that the occurrence of certain feature is independent of the occurrence of other features. Naive Bayes algorithm is the algorithm that learns the

probability of an object with certain features belonging to a particular group/class. Naive Classifiers assumes that the effect of an attribute value on the given class is independent of the values of other attributes. Naive Bayesian Classifiers depends upon the Baye's theorem. An advantage of the Naive Baye's classifier is that it requires a small amount of data to estimate the parameters needed for classification. It performs superior in numerous complex real world situations like weather forecasting, Medical Diagnosis, spam classification. It is suitable when dimensionality of input is high.

RELATED WORK

Kritharth Pendyala et.al (2017) proposed a system that detects and predicts breast cancer. They have applied classification algorithm to data using Naïve Bayes and SVM algorithm. The proposed hybrid algorithm gives more accuracy and reduces the time complexity.

Subrata Kumar Mandal (2017) has applied various techniques such as data cleaning, feature selection, feature extraction, data discretization and classification for predicting breast cancer as accurately as possible. They states that Logistic Regression Classifier has given maximum accuracy.

N.V. Ramana Murty and Prof. M.S. Prasad Babu (2017) analyzed lung cancer data set using WEKA tool with various data mining classification techniques. It is recognized that the Navie Bayes classification algorithms.

Keerti Yeulkar et. al. (2017) classified the SEER dataset in to malignant tumor and benign tumor using

C4.5 and Naïve Bayes algorithm. They have also applied decision tree classification technique.

Younus Ahmad Malla et.al (2017) performed analytical evaluation of machine learning algorithm such as Naïve Bayes, Logistic Regression and Random Forest. WEKA tool is used for running different data mining algorithms for breast cancer data set.

Rashmi M, Usha K Patil(2016) have used Navie Bayes algorithm to diagnose cancer using Gene data set. They stated that data mining can be used to classify cancer patient by analyzing the data.

Animesh Hazra et.al. (2016) have studied various classification techniques and compared them in terms of accuracy and time complexity. They have concluded that Navie Bayes gives better accuracy and it is best classification technique.

Shweta Kharya , Sunita Soni (2016) have developed a predictive model using Navie Bayes algorithm. They have applied weighted concept on Navie Bayes Classification for breast cancer diagnosis. They developed model is more readable, modifiable, efficient and can be handled easily.

Dr. Vani Prumal et.al. (2016) have proposed methodology that uses Naïve Bayes classifier to predict stomach cancer. They stated that Naïve Bayes algorithm is the most appropriate algorithm for the prediction of stomach cancer. It gives more accurate result.

Dr. P. Indra Muthu Meena (2016) implemented C4.5 and Naïve Bayes algorithm for prediction of stomach cancer. They have also analyzed and compared the performance of both algorithms.

Kathija, Shajun Nisha (2016) proposed an algorithm by using support vector machine and Naïve Bayes algorithm. They evaluated performance of the proposed algorithm using confusion matrix. After comparing both algorithms, it is found that Naïve Bayes algorithm gives highest accuracy.

Samuel Giftson Durai et.al (2016) have done survey and shown that decision tree is the best suitable classification algorithm for prediction of breast cancer

and experiments have been conducted using MATLAB software.

B. M. Gayathri, C. P. Sumathi (2016) applied Naïve Bayes algorithm and designed GUI to enter details of the patient for predicting whether given data is benign or malignant. Proposed work has shown that Naïve Bayes classifier predicts breast cancer with less attributes.

P. Saranya, B. Satheeskumar (2016) analyzed that data mining techniques helps to recognize data to predict the cancer at an early stage. They have studied various data mining algorithms such as Naïve Bayes, ANN, Support Vector Machine and decision tree.

Monika Hedawoo et. al. (2016) proposed method to classify mammogram using different classification algorithms for predicting tumor cells in breasts. It has been revealed that Naïve Bayes algorithm is more sensitive.

Animesh Hazra et. al. (2016) applied data cleaning, feature selection, feature extraction, data discretization and classification techniques to predict breast cancer correctly. In proposed method Naïve Bayes algorithm gives maximum accuracy.

V. Kirubha, S. Manju Priya (2015) proposed method to analyze risk factors of lung cancer. They have used classification algorithm like Naïve Bayes, Random Forest, REP Tree and Random tree. It is proved that REP Tree provides better result.

S. Kalaivani et.al (2015) studied three classifiers Naïve Bayes, Random Tree and Support Vector Machine. They have calculated the performance of these algorithms and it is observed that Naïve Bayes algorithm gives more accurate results.

Mandeep Rana et.al (2015) has applied different algorithms on dataset such as KNN, logistic regression and Naïve Bayes algorithm. It is observed that Naïve Bayes and logistic regression gives accurate results for diagnosing breast cancer.

PROPOSED WORK

Introduction

For better perceptive of approach, we take the paradigm which is shown below in the following Tables:

Symptoms of Lung Cancer
Cough that won't quit
Pain in the chest area
Wheezing
Raspy, hoarse voice
Drop in weight
dyspnea (shortness of breath with activity)
dysphasia(difficulty Swallowing)
Shortness of breath

Table No. 1 Lung Cancer

Symptoms of Breast Cancer
Hard Lump
Changes in the shape or size of the breast
Changes to the nipple(inverted nipple)
Discharge that comes out of the nipple without squeezing
Breast pain

Table No. 2 Breast Cancer

Symptoms of Oral Cancer
Lump in the neck
Unexplained bleeding in the mouth
Losing in the teeth
Pain in chewing, swallowing or moving tongue

Table N. 3 Oral Cancer

Symptoms of Ovarian Cancer
Irregular menstrual
Painful intercourse
Increase in urination
Changes in bowel movement
Abnormal bleeding
Pain in the abdomen
Bloating

Table No. 4 Ovarian Cancer

Symptoms of Stomach Cancer
Indigestion or heartburn
Vomiting and nausea
Pain in the abdomen
Constipation or diarrhea
Bloating Stomach

Table No. 5 Stomach Cancer

Naïve Bayes classifiers are a group of basic probabilistic classifiers based on applying Bayes theorem with independence assumptions between the features. Using Bayesian classifiers, the method can determine the hidden facts related with diseases from symptoms of the patients. According to naïve Bayesian classifier the occurrence (or unoccurrence of a particular feature of a class is considered as independent to the presence (or absence) of any other feature. When the dimension of the inputs is high and more efficient result is expected. Naïve Bayes model identifies the physical characteristics and symptoms of patients suffering from cancer disease. Naive Bayes algorithm is based on Bayes theorem. There are two types of probabilities

- Posterior Probability $[P(H/X)]$
- Prior Probability $[P(H)]$

where X is data tuple and H is some hypothesis. According to Bayes' Theorem, $P(H/X) = \frac{P(X/H)P(H)}{P(X)}$.

System Architecture:

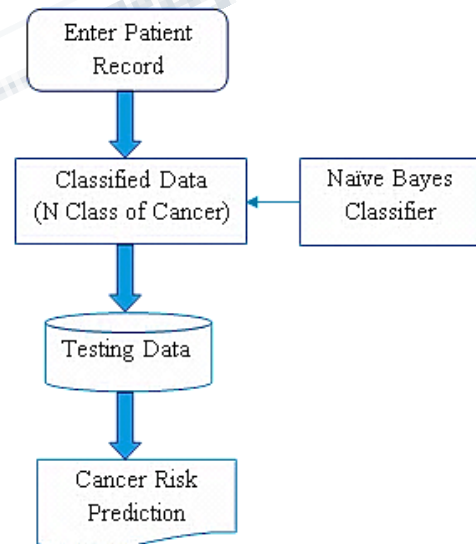


Figure 1 System Architecture

Implementation of an Algorithm is as given below:

Suppose that there are n classes of cancer, C_1, C_2, \dots, C_n and each data sample is represented by n dimensional symptoms of cancer, $S = \{S_1, S_2, \dots, S_n\}$.

Step 1: To calculate the probability of type of cancer. First recognize the symptoms of the specific cancer. $P(C_1|S_1, S_2, \dots, S_n)$.

Step 2: Calculate the probability of each symptoms. Probability of each Symptom = $P(S_i|C_1)$

Where $i = \{1, 2, \dots, n\}$

Step 3: Multiply each probability of symptoms together Probability of $C_1 =$

$$P(S_1|C_1) * P(S_2|C_1) * \dots * P(S_n|C_1)$$

Step 4: Repeat step 1 to Step 3 for each cancer class.

Step 5: Predict that S belongs to the type of cancer which is having the highest posterior probability.

$$P(C_i|S) > P(C_j|S) \text{ for all } 1 \leq j \leq n \text{ and } j \neq i$$

Display the risk of one of the type of cancer from predefined set of class.

The Naïve Base algorithm is playing a vital role in mining the necessary information. Based on the above mentioned algorithm and computed the risk of cancer, some tests will be prescribed to confirm the presence of cancer.

CONCLUSION

In this paper, we have used Naïve Bayes classifier for predicting cancer by taking into consideration the related symptoms. It is essential to work on different symptoms and discover the probability of cancer in patient. Prediction of cancer at an early stage can be helpful for better treatment and survival of the patient. In this paper, we compute the probability of risk of the cancer based on symptoms. And then gives a warning of which type of cancer patient may have. The goal of this paper is to give the prior warning to the patient. This will eventually spare the time and decrease the cost of treatment and conjointly will increase the possibility of survivability. Future work can be directed on the prediction of cancer by integrating the various dimensions such as cancer causing factors, contaminated food items and symptoms.

REFERENCES

1. Rashmi M, Usha K Patil "Cancer Diagnosis Using Naive Bayes Algorithm" International Journal of Recent Trends in Engineering & Research (IJRTER)

Volume 02, Issue 05; May - 2016 [ISSN: 2455-1457].

2. Animesh Hazra, Subrata Kumar Mandal, Amit Gupta "Study and Analysis of Breast Cancer Cell Detection using Naïve Bayes, SVM and Ensemble Algorithms" International Journal of Computer Applications (0975 – 8887) Volume 145 – No.2, July 2016.

3. Shweta Kharya, Sunita Soni "Weighted Naive Bayes Classifier: A Predictive Model for Breast Cancer Detection" International Journal of Computer Applications (0975 – 8887) Volume 133 – No.9, January 2016.

4. Dr. Vani Prumal, Shibu Samuel, Dr. P. Indra Muthu Meena "Application of Training Dataset using Naïve Bayes Classifier for Prediction of Stomach Cancer in Female Population" International Journal of Scientific Engineering and Technology Research Volume.05, IssueNo.45, November-2016, Pages: 9230-9234

5. Dr. P. Indra Muthu Meena, Dr. Vani Perumal "Performance of C4.5 and Naïve Bayes Algorithm to Predict Stomach Cancer - An analysis" International Journal of Advanced Research in Computer and Communication Engineering ISO 3297:2007 Certified Vol. 5, Issue 11, November 2016

6. Kathija, Shajun Nisha "Breast Cancer Data Classification Using SVM and Naïve Bayes Techniques" International Journal of Innovative Research in Computer and Communication Engineering (An ISO 3297: 2007 Certified Organization) Website: www.ijrcce.com Vol. 4, Issue 12, December 2016

7. Subrata Kumar Mandal "Performance Analysis Of Data Mining Algorithms For Breast Cancer Cell Detection Using Naïve Bayes, Logistic Regression and Decision Tree" International Journal Of Engineering And Computer Science ISSN: 2319-7242 Volume 6 Issue 2 Feb. 2017, Page No. 20388-20391

8. Kritharth Pendyala., Divyesh Darjee., Vaishnavi Adhyapak and Vaishali Gaikwad "Cancer Detection & Prediction Using Dual Hybrid Algorithm" International Journal of Recent Scientific Research Vol. 8, Issue, 3, pp. 15777-15789, March, 2017.

9. Younus Ahmad Malla, Mohammad Ubaidullah Bokari "A Machine Learning Approach for Early

Prediction of Breast Cancer” International Journal Of Engineering And Computer Science ISSN:2319-7242 Volume 6 Issue 5 May 2017, Page No. 21371-21377

10. V. Kirubha, S. Manju Priya “Comparison of Classification Algorithms in Lung Cancer Risk Factor Analysis” International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064 Index Copernicus Value (2015): 78.96 | Impact Factor (2015): 6.391

11. Samuel Giftson Durai, S. Hari Ganesh and A. Joy Christy “Prediction of Breast Cancer through Classification Algorithms: A Survey” IJCTA,9(27),2016,pp.359-365© International Science Press.

12. B. M. Gayathri, C. P. Sumathi “An Automated Technique using Gaussian Naïve Bayes Classifier to Classify Breast Cancer” International Journal of Computer Applications (0975 – 8887) Volume 148 – No.6, August 2016.

13. P. Saranya, B. Satheskumar “A Survey on Feature Selection of Cancer Disease Using Data Mining Techniques” International Journal of Computer Science and Mobile Computing, Vol.5 Issue.5, May- 2016, pg. 713-719.

14. Animesh Hazra, Subrata Kumar Mandal, Amit Gupta “Study and Analysis of Breast Cancer Cell Detection using Naïve Bayes, SVM and Ensemble Algorithms” International Journal of Computer Applications (0975 – 8887) Volume 145 – No.2, July 2016.

15. Keerti Yeulkar, Dr. Rahila Sheikh “Utilization of Data Mining Techniques for Analysis of Breast Cancer Dataset Using R” www.ijraset.com, Volume 5 Issue III, March 2017 IC Value: 45.98 ISSN: 2321-9653.

16. Monika Hedawoo, Abhinandan Jaiswal, Nishita Mehta “Comparison of Data Mining Algorithms For Mammogram Classification” International Journal Of Electrical, Electronics And Data Communication, ISSN: 2320-2084 Volume-4, Issue-7, Jul.-2016.

17. S. Kalaivani, S. Gandhimathi “An Efficient Bayes Classification Algorithm for analysis of Breast Cancer Dataset using Cross Validation Parameter” International Journal of Advanced Research in Computer Science and Software Engineering 5(10), October- 2015,

pp. 512-515.

18. Mandeep Rana, Pooja Chandorkar, Alishiba Dsouza, Nikahat Kazi “Breast Cancer Diagnosis and Recurrence Prediction Using Machine Learning Techniques” International Journal of Research in Engineering and Technology eISSN: 2319-1163 | pISSN: 2321-7308 Volume: 04 Issue: 04 | Apr-2015, <http://www.ijret.org>

19. N.V. Ramana Murty and Prof. M.S. Prasad Babu “A Critical Study of Classification Algorithms for Lung Cancer Disease Detection and Diagnosis” International Journal of Computational Intelligence Research ISSN 0973-1873 Volume 13, Number 5 (2017), pp. 1041-1048

Biography:



Mrs. Pallavi Mirajkar passed M.Sc. (Computer Science) from North Maharashtra University, Jalgaon. Currently she is pursuing her Ph.D. from Pacific University, Udaipur.



Dr. G. Prasanna Lakshmi, Guide.
Associate Professor, Andhra University.