Original Article

DOI: 10.7860/JCDR/2023/59472.17598

Statistics Section

An Auxiliary Approach to Prediction of Binary Outcome with Bayesian Network Model: Exploration with Data for Recurrence of Breast Cancer

SACHIT GANAPATHY[1], KT HARICHANDRAKUMAR[2], KADHIRAVAN TAMILARASU[3], PRASANTH PENUMADU[4], N SREEKUMARAN NAIR[5]

ABSTRACT

Introduction: Logistic regression is the classical statistical model that is incorporated to predict a binary outcome variable. These models have theoretical assumptions of independence of predictor variables and linearity of association with the outcome in the logarithmic scale. Alternative models developed in the machine learning context like Naïve Bayes model with similar assumptions and Bayesian Network (BN) model can be used for binary prediction.

Aim: To compare the predictive performance of logistic regression, Naïve Bayes and BN model in predicting the recurrence of Breast cancer.

Materials and Methods: The dataset was procured from UCI Machine Learning repository on recurrence of breast cancer. The study was done on retrospective data from December 2021 to July 2022. The sample size was boosted with the bootstrapping with logistic regression model. The dataset was split into training (70%) and testing (30%) dataset for internal validation. The effect estimates of the potential prognostic variables were estimated using multiple logistic regression model. Naïve Bayes and BN model was also learnt from the training dataset. The indices of predictive accuracy were estimated for the models in both training and testing dataset.

Results: Degree of malignancy and side of affected breast were found to be significant predictors of recurrence of breast cancer. BN model had the least misclassification rate and the best sensitivity in comparison to other models in spite of imbalance in outcome variable.

Conclusion: BN model performed the best in comparison to logistic regression model when the assumptions of logistic regression model were violated and there is imbalance in proportion of outcome.

Keywords: Binary prediction, Naïve Bayes model, Predictive accuracy

INTRODUCTION

Statistical models in health care have been extensively developed to help in medical decision-making [1]. They assist at the process of making important decisions to archive specific clinical outcomes and also in managing resources to be allocated. Prognostic modeling has had immense application in the field of medicine [2]. Prognostic models estimate the probability of an outcome of a condition and also explore the relationship of factors affecting this outcome. Unlike other models which incorporate a single explanatory variable and consider other variables as confounders, prognostic models focus on incorporating the combined effect of variables to predict the outcome. They are particularly important in selecting the right treatment and managing resources [2].

When the outcome variable is binary, logistic regression model is preferred for the prognosis of disease outcome [3]. Binary logistic regression model encompasses the effect of predictor variables on the dependent binary variable by linearising the relationship using a log link function. Although the performance of logistic regression as a prognostic model has been good, practically, various assumptions are violated [4]. One of the most important assumptions of logistic regression is that the predictor variables are independent of one another. This assumption is almost never true in medical research, especially in the prognostic model [5]. Regression models which are developed in the frequentist context have the assumption of normality for the error term and homoscedasticity for each level of the independent variable in the model. In spite of these assumptions being violated, logistic regression is widely used. There are some alternative predictive models suggested in literature which can be used as an alternative to logistic regression model which can overcome these assumptions [6]. BN model are graphical representations which consists of Directed Acyclic Graphs (DAG) with nodes and edges which can be used to query a binary outcome variable [7]. Naïve Bayes models are simple classifiers which are a subset of BN models which considers conditional independence between the set of independent variables to predict the outcome variable [8]. These are some alternative models that can be explored for the prediction of binary outcome variable.

Breast cancer is one of the most prominent cancer affecting women around the world [9]. Although, recently, there have been advances that has improved the survival outcomes like mortality, recurrence of breast cancer still persists to be around 8-11% after different treatment modalities in India [10]. It has been established in literature that some of the most common prognostic factors associated with recurrence of breast cancer includes age, menopausal status, pathological N stage, pathological T stage, treatment modality, HER2, eGFR, oestrogen and progesterone receptors [11].

The prognosis of medical condition such as cancer is dependent on multiple factors which are correlated to one another. Clinical, sociodemographic and treatment modalities given play a crucial role in the progression of breast cancer. Several statistical and machine learning models have been implemented in the prediction of recurrence of breast cancer that has proven to be excellent in their predictive ability [12,13]. Although they have proven to be good, it is imperative that we consider incorporating the expert opinion into these models which can bring in a better insight into the practical use of the models [14]. This is the gap between clinical and model

experts that needs to be bridged. BN models are an alternative approach which can incorporate the dependency between the factors with supervised learning from data and expert opinion. Data have also shown that hybrid BN models have good predictive accuracy and intuitive explanation ability [15]. In this study, our objective was to assess the predictive ability of Naïve Bayes model and BN model compared to logistic regression model in predicting the recurrence of breast cancer.

MATERIALS AND METHODS

The present exploratory study from a retrospective secondary data of breast cancer cases was conducted from December 2021 to July 2022 in Jawaharlal Institute of Post Graduate Medical Education and Research, Puducherry.

Models: The Naïve Bayes Model-Naïve Bayes classifier are probabilistic classifiers that is based on Bayes theorem which uses the properties of conditional independence to compactly represent high-dimensional probability distribution [16]. The variables are not completely marginally independent in the case of this classifier model. The Naïve Bayes classifier model can be constructed for an outcome variable Y with possible distinct classes $\{c^1,c^2\ldots c^k\}$ which are mutually exclusive and exhaustive. Naïve Bayes model, though, makes a very strong assumption about the independent variables. In the presence of *n* independent variables $X_1,X_2\ldots X_n$ which are potential factors affecting the outcome variable Y, the Naïve Bayes assumption states that Xi's are conditionally independent of each other given the outcome of the individual. Formally, it is represented as:

$$(X_i \perp X_{-i} | Y) \text{ for all i}$$

Naïve Bayes model can be represented as a BN model although the assumptions of independency are strong and generally not true practically. The joint probability distribution of Naïve Bayes model accounting for the assumption is given by

$$P(Y, X_1, \ldots, X_n) = P(Y)\prod_{i=1}^{n} P(X_i|Y)$$

Bayesian Network (BN) model: BN models are graphical representation of the interdependencies between variables represented by a DAG and conditional probabilities. Let 'G' be a DAG, then it consists a set of variables, '*X*' and a set of directed edges, '*E*' connecting these set of variables represented by nodes [17]. In BN models, a node without a parent node is parametrised by the assumed prior distribution, whereas those with parent nodes are parametrised by conditional probability given by P(X|parent(X)). The joint conditional probability of all the variables in the BN model is given by:

$$P(x1,x2,\ldots,xp)=i=1pP(xi|Parent(xi))$$

Building a BN model includes steps of variable selection, structure learning and parameter learning, which can be undertaken by supervised learning from the data including expert opinion.

Dataset: The dataset for building the Naïve Bayes model was procured from an online database, UCI Machine Learning Repository [18]. The data was with reference to a Breast cancer study to predict the recurrence of event based on certain attributes. The total sample size in the dataset was 286. There were a total of nine variables in the dataset including age, menopause status, tumour size, number of nodes involved, presence of node caps, degree of malignancy, breast, breast quadrant and status of irradiation. The dataset was sourced from Institute of Oncology, University Medical Center, Ljubljana, Yugoslavia by M. Zwitter and M. Soklic in 1988 available from: https://archive.ics.uci.edu/ml/datasets/breast+cancer. The dataset obtained was inflated to a sample size of 1000 with the help of logistic regression equation with all the variables in the existing dataset as predictor variables for recurrence as the outcome. The total effective sample size used in the current manuscript was 1000 after inflation.

Variables in the model: The dataset depicted the multivariable classification of the patients for the prognosis of Breast cancer. The event of interest here was the recurrence of the disease. The dataset contained the information for all the samples. The variables in the model were defined and categorised based on the criterion from the 8th edition of AJCC Cancer Staging Form Supplement [19]. The variables in the model are defined and the recategorisation is given below:

1. Age of the patients at the time of diagnosis:
   a. 10-39 years
   b. 40-49 years
   c. 50-59 years and
   d. ≥60 years.
2. Whether the patient was pre-or post-menopausal at the time of the diagnosis:
   a. <40 years
   b. ≥40 years and
   c. premenopausal
3. The greatest diameter of the excised tumour. Based on the tumour size chart, they were categorised as
   a. T1 (0-2 cm),
   b. T2 (2-5 cm) and
   c. T3 (>5 cm).
4. The number of axillary lymph nodes that contain metastatic breast cancer visible on histological examination:
   a. 0-2,
   b. 3-9 and
   c. >10
5. The presence of tumour as a capsule of the lymph node, which over time with more aggressive disease, tumour may replace the lymph node.
6. The histological grade of the tumour.
   - 1,
   - 2 and
   - 3 where Grade 1 predominantly consists of cells that retain their usual characteristics and Grade 3 predominantly consists of cells that are highly abnormal.
7. The side of the affected breast.
8. The breast was also divided into five quadrants using nipple as a central point; categorised as
   - left-up
   - left-down
   - right-up
   - right-low and
   - centre
9. Whether radiation therapy, was given or not.

STATISTICAL ANALYSIS

The dataset was classified into two parts as training and testing dataset. Approximately, 70% of the data was used for training the model and the rest of the 30% of the data was used for testing the classification accuracy of the model. The distribution of the prognostic variables across the binary outcome of recurrence was assessed in the training, testing and the entire dataset. The univariate logistic regression was performed initially and with p-value <0.15 as the cut-off, the potential factors were used to build the multiple logistic regression model. A p-value <0.05 was considered to be statistically significant in the final model.

All the models were trained using training dataset and then tested using both training and testing dataset. Logistic regression model

was built with all the potential prognostic variables. The predicted probabilities were estimated from the model. Naïve Bayes model with Laplace smoothing was used to develop the model. BN model was built with two important steps. The structure learning of the BN model was carried out based on the Tree Augmented Network (TAN) method [20]. Conditional probabilities associated with each node was estimated using Expectation-Maximisation (EM) method [21]. Misclassification rate, sensitivity, specificity, Positive Predictive Value (PPV) and Negative Predictive Value (NPV) were estimated in both training and testing dataset. All the statistical analysis was performed in R Studio Version 1.2.1335 and Netica 6.09 for Bayes nets. The Naïve Bayes model was built using the *naivebayes* package.

RESULTS

The distribution of all the factors in the model across both the outcome category in both training and testing dataset is given in [Table/Fig-1]. Logistic regression model was used and the effect estimates from univariate and multiple logistic regression estimates were obtained and the results are shown in [Table/Fig-2]. It was found that degree of malignancy and the side of the breast were the two variables which significantly contributed in the prediction of recurrence of breast cancer from multiple logistic regression model. BN model developed from the TAN method for structure learning and EM method for parameter learning is given as [Table/Fig-3]. The probability distribution associated with each variable is given in the network model.

In the training dataset, it was found that logistic regression had a misclassification rate of 33.52%, BN model with 31.09% whereas it was estimated to be 33.38% for Naïve Bayes classifier as given in [Table/Fig-4]. When the same model was used to classify the recurrence status in testing dataset, logistic regression had a misclassification rate of 35.1%, BN model had 36.42% whereas it was 34.77% for Naïve Bayes classifier. The sensitivity was poor for all the models. Specificity was excellent for all the models, 96.96% for LR model, 91.52% for BN model and 97.83% for NB model in training dataset. In the testing dataset it was estimated to be 91.83% for LR model, 87.02% for BN model and 92.31% for NB model in testing dataset. PPV was estimated to be 56.25% for LR model, 60% for NB model and 60.6% for BN model in training dataset. In testing dataset, it was estimated to be 22.73% for LR model, 23.81% for NB model and 30.56% for BN model. NPV was estimated to be 66.97% for LR model, 66.86% for NB model and 70.28% for BN model in training dataset. In testing dataset, it was estimated to be 68.21% for LR model, 68.33% for NB model and 65.56% for BN model.

DISCUSSION

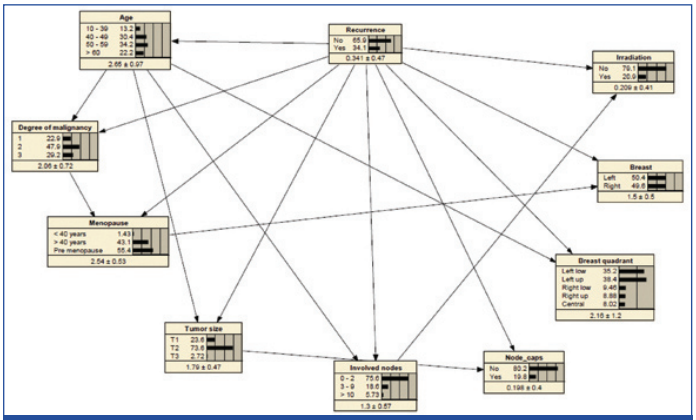
In the present study, the prognostic factors associated with recurrence of breast cancer were determined. It was found that degree of malignancy and side of the affected breast had an impact on the outcome. A study has shown that tumour size, grade of the cancer, nodal status and hormonal factors along with smoking status to have significant association with recurrence of breast cancer [22]. A study have also pointed out that receiving neoadjuvant chemotherapy reduced the risk of recurrence for breast cancer [23]. The current dataset had variables related to the disease status and not with lifestyle characteristics. The primary objective of this study was to compare the predictive ability of BN, Naive Bayes and Logistic regression model. It was found that even with imbalance in the proportion of outcome variable, BN model outperformed the other models overall. The misclassification rate was least for BN model and it provided a better ability in predicting the recurrence of breast cancer with better sensitivity, which is the key in these models.

Variables		Total		Training		Testing	
		Recurrence (n=332)	No recurrence (n=668)	Recurrence (n=238)	No recurrence (n=460)	Recurrence (n=94)	No recurrence (n=208)
Age (years)	10-39	41 (12.3)	85 (12.7)	30 (12.6)	62 (13.5)	11 (11.7)	23 (11.1)
	40-49	98 (29.6)	211 (31.6)	66 (27.7)	146 (31.7)	32 (34)	65 (31.2)
	50-59	123 (37)	229 (34.3)	91 (38.3)	148 (32.2)	32 (34)	81 (38.9)
	≥60	70 (21.1)	143 (21.4)	51 (21.4)	104 (22.6)	19 (20.3)	38 (18.8)
Menopause	<40 years	5 (1.5)	15 (2.2)	5 (2.1)	5 (1.1)	0 (0)	10 (4.8)
	≥40 years	151 (45.5)	283 (42.4)	111 (46.6)	190 (41.3)	40 (42.6)	93 (44.7)
	Premenopause	176 (53)	370 (55.4)	122 (51.3)	256 (55.6)	54 (57.4)	105 (50.5)
Tumour size (cm)	T1	81 (24.4)	158 (23.7)	54 (22.7)	111 (24.1)	27 (28.7)	47 (22.6)
	T2	246 (74.1)	486 (72.8)	180 (75.6)	334 (72.6)	66 (70.2)	152 (73.1)
	T3	5 (1.5)	24 (3.6)	4 (1.7)	15 (3.3)	1 (1.1)	9 (4.3)
Involved nodes	0-2	261 (78.6)	497 (74.4)	187 (78.6)	341 (74.1)	74 (78.7)	156 (75)
	3-9	48 (14.5)	128 (19.2)	36 (15.1)	94 (20.4)	12 (12.8)	34 (16.3)
	≥10	23 (6.9)	43 (6.4)	15 (6.3)	25 (5.5)	8 (8.5)	18 (8.7)
Node caps	Yes	75 (22.6)	125 (18.7)	52 (21.8)	86 (18.7)	21 (22.3)	45 (21.6)
Degree of malignancy	1	61 (18.4)	165 (24.7)	40 (16.8)	120 (26.1)	32 (34.1)	61 (29.4)
	2	161 (48.5)	316 (47.3)	120 (50.4)	214 (46.5)	23 (24.5)	39 (18.7)
	3	110 (33.1)	187 (28)	78 (32.8)	126 (27.4)	71 (75.5)	169 (81.3)
Breast	Left	159 (47.9)	360 (53.9)	105 (44.1)	247 (53.7)	54 (57.4)	113 (54.3)
	Right	173 (52.1)	308 (46.1)	133 (55.9)	213 (46.3)	40 (42.6)	95 (45.7)
Breast quadrant	Left-up	123 (37)	232 (34.7)	89 (37.4)	157 (34.1)	34 (36.2)	75 (36.1)
	Left-low	127 (38.4)	259 (38.8)	83 (34.9)	185 (40.2)	44 (46.8)	74 (35.6)
	Right-up	33 (9.9)	64 (9.6)	24 (10.1)	42 (9.1)	9 (9.5)	22 (10.6)
	Right-low	23 (6.9)	58 (8.7)	19 (8)	43 (9.3)	4 (4.3)	15 (7.2)
	Central	26 (7.8)	55 (8.2)	23 (9.7)	33 (7.2)	3 (3.2)	22 (10.6)
Irradiation	Yes	69 (20.8)	141 (21.1)	52 (21.8)	94 (20.4)	17 (18.1)	47 (22.6)

[Table/Fig-1]: Distribution of all the potential prognostic factors for the recurrence of Breast cancer across the outcome.

Variables		Unadjusted OR (95% CI)	Adjusted OR (95% CI)	p-value
Age (years)	10-39	0.99 (0.57, 1.71)		
	40-49	0.92 (0.59, 1.44)		
	50-59	1.25 (0.82, 1.92)		
Menopause	<40	2.17 (0.62, 7.64)	2.49 (0.69, 8.98)	0.164
	≥40	1.27 (0.92, 1.74)	1.33 (0.96, 1.83)	0.088
Tumour size (cm)	T3	0.55 (0.17, 1.73)		
	T2	1.11 (0.76, 1.61)		
Involved nodes	≥10	1.09 (0.56, 2.13)	0.94 (0.48, 1.86)	0.867
	3-9	0.70 (0.46, 1.07)	0.70 (0.45, 1.07)	0.101
Node caps	Yes	1.22 (0.83, 1.79)		
Degree of malignancy	3	1.86 (1.18, 2.93)	1.89 (1.19, 2.99)	0.007
	2	1.68 (1.10, 2.57)	1.71 (1.12, 2.63)	0.014
Breast	Left	1.47 (1.07, 2.01)	0.67 (0.48, 0.92)	0.014
Breast quadrant	Left-up	0.81 (0.45, 1.47)		
	Left-low	0.64 (0.36, 1.16)		
	Right-up	0.82 (0.40, 1.70)		
	Right-low	0.63 (0.30, 1.35)		
Irradiation	Yes	1.09 (0.74, 1.60)		

[Table/Fig-2]: Effect estimates of the potential prognostic factors in the recurrence of Breast cancer using Logistic regression.



[Table/Fig-3]: Bayesian Network model for prediction of recurrence of Breast cancer.

Indices	Training			Testing		
	LR model	NB model	BN model	LR model	NB model	BN model
Misclassification rate	33.52%	33.38%	31.09%	35.10%	34.77%	36.42%
Sensitivity	7.56%	6.30%	25.21%	5.32%	5.32%	11.70%
Specificity	96.96%	97.83%	91.52%	91.83%	92.31%	87.02%
PPV	56.25%	60.00%	60.60%	22.73%	23.81%	30.56%
NPV	66.97%	66.86%	70.28%	68.21%	68.33%	65.56%

[Table/Fig-4]: Comparison of the predictive performance of Naive Bayes, Logistic regression and Bayesian Network (BN) model.
PPV: Positive predictive value; NPV: Negative predictive value; LR: Logistic regression; NB: Naïve bayes, BN: Bayesian network

Naïve Bayes model and logistic regression have already been applied for predicting the recurrence of breast cancer and has proven to have performed considerably well [24]. Naïve Bayes classifier offers a novel approach for categorising patients and offers good performance with low algorithmic cost and high speed of computation. Another study has shown that Naïve Bayes model performs as well as other equivalent machine learning techniques [25]. With just seven prognostic factors, nomogram based on Naïve Bayes model gave 80% accuracy suggesting the model can be translated to practical use. Bayesian classifiers have gained importance in classification problem in health care studies and have performed better than classical approach to prognostic modeling [26]. Even amongst the Bayesian classifiers, Naïve Bayes model with tree augmented structure and gradient boosting has shown to perform well in predictive accuracy [27]. A study by Choi J et l., has showed that hybrid BN models have excellent predictive ability in comparison to any other machine learning algorithms in predicting breast cancer prognosis [15]. It was seen that hybrid BN models had AUC of 0.935 as compared to 0.930 and 0.813 for artificial neural network and classical BN model. BN models have also been applied in the prediction of risk of triple negative breast cancer with epidemiological factors and has shown to perform well [28]. Studies have compared the predictive accuracy of BN model with other machine learning algorithms like support vector machine and artificial neural network for a binary outcome, and have proven that they are better or comparable at handling missing data and predictive accuracy [29,30]. BN model has further illustrated that it can incorporate complex interactions of prognostic factors and individualising patient care in oncology [31]. This suggests that we have to try to translate the machine algorithms such as BN model as a more viable option for clinicians to use.

Witteveen A et al., on the other hand has also reported that conventional logistic regression models have outperformed BN model in predictive accuracy related to breast cancer [32]. Although BN model performed better in the development cohort, on validation, it was seen that LR models had a C-statistic of 0.71 whereas it was 0.67 for BN model. The difference observed in the overall predictive ability between the models is not high. Generally, it is seen that the difference in the AUC or C statistic was seen to be less than 0.05 in studies [33,34]. A study by Holm CE et al., has also shown that proper internal and external validation is unaccounted for BN models [35].

Limitation(s)

Our study was limited to the factors that were a part of the source of secondary data which did not include some important established prognostic factors in recurrence of breast cancer. Variables such as Her2, oestrogen receptors, progesterone receptors and eGFR values could have improved the predictive ability of the models. The proportion of outcome had imbalance and therefore, a Synthetic Minority Oversampling Technique (SMOTE) for imbalanced classification can further strengthen the predictive accuracy of the models. External validation was not performed in the study with an independent dataset for generalisability of the model. Other estimates could have also been estimated for showing the predictive accuracy of models, such as AUC, Gini coefficient and C-index which suggests the overall discriminatory ability of the model but this study was with the intention of suggesting alternative techniques for predicting a binary outcome.

CONCLUSION(S)

BN model can be used as an alternative model for predicting a binary outcome in the recurrence of breast cancer. The predictive ability of BN model was found to be better and it can handle imbalanced classification better. They also provide with a visually intuitive model with lesser assumptions. With further improving the model, they can provide a better predictive model to be used bed-side for clinicians.

Acknowledgement

Dr. P. Venkatesan for his contribution in helping to understand the models that were used in the application in this study.

REFERENCES

[1] Malehi AS, Pourmotahari F, Angali KA. Statistical models for the analysis of skewed healthcare cost data: A simulation study. Health Econ Rev. 2015;5(1):11.
[2] Vogenberg FR. Predictive and prognostic models: Implications for healthcare decision-making in a modern recession. Am Health Drug Benefits. 2009;2(6):218-22.
[3] Steyerberg EW, Eijkemans MJ, Harrell FE Jr, Habbema JD. Prognostic modeling with logistic regression analysis: In search of a sensible strategy in small data sets. Med Decis Making. 2001;21(1):45-56. Doi: 10.1177/0272989X0102100106. PMID: 11206946.

[4] Schreiber-Gregory D, Bader K. Logistic and linear regression assumptions: Violation recognition and control. Proc Midwest SAS User Group. 2018;01-21.
[5] Senaviratna NAMR, Cooray TMJA. Diagnosing multicollinearity of logistic regression model. Asian J Probab Stat. 2019;5(2):01-09.
[6] Westreich D, Lessler J, Funk MJ. Propensity score estimation: Neural networks, support vector machines, decision trees (CART), and meta-classifiers as alternatives to logistic regression. J Clin Epidemiol. 2010;63(8):826-33.
[7] Cobb BR, Rumí R, Salmerón A. Bayesian network models with discrete and continuous variables. Advances in probabilistic graphical models. 2007:81-102.
[8] Leung KM. Naive Bayesian Classifier [Internet]. 2007; Polytechnic University. Available from: https://cse.engineering.nyu.edu/~mleung/FRE7851/f07/naiveBayesianClassifier.pdf
[9] Global Burden of Disease Cancer Collaboration, Fitzmaurice C, Abate D, Abbasi N, Abbastabar H, Abd-Allah F, et al. Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 29 cancer groups, 1990 to 2017: A systematic analysis for the global burden of disease study. JAMA Oncol. 2019;5(12):1749-68.
[10] Rangarajan B, Shet T, Wadasadawala T, Nair NS, Sairam RM, Hingmire SS, et al. Breast cancer: An overview of published Indian data. South Asian J Cancer. 2016;5(3):86-92.
[11] Kim JY, Lee YS, Yu J, Park Y, Lee SK, Lee M, et al. Deep learning-based prediction model for breast cancer recurrence using adjuvant breast cancer cohort in tertiary cancer center registry. Front Oncol [Internet]. 2021 [cited 2022 Jul 16];11. Available from: https://www.frontiersin.org/articles/10.3389/fonc.2021.596364
[12] Kim W, Kim KS, Lee JE, Noh DY, Kim SW, Jung YS, et al. Development of novel breast cancer recurrence prediction model using support vector machine. J Breast Cancer. 2012;15(2):230-38.
[13] Ahmad LG, Eshlaghy AT, Poorebrahimi A, Ebrahimi M, Razavi AR. Using Three machine learning techniques for Predicting breast cancer recurrence. J Health Med Inform. 2013;4:124.
[14] Štrumbelj E, Bosnić Z, Kononenko I, Zakotnik B, Grašič Kuhar C. Explanation and reliability of prediction models: The case of breast cancer recurrence. Knowl Inf Syst. 2010;24(2):305-24.
[15] Choi JP, Han TH, Park RW. A hybrid Bayesian network model for predicting breast cancer prognosis. J Kor Soc Med Informatics. 2009;15(1):49-57.
[16] Mitchell TM. Machine learning. International ed., [Reprint.]. New York, NY: McGraw-Hill; 20. 414 p. (McGraw-Hill series in computer science).
[17] Bayesian Networks and Decision Graphs [Internet]. [cited 2022 Jul 16]. Available from: https://link.springer.com/book/10.1007/978-1-4757-3502-4.
[18] UCI Machine Learning Repository: Breast Cancer Data Set [Internet]. [cited 2020 Nov 23]. Available from: http://archive.ics.uci.edu/ml/datasets/Breast+Cancer?ref=datanews.io.
[19] Zanoni DK, Patel SG, Shah JP. Changes in the 8th Edition of the American Joint Committee on Cancer (AJCC) Staging of Head and Neck Cancer: Rationale and Implications. Curr Oncol Rep. 2019;21(6):52.
[20] Friedman N, Geiger D, Goldszmidt M. Bayesian Network Classifiers. Mach Learn. 1997;29(2):131-63.
[21] Ji Z, Xia Q, Meng G. A review of parameter learning methods in Bayesian Network | SpringerLink. In: Advanced Intelligent Computing Theories and Applications [Internet]. Cham: Springer; 2015 [cited 2022 Oct 27]. pp. 03-12. Available from: https://link.springer.com/chapter/10.1007/978-3-319-22053-6_1.
[22] Lafourcade A, His M, Baglietto L, Boutron-Ruault MC, Dossus L, Rondeau V. Factors associated with breast cancer recurrences or mortality and dynamic prediction of death using history of cancer recurrences: The French E3N cohort. BMC Cancer. 2018;18(1):171.
[23] Stankov A, Bargallo-Rocha JE, Silvio AÑS, Ramirez MT, Stankova-Ninova K, Meneses-Garcia A. Prognostic factors and recurrence in breast cancer: Experience at the national cancer institute of Mexico. ISRN Oncol. 2012;2012:825258.
[24] Kim W, Kim KS, Park RW. Nomogram of Naive bayesian model for recurrence prediction of breast cancer. Healthc Inform Res. 2016;22(2):89-94.
[25] Dumitru D. Prediction of recurrent events in breast cancer using the Naive bayesian classifcation. Analele Univ Din Craiova Ser Mat Informatică. 2009;36.
[26] Al-Aidaroos KM, Bakar AA, Othman Z. Medical data classification with Naive bayes approach. Information Technology Journal. 2012;11(9):1166-74.
[27] Banu AB, Thirumalaikolundusubramanian P. Comparison of Bayes classifiers for breast cancer classification. Asian Pac J Cancer Prev. 2018;19(10):2917-2920. Doi: 10.22034/APJCP.2018.19.10.2917. PMID: 30362322; PMCID: PMC6291060.
[28] Huang Y, Zheng C, Zhang X, Cheng Z, Yang Z, Hao Y, et al. The usefulness of bayesian network in assessing the risk of triple-negative breast cancer. Acad Radiol. 2020;27(12):e282-91.
[29] Jayasurya K, Fung G, Yu S, Dehing-Oberije C, De Ruysscher D, Hope A, et al. Comparison of bayesian network and support vector machine models for two-year survival prediction in lung cancer patients treated with radiotherapy. Med Phys. 2010;37(4):1401-07.
[30] Correa M, Bielza C, Pamies-Teixeira J. Comparison of bayesian networks and artificial neural networks for quality detection in a machining process. Expert Syst Appl. 2009;36(3):7270-79.
[31] Reijnen C, Gogou E, Visser NCM, Engerud H, Ramjith J, van der Putten LJM, et al. Preoperative risk stratification in endometrial cancer (ENDORISK) by a bayesian network model: A development and validation study. PLoS Med. 2020;17(5):e1003111.
[32] Witteveen A, Nane GF, Vliegen IMH, Siesling S, IJzerman MJ. Comparison of logistic regression and bayesian networks for risk prediction of breast cancer recurrence. Med Decis Mak Int J Soc Med Decis Mak. 2018;38(7):822-33.
[33] Cho SM, Austin PC, Ross HJ, Abdel-Qadir H, Chicco D, Tomlinson G, et al. Machine learning compared with conventional statistical models for predicting myocardial infarction readmission and mortality: A systematic review. Can J Cardiol. 2021;37(8):1207-14.
[34] Clark DO, Stump TE, Tu W, Miller DK. A comparison and cross-validation of models to predict basic activity of daily living dependency in older adults. Med Care. 2012;50(6):534-39.
[35] Holm CE, Grazal CF, Raedkjaer M, Baad-Hansen T, Nandra R, Grimer R, et al. Development and comparison of 1-year survival models in patients with primary bone sarcomas: External validation of a Bayesian belief network model and creation and external validation of a new gradient boosting machine model. SAGE Open Med. 2022;10:20503121221076388.

PARTICULARS OF CONTRIBUTORS:
1. PhD Scholar, Department of Biostatistics, Jawaharlal Institute of Post Graduate Medical Education and Research, Puducherry, India.
2. Assistant Professor, Department of Biostatistics, Jawaharlal Institute of Post Graduate Medical Education and Research, Puducherry, India.
3. Professor, Department of Medicine, Jawaharlal Institute of Post Graduate Medical Education and Research, Puducherry, India.
4. Additional Professor, Department of Surgical Oncology, Jawaharlal Institute of Post Graduate Medical Education and Research, Puducherry, India.
5. Professor and Head, Department of Biostatistics, Jawaharlal Institute of Post Graduate Medical Education and Research, Puducherry, India.

NAME, ADDRESS, E-MAIL ID OF THE CORRESPONDING AUTHOR:
N Sreekumaran Nair,
Professor and Head, Department of Medicine, Jawaharlal Institute of Post Graduate Medical Education and Research, Puducherry, India.
E-mail: nsknairmanipal@gmail.com

PLAGIARISM CHECKING METHODS: [Jain H et al.]
- Plagiarism X-checker: Aug 06, 2022
- Manual Googling: Nov 01, 2022
- iThenticate Software: Nov 10, 2022 (7%)

ETYMOLOGY: Author Origin

AUTHOR DECLARATION:
- Financial or Other Competing Interests: None
- Was Ethics Committee Approval obtained for this study? No
- Was informed consent obtained from the subjects involved in the study? Yes
- For any images presented appropriate consent has been obtained from the subjects. NA

Date of Submission: **Aug 05, 2022**
Date of Peer Review: **Oct 13, 2022**
Date of Acceptance: **Nov 11, 2022**
Date of Publishing: **Mar 01, 2023**