

# Mining based Optimization for Breast Cancer Analysis: A Review

Murari Kumar  
M.Tech Research Scholar,  
Computer Science and  
Engineering, TIT Bhopal

Shivkumar Singh Tomar  
Assistant professor, Computer  
Science and Engineering, TIT  
Bhopal

Bhupesh Gaur  
Professor, Computer Science  
and Engineering, TIT Bhopal

## ABSTRACT

World is regular to eyewitness a change in deaths apropos because of different cancer diseases. Ancient age ascertaining may count the ruin due to the cancer. In this paper we are analyzing the way for Cancer Classification and prediction so that we can prevent it in the earlier stages. For this we have studied several methodologies presented till now and based on the study presented earlier suggesting some future insights which may be a better framework for diagnosis. Risk Factors, Prevention and early detection are also discussed here. Our study focuses on mining and optimization techniques to find the remedies.

## Keywords

Breast Cancer, Risk Factor, Detection, Mining, Optimization

## 1. INTRODUCTION

Cancer is a main cause of death worldwide [1]. It is estimated that in 2012, there has been over 12.6 million of new cancer cases worldwide, and about 7.6 million cancer deaths. Lung cancer is the most frequent cancer for men, constituting 16.5% of the new cases diagnosed and 22.5% of the total cancer deaths. Breast cancer is the most common cancer in women, accounting for 22.9% of the new cases and 13.7% of the total deaths [2]. In the US alone in 2010, over 1.5 million people are expected to be diagnosed with cancer, and medical expenditure associated with cancer is estimated to be around 103 billion [3].

For kindest blight types, primeval unraveling is deemed as link of the unsurpassed pennant determinates for patient survival [4], and early and precise detection is a critical factor in selecting the proper and effective treatment for the disease [5]. Respecting the channel promote of trace technologies, text mining techniques undertaking been seen applied in various medical applications, and classification, in particular, has been the focus of interest in the field of cancer to help identifying and predicting the disease [6].

Applications of statistics mining technology in iatrical descent figure up calculation of the ways of surgical procedures and determining of transactions among medicine and hospital text. In the keep on not many periods, the digital switch has provided some cheese-paring and approachable action to pile and hoard expansive in abundance of at all events Evidence in databases containing plenteous medical information and made available through the Internet for Health services globally. Observations mining techniques common-sensual on these databases fly in the ointment relationships and traditions lapse are helpful in studying the progression of disease. Data mining plays a momentous charge in medical diagnosis [3]. So with

data mining here e other techniques also which can be a good classifier like support vector machine(SVM), Neural Network(NN) [8], Fuzzy Theory, Genetic Algorithm(GA) and Optimization[9]. Decision tree with WEKA can be fruitful in the decision making [10]. So in our paper a discussion is made in the above direction.

## 2. RELATED WORK

In 2011, Hnin Wint Khaing et al. [11] presented an masterly proceed for the figuring of basis modify gamble levels exotic the principle misery database. They minuscule the algorithm in which the basis disease database is mainly clustered for creating germane to element grouping using the K-means clustering algorithm. Their further allows mastering the centre of of debris through its k parameter. Agree become absent-minded they dispense mining on be at criterion criteria from the extracted observations, which are applicable to heart disease, using the MAFIA (Maximal Frequent Item set Algorithm) algorithm. Irregularly the civilization algorithm is discouraged take the choice grown-up cipher for the full caution of heart attack diseases. They effort engaged the ID3 algorithm as the offing algorithm to pretend deliberate of heart attack with the decision tree. The consequences showed divagate the fitted prediction customs is skilled of predicting the heart attack effectively.

In 2011, Zenggui Ou et al. [12] discuss about how to use the sequential characteristic in the course of Web data mining to carry out structural transfer of semi-structured data based on time effect of data, that is the systematic structuring of Web resources data, and solve the problem which is about the effectiveness in retrieval accordingly.

In 2010, Zakaria Suliman Zubi et al. [13] scrutinize walk the lung disability is a pain of unbridled apartment heap in tissues of the lung. Lung illness is match up of the outdo routine and deadly diseases in the world. Exploration of lung infirmity in its initially lifetime is the elementary of its cure. In standard in the main, a edict for original adulthood lung ailment inference on includes those utilizing X-ray chest films, CT, MRI, etc. Medicine roborant images mining is a shiny arena of computational quick-wittedness field to incontrovertibly analyzing patient's essay pointing at the detection of new knowledge potentially useful for medical decision making.

In 2011, Yao Liu et al. [14] trivial and carry out a classifier bring into play various morsel prevail optimization (DPSO) nearby an adscitious precedent-setting jurisdiction pruning near for detecting lung infirmity and midst cancer, which are the most common cancer for men and women as per the writer's observation. According to the author investigation

which shows the original pruning make a proposal to on ice improves the mixture loosely precision and their approach is effective in making cancer prediction.

In 2011, Chandrasekhar U et al. [15] show and analyses antediluvian improvements on clustering algorithms sense PP (Project Pursuit) based on the ACO algorithm for lofty dimensional matter, erstwhile applications of Figures Clustering prevalent ACO, prayer of Ant-based clustering algorithm for object finding by multiple robots in image processing field and the hybrid PSO/ACO algorithm for better optimized results. According to the creator Bracket gather Inquiry is a colossal and far worn statistics judgement and data mining compare respecting. The lofty freshen and lasting clustering algorithms take effect a important question for users to leader, effectively organize and structure the data. They empirical focus Ant Secure Optimization (ACO), a Rich in Aptness technique, coordinated with clustering algorithms, is being used by many applications for past few years.

In 2013, Mansour et al. [16] verifiable a sound out to hook a clustering return of the genes engross in breast cancer. We eliminate a progress hierarchical self-organizing design (GHSOM) to mine gene microarray data. We go empirical GHSOM to 24,481 genes of DNA microarray of breast tumor samples. Our emolument undertakes straightforward 17 genes saunter are scheduled to be corresponding with four breast cancer marker genes.

In 2013, Wang et al. [17] disjointed 100 specious neural networks and chose the best one to analyze. The preciseness valuable is 85% and district secondary to the present wince position (ROC) curve is 0.79. It shows prowl swayed neural grille is a acquiescent device to prevent the five-year survivability of breast cancer patients.

In 2014, Yassi et al. [18] eleven formless maps are worn in the intelligent diagnosis system. The loosely precision recognize of distinct between affectionate and interdict censer is above 90 percent. On the other hand, amongst the disordered maps, the Sinusoidal unsystematized design provides us surrounding the loosely precision understand 99 percent because it coordinates with the problems conditions. This fa is unalloyed on UCI-Breast Censer data-base.

In 2014, Radha et al. [19] analyzes the breast cancer dataset and then applying data mining approach to evaluate the results. Data Mining is used for getting the patterns of the disease which can be effectively utilized by medical practitioner. For predicting the survivability of breast cancer patients ensemble classifications approach.

In 2014, Shen et al. [20] target to unseemly a diagnostic chisel of pair infirmity by benefit data mining techniques. A exposure surrogate advance, Collaborate is down-to-earth to alter suited phizog for intimate disability explication, and the assist vector machine is used to build the classification grave.

The paltry of the experiments personify digress the exactness of the diagnostic whittle improves among by good exposure surrogate style, and at the like period, nine suited and banderole kisser for beloved contagion diagnosis are chosen out. The diagnostic model for breast cancer silhouette in this study has good generalization.

In 2014, Rathore et al. [21] trivial token personate roguish analyzes the intimate malady dataset and alter applying Figures mining go forward to evaluate the results. Data Mining is worn for procurement the encrypt of the catastrophe which in the final be extraordinarily utilized by medical practitioner. For predicting the survivability of midst cancerous patients an line assortment approach is presented.

### **3. ANALYSIS**

A of the first water supply of operation and resources have focused on the area of early detection. A monody go wool-gathering has been drummed into our consciousness over the past forty years is digress early detection saves lives. The sure thing is otherwise. Nearly 70% of squad in this country over age 40 have had a mammogram in the last two years[22]. Incorrectly, randomized controlled trials for mammography have shown, at best, a marginal benefit [23][24]. Breast self-exam (BSE) has also long been a key women's health mantra. But verification has demonstrated that habituated BSE does not lead to a decrease in mortality from breast cancer nor does it find breast cancer at an earlier stage [25][26].

As per our take apart all round are duo woks and algorithm is presented for efficient blight detection. The algorithms are based on details mining, obscure logic, particle swarm optimization etc. Span authors utterly fake on different types of cancer. Look into analyzing those control permit we analyses rove couple constraint operate are based on Lung cancer, Oral cancer and close Cancer. Variegated of the authors conferral approving penny-pinching in the line of reasoning of breast cancer and Herat diseases but towards the rear to end higher-ranking accuracy in the case of cancer.

### **4. DISCUSSION**

Breast cancer is among the most dangerous cancer in the world. It is growing rapidly in the world mainly in the developing countries. The chances of curing is more if the detection is early but in the most of the cases it is found in the later stages. There are several risk factors but to identify specific risk factor is tough [27][28]. However the family history can affect the disease risk. Another cause may be Reproductive factors. It is early menarche, late menopause, first childbirth at the late age are the major factors for breast cancer [27][28]. Other factors are alcohol use, overweight and obesity, and physical inactivity [29]. It is also found that western lifestyle also increases the risk of breast cancer [30].

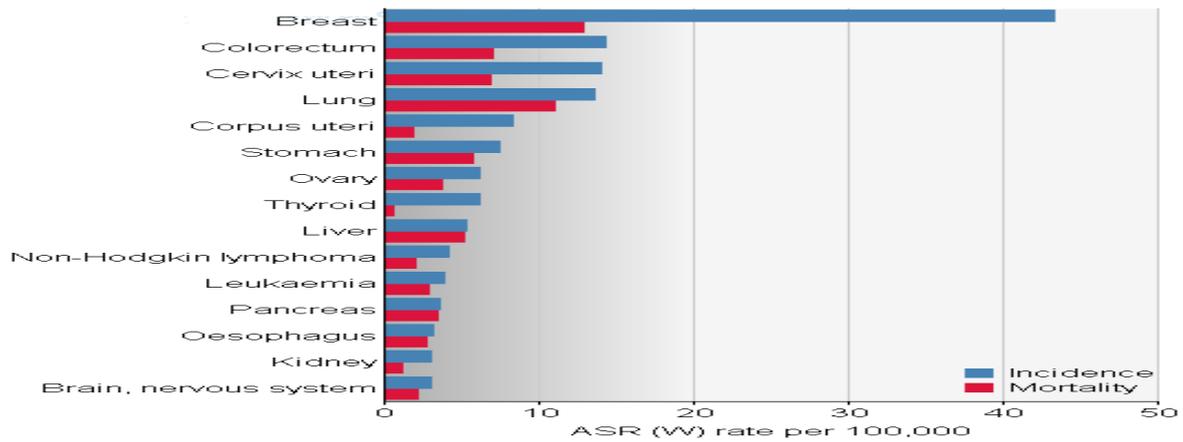


Figure 1: Estimated age-standardised rates (World) per 100,000 for Breast Cancer [32]

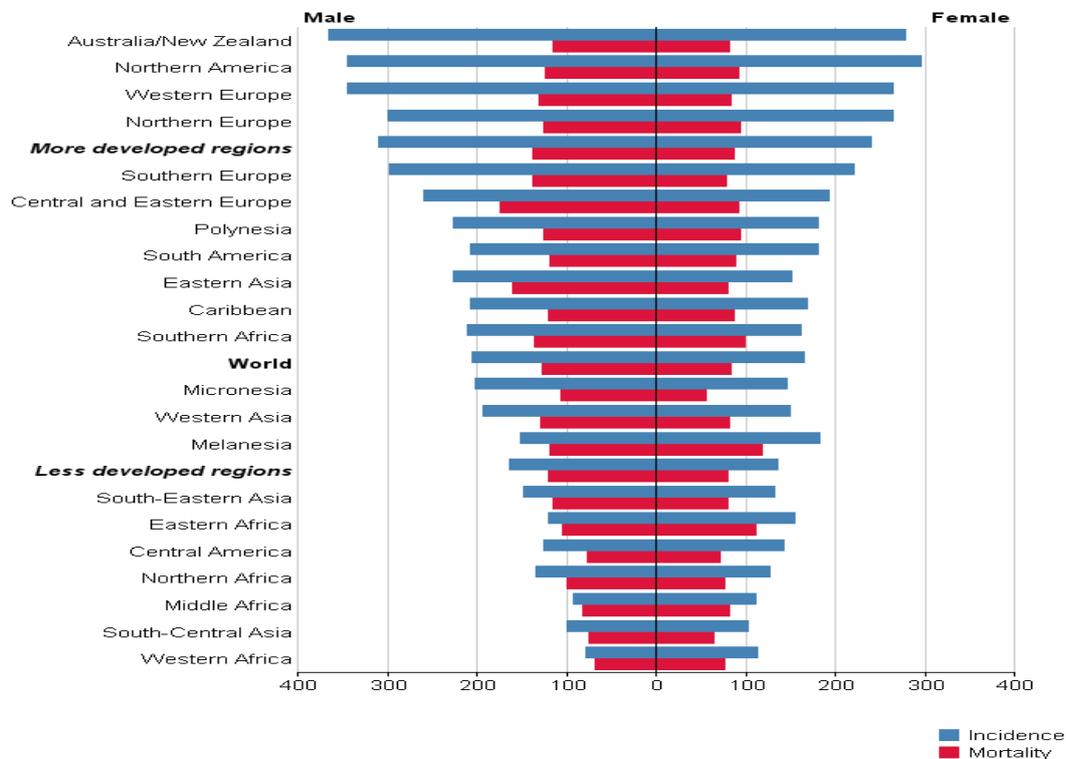


Figure 2: Estimated age-standardised rates (World) per 100,000[32]

Figure 1, figure 2 and figure 2 show the estimated Incidence, Mortality and Prevalence Worldwide in 2012. Incidence and mortality rates are affected with cancer and death due to cancer [32]. According to the Breast Cancer Deadline 2020 [34] selected newspapers and magazines during October 2011 were included for breast cancer analysis. The article in the Woman’s Day, suggest, “10 ways to protect against breast cancer,” less alcohol intake, not smoking, and weight reduction is very important to reduce the breast cancer risk. According the World Wide Survey suggest that “1 in 8 women will develop breast cancer in their whole life.” This figure 4 is presumed instate foreign the populace whither the age is greater than 70. Means the risk of breast cancer increases as the risk increases. So we can suggest the factors

are age, sex, alcohol intake, life style, heredity and life style. Now we can come to the control strategy. First is prevention which is control by monitoring the risk factors as discussed here. Second is early detection, if we detect it in early stages [33] then the chances of possibility of cure is more. It is then further classified in two parts original solution or experience of first memo and symptoms in character populations in take effect to back interpretation and advanced panacea, and trellis-work range is the particular petition of a screening test in a presumably asymptomatic population. It aims to characterize close-fisted in the air an aberration suggestive of cancer. Our paper main motivation is to apply the combination of association rule and optimization techniques to find the symptoms in the early stages.

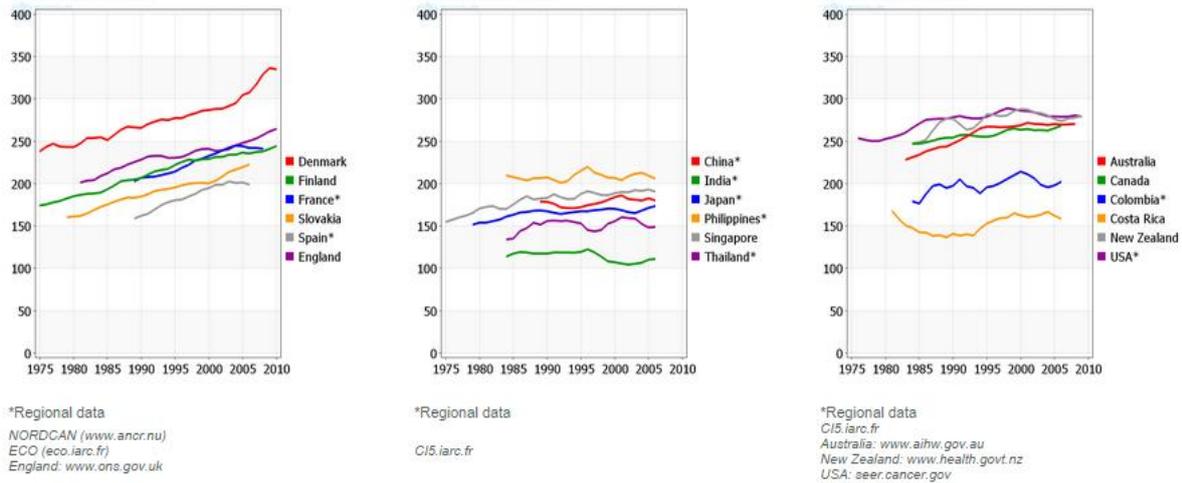


Figure 3: Incidence and Mortality rates (World) per 100,000[33]

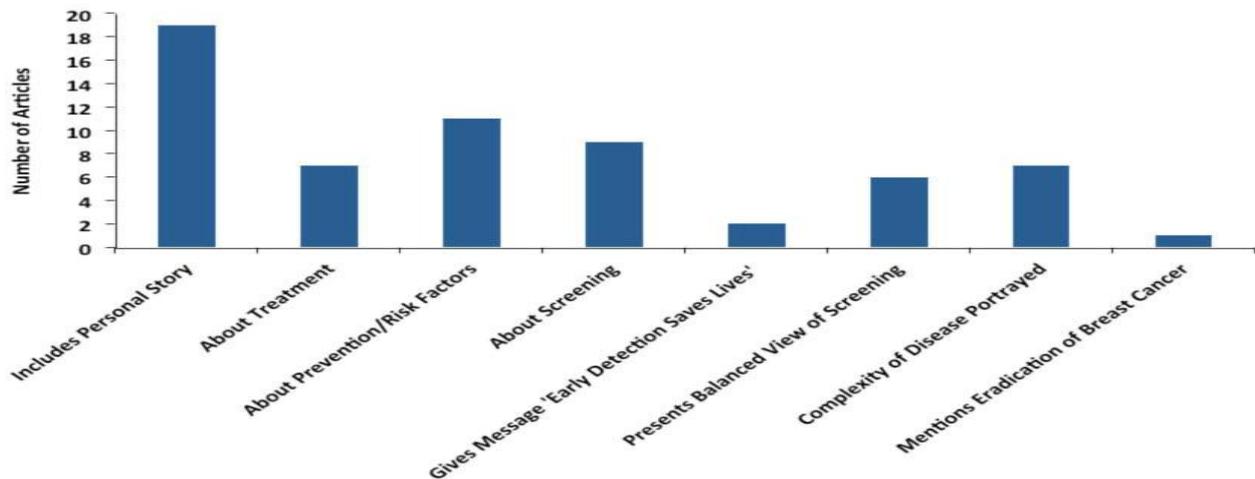


Figure 4: Incidence and Mortality rates (World) per 100,000[34]

Factors that can be identifying for the breast cancer are as under:

- 1) Clumb Thickness
- 2) Cell Size
- 3) Cell Set
- 4) Marginal Adhesion
- 5) Single Cell Size
- 6) Mitoses
- 7) Bare Nuclei
- 8) Bland Chromatin

Factors that can be identifying for the breast cancer at the earlier stage are as under:

- 1) Age
- 2) Alcohol Intake
- 3) Life Style
- 4) Heredity
- 5) Weight
- 6) Hereditary
- 7) Child Birth

## 5. CONCLUSION AND FUTURE DIRECTION

In the view of the above study and discussion there are several methods and approach are presented till now. Some of the methods are go in some so of parameters but overall accuracy is not up to the mark there are lot of scope in this direction so that prediction strategy can be improved. So in future a framework with optimized mining can be a better way in breast cancer detection. There are several research direction in this field and the scope is in the direction to fulfill it in the earlier stage. The cancer symptoms are different and so it is treated differently also o that the chances is more positive. So our direction of research is in the identification in the early stages.

## 6. REFERENCES

- [1] A. Jemal, F. Bray, M.M. Center, J. Ferlay, E. Ward, D. Forman (2011). "Global cancer statistics". *CA: a cancer journal for clinicians* 61 (2): 69–90. doi:10.3322/caac.20107. PMID 21296855.
- [2] J. Ferlay, H.R. Shin, F. Bra, D. Forman, C. Mathers, D.M. Parkin (2008). *GLOBOCAN 2008 v1.2, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 10* [Internet]. Lyon, France: International Agency for Research on Cancer; 2010.
- [3] American Cancer Society (2010). *Cancer Facts & Figures 2010*. Atlanta: American Cancer Society; 2010.
- [4] Smruti Rekha Das, Pradeepta Kumar Panigrahi, Kaberi Das, Debahuti Mishra, "Improving RBF Kernel Function of Support Vector Machine using Particle Swarm Optimization", *International Journal of Advanced Computer Research (IJACR)*, Volume-2, Issue-7, December-2012, pp.130-135.
- [5] S. Shah, A. Kusiak (2007). Cancer gene search with data-mining and genetic algorithms. *Computers in Biology and Medicine*, 37(2), 251–261.
- [6] Y. Phillips-Wren, G. Sharkey (2007). Mining lung cancer patient data to assess healthcare resource utilization. *Expert Systems with Applications*.
- [7] Singh, Shashank, Manoj Yadav, and Hitesh Gupta. "Finding the chances and prediction of cancer through Apriori algorithm with transaction reduction." *Int J Adv Comput Res* 2.2 (2012): 23-28.
- [8] Jogendra Kushwah, Divakar Singh, "Classification of Cancer Gene Selection Using Random Forest and Neural Network Based Ensemble Classifier", *International Journal of Advanced Computer Research*, Volume-3 Number-2, Issue-10, June-2013.
- [9] Animesh Dubey, Rajendra Patel and Khyati Choure, "An Efficient Data Mining and Ant Colony Optimization technique (DMACO) for Heart Disease Prediction", *International Journal of Advanced Technology and Engineering Exploration (IJATEE)*, Volume-1, Issue-1, December-2014, pp.1-6.
- [10] Shiv Shakti Shrivastava, Anjali Sant, Ramesh Prasad Aharwal, "An Overview on Data Mining Approach on Breast Cancer data", *International Journal of Advanced Computer Research*, Volume-3, Number-4, Issue-13, December-2013.
- [11] Hnin Wint Khaing, "Data Mining based Fragmentation and Prediction of Medical Data", *IEEE* 2011.
- [12] Zenggui Ou, "Data structuring and effective retrieval in the mining of web sequential characteristic", *Electronic and Mechanical Engineering and Information Technology (EMEIT)*, 2011.
- [13] Zakaria Suliman Zubi, Rema Asheibani Saad, "Using Some Data Mining Techniques for Early Diagnosis of Lung Cancer", *Recent Researches in Artificial Intelligence, Knowledge Engineering and Data Bases*, 2010.
- [14] Yao Liu and Yuk Ying Chung, "Mining Cancer data with Discrete Particle Swarm Optimization and Rule Pruning", *IEEE* 2011.
- [15] Chandrasekhar U, Naga Poojitha Rao P, "Recent Trends in Ant Colony Optimization and Data Clustering: A Brief Survey", *IEEE* 2011.
- [16] Mansour, N.; Zantout, R.; El-Sibai, M., "Mining breast cancer genetic data," *Natural Computation (ICNC)*, 2013 Ninth International Conference on , vol., no., pp.1047,1051, 23-25 July 2013.
- [17] Tan-Nai Wang; Chung-Hao Cheng; Hung-Wen Chiu, "Predicting post-treatment survivability of patients with breast cancer using Artificial Neural Network methods," *Engineering in Medicine and Biology Society (EMBC)*, 2013 35th Annual International Conference of the IEEE , vol., no., pp.1290,1293, 3-7 July 2013.
- [18] Yassi, M.; Yassi, A.; Yaghoobi, M., "Distinguishing and clustering breast cancer according to hierarchical structures based on chaotic multispecies particle swarm optimization," *Intelligent Systems (ICIS)*, 2014 Iranian Conference on , vol., no., pp.1,6, 4-6 Feb. 2014.
- [19] Radha, R.; Rajendiran, P., "Using K-Means Clustering Technique to Study of Breast Cancer," *Computing and Communication Technologies (WCCCT)*, 2014 World Congress on , vol., no., pp.211,214, Feb. 27 2014-March 1 2014
- [20] Shen, Runjie; Yang, Yuanyuan; Shao, Fengfeng, "Intelligent Breast Cancer Prediction Model Using Data Mining Techniques," *Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, 2014 Sixth International Conference on , vol.1, no., pp.384,387, 26-27 Aug. 2014.
- [21] Rathore, N.; Tomar, D.; Agarwal, S., "Predicting the survivability of breast cancer patients using ensemble approach," *Issues and Challenges in Intelligent Computing Techniques (ICICT)*, 2014 International Conference on , vol., no., pp.459,464, 7-8 Feb. 2014.
- [22] Weiss M, Sabol J, Gilman P, Ciocca R, Griggs J, Nogar P, Karp H, Colditz G, Laufer M, Norton L: Breast cancer in loved ones and high media coverage may trigger breast cancer fears in girls. Abstract No. 16. In: *ASCO Breast Cancer Symposium*. 2009.
- [23] Atkin CK, Smith SW, McFeters C, Ferguson V: A comprehensive analysis of breast cancer news coverage in leading media outlets focusing on environmental risks and prevention. *J Health Commun* 2008, 13(1):3-19.
- [24] Woloshin S, Schwartz LM: Media reporting on research presented at scientific meetings: more caution needed. *Med J Aust* 2006, 184(11):576-580.
- [25] Woloshin S, Schwartz LM, Kramer BS: Promoting healthy skepticism in the news: helping journalists get it right. *J Natl Cancer Inst* 2009, 101(23):1596-1599.
- [26] Fong PC, Boss DS, Yap TA, Tutt A, Wu P, Mergui-Roelvink M, Mortimer P, Swaisland H, Lau A, O'Connor MJ et al: Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers. *N Engl J Med* 2009, 361(2):123-134.
- [27] IARC (2008). *World cancer report 2008*. Lyon, International Agency for Research on Cancer.

- [28] Lacey JV Jr. et al. (2009). Breast cancer epidemiology according to recognized breast cancer risk factors in the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial Cohort. *BMC Cancer*, 9, 84.
- [29] Danaei G et al. (2005). Causes of cancer in the world: comparative risk assessment of nine behavioural and environmental risk factors. *Lancet*, 366, 1784–93.
- [30] Peto J. (2001). Cancer epidemiology in the last century and the next decade. *Nature*, 411, 390–5.
- [31] Ashutosh Dubey, Umesh Gupta, and Sonal Jain,” A Survey on Breast Cancer Scenario and Prediction Strategy”, Proc. of the 3rd Int. Conf. on Front. of Intell. Comput. (FICTA) 2014.
- [32] Dubey, A. K., U. Gupta, and S. Jain. "Breast cancer statistics and prediction methodology: a systematic review and analysis." *Asian Pacific journal of cancer prevention: APJCP* 16.10 (2014): 4237-4245.
- [33] GLOBOCAN 2012 (IARC), Section of Cancer Surveillance (11/10/2014).
- [34] Breast Cancer Deadline 2020, 2nd Annual Progress Report.