

Automatic Tool for Prediction of Type of Cancer Risk and Recommendations

Pallavi Mirajkar¹, Dr. G. Prasanna Lakshmi², Dr. Ritu Khanna³

¹Research Scholar, Faculty of Computer Science, Pacific Academy of Higher Education and Research University, Udaipur, Rajasthan, India

²(WOS-A) Andhra University, AU North Campus, Visakhapatnam, Andhra Pradesh, India

³Professor & Head, Basic Science, Faculty of Engineering, Pacific University, Udaipur, Rajasthan, India

ABSTRACT

Cancer can begin in any part of the body and can spread to other parts also. It is uncontrollable and it has many types. In the proposed thesis research paper, a tool for prediction of type of cancer risk with five different cancer diagnosis and recommendations is presented. For recognizing cancer disease number of tests ought to be required from the patient. But using data mining techniques these test can be diminished. Indeed, an accurate prediction of cancer is very difficult task for medical practitioner and it is also high concern to the patients so that better treatment can be given and it will also increase the survival time of the patients. Our findings suggested that suitable prediction tool can effectively reduce the several tests for diagnosing cancer and prediction accuracy thereby increasing the technical possibility of early detection of cancer. The main features of the tool comprise a balance between the number of necessary inputs and prediction performance, being portable, and it empowers the automatic development of the cancer risk prediction tool in cancer disease.

Keywords : Prediction Tool, Cancer, Data Mining, Automation, Integration.

I. INTRODUCTION

The rising high-performance computing has benefited numerous disciplines in finding realistic solutions to their issues. Our health services are no special case to this. Data mining tools have been created for useful investigation of medical information, to help oncologist in improving determination for treatment purposes. In cancer disease research, data mining technique have played out a noteworthy role. Cancer disease categorization contributes the unsafe reason for the treatment of patients.

The aim of thesis work is to present an easy to use tool that provides predictions of cancer risk in patients or individual. Due to the internet facility available everywhere and the ease with which one is able to

consult with proposed prediction tool, we chose to develop cancer prediction tool as an online system.

Here the scope of the prediction tool is that integration of various risk factors that causes cancer, with computer-based patient records could reduce medical errors, enhance patient safety, improves the prediction of cancer risk in practice variation, and improve patient survival rate. The application is fed with different details to help medical practitioner to predict risk of type of cancer. The application permits user to share their health connected issues. It then processes user's particular details to determine for varied cancer disease that might be related to it. Here we tend to utilize some intelligent data mining techniques to figure the correct risk level of cancer that may be

identified with patient's details. Based on result, system automatically demonstrates the result to specific doctors for specific test/ treatment.

II. LITERATURE REVIEW

[Fan Wang et.al \(2018\)](#) worked on predictor to precisely predict six breast cancer cell types using gene expression. They identified 308 predictors of breast cancer cell types using machine learning. They found that mixture of protein coding genes and long non-coding RNAs are superior predictors compared to when the two sets of transcripts are treated independently. They have also shown that predictors constrained to a particular cell type provide as better prognostic markers for the relevant patient subtype. Author has stated that the breast cancer cell type predictors are also related with patient survivability and therefore have clinical significance.

[Subrata Kumar Mandal \(2017\)](#) have applied different techniques such as data cleaning, feature selection, feature extraction, data discretization and classification for predicting breast cancer as perfectly. They stated that Logistic Regression Classifier gives the maximum accuracy with reduced subset of features and time complexity of proposed algorithm is least compared to other two classifiers.

[R.Kaviarasi, Dr.A.Valarmathi \(2017\)](#) has applied two types of clustering. The hierarchical clustering is produced dendrogram results are produced using Euclidean distance and Ward.D linkage. The K-Means clustering are produced WSS values against number of cluster K values. The proposed method finally validate to the two type of clustering fit values. The validation measurement result is helped to the distance are measured in two clustering values. The results are helped to at the beginning of Non Small Cell Lung Cancer prevention through human way of life handled risk factors characteristics. The proposed paper work

is very helped to the cancer research center and hospitals to prevent the Non Small Cell Lung Cancer.

[Arpit Bansal et, al.\(2017\)](#) have proposed technique for a modification in K-Means Clustering Algorithm. In proposed modification, the K-Means clustering will vanish off the two most important drawbacks of K-Means clustering that are accuracy level and calculation time consumed in clustering the dataset. Although when they have also used small datasets these two factors accuracy level and calculation time may not matter much but when they used large datasets that have trillions of records, then little dispersion in accuracy level will matter a lot and can lead to a disastrous situation, if not handled properly, it can be stated that proposed modification can be more extended to achieve the full accuracy level up to 100%, with very little time and with more quality clusters.

[R.Senkamalavalli and Dr.T.Bhuvaneswari \(2017\)](#) have proposed novel algorithm was experimented on the Breast cancer database. It has been proved that proposed approach achieved a very high accuracy rate than the existing methods. They also demonstrated a certain level of accuracy in the classifier, and for finding accurate results there must be sufficient preprocessing of data done. They have also demonstrated accuracy in diagnosing breast cancer disease using the K-means classifier, adaboost and Support Vector Machines.

[Sumalatha.G, Archana.S \(2017\)](#) has analyzed cancer patient data using ZeroR method and J48 algorithm of data mining techniques. They stated that proposed cancer risk prediction system may give easy and a cost effective approach for screening cancer and may play a significant role in earlier diagnosis process for different types of cancer. Proposed system can also be used as a source of record with detailed patient history in hospitals as well as help doctors to premeditate on particular therapy for any patient.

Megha Rathi et.al. (2016) have developed a software tool for the prediction of disease which helps in decision making for the treatment method. Proposed tool will be helpful in diagnosing the disease type and to help out for decision support in medical system. They studied hybrid classification technique is used for classification of medical data sets and is applicable in healthcare domain. SVM and bootstrap is integrated to improve classification accuracy. Author expresses that prediction tool will help doctors or patient to decide in a short time whether the person is suffering from disease and is generic to all types of disease.

Prabhakar Chalise et.al.,(2014) have proposed Cluster analysis aims to highlight meaningful patterns or groups inherent in the data that will be useful in identifying the subtypes of the diseases. Different types of clustering algorithms have been proposed that use various assays of molecular variation of cells most of which are designed for one type of data at a time. These types of methods have been successfully implemented in several disease classification studies.

Rahul Patil et.al (2016) has used various clustering techniques of data mining such as partition based method, hierarchical based clustering. They stated that hybrid method will help society to analyze and understand disease and their health status.

Tanu Minhas and Nancy Sehgal (2017) concluded that prediction analysis is an efficient technique for the complex data analysis. They have applied k-means clustering algorithm with Boltzmann algorithm to increase accuracy of data clustering. They stated that the performance of proposed algorithm is tested in MATLAB and it has been analyzed that accuracy is increased up to 20 percent.

genuine reason and complete treatment of cancer is not invented. Different tests are available for predicting cancer, however detecting cancer in earlier stage is troublesome, yet prior discovery of cancer disease is treatable. With the help of data mining we endeavor to predict the risk of cancer in earlier stage. We build up a system called the cancer prediction tool which predicts specific cancer risks. Specifically, Cancer prediction tool estimates the risk of the breast, ovarian, oral, stomach, and lung cancers by analyzing a number of user-provided genetic and non-genetic factors. The major objective of prediction tool to provide the earlier warning to the users, to make a precaution based on their risk status.

The proposed prediction tool assists the people to know their chance of cancer risk with ease and additionally, it helps the people to take the proper decision based on their cancer risk status. Once the user enters into the cancer prediction system, they need to answer the questionnaire, related to cancer causing factors and symptoms. After that the prediction system assigns the risk value based on the user responses. Once the risk value is predicted, the range of the risk can be computed by the prediction system. We have three levels of risk low level, medium level, and high level. The result can be given away to the user through data base. The above mentioned technique can be successfully applied to the data sets for any cancer (such as breast cancer, lung cancer, oral, stomach cancer), as it is effectively checked on the different types of cancer. Once the risk is calculated the data given by the user is stored in the data base. The result will be displayed to the user through the database.

III. PROPOSED WORK

Diagnosis of cancer is still challenging for the specialists in the field of medicine. Even now the

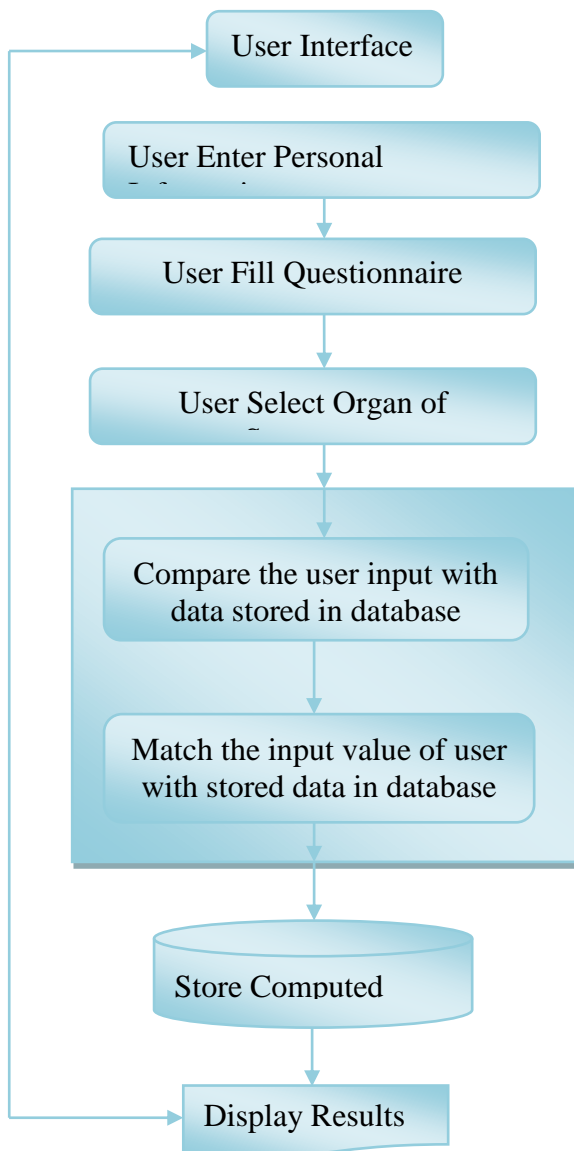


Fig. 1 Working of Cancer prediction system

IV. METHODS AND MATERIAL

4.1 Methodology

Collected data is prepared to fit into the developed tool appropriately. Here converse about the entire procedure (collection of data to significant pattern) of prediction of different types of cancer risk level.

4.2 Data Collection

Collection of data for the developed prediction tool has been done from different diagnostic centers which contain both male female data of different ages and from the various non-patients those who want to

check their risk of cancer. Collected data also contains both patient and non patient data. From the previous studies total 18 risk factors were considered for risk level cancer assessment, which includes- age, gender, hereditary, occupational hazard, smoking, chewing tobacco, alcohol, daily exercise, intake of vegetables, weight loss, weight gain, soft drink, contraceptive pills and various symptoms of different types of cancer.

4.3 Data Preprocessing

Sometimes it happens that collected data consists of repetition of information or more information of same person or missing any values of information. Therefore, data preprocessing is vital task which is necessary in data mining. The main objective of data pre-processing is making an appropriate analysis of collected data. Data preprocessing allows avoiding the double data and adds the missing values according to the need of tool. It also helps to reduce the memory and normalizes the values that are stored into the database.

4.4 System Architecture

The proposed cancer risk prediction tool is developed using various web technologies such as HTML, CSS and Java Script. It is suitable for several platforms and is distributable through an application store, just like other applications. The proposed tool is developed using Angular JS which wraps the HTML/JavaScript app into a native container which can access the device functions of multiple platforms. When the medical practitioner submits the values for multiple queries, the data is sent over the HTML POST method. The data is inserted into a Postgre SQL innodb database and, for each new registration. The proposed tool generates new XML file for each type of cancer. For the health care industry, cancer prediction applications yielded new boundaries in providing better care and treatments to patients. Furthermore, it is building a revolution in the way data is managed and made available [21]. The portability of proposed

prediction applications can enhance the productivity of health care professionals.

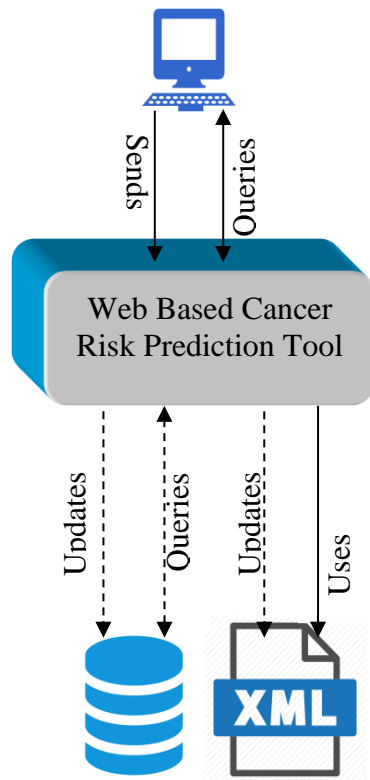


Fig. 2 Architecture of Cancer Risk Prediction Tool

Figure 2(a) shows the first screen that appears when the prediction tool is initiated. User has to login and then clicking on the *Test* Fig. 2, all tabs containing questionnaire available in the application become visible. User has to answer each question as shown in fig. 4, fig. 5 fig. 6, fig 7and fig. 8.

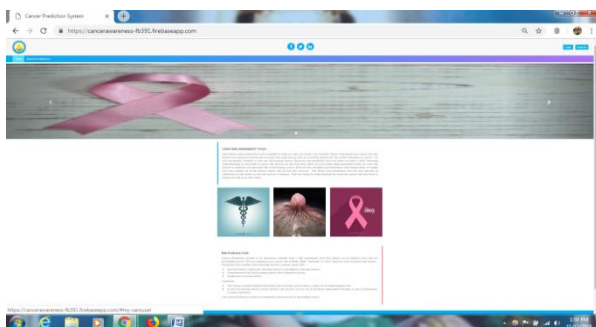


Fig. 3 Home Page

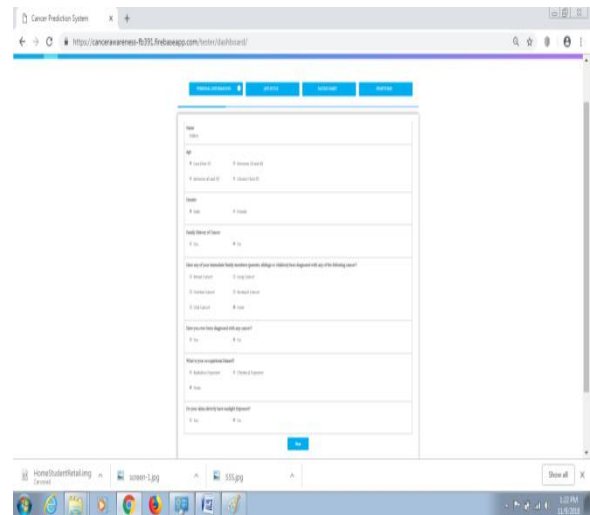


Fig.4 Personal Information Tab

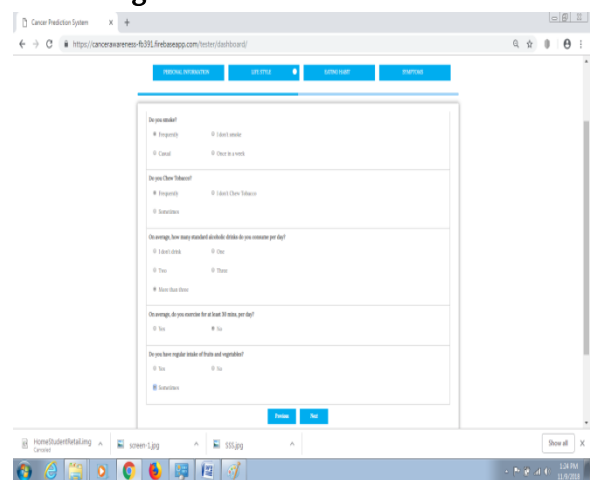


Fig. 5 Lifestyle Tab

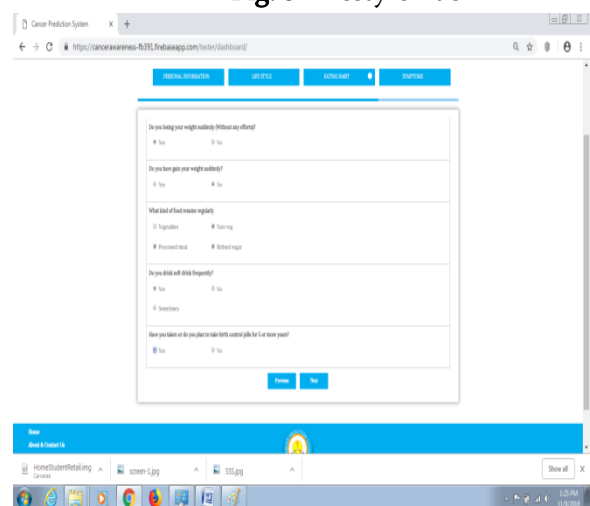


Fig.6 Eating Habit Tab

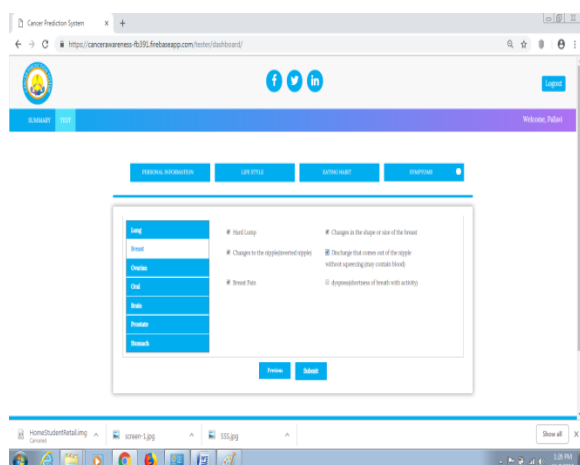


Fig. 7 Symptoms of different types of cancers

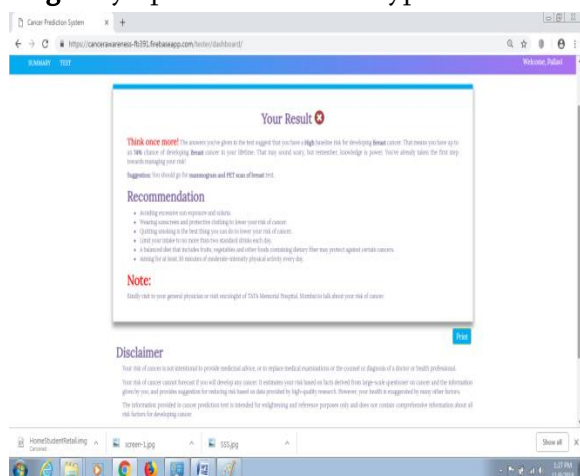


Fig. 8 Display of Result

V. RESULTS AND DISCUSSION

CANCER PREDICTION SYSTEM — POPULATION AND SAMPLE RESULTS

- To find out the performance of the Data Mining Based Cancer Prediction System among the users based on cancer prediction.
- To identify cancer risk level of the user using newly developed Data Mining Based Cancer Prediction System based on various risk factors.

To find the effectiveness of data mining based cancer prediction system, the proposed system was implemented on web based technology. Around 169 responses have been collected during Jun to Dec 2018. Details of the responses given in the table No. 1

Gender	Responses
Male	76
Female	93
Total	169

Table 1

VI. CONCLUSION

The thesis paper in which I developed cancer prediction system based on the online discourse of patient's information. It was shown that relying on cancer prediction system will give warning to the user about cancer risk level. Combining the different approaches boosts results, as content and interaction seem to model complementary aspects of patient discourse. The proposed prediction tool can be helpful to health industry who wants to diagnose patient online. The proposed system is web based, easy to use, versatile and reliable that can be implemented in remote areas like rural regions in Primary Health center, to mimic like human diagnostic expertise for finding the chances of having cancer. The main aim of the proposed prediction tool is to provide the earlier warning to the users, and it also saves cost and time of the user. It predicts specific cancer risks based on the symptoms found in the patient. Specifically, Cancer prediction system estimates the risk of the breast, Oral, Stomach, Ovarian, and Lung cancers by examining a numerous user-provided genetic and non-genetic factors. The proposed prediction system is easily accessible; people can easily check their risk and take appropriate action based on their risk status.

Future Scope on the prediction tool can include the development of different models with high end technologies to prevent cancer.

VII. REFERENCES

- [1]. Fan Wang, Zachariah Dohogne , Jin Yang , Yu Liu and Benjamin Soibam “Predictors of breast cancer cell types and their prognostic power in breast cancer patients” *BMC Genomics* (2018) 19:137 <https://doi.org/10.1186/s12864-018-4527-y>.
- [2]. Dave Smith, “Data Mining in the Clinical Research Environment”. Available at. <http://www.sas.com/>
- [3]. Damtew A., “Designing a predictive model for heart disease detection using data mining Techniques” A Thesis Submitted to the School of Graduate Studies of Addis Ababa University, 2011.
- [4]. <https://www.cancer.gov/>
- [5]. 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.
- [6]. R. Kaviarasi, Dr. A.Valarmathi “Near The Beginning of Non Small Cell Lung Cancer Avoidance in Human Way of Life Risk Factors Classification Using Clustering Algorithm in the R Environment” *International Journal of Advanced Research in Computer Science*, 8 (5), May-June 2017,1023-1026.
- [7]. Arpit Bansal, Mayur Sharma “Improved k-mean clustering algorithm for prediction analysis using classification technique in data mining”, *IJCA* (0975 – 8887) Vol. 157 – No 6, January 2017.
- [8]. Prabhakar Chalise, Devin C.Koestler, Milan Bimali, Qing Yu, Brooke L.Fridley to “Integrative clustering methods for high – dimensional molecular data”, *Transl Cancer Res* 2014.
- [9]. Ahmad LG, Eshlaghy AT, Poorebrahimi A, Ebrahimi M, Razavi AR (2013) “Using Three Machine Learning Techniques for Predicting Breast Cancer Recurrence”. *J Health Med Inform* 4: 124. doi:10.4172/2157-7420.1000124
- [10]. V.Kirubha , S.Manju Priya “Survey on Data Mining Algorithms in Disease Prediction” *International Journal of Computer Trends and Technology (IJCTT) – Volume 38 Number 3 - August 2016*.
- [11]. P. Saranya , B. Satheeskumar “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.
- [12]. R.Senkamalavalli and Dr.T.Bhuvaneswari “Improved Classification Of Breast Cancer Data Using Hybrid Techniques” *International Journal of Advanced Research in Computer Science* Volume 8, No. 8, September-October 2017 ISSN No. 0976-5697.
- [13]. Sumalatha.G, Archana.S “A Study on Early Prevention and Detection of Breast Cancer using Data Mining Techniques” *International Journal of Innovative Research in Computer and Communication Engineering (An ISO 3297: 2007 Certified Organization)* Website: www.ijircce.com Vol. 5, Issue 6, June 2017.
- [14]. Megha Rathi , Vikas Pareek “Disease prediction tool: an integrated hybrid data mining approach for healthcare” *IRACST - International Journal of Computer Science and Information Technology & Security (IJCSITS)*, ISSN: 2249-9555 Vol.6, No.6, Nov-Dec 2016.
- [15]. Rahul Patil, Pavan Chopade, Abhishek Mishra, Bhushan Sane, Yuvraj Sargar “Disease Prediction System using Data Mining Hybrid Approach” *Communications on Applied Electronics (CAE) – ISSN : 2394-4714* Foundation of Computer Science FCS, New York, USA Volume 4 – No.9, April 2016.

- [16]. Tanu Minhas , Nancy Sehgal “Prediction Analysis Technique using Hybrid Clustering and SVM Classification” International Journal of Innovative Research in Science, Engineering and Technology (An ISO 3297: 2007 Certified Organization) Website: www.ijirset.com Vol. 6, Issue 7, July 2017.
- [17]. K. Suneetha “Early Prediction and Detection of Lung Cancer using Data Mining” International Journal of Advanced in Management, Technology and Engineering Sciences Volume 7, Issue 12, 2017 ISSN NO : 2249-7455.
- [18]. Chih-Jen Tseng, Chi-Jie Lu, Chi-Chang Chang, Gin-Den Chen, Chalong Cheewakriangkrai “Integration of data mining classification techniques and ensemble learning to identify risk factors and diagnose ovarian cancer recurrence” Contents lists available at Science Direct Artificial Intelligence in Medicine journal homepage: www.elsevier.com/locate/aiim/ Artificial Intelligence in Medicine 78 (2017) 47–54.
- [19]. Huang M-W, Chen C-W, Lin W-C, Ke S-W, Tsai C-F (2017) “ SVM and SVM Ensembles in Breast Cancer Prediction” PLoS ONE 12(1):e0161501.
doi:10.1371/journal.pone.0161501.
- [20]. Florije Ismaili , Luzana Shabani , Bujar Raufi , Jaumin Ajdari , Xhemal Zenuni “Enhancing breast cancer detection using data mining classification techniques” 2nd World Conference on Technology, Innovation and Entrepreneurship May 12- 14, 2017.
- [21]. Ana Silva, Tiago Oliveira, Vicente Julian, Jose Neves, and Paulo Novais “A Mobile and Evolving Tool to Predict Colorectal Cancer Survivability” Published by Springer International Publishing Switzerland 2016. All Rights Reserved L. Iliadis and I. Maglogiannis (Eds.): AIAI 2016, IFIP AICT 475, pp. 14–26, 2016.

Author’s :

Mrs. Pallavi Mirajkar passed M.Sc. (Computer Science) from North Maharashtra University, Jalgaon. Currently she is pursuing her Ph.D. from Pacific University, Udaipur, Rajasthan, India
Dr. G.Prasanna Lakshmi, Guide.

Dr. G. Prasanna Lakshmi : Women Scientist (WOS-A)
Andhra University, Andhra Pradesh, India

Dr. Ritu Khanna : Professor & Head, Basic Science,
Faculty of Engineering, Pacific University, Udaipur,
Rajasthan, India

Cite this article as :

Pallavi Mirajkar, Dr. G. Prasanna Lakshmi, Dr. Ritu Khanna, "Automatic Tool for Prediction of Type of Cancer Risk and Recommendations", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 5 Issue 1, pp. 01-08, January-February 2019.

Available at doi

: <https://doi.org/10.32628/CSEIT1838116>

Journal URL : <http://ijsrcseit.com/CSEIT1838116>