

A Survey on Latest Academic Thinking of Breast Cancer Prognosis

¹ Ravi Aavula and ²Dr. R. Bhramaramba

¹Associate Professor, Department of Computer Science and Engineering,
Guru Nanak Institutions Technical Campus, Hyderabad, Telangana State, India.

²Associate Professor Department of Information Technology,
GITAM, Visakhapatnam, Andhra Pradesh, India.

Abstract

Rapid growth of genomics and proteomics in biology has resulted in exponential growth of data which needs sophisticated computational analysis to discover intelligence. Research on computational biology or bioinformatics with application of data mining focuses on leveraging bioinformatics to address many real world problems in healthcare domain. Breast cancer is the second most lethal type of cancer causing death of woman. Many researchers contributed towards early detection, prognosis and better treatment of breast cancer in the last two decades causing decline of mortality rate. However, the breast cancer problem is still alarming and needs further research in the area of betterment of detection and prediction besides methods for treating it. In this paper, we reviewed the present state of the art of breast cancer prognosis which includes breast cancer diagnosis, breast cancer risk prediction and breast cancer survivability prediction. Different datasets were used by the researchers to evaluate their techniques. The results revealed that the presence of various algorithms, their advantages and limitations. It also provides insights on the mortality and incidence of breast cancer in different countries.

Keywords: Breast cancer, risk prediction, survivability prediction, recurrence prediction

INTRODUCTION

Bioinformatics is the application of computer technology for managing information pertaining to biology. DNA sequencing and mapping approaches come under bioinformatics. Paulien Hogeweg in 1979 coined the term bioinformatics for the study of genetics and genomics with large scale DNA sequencing. Bioinformatics can be understood as the science which deals with storing, extracting, organizing, analyzing, interpreting and utilizing information that is related to molecules and biological sequences. Of late there was much focus on this area as it helps humanity in many respects. For instance it can be used to understand biological processes and bring about dramatic changes in biology and healthcare domains.

Data mining is the process of analyzing huge amount of data in order to extract trends or patterns from the data. These trends or patterns are hidden otherwise. They can be interpreted to discover actionable knowledge or business intelligence (BI). The data mining is actively used in the area of bioinformatics. Data mining techniques are being

developed for solving biological problems. The biological data is in the form of large datasets that needs to be processed in order to have useful knowhow. Data mining is the branch of computer science that can be used with bioinformatics for analysing biological data such as protein structure, DNA sequences, gene classification and so on. The classification, clustering and other data mining methods can be exploited in combination with bioinformatics to leverage understanding of biological processes. The research areas in bioinformatics where data mining can be employed include sequence alignment, multiple sequence alignment, gene finding, protein domain analysis, pattern identification, genomic analysis, and motif finding.

According to WHO, breast cancer is the second lethal cancer which needs to be addressed in the real world. It has 14.1% of all cancer cases in the world. It is really alarming that women are killed with breast cancer across the globe. The technological advances, fortunately, and innovative treatment survival rate of women who are diagnosed breast cancer is more. Nevertheless, breast cancer is still the problem to be addressed efficiently. In this research the focus is to investigate on data mining techniques for bioinformatics in order to bring about efficient prognosis method. This paper throws light into the review of data mining techniques used in breast cancer prognosis, breast cancer risk assessment, breast cancer diagnosis and breast cancer survival prediction.

DATA MINING TECHNIQUES USED FOR BREAST CANCER RESEARCH

Kulkarni and Bhagwat [5] explored data mining techniques to investigate breast cancer recurrence. Different data mining techniques are evaluated. The classifiers they evaluated include BayesNet classifier, Naive Bayesian Multinomial Text classifier, IBM classifier, Naive Bayes classifier, KStar, LWL classifier, jRip classifier, OneR, InputMapped classifier, Decision Table classifier, PART classifier, ZeroR classifier, LMT classifier, Decision Stump classifier, Random Forest, J48, RandomTree, REPTree, KNN-Hamming Distance classifier. Out of them maximum accuracy is provided by jRip and KNN-Hamming Distance. Tomar and Agarwal [8] reviewed data mining techniques used for healthcare data analysis. They include classification, K-Nearest Neighbour, Decision Tree (DT), Support Vector Machine (SVM), Neural Network (NN), Bayesian Methods, Regression, clustering, partitioned clustering, hierarchical clustering; density based

clustering, and association rule mining.

Ahmed and Hasan [11] proposed a hybrid model to detect breast cancer at initial stages of the disease. They used UCI breast cancer dataset to evaluate their model. They generated rules using J48 algorithm and then subjected them to adjustable autonomy coordinator which resulted in the predication of breast cancer. Rani et al. [14] focused on the investigation of the importance of feature extraction process for breast cancer classification. They found that feature extraction is very useful for breast cancer classification. Shrivatsava and Sant [16] used data mining technique such as feature extraction for detection of breast cancer. Khalkhaji et al. [19] focused on finding the survival dynamic of breast cancer patients using cohort study. They used data mining technique known as Classification and Regression Trees (CART) to extract latent features and understand the survival of breast cancer patients. Senturk and Kara [20] used data mining techniques for breast cancer diagnosis. They used techniques like discriminant analysis, Artificial Neural Network (ANN), DT, Logistic Regression, SVM, Naive Bayes, and K-Nearest Neighbourhood.

Abreu et al. [29] proposed a methodology for breast cancer recurrence. The methodology involves data mining techniques

such as DT, Naive Bayes, Logistic Regression, K-Means, Bagging, Boosting, Linear Discriminant Analysis, SVM, KNN, association rule learning, isotonic separation, random forests, neural networks, self organized maps (SOM), classification restricted Boltzmann machines, genetic algorithms, sampling strategies, and evaluation metrics. Bochare et al. [33] explored supervised machine learning approach for risk prediction of breast cancer. They used classification algorithms like DT, Naive Bayes, Bayesian Networks, and SVM. Cheng et al. [42] proposed a methodology based on data mining using Cancer Genome Atlas (CGA). It was meant for improving the area in healthcare such as precision cancer medicine. Kumar et al. [55] explored the relation between obesity and breast cancer. Interestingly they found that obesity is associated with breast cancer. They used PubMed dataset and found a gene which has links with breast cancer. Ramadevi et al. [60] evaluated different data mining algorithms used for breast cancer research. They used resampling approach for evaluation. The algorithms investigated are random forest, logistic regression, C4.5, SVM, and KNN. They found that KNN and random forest are the two techniques preferred for breast cancer diagnosis. Chaurasia and Pal [67] used data mining techniques to find breast cancer survivability. They found that simple logistic classification provided better performance.

SUMMARY OF DATA MINING TECHNIQUES

Table 1: Summary of data mining techniques used for breast cancer research

Author and Year	Technique Used	Advantages	Limitations	Remarks
Siddhant Kulkarni, 2015 [5]	Data Mining Techniques like DT, Naïve Bayess, LWL etc.	Prediction of breast cancer recurrence	More methods need to be investigated	UCI breast cancer dataset is used.
Divya Tomar and Sonali Agarwal, 2013 [8]	Classification, KNN, DT, SVM, NN etc.	Healthcare services.	Standardization of approaches needs to be made.	It is a survey of data mining approaches
Al-Imran Ahmed, 2014 [11]	J48 algorithm and adjustable autonomy coordinator.	Detection of breast cancer	User and autonomous agents' role is yet to be investigated	UCI breast cancer dataset is used.
K. Usha Rani, 2015 [14]	Feature extraction with SVM, KNN, LR etc.	Classification of breast cancer	Classification accuracy is to be improved further	WDBC and WPBC datasets are used.
Shiv Shakti Shrivastava, 2014 [16]	General data mining techniques	Prediction of cancer	Prototype needs further improvement	UCI breast cancer dataset is used.
Hamid Reza Khalkhali, 2016 [19]	Classification and Regression Tree (CART)	Prediction of breast cancer survival	Limited dataset size.	Cohort study
Zehra Karapinar Senturk, 2014 [20]	Discriminant analysis, ANN, DT, LR, SVM etc.	Breast cancer diagnosis	Combination of techniques can be explored.	RapidMiner tool is used.
PEDRO HENRIQUES ABREU, 2016 [29]	Machine learning methods including SOM, genetic algorithms etc.	Prediction of breast cancer recurrence.	Standardization issues are to be investigated	UCI breast cancer dataset is used.
Aniket Bochare, Aryya Gangopadhyay, 2014 [33]	Integration of domain knowhow in Supervised machine learning methods like SVM.	Prediction of breast cancer risk.	-	dbGap dataset is used.
Phil Cheng, 2015 [42]	Data mining on caner genome atlas	Precision cancer medicine	Integration analysis yet to be done.	TCGA dataset
Dr. Ashok Kumar, Priyanka Thakur, 2015 [55]	Text Mining	Finds relation between obesity and cancer.	Further research is needed for validation.	PubMed database is used.
G. NAGA RAMADEVI, Dr. K. USHA RANI, 2015 [60]	Resampling	Evaluation of classifiers	-	Election techniques on the same data sets.
Vikas Chaurasia, Saurabh Pal, 2014 [67]	Simple logistic, RBF network and RepTree.	Prediction of breast cancer survival.	Ensemble approach can be explored.	Data provided by UMC university.

Breast Cancer Diagnosis

Ahmed and Hasan [11] proposed a hybrid model to detect breast cancer at initial stages of the disease. They used UCI breast cancer dataset to evaluate their model. They generated rules using J48 algorithm and then subjected them to adjustable autonomy coordinator which resulted in the predication of breast cancer. Singh [12] employed rule based classification technique for breast cancer prediction. They proposed a Genetic Algorithm (GA) based classifier that is used to classify breast cancer data. Risk factors and symptoms are considered by the method and it finally classifies the items into safe classes and risky classes. Evaluation is made using true positives, true negatives, false positives and false negatives. Protein sequences data is used to predict breast cancer. Senturn and Kara [20] explored seven data mining algorithms for breast cancer prediction. They include discriminant analysis, ANN, DT, LR, SVM, Naive Bayes, and KNN. They used UCI dataset with RapidMinor tool. They found that SVM performed better than other algorithms.

Faradmal et al. [21] compared Logistic Regression and ANN for prediction of breast cancer. They found that ANN could have high prediction accuracy than LR. Devi and Indra Devi [27] used J48 classification algorithm and outlier detection algorithm for breast cancer diagnosis. Two datasets such as WBCD and WDBC are used to investigate the algorithms on

the desired purpose. They exhibited more than 99% accuracy in prediction. Dubey et al. [30] made a good review of prediction methodologies used for breast cancer diagnosis. They studied the breast cancer prediction with data of different countries. The highest death rate was reported in USA. Mathieson et al. [34] used reduction rules and optimization techniques in order to have better cancer diagnosis. They used the problem of k-feature set for modelling prediction of breast cancer.

Zheng et al. [36] explored a hybrid approach using data mining techniques like K-Means and SVM. The combined method was named as K-SVM. WDBC dataset is used for evaluating breast cancer diagnosis. Their methodology showed 97.38% accuracy. They used training data for SVM and performed classification. Before applying classification, they employed clustering algorithm as pre-processing technique. This could improve performance of the method. Roy et al. [40] employed Back Propagation Neural Network (BPNN) for breast cancer detection. It is a classification approach using data mining. WBCD dataset was used for prediction of breast cancer. Acharya et al. [50] used breast thermograms for automated identification of breast cancer uMising higher order spectra analysis. They compared different algorithms like ANN and SVM. They found that ANN outperformed SVM. Dubey et al. [61] made a survey of predication strategies for breast cancer diagnosis.

Table 2: Summary of breast cancer diagnosis

Author and Year	Technique Used	Advantages	Limitations	Remarks
Al-Imran Ahmed, 2014 [11]	Ensemble method with J48 and adjustable autonomy coordinator	Prediction of breast cancer.	Role of user and autonomous agent are to be investigated further	UCI breast cancer dataset was used.
Nagendra Kumar Singh, 2015 [12]	Rule based classification with GA	Breast cancer diagnosis	-	UCI machine learning dataset was used.
Zehra Karapinar Senturk, 2014 [20]	ANN, DT, LR, SVM, Naive Bayes, and KNN.	Breast cancer diagnosis.	Further research is needed to improve prototype.	RapidMiner tool was used with UCI breast cancer dataset.
Javad Faradmal, 2014 [21]	Logistic Regression and ANN	Prediction of breast cancer.	Ensemble methods need to be investigated	ANN showed high prediction accuracy.
RDelshi Howsalya Devi, Dr. M Indra Devi, 2016 [27]	J48 classification algorithm and outlier detection algorithm	Breast cancer diagnosis	Outlier detection needs to be improved further.	WBC and WDBC datasets were used.
Ashutosh Kumar Dubey, 2015 [30]	Review of prediction methodologies	Breast cancer prediction models.	Models need improvement	Microarray datasets were used.
Luke Mathieson, 2017 [34]	Reduction rules and optimization techniques	Better cancer diagnosis	Ensemble methods need to be explored.	Optimal feature sets are used.
Bichen Zheng, Sang Won Yoon, 2014 [36]	K-Means and SVM. The combined method was named as K-SVM.	Breast cancer prediction.	Large scale sparse dataset needs to be evaluated further	WDBC dataset was used.
Payel Roy, 2015 [40]	BPNN classification	Breast cancer prediction	BPNN classification needs to be evaluated with other algorithms.	WDBC dataset was used.
U. Rajendra Acharya, 2012 [50]	Breast thermographs with ANN and SVM.	Breast cancer diagnosis	A tool needs to be built based on breast thermography.	MIAS dataset was used.
Ashutosh Kumar Dubey, 2014 [61]	Breast cancer prediction strategies	Breast cancer prediction	Hybrid framework needs to be realized.	UCI breast cancer dataset

Breast Cancer Recurrence Prediction

Kulkarni and Bhagwat [5] explored data mining techniques to investigate breast cancer recurrence. Different data mining techniques are evaluated. The classifiers they evaluated include BayesNet classifier, Naive Bayesian Multinomial Text classifier, IBM classifier, Naive Bayes classifier, KStar, LWL classifier, jRip classifier, OneR, InputMapped classifier, Decision Table classifier, PART classifier, ZeroR classifier, LMT classifier, Decision Stump classifier, Random Forest, J48, RandomTree, REPTree, KNN-Hamming Distance classifier. Out of them maximum accuracy is provided by jRip and KNN-Hamming Distance. Abreu et al. [29] proposed a methodology for breast cancer recurrence. The methodology involves data mining techniques such as DT, Naive Bayes, Logistic Regression, K-Means, Bagging, Boosting, Linear Discriminant Analysis, SVM, KNN, association rule learning,

isotonic separation, random forests, neural networks, self organized maps (SOM), classification restricted Boltzmann machines, genetic algorithms, sampling strategies, and evaluation metrics.

Ahmed et al. [49] used three machine learning techniques for prediction of breast cancer recurrence. They are known as C4.5, ANN and SVM. SVM showed high classification accuracy and least error rate. Pritom et al. [59] used UCI breast cancer dataset for experimentation. They algorithms they employed include SVM, DT, C4.5, SMO and Naive Bayes. These algorithms are evaluated for their prediction accuracy in recurrence prediction. The evaluation criteria used are root relative squared error (RRSE), relative absolute error (RAE), root mean squared error (RMSE), mean absolute error (MAE) and Kappa Statistic (KS). SMO exhibited best cancer recurrence prediction capabilities.

Table 3: Breast cancer recurrence prediction summary

Author and Year	Technique Used	Advantages	Limitations	Remarks
Siddhant Kulkarni, 2015 [5]	BayesNet classifier, Naive Bayesian Multinomial Text classifier etc.	Breast cancer recurrence prediction	Other machine learning techniques are to be explored.	UCI breast cancer dataset was used.
PEDRO HENRIQUES ABREU, 2016 [29]	DT, Naive Bayes, Logistic Regression, K-Means, Bagging etc.	Recurrence prediction of breast cancer	Non standardization needs to be investigated in future.	UCI breast cancer and SEER datasets are used
Ahmad LG, 2013 [49]	C4.5, ANN and SVM	Breast cancer recurrence prediction.	Further research is needed to improve performance.	ICBC dataset was used.
Ahmed Iqbal Pritom, 2016 [59]	Feature selection and classification	Recurrent prediction of breast cancer	New algorithms with better feature selection are to be defined.	UCI breast cancer dataset was used.

Table 4: Breast cancer risk prediction summary

Author and Year	Technique Used	Advantages	Limitations	Remarks
Al-Imran Ahmed, 2014 [11]	A hybrid approach with J48 algorithm and adjustable autonomy coordinator.	Breast cancer risk prediction.	The role of user and autonomous agent are to be investigated further.	UCI breast cancer dataset.
Aniket Bochara, Aryya Gangopadhyay, 2014 [33]	DT, Naive Bayes, Bayesian Networks, and SVM	Breast cancer risk prediction.	Needs to evaluate ensemble methods	NHS dataset was used.
Anothaisinatawee et al., 2012 [71]	Model performance comparison review.	Different models for breast cancer risk prediction.	Further research is needed to improve models.	A good review of risk prediction models.

Breast Cancer Risk Prediction

Ahmed and Hasan [11] proposed a hybrid model to detect breast cancer at initial stages of the disease. They used UCI breast cancer dataset to evaluate their model. They generated rules using J48 algorithm and then subjected them to adjustable autonomy coordinator which resulted in the prediction of breast cancer. Bochara et al. [33] explored supervised machine learning approach for risk prediction of breast cancer. They used classification algorithms like DT, Naive Bayes, Bayesian Networks, and SVM. They also used ten-fold cross validation for measuring performance of the methods. Besides they used ROC plots for graphically showing the performance of algorithms. Sensitivity and

specificity are the two statistical measures used to evaluate performance of binary classifiers. Sensitivity shows proportion of positive instances that have been classified correctly while the specificity shows the proportion of negative instances that have been classified correctly. With classification, feature selection approach is used in order to have better classification performance. More accuracy is exhibited by LibSVM which is a variant of SVM. Anothaisinatawee et al. [71] explored PubMed dataset for exploring risk prediction techniques pertaining to breast cancer. They opined that there was need for ensemble method that combined multiple models in order to have best risk prediction capabilities.

Breast Cancer Survivability Prediction

Khalkhaji et al. [19] focused on finding the survival dynamic of breast cancer patients using cohort study. They used data mining technique known as Classification and Regression Trees (CART) to extract latent features and understand the survival of breast cancer patients. Safe et al. [23] studied two techniques for breast cancer survival prediction. The techniques used here include Random Survival Forest, and model based recursive partitioning. They found the use of complex models that caused loss of interpretability. Huang [26] proposed a methodology for web based TCGA data analysis and visualization for identification of gene expression patterns. They used tools such as genomic tools and TCGA datasets in order to have prediction of breast cancer survivability. Kim and Shin [37] focused on survivability prediction of breast cancer. They used different kinds of data such as labelled, unlabelled and pseudo-labelled data with a semi-supervised learning approach known as SSL co-training. It was a graph based approach for prediction of breast cancer survivability.

Garcia-Laencina et al. [43] proposed a method for 5 year survival prediction. They used methods like expectation maximization, KNN, SVM, logistic regression, and SVM. They found best results with KNN. Rathore et al. [44] followed an ensemble approach for prediction of survivability of breast cancer patients. They used SEER dataset. Ensemble method is the method which combines multiple models in

order to have better results in prediction. They used classification algorithms like DT, Naive Bayes, and association based classification. They proposed a methodology with voting strategy that compares performance of different algorithms. Risk mapping is performed with predicted classes. The risk classes include low risk, medium risk and high risk.

Abreu et al. [57] also used ensemble method for predication of breast cancer survivability. They used three ensemble methods such as Subspace, LPBoost, and TreeBagger besides k-NN algorithm for Brest cancer survival prediction. Five steps are followed in the survivability prediction. They include data collection, database creation, finding and getting rid of missing data, model training, and data classification. Chaurasia and Pal [67] used data mining techniques to find breast cancer survivability. They found that simple logistic classification provided better performance. Jordan et al. [69] proposed another prediction technique for breast cancer survivability known as conformal prediction technique. They used SEER breast cancer dataset. They used data mining algorithms like SVM, DT, KNN and ANN. Their approach focuses on computing non-conformability score and using it for prediction of breast cancer survivability. Over all accuracy of DT was more when compared with KNN, SVM and ANN. Shin and Nam [70] proposed a coupling approach for breast cancer survivability prediction using SEER dataset and using semi-supervised learning co-training algorithm.

Table 5: Breast cancer survivability prediction summary

Author and Year	Technique Used	Advantages	Limitations	Remarks
Hamid Reza Khalkhali, 2016 [19]	CART	Breast cancer survivability prediction.	Its needs to be tested with large side datasets.	SEER dataset is used.
Mozhgan Safe, 2017 [23]	Random Survival Forest, and model based recursive partitioning	Breast cancer survivability prediction.	-	Iranian breast cancer dataset.
Zhenzhen Huang and Huilong Duan, 2015 [26]	TCGA data analysis and visualization	Prediction of breast cancer survival.	Further research needed to validate the approach.	TCGA dataset
Juhyeon Kim, Hyunjung Shin, 2013 [37]	semi-supervised learning approach known as SSL co-training.	Breast cancer survivability prediction.	Confidence of labeling needs to improved.	SEER dataset was used.
Pedro J. García-Laencina, 2015 [43]	Expectation maximization, KNN, SVM, logistic regression, and SVM	Breast cancer survival prediction	-	IPO breast cancer dataset
Neha Rathore, 2014 [44]	Ensemble method using DT, Naïve Bayes, and association based classification.	Prediction of survivability of breast cancer patients.	Prototype needs to be improved further	SEER dataset was used.
Pedro Henriques Abreu, 2013 [57]	Ensemble methods with Subspace, LPBoost, and TreeBagger besides k-NN algorithm	Better prediction of breast cancer survivability.	Further research is required to validate ensemble method	SEER dataset was used.
Vikas Chaurasia, Saurabh Pal, 2014 [67]	Simple logistic, RepTree and RBF network.	Breast cancer survivability	Accuracy of prediction needs to be improved further.	Dataset was collected from University Medical Centre of Yugoslavia.
Loai Alnemer, 2016 [69]	Conformal prediction technique with data mining techniques such as SVM, DT, KNN and ANN.	Suvivablity prediction of breast cancer patients.	-	SEER dataset was used.
Shin and Nam, 2014 [70]	Semi-supervised learning co-trainigi algorithm.	Prediction of survivability of breast cancer patients.	It is general and needs to be validated with further research.	SEER dataset was used.

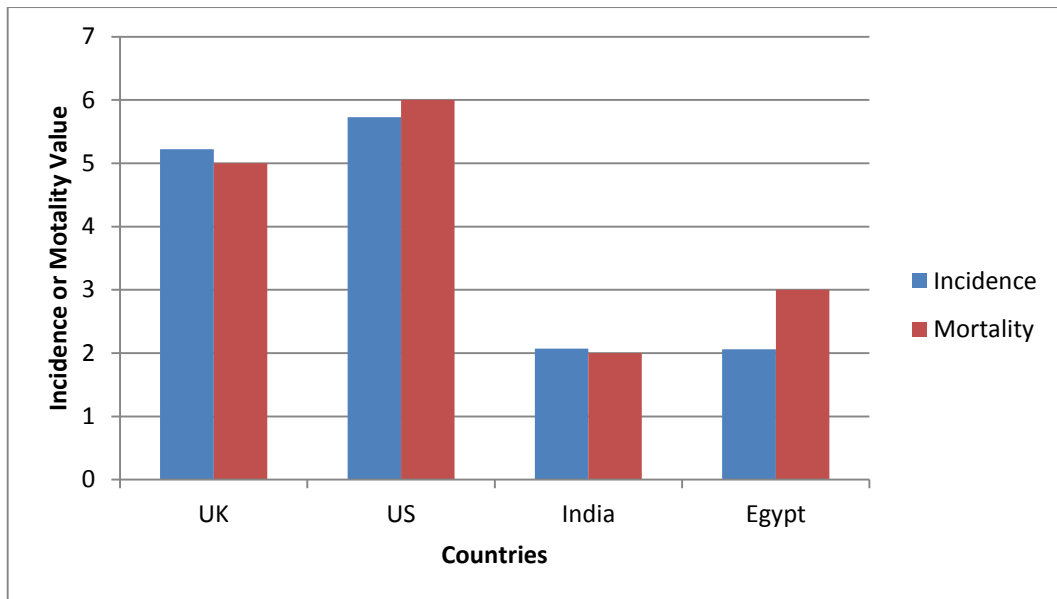


Figure 1: Mortality and incidence rates of breast cancer in different countries

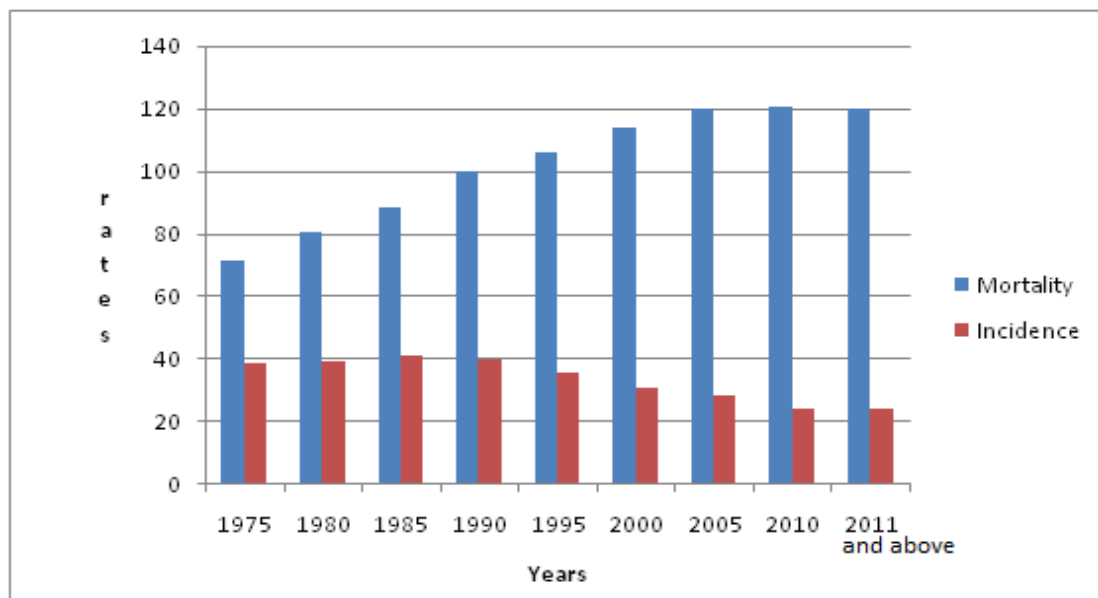


Figure 2: Mortality and incidence dynamics chronologically

As shown in Figure 1, it is evident that the incidence and mortality of breast cancer details are presented. USA showed high incidence and mortality while India shows least rate of mortality and incidence.

As shown in Figure 2, the breast cancer has been growing in different countries since 1975. The mortality is increased gradually while the incidence is decreased from 1995.

CONCLUSIONS AND FUTURE WORK

When cells in human body grow out of control, then we call it cancer. Cancer can start in any part of human body. When cancer is in place, the body cannot function as it is supposed

to. Cancer is not a single disease. In fact it is a collection of diseases as cancer can start in the colon, breast, lungs and even blood. Different cancers grow differently. The most common cancer type in woman is breast cancer. Breast cancer, as said earlier about cancer, starts when cell in woman's breast start growing abnormally. Thus a tumour is formed and that can be detected with a simple x-ray or it can be felt as a lump. The tumour is said to be malignant when it grows into surrounding tissues and distant areas of human body. As breast cancer cases are still alarming in the world, this paper throws light into the present academic thinking of breast cancer prognosis methods and their advantages and limitations. It provides insights into different methods of prognosis including diagnosis, risk prediction and survival

prediction. From this survey it is understood that a comprehensive framework is still desired with placeholders for different techniques in order to have breast cancer prognosis in a better way. This is the research on which we focus in future.

REFERENCES

- [1]. Alfonso Gonz´alez-Briones, Juan Ramos, Juan F. De Paz and Juan M. Corchado. (2015). Multi-agent System for Obtaining Relevant Genes in Expression Analysis between Young and Older Women with Triple Negative Breast Cancer. *Journal of Integrative Bioinformatics*, p1-15.
- [2]]DiptaSengupta,MoonmoonDeb , SandipKumarRath,SwayamsiddhaKar, Sabnam Parbin,NibeditaPradhan AND SamirKumarPatra . (2016). DNAmethylationandnotH3K4trimethylationdictatest heexpression status ofmiR-152genewhichinhibitmigrationofbreastcancercells via DNMT1/CDH1loop. *elsevier*, p176-182.
- [3] Chun-Ting Cheng^{1,2}, Ching-Ying Kuo¹, Ching Ouyang³, Chien-Feng Li^{4,5}, Yiying Chung¹, David C. Chan⁶, Hsing-Jien Kung and David K. Ann., (2016). Metabolic stress-induced phosphorylation of KAP1 Ser473 blocks mitochondrial fusion in breast cancer cells, P1-36.
- [4] Lifenge ZHOU AND Hong Wang. (2016). Chun-Ting Cheng^{1,2}, Ching-Ying Kuo¹, Ching Ouyang³, Chien-Feng Li^{4,5}, Yiying Chung¹, David C. Chan⁶, Hsing-Jien Kung and David K. Ann., (2016). Metabolic stress-induced phosphorylation of KAP1 Ser473 , P1-16.
- [5] Siddhant Kulkarni and Mangesh Bhagwat. (2015). Predicting Breast Cancer Recurrence using Data Mining Techniques. *International Journal of Computer Applications*. 122 , p1-6.
- [6] Jie Liu, MS, David Page, PhD, Peggy Peissig, PhD, Catherine McCarty, PhD³, Adedayo A. Onitilo, MD, MSCR, FACP and Amy Trentham-Dietz, PhD (2016) New Genetic Variants Improve Personalized Breast Cancer Diagnosis, p1-7.
- [7] Laura Lehtinen, Kirsi Ketola, Rami Mäkelä, John-Patrick Mpindi, Miro Viitala, Olli Kallioniemi and Kristiina Iljin. (2012)High-throughput RNAi screening for novel modulators of vimentin expression identifies MTHFD2 as a regulator of breast cancer cell migration and invasion.4, p1-6.
- [8] Jun Fan and Yirong Wu. (2016). Laura Lehtinen, Kirsi Ketola, Rami Mäkelä, John-Patrick Mpindi, Miro Viitala, Olli Kallioniemi and Kristiina Iljin. (2012). High-throughput RNAi screening for novel modulators of vimentin expression i. *Journal of Machine Learning Research*, p1-15.
- [9] Divya Tomar and Sonali Agarwal. (2013). A survey on Data Mining approaches for Healthcare5 (.), p1-26.
- [10] Latha K.C., Madhu B., Ayesha S., Ramya R., Sridhar And Balasubramanian . (2013). Visualization Of Risk In Breast Cancer Using Fuzzy Logic In Matlab Environment, 4 , p1-4.
- [11] Al- Imran Ahmed and Md Mahmudul Hasa. (2014). A Hybrid Approach for Decision Making to Detect Breast Cancer using Data Mining and Autonomous Agent based on Human Agent Teamwork, p1-14
- [12] Nagendra Kumar SINGH. (2015). Prediction of Breast Cancer using Rule Based Classification. 37 , p1-12.
- [13] B.-L. ZHONG, L.-J. BIAN, G.-M. WANG, Y.-F. ZHOU, Y.-Y. CHEN. PENG. (2016). Identification of key genes involved in HER2-positive breast cancer 20 , p1-9.
- [14] K. Usha Rani. (2015). Importance of Feature Extraction for Classification of Breast Cancer Datasets – A Study, p1-7.
- [15]. Shomona Gracia Jacob , R. Geetha Ramani. (2012). Efficient Classifier for Classification of Prognostic Breast Cancer Data through Data Mining Techniques. *Proceedings of the World Congress on Engineering and Computer Science* ,1 p1-6.
- [16] Shiv Shakti Shrivastava, Dr. Anjali Sant. (2014). Data Mining Techniques in Cancer Research Area., p1-4.
- [17] Andra´s La´nczky ´m Nagy, Giulia Bottai Gyo¨ngyi Munka´csy, Andra´s Szabo Libero Santarpia³ and Bala´zs Gy}orffy. (201`6). miRpower: a web-tool to validate survival-associated miRNAs utilizing expression data from 2178 breast cancer patients, p1-8.
- [18] Samuel Branders. (2014). A generic cycling hypoxia-derived prognostic gene signature: Application to breast cancer profiling., p1-18.
- [19] Hamid Reza Khalkhali (PhD)^a, Hadi Lotfnezhad Afshar (PhD)^b, Omid Esnaashari (MD)^c, and Nasrollah Jabbari (PhD)^d. (2016). Applying Data Mining Techniques to Extract Hidden Patterns about Breast Cancer Survival in an Iranian Cohort Study, p1-5.
- [20] Zehra Karapinar Senturk and Resul Kara. (2013). BREAST CANCER DIAGNOSIS VIA DATA MINING: PERFORMANCE ANALYSIS OF SEVEN DIFFERENT ALGORITHMS⁴ , p1-12.
- [21] Javad Faradmali andAli Reza Soltanian. (2014). Comparison of the Performance of Log-logistic Regression and Artificial Neural Networks for Predicting Breast Cancer, p1-7.
- [22] Chao Liu§, Chi Zhang, Jing Su, Dongsheng

- Zhangand Sha Cao. (2013). Stresses drive a cancer's initiation, progression and metastasis: Critical comments on the book \Cancer Bioinformatics", p1-11.
- [23] Mozghan Safe, Hossein Mahjub & Javad Faradmal. (2017). A Comparative Study for Modelling the Survival of Breast Cancer Patients in the West of Iran 9 p-8.
- [24] Li He Yuelong Wang Yongning Yang, Liqiu Huang, and Zhining Wen. (2014). Identifying the Gene Signatures from Gene-Pathway Bipartite Network Guarantees the Robust Model Performance on Predicting the Cancer Prognosis, p1-111.
- [25] Haofan Yang and Yi-Ping Phoebe Chen . (2015). Expert Systems with Applications. elsevier, p1-3.
- [26] Zhenzhen Huang, Huilong Duan, and Haomin Li. (2015). Identification of Gene Expression Pattern Related to Breast Cancer Survival Using Integrated TCGA Datasets and Genomic Tools., p1-11.
- [27] Zhenzhen Huang, Huilong Duan, and Haomin Li. (2015). Identification of Gene Expression Pattern Related to Breast Cancer Survival Using Integrated TCGA Datasets and Genomic Tools., p1-11.
- [28] R Delshi Howsalya Devi1, Dr. M Indra Devi. (2016). OUTLIER DETECTION ALGORITHM COMBINED WITH DECISION TREE CLASSIFIER FOR EARLY DIAGNOSIS OF BREAST CANCER, p1-16.
- [29] Duoqiao Wu and Xiangdong Wang. (2015). Application of clinical bioinformatics in lung cancer-specific biomarkers , p1-8.
- [30] Pedro Henriques Abreu and Miriam Seoane Santos. (2016). Predicting Breast Cancer Recurrence Using Machine Learning Techniques: A Systematic Review. 49 , p1-41.
- [31] Lance D Miller, Jeff W Chou, Michael A Black, Cristin G Print, Eric Jimenez, Julia Chifman and Angela Alistar. (2014). Immune gene signatures and tumor intrinsic markers delineate novel immunogenic subtypes of breast cancer, p1-3.
- [32] Mitochondrial Ca²⁺ uniporter is critical for store-operated Ca²⁺ entry-dependent breast cancer cell migration. (2015). Biochemical and Biophysical Research Communications, p1-9.
- [33] Aniket Bocharé; Aryya Gangopadhyay (2014) Integrating domain knowledge in supervised machine learning to assess the risk of breast cancer .IJMEL,6,p87-99
- [34] Mathieson , Mendes (2017), Computer-Aided Breast Cancer Diagnosis with Optimal Feature Sets: Reduction Rules and Optimization Techniques. Methods Mol Biol.1526,p299-325
- [35] Hanaa M. Hussain, Khaled Benkrid, Huseyin Seker, Ahmet T. Erdogan. (2011). FPGA Implementation of K-means Algorithm for Bioinformatics Application: An Accelerated Approach to Clustering Microarray Data. IEEE. p1-8.
- [36] Huan Liu, Rudy Setiono, Zheng Zhao and Hiroshi Motoda. (2010). Feature Selection: An Ever Evolving Frontier in Data Mining. *Indian Journal of Computer Science and Engineering.* , p1-10.
- [37] Juhyeon Kim, Hyunjung Shin. (2013). Breast cancer survivability prediction using labeled, unlabeled, and pseudo-labeled patient data. *computer science0* p1-6.
- [38] Rashmi Nagpal, Rashmi Shrivastava (2015) Cancer Classification Using Elitism PSO Based Lezy IBK on Gene Expression Data, p1-5.
- [39] Judith Patricia Doherty and Lily I Huschtscha. (2014). STC1 expression is associated with tumor growth and metastasis in breast cancer, p1-14.
- [40] Sudeep Roy, AK Kukreja (2013) Application of Artificial Neural Networks for Predicting Maximum in vitro Shoot Biomass Production of Safed Musli, *scientificreports.1*, p464
- [41] Xinan (Holly) Yang. (2014). Computational prognostic indicators for breast cancer, p1-13.
- [42] Phil Cheng and Mitch Levesque. (2015). Data mining The Cancer Genome Atlas in the era of precision cancer medicine p1-6.
- [43] Pedro J. García-Laencina , Pedro Henriques Abreu , Miguel Henriques Abreu , Noémia Afonso. (2015) Missing data imputation on the 5-year survival prediction of breast cancer patients with unknown discrete values, *Computers in Biology and Medicine*, 59, p125-133
- [44] Neha Rathore, Divya and Sonali Agarwal. (2015). Predicting the Survivability of Breast Cancer Patients using Ensemble Approach, p1-7.
- [45] D. Lavanya and Dr. K. Usha Rani. (2012). ENSEMBLE DECISION TREE CLASSIFIER FOR BREAST CANCER DATA. *International Journal of Information* p1-8.
- [46] Hossein Ghayoumi Zadeh, Omid Pakdelazar, Javad Haddadnia, Gholamali Rezai-Rad, Mohammad Mohammad-Zadeh. (2012). Diagnosing Breast Cancer with the Aid of Fuzzy Logic Based on Data Mining of a Genetic Algorithm in Infrared Images. *Middle East Journal of Cancer* p1-11.
- [47] CASEY S. GREENE JIE TAN, MATTHEW UNG, JASON H. MOORE AND CHAO CHENG (2014) Big Data Bioinformatics, p1-5.
- [48] Fuduan Peng and Haihai Liang, Wenyuan Zhao and Chenguang Wang. (2015) Autophagy-related prognostic signature for breast cancer, p1-5.
- [49] Ahmad LG, Eshlaghy AT, Poorebrahimi A and Ebrahimi M and Razavi AR (2013) Using Three

- Machine Learning Techniques for Predicting Breast Cancer Recurrence, p1-3.
- [50] Muthu Rama Krishnan Mookiah, Acharya, U Rajendra ,(2012),Data mining technique for breast cancer detection in thermograms using hybrid feature extraction strategy, Quantitative InfraRed Thermography Journal9,p151-165.
- [51] Shweta Kharya . (2012). U SING D ATA M INING T ECHNIQUES FOR DIAGNOSIS AND PROGNOSIS OF CANCER DISEASE. International Journal of Computer Science. 2 (2), p1-12.
- [52] Dan Wang , Ping Lu , Hao Zhang , Minna Luo, Xin Zhang, Xiaofei Wei, Jiyue Gao, Zuowei Zhao, Caigang Liu. (2014). Oct-4 and Nanog promote the epithelial-mesenchymal transition of breast cancer stem cells and are associated with poor prognosis in breast cancer patients. Oncotarget. p1-13.
- [53] Yang Yang, Lu Chen, Xinyue Xu, Zhenlin Fan, Wenjing Li, Chunfeng Niu(2016)Identification of therapeutic targets of breast cancer, p1-10.
- [54] Andreas Holzinger. (2014). Extravaganza Tutorial on Hot Ideas for Interactive Knowledge Discovery and Data Mining in Biomedical Informatics. *Springer International Publishing Switzerland*. p1-15
- [55] Dr. Ashok Kumar,a, Priyanka Thakur,b, Kanika Gupta,c and Dr. Amit Pal,d(2015)Text Mining Approach to analyse the relation between obesity and breast cancer data, p1-10.
- [56] P. Mahanta, H. A. Ahmed,D. K. Bhattacharyya and Jugal K. Kalita. (2011). Triclustering in Gene Expression Data Analysis: A Selected Survey. *IEEE*. p1-6.
- [57] DSVGK Kaladhar, B. Chandana and P. Bharath Kumar. (2011). Predicting Cancer Survivability Using Classification Algorithms. *International Journal of Research and Reviews in Computer Science*. p1-5.
- [58] BRENDAN MONROE. (2014)Big data versus the big C, p1-2.
- [59] KHALID RAZA. (2010). APPLICATION OF DATA MINING IN BIOINFORMATICS. Indian Journal of Computer Science and Engineering. 1 (2), p1-5.
- [60] G. NAGA RAMADEVI, Dr. K. USHA RANI, Dr.D.LAVANYA. (2015). Evaluation of Classifiers Performance using Resampling on Breast cancer Data 6 , p1-8.
- [61] Ashutosh Kumar Dubeyand Umesh Gupta. (2014). A Survey on Breast Cancer Scenario and Prediction Strategy., p1-10.
- [62] Hyunjung Shin, Yonghyun Nam. (2014). A coupling approach of a predictor and a descriptor for breast cancer prognosis. Shin and Nam BMC Medical Genomics p1-12.
- [63] Jaree Thongkam,guandong xu,yanchun zhang and fuchun huang. (2016). Breast Cancer Survivability via AdaBoost Algorithms. *Indian Journal of Computer Science and Engineering*. p1-11.
- [64] Masahiro Sugimoto, Masato Kawakami, Martin Robert, Tomoyoshi Soga, and Masaru Tomita. (2012). Bioinformatics Tools for Mass Spectroscopy-Based Metabolomic Data Processing and Analysis. *Bentham Science Publishers* p1-13.
- [65] Clare Stirzaker,, Phillippa C. Taberlay, Aaron L. Statham, and Susan J. Clark. (2013). Mining cancer methylomes: prospects and challenges, p1-11.
- [66] J. Kacprzyk. (2010). Advances in Intelligent and Soft Computing. *Springer Science+Business Media*. p1-244.
- [67] Vikas Chaurasia and Saurabh Pal. (2014). Data Mining Techniques: To Predict and Resolve Breast Cancer Survivability 3 , p1-14.
- [68] Randall Matignon. (2007). Data Mining Using SAS Enterprise Miner. *United States of Amenca*. p1-58.
- [69] Loai M. Alnemer, Lama Rajab and Ibrahim Aljarah. (2016). Conformal Prediction Technique to Predict Breast Cancer Survivability, p1-11.
- [70] Hyunjung Shinn and Yonghyun Nam. (2013). A coupling approach of a predictor and a descriptor for breast cancer prognosis. p1-12.
- [71] Anothaisintawee, T., Teerawattananon, Y., Wiratkapun, C., Kasamesup, V., and Thakkinian, A. (2012). Risk prediction models of breast cancer: a systematic review of model preferences.