

Hierarchical clinical decision support for breast cancer care empowered with Bayesian networks

Clinical
decision
support for
breast cancer

163

Omran Alomran

*Department of Industrial and Manufacturing Engineering, Penn State University,
University Park, Pennsylvania, USA*

Robin Qiu

*Department of Information Science, Pennsylvania State University, Malvern,
Pennsylvania, USA, and*

Hui Yang

*Department of Industrial and Manufacturing Engineering, Penn State University,
University Park, Pennsylvania, USA*

Received 20 November 2022
Revised 20 December 2022
Accepted 23 December 2022

Abstract

Purpose – Breast cancer is a global public health dilemma and the most prevalent cancer in the world. Effective treatment plans improve patient survival rates and well-being. The five-year survival rate is often used to develop treatment selection and survival prediction models. However, unlike other types of cancer, breast cancer patients can have long survival rates. Therefore, the authors propose a novel two-level framework to provide clinical decision support for treatment selection contingent on survival prediction.

Design/methodology/approach – The first level classifies patients into different survival periods using machine learning algorithms. The second level has two models with different survival rates (five-year and ten-year). Thus, based on the classification results of the first level, the authors employed Bayesian networks (BNs) to infer the effect of treatment on survival in the second level.

Findings – The authors validated the proposed approach with electronic health record data from the TriNetX Research Network. For the first level, the authors obtained 85% accuracy in survival classification. For the second level, the authors found that the topology of BNs using Causal Minimum Message Length had the highest accuracy and area under the ROC curve for both models. Notably, treatment selection substantially impacted survival rates, implying the two-level approach better aided clinical decision support on treatment selection.

Originality/value – The authors have developed a reference tool for medical practitioners that supports treatment decisions and patient education to identify patient treatment preferences and to enhance patient healthcare.

Keywords Clinical decision support, Bayesian network, Machine learning, Breast cancer, Data-driven modeling

Paper type Research paper

© Omran Alomran, Robin Qiu and Hui Yang. Published in *Digital Transformation and Society*. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licenses/by/4.0/legalcode>

This research is funded in part by the NSF I/UCRC Center for Healthcare Organization Transformation (CHOT) award 1624727 and in part by Susan G. Komen. The authors gratefully acknowledge the valuable contributions and suggestions from Dr. Jerome Jourquin and Jessica Epps for this research paper. Any opinions, findings or conclusions found in this paper are those of the authors and do not necessarily reflect the views of the sponsors.



1. Introduction

Cancer is the second leading cause of death in the United States and a significant public health problem. Breast cancer is the most common cancer and accounts for 31% of new cancer incidents in females in the United States. According to the American Cancer Society, one in eight women will develop breast cancer in her lifetime (Siegel, Miller, Fuchs, & Jemal, 2022). Unlike patients with other types of cancer, breast cancer patients have about a 90% chance of surviving more than five years (Islami *et al.*, 2022). Modeling studies show that early detection and effective treatment plans improve survival rates (Berry *et al.*, 2005). To that end, researchers have adapted various machine learning (ML) tools to propose prognostic and diagnostic models utilizing medical datasets. Because the five-year survival rate is commonly used, better understanding of treatments for longer survival periods should be included in breast cancer studies as a necessity.

In the past decade, many medical organizations adopted electronic health records (EHRs) in the United States after the Meaningful Use initiative in 2009 (Evans, 2016). The use of EHRs made patient information easier to read and to be remotely accessed. As a result, clinical decision support systems received wide attention on account of their potential to improve the quality of healthcare (Murphy, 2014).

Decision systems with classification or prediction purposes typically employ ML models. Nevertheless, most of these ML methods are similar to a black-box model, and it can be hard or even impossible to explain how outcomes were identified. For that reason, Bayesian networks (BNs) are more attractive for medical applications. The graphical representation of the structure with the conditional probability distribution (CPD) for each node (variable) makes BNs highly interpretable models that are easy to comprehend. Interpretability, especially in the healthcare domain, helps provide medical practitioners insights and thus make proper therapy decisions with high confidence. Using ML and BN together can be a powerful approach for individualized treatment recommendations, as it allows to take advantage of the strengths of both approaches.

This paper presents a novel two-level framework to provide data-driven clinical decision support on breast cancer treatment selection. In theory, personalized treatments make a difference in prolonging individual survival periods. Knowing an individual's personal survivability should be an important step before her personal treatment recommendation can be made. Therefore, in this study, the first level classifies patients into different survival periods using ML methods. Then, the second level derives probabilistic inferences of prognostic outcomes using BNs.

The remainder of the paper is organized as follows. Section 2 gives a brief overview of related works on breast cancer prognosis with an emphasis on BNs. Section 3 describes our data collection and preprocessing techniques. A new two-level methodology structure is described in Section 4. Section 5 exhibits the experimental results of the proposed approach. Section 6 discusses the benefits of the new methodology and possible applications. Lastly, Section 7 provides conclusions and future directions.

2. Background

A vast amount of literature discusses the adoption of ML techniques for prognosis prediction in breast cancer. Two of the most frequently used ML methods to predict the survivability of breast cancer patients are artificial neural network (ANN) and support vector machine (SVM) (Li *et al.*, 2021). Delen, Walker and Kadam (2005) developed an ANN model to predict the five-year life expectancy of breast cancer patients using the Surveillance, Epidemiology and End Results database, and it achieved high-performance measures. However, Shin and Nam (2014) report that SVM has better performance measures than ANN over ten datasets. Most survivability prediction studies in breast cancer focus on five-year relative survival, as it

indicates treatment success for many cancer types. Nevertheless, the risk of distant recurrence can reach 41% after five years of survival, depending on different factors (Pedersen *et al.*, 2022). In addition, the rate of breast cancer recurrence is high for patients between the age of 20 and 50 (Imani, Chen, Tucker, & Yang, 2019). Thus, involving longer survival rates in prognosis prediction and treatment selection models provides a deeper understanding of the effectiveness of any intervention.

Current research on breast cancer using BNs is primarily related to medical diagnosis, risk evaluation and prognostic applications. Cruz-Ramírez, Acosta-Mesa, Carrillo-Calvet, Alonso Nava-Fernández and Barrientos-Martínez (2007) developed and evaluated seven BNs to diagnose breast cancer using two databases that contain information derived from fine-needle aspiration, whereas Kahn, Roberts, Shaffer and Haddawy (1997) built BNs using features obtained from mammographic findings to detect breast malignancy. In a study published in 2018, Witteveen, Nane, Vliegen, Siesling and IJzerman (2018) designed different BNs to predict the risk of locoregional recurrence and second primary breast cancer. Gevaert, Smet, Timmerman, Moreau and Moor (2006) established prediction methods using BNs toward classifying breast cancer patients into poor or good prognosis groups. Gevaert *et al.* integrated clinical and microarray data in three separate ways. Nonetheless, survivability is part of prognosis, and few survivability studies apply BNs.

In relevant BN applications on survival prediction, Forsberg, Eberhardt, Boland, Wedin and Healey (2011) estimated 3-month and 12-month life expectancy of patients with operable skeletal metastases. Jayasurya *et al.* (2010) and Sesen, Nicholson, Banares-Alcantara, Kadir and Brady (2013) created BNs to predict lung cancer patient survival by focusing only on short-term survival. In addition, a study of colon cancer built BNs to perform individualized survival prediction (Stojadinovic *et al.*, 2013). Like these examples, mid- and long-term survival rates are usually neglected.

Focusing on the application of BNs in breast cancer survivability, Choi, Han and Park (2009) developed three models with the aim of predicting five-year survival: two BNs and one hybrid BN model that combined ANN and BN. Also, Endo, Shibata and Tanaka (2008) and Lotfnezhad Afshar, Ahmadi, Roudbari and Sadoughi (2015) applied ML methods, including BNs, to predict five-year survival and compared their performance. Mainly BNs are employed in breast cancer survivability models for prediction or variable selection. At the same time, probability inference is often ignored, which could help answer essential questions, such as “How do different treatment decisions affect the probability of survival for a patient?”. Therefore, this study seeks to fill the highlighted gaps by considering different survival targets and providing treatment decision support.

3. Data

3.1 Data collection

We used EHR data provided by TriNetX Research Network. TriNetX allows access to de-identified patient records from around 60 different healthcare organizations (HCOs). Also, it is compliant with the Health Insurance Portability and Accountability Act. The data comprise patients’ clinical information, such as demographics, diagnosis, tumor properties and genomics. Each of these patients’ information is represented in a different table and could be mapped using key features, for instance, patient ID and encounter ID.

For this study, we created one dataset that contains patient information from the following tables: demographics, tumor, tumor properties, oncology treatment and diagnosis. We used ICD-10 codes starting with C50 to identify breast cancer patients and cross-checked them with the tumor registry table to reduce the chance of misdiagnosis and missing values. We included only one record with the earliest diagnosis date derived from HCOs’ cancer registry for each patient. After the patient demographic was aggregated with the tumor table based on the

unique ID of the patient, the tumor properties table was combined with them when similar variables/features in the tumor table matched, such as patient ID, diagnosis date and tumor site.

Similarly, we merged the oncology treatment table to assemble the needed information in one table. In accordance with the American Cancer Society, breast cancer treatment can be divided into two primary categories: local and systematic. Local treatments include surgery and radiation, whereas systemic treatments include chemotherapy and hormone therapy. However, we only obtained radiation (RTx), chemotherapy (chemo) and hormone therapy (HT) treatments as binary variables. Furthermore, the three treatments were combined into one variable to reduce the dimensionality of the BN structure and to better assess treatment recommendations.

Two variables were internally computed: age at diagnosis and survival. The first was obtained from the difference between the earliest diagnosis date of malignant neoplasm of the breast and the birth date. The latter was computed as the interval between the date of the patient's death and the diagnosis date. In addition, we included five of the most common health conditions in breast cancer patients based on the cohort analysis provided by TriNetX plus personal history of malignant neoplasm of the breast. The health conditions include essential hypertension (ICD-10: I10), heart failure (I50), chronic ischemic heart disease (I25), diabetes type I (E11), and acute kidney failure and/or chronic kidney disease (N17–19). The health conditions were binarized variables that indicated presence or absence. [Table 1](#) shows the list of variables obtained in this study and their corresponding table location.

3.2 Data preprocessing

Data preprocessing is necessary for any ML model, as data quality is vital for a reliable predictive model ([Kotsiantis, Kanellopoulos, & Pintelas, 2006](#)). Therefore, we performed the following preprocessing steps.

#	Variables	Original tables	Values
1	Marital status	Patient demographics	Married; Single; Unknown
2	Tumor site code	Tumor	C50.0; C50.1; C50.2; C50.3; C50.4; C50.5; C50.6; C50.8; C50.9
3	Tumor stage	Tumor	0; I; II; III; IV; Unknown
4	Tumor size (T)	Tumor	T0; T1; T2; T3; T4; Tis; TX
5	Number of lymph nodes (N)	Tumor	N0; N1; N2; N3; NX
6	Metastatic (M)	Tumor	M0; M1; Unknown
7	Estrogen receptor (ER)	Tumor properties	Positive; Negative; Unknown
8	Progesterone receptors (PR)	Tumor properties	Positive; Negative; Unknown
9	HER2	Tumor properties	Positive; Negative; Unknown
10	Essential hypertension	Diagnosis	Yes; No
11	Personal history of malignant neoplasm of breast	Diagnosis	Yes; No
12	Heart failure	Diagnosis	Yes; No
13	Essential hypertension	Diagnosis	Yes; No
14	Chronic ischemic heart disease	Diagnosis	Yes; No
15	Diabetes type I	Diagnosis	Yes; No
16	Age at diagnosis	None; internally calculated	<50; 50–60; 60–70; >70
17	Therapy	Oncology treatment	RTx; chemo; HT; RTx and chemo; RTx and HT; chemo and HT; all; None
18	Survival	None; internally calculated	<= 5yr; 5–10 yr; >10 yr

Table 1.
List of variables and their original tables

First, we removed patients whose sex was male or unknown since the study focused on female breast cancer. Also, we deleted incidences with no birth or death date since age at diagnosis and survival could not be obtained without them. In addition, records with contradictory information were removed, for example, when the diagnosis date was before the birthdate or after the death date.

Secondly, we discretized “age at diagnosis”, which was the only non-categorical variable in the database, since most available BNs packages are applicable only to categorical variables. This variable was discretized into four intervals (<50; 50–60; 60–70; >70), achieving an almost balanced distribution, as shown in [Figure 1](#).

Thirdly, we only allowed up to three values missing among tumor stage, tumor size (T), number of lymph nodes (N), metastatic (M), estrogen receptor (ER), progesterone receptors (PR) and HER2, as they are valuable in determining the case severity and treatment plan, according to the American Cancer Society. Therefore, we replaced the null observations in the dataset with the “Unknown” string. Altogether, 6,375 patients were included in this study.

Lastly, we selected “5_yr ≥”, “5–10_yr” and “>10_yr” as our multi-class definitions for the classification model. [Figure 2](#) reveals the distribution of patients across the three proposed classes. To address the imbalance in the dataset, we performed SMOTE-N to up-sample the minority classes (5–10_yr and >10_yr) for model tuning and evaluation ([Chawla, Bowyer, Hall, & Kegelmeyer, 2002](#)). After the up-sampling, we created two BN models, each of which used a binary variable indicating whether the patient survived for at least five years or at

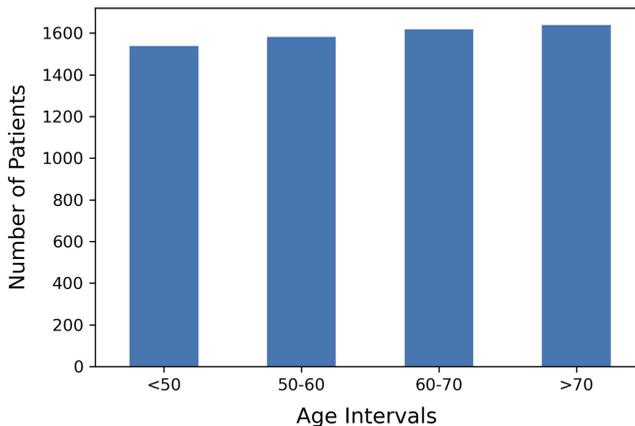


Figure 1.
Age at diagnosis
distribution

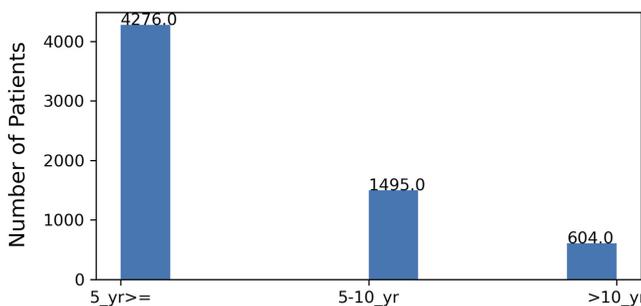


Figure 2.
Patients distribution
across all classes

least ten years. The dataset was randomly partitioned into training and test sets with a ratio of 80:20, and we performed 10-fold cross-validation on the training set to tune the hyperparameters for ML algorithms. We used the same training and testing sets to train and evaluate both ML and BN models, to ensure consistency in the building of the models.

4. Research methodology

In this study, we used a two-level architecture to predict and make treatment recommendations for breast cancer patients. In the first level, we used ML algorithms to classify patients into different survival categories, which allowed us to identify subgroups of patients with specific survival categories (e.g. “>10_yr”). In the second level, we used BN to make treatment recommendations for the identified subgroups based on the patient’s survival category and other relevant factors. Finally, using BNs allowed for the consideration of uncertainty and dependencies between the different variables and facilitated more informed treatment decisions.

To analyze the data with limited patient numbers, this study used three classes for ML and two survival rates for BN analysis. Figure 3 shows the outline of the proposed methodology. In the first level, ML classification algorithms predict whether the patient belongs to “5_yr ≥”, “5–10_yr” or “>10_yr” without including the therapy variable. For example, if the classifier

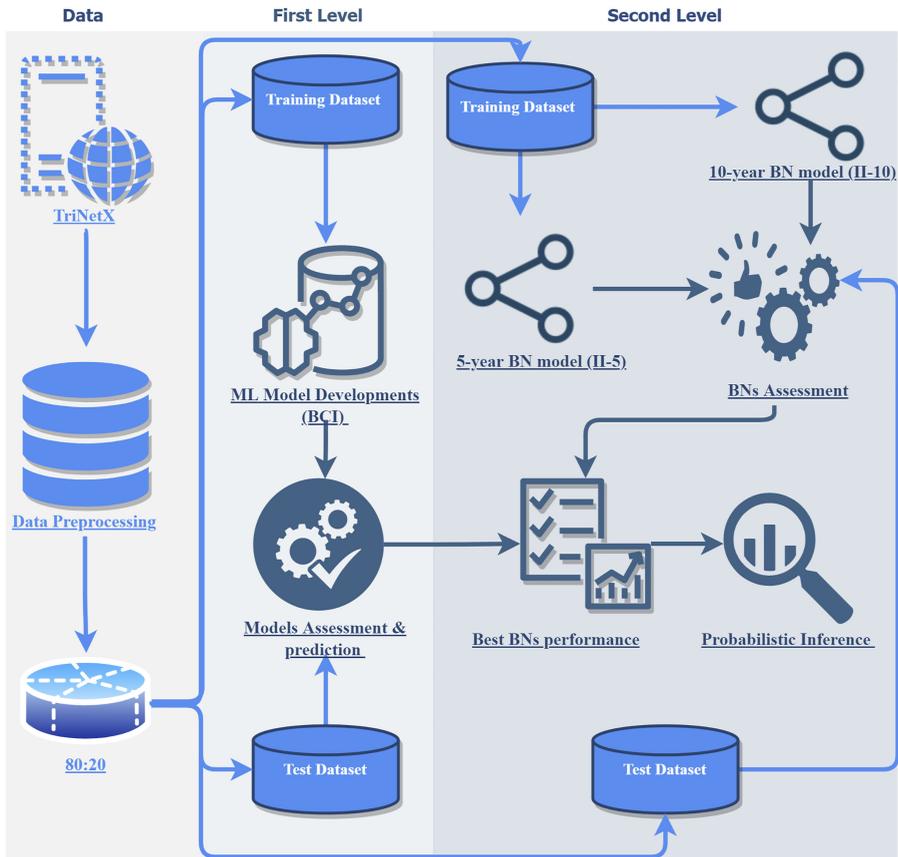


Figure 3.
The outline of the
proposed methodology

predicts that the patient belongs to the “5–10_yr” class, then the second level is responsible for emanating inference from the BN of the five-year survival rate model for the given patient. Therefore, three models were prepared: a first-level classification model (BCI), a second-level BN model for the five-year survival rate (II-5) and a second-level BN model for the ten-year survival rate (II-10). The survival variable (target variable) has three classes in BCI and is binary in II-5 and II-10. For the construction of II-5, patients were labeled with one if they survived more than five years and zero otherwise. In contrast, patients were labeled with one if they survived more than ten years and zero otherwise for the construction of II-10.

4.1 First level

In the first level of the approach, the goal is to assign each patient (i.e. instance) to one of the several predefined categories (e.g. survival periods). To do this, we used ML algorithms designed for classification tasks with multiple classes. Five traditional ML algorithms were applied: logistic regression, random forest, SVM, ANN and naïve Bayes. According to the previous literature on survival prediction, no single traditional classification ML algorithm performs consistently better in all experiments. Thus, these widely applied ML algorithms in survival analysis were used to investigate which one produces better performance results on the gathered dataset. In addition, these algorithms vary in terms of their complexity and flexibility.

4.2 Second level

In the second level of this approach, two BN models were constructed with different survival targets to investigate variations in variable dependencies between ten-year and five-year survivals. The ability to observe differences in the topology and the conditional probability of the models may provide more personalized treatment recommendations based on the specific dependencies between variables for different survival periods. Using multiple models with different survival targets may also facilitate a more comprehensive analysis of the data. In addition, BNs are suitable tools for determining several probabilistic inferences that aid clinical decision-making. However, we focused on the inference of survival given different treatments and observed evidence on patient variables. Mainly, there are two steps for developing a BN: define the network structure and specify a conditional probability table (CPT) for each node. What follows is a brief description of the BN and the two steps (structure and parameter learning) for creating a BN.

4.3 Bayesian networks

A BN is formally defined as a pair (G, Ω) that encodes a joint probability distribution over a finite set of categorical variables (Pearl, 1988). The first component, G , is a directed acyclic graph (DAG) whose nodes resemble the random variables in the dataset and arcs represent direct dependencies between variables. The latter component, Ω , represents CPDs that define each variable behavior given its parents. In addition, the BN has a Markov property, since each variable is conditionally independent of its non-descendant given its parents. Function 1 shows a unique representation of the joint probability distribution.

$$p(X_1, \dots, X_n) = \prod_{i=1}^n p(X_i | P_{X_i}) \quad (1)$$

(1) Structure learning

Mainly, there are three general methods to obtain a DAG structure: (1) manual construction, (2) automatic structure learning and (3) hybrid learning. The first method requires access to human knowledge experts in each development stage. Therefore, we have not included it in

this study – this work emphasizes the second and third methods to learn the network structure.

Automatic learning obtains the structure of a DAG purely from the data. Several methods are adopted to learn BN structure from data, and each may provide a different structure. They can generally be categorized into (1) score-based algorithms, which explore the search space for the DAG with maximum score function and (2) constraint-based algorithms that link nodes based on conditional independence constraints.

Two score-based algorithms were used in this paper. The first is hill-climbing (HC), a greedy search that starts exploring with disconnected DAG by performing a single-arc operation (addition, removal and reversals) to maximize the structure score (Gómez, Mateo, & Puerta, 2011). The algorithm ends when a local maximum is found. The second is tree-augmented naïve Bayes (TAN), which relaxes the naïve Bayes assumption of independency and permits each variable to depend on another variable as well as the target class (Friedman, Geiger, & Goldszmidt, 1997). TAN is a tree-based approach and can learn the structure in polynomial time. We used Bayesian Dirichlet equivalent uniform (Bdeu) and Bayesian Information Criterion (BIC) as scoring functions in both methods. Therefore, we have four approaches for learning the network structure entirely from the dataset: HC(Bdeu), HC(BIC), TAN(Bdeu) and TAN(BIC).

Lastly, we explored hybrid structure learning using Causal Minimum Message Length (CaMML) (Wallace & Korb, 1999), which allows experts to specify prior information to be incorporated with automatic learning. This information can include tier information, which allows the order of variables to be specified ($A < B$; A happens before B, A can be a parent of B, but B cannot be a parent of A), and direct connection, which indicates direct influence ($A \rightarrow B$). Thus, we have five methods to learn the topology of BNs for II-5 and II-10.

(2) Parameter learning

After learning the structure, we used maximum likelihood estimation to estimate the parameters and represent the CPTs in all BN experiments.

4.4 Architecture setting

We utilized the bnlearn package (Taskesen, 2020) in python to learn the structure and the parameters for TAN and HC algorithms. However, we used the BI-CaMML (Wallace, 2014), developed at Monash University, for the hybrid structure learning. We built ML models using the Scikit-learn package in python.

4.5 Performance metrics

Accuracy, precision, F1 score, recall and area under the ROC curve (AUC) were selected as the evaluation metrics for the models built in this study. Equations for the evaluation metrics are defined as follows:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

$$F1 \text{ score} = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (5)$$

5. Experimental results

In this section, we report the experimental results of both level settings mentioned in the previous section regarding accuracy, precision, recall and *F1* score. Also, it presents a comparison between II-5 and II-10 models on probabilistic inference findings when the same evidence is observed.

5.1 Experimental results of the first-level classification models

For this level, we evaluated the performance of five ML algorithms to classify patients into predetermined survival classes. These algorithms were trained using the training dataset and evaluated using the test dataset. We found that the random forest classifier achieved the highest accuracy at 0.85 as well as the highest precision, recall and *F1* score across the three classes. Next, the ANN had an accuracy score of 0.8 with close performance measures to the random forest classifier for the long survival class “>10_yr”. On the other hand, naïve Bayes and logistic regression had the lowest performance measures. The long survival period mostly achieved a better performance than the other survival classes. Additionally, “5_yr ≥” had better performance than the mid-survival class “(5–10_yr)” in the majority of the ML algorithms proposed. Table 2 displays the complete set of results in a tabular format.

As the mission of the first level was to determine which model to use in the second level, a classifier algorithm with the strongest performance results had to be selected. Thus, the random forest classifier was the best fit for the BCI model.

5.2 Experimental results of the second-level BN models

We developed two BNs: one for the five-year survival rate and the other for the ten-year survival rate. Both models used the same dataset except for the target variable for training and testing. In order to discover a network structure for BNs, we tried several network structure learning approaches. At the outset, we used four approaches that automatically learn the network structure from the data and one with a hybrid approach. Therefore, ten experiments were done at this level (2 models * 5 approaches for building the structure). To construct hybrid learning, we used CaMML to utilize the data and prior information to discover the structure. Then, through knowledge gained from previous research, we obtained the prior information used in hybrid learning. Finally, the following set of rules was applied as an expert prior in CaMML.

Model	Class	# Instances		Accuracy	Precision	Recall	F1 score
		Train	Test				
Logistic Regression	5_yr ≤	3,420	856	0.64	0.65	0.67	0.66
	5–10_yr	3,419	857		0.52	0.50	0.51
	>10_yr	3,423	853		0.75	0.76	0.76
Random Forest	5_yr ≤	3,420	856	0.85	0.80	0.81	0.81
	5–10_yr	3,419	857		0.81	0.80	0.80
	>10_yr	3,423	853		0.93	0.93	0.93
SVM	5_yr ≤	3,420	856	0.79	0.76	0.73	0.74
	5–10_yr	3,419	857		0.72	0.71	0.72
	>10_yr	3,423	853		0.89	0.91	0.90
ANN	5_yr ≤	3,420	856	0.81	0.79	0.74	0.76
	5–10_yr	3,419	857		0.75	0.78	0.77
	>10_yr	3,423	853		0.90	0.93	0.92
Naïve Bayes	5_yr ≤	3,420	856	0.56	0.73	0.3	0.43
	5–10_yr	3,419	857		0.4	0.43	0.42
	>10_yr	3,423	853		0.54	0.81	0.65

Table 2.
ML model performance

Table 3 displays the performance in predicting the five-year and ten-year outcomes achieved by the different learning methods. For BNs with the five-year survival model, the network structure learned by CaMML achieved the highest performance measures with an accuracy of 0.798 and AUC of 0.757. Similarly, for the ten-year survival model, the network structure discovered by CaMML had the best performance results of 0.857 and 0.84 for accuracy and AUC, respectively. Despite applying two scoring functions in the automatic approach, they provided the same network structure in the TAN algorithm. On the other hand, the learned network structure determined using the Bdeu scoring function in the HC algorithm had a higher performance than the BIC function for both survival targets. Furthermore, the TAN algorithm for BN with the five-year survival had the second-highest performance. However, the HC algorithm learned via Bdeu had slightly lower performance measures than CaMML for ten-year survival. Thus, we chose the structures learned via CaMML for both survival targets for making the inference. CaMML enables the specification of prior knowledge, such as the assignment of variables to tiers, to improve the accuracy of causal models. For instance, by placing the variable “Survival” in a higher tier, the model will be constrained to not allow it to influence other variables, thus reducing the risk of identifying inaccurate arcs. **Figures 4 and 5** show the selected BN structure for the II-5 and II-10 models, respectively.

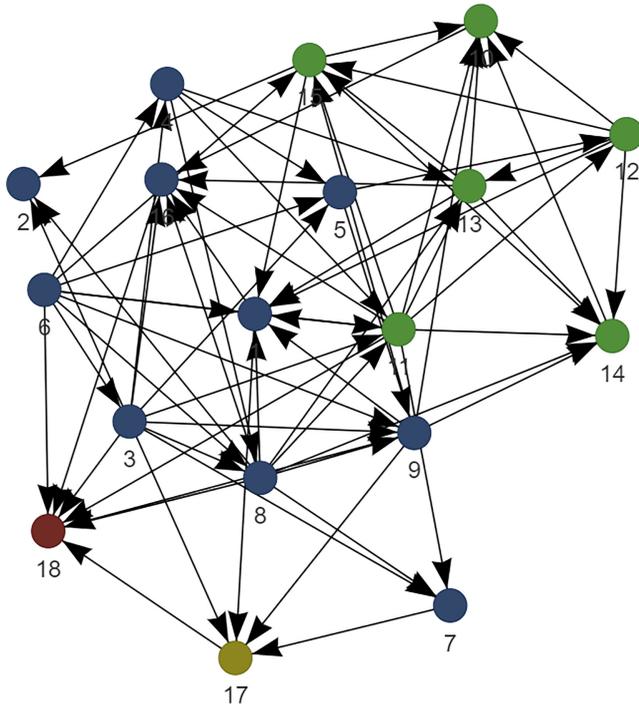
5.3 Inference results

After selecting the BN structure for the II-5 and II-10 models, we explored the influence of different treatments on the probability of survival using these models. For illustration, we explored two cases: (1) $P(>5_yr?/>10_yr? = 1 \mid \text{Treatment, ER} = \text{Positive, Age} = >50, \text{Tumor_stage} = \text{III})$ and (2) $P(>5_yr?/>10_yr? = 1 \mid \text{Treatment, ER} = \text{Negative, Age} = 50\text{--}60, \text{Tumor_stage} = \text{I})$. In **Figures 6 and 7**, the orange bars represent the probability from the II-5 model and the blue bars represent the probability from the II-10 model.

Figure 6 shows that patients who received all three types of treatment (RTx, Chemo and HT) had the highest probability of surviving more than five years, but patients who received only RTx had a better probability of surviving more than ten years. Also, in the second case (**Figure 7**), patients who received all treatments had maximal probability for the five-year survival but lower probability of ten-year survival. Nonetheless, patients who received a single type of treatment had similar outcomes for both survival targets in the second case.

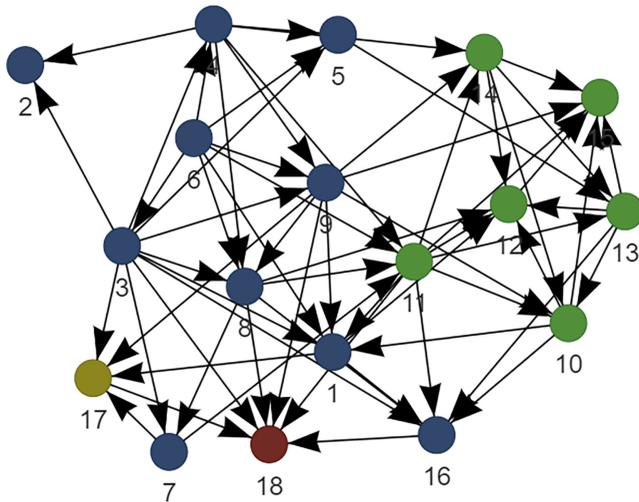
Survival target models	Learning algorithms	# Instances		Accuracy	Precision	Recall	F1 score	AUC
		Train	Test					
Five-year	TAN (BIC)	5_yr = 0		0.785	0.813	0.884	0.847	0.731
	TAN (Bdeu)	3,420	856	0.785	0.813	0.884	.847	0.731
	HC (BIC)	5_yr = 1		0.756	0.779	0.890	0.831	0.684
	HC (Bdeu)	6,842	1,710	0.779	0.809	0.880	0.843	0.724
	CaMML			0.798	0.825	0.881	0.852	0.757
Ten-year	TAN (BIC)	10_yr = 0		0.826	0.739	0.704	0.721	0.794
	TAN (Bdeu)	6,839	1,713	0.826	0.740	0.703	0.721	0.794
	HC (BIC)	10_yr = 1		0.809	0.703	0.695	0.699	0.779
	HC (Bdeu)	3,423	856	0.849	0.761	0.765	0.763	0.826
	CaMML			0.857	0.765	0.796	0.780	0.84

Table 3.
BNs performance comparison



Note(s): Red node; survival variable, yellow node; treatments, green; health conditions, blue; cancer conditions and marriage status

Figure 4.
CaMML structure for
five-year survival



Note(s): Red node; survival variable, yellow node; treatments, green; health conditions, blue; cancer condition and marriage status

Figure 5.
CaMML structure for
ten-year survival

DTS
2,2

174

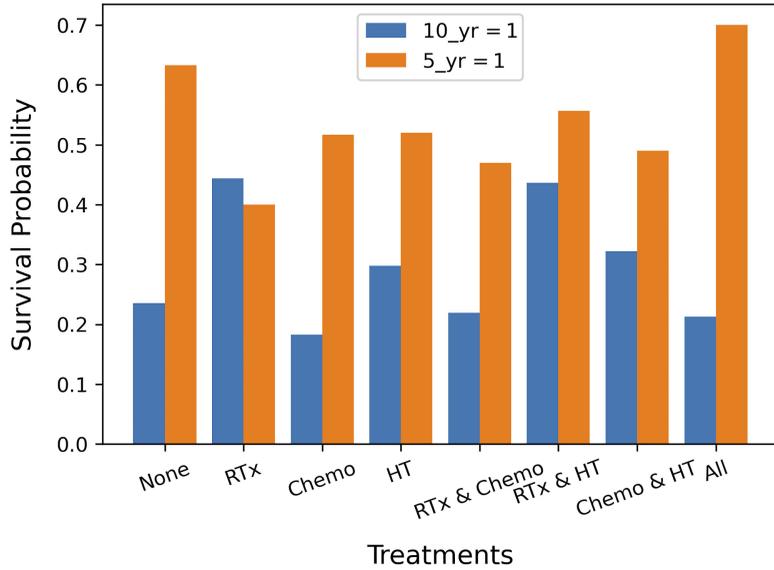


Figure 6.
First case of inference
result

Note(s): $P(> 5_yr? / > 10_yr? = 1 \mid \text{Treatment, ER = Positive, Age} = < 50, \text{Tumor_stage} = \text{III})$

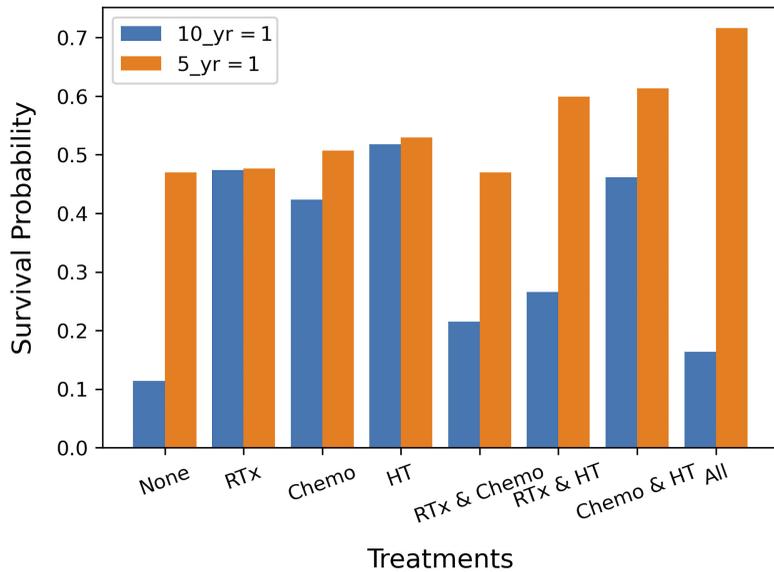


Figure 7.
Second case of
inference result

Note(s): $P(> 5_yr? / > 10_yr? = 1 \mid \text{Treatment, ER = Negative, Age} = 50 - 60, \text{Tumor_stage} = \text{I})$

6. Discussion

Clinical decision support tools aid clinical decision-making (Hunt, Haynes, Hanna, & Smith, 1998) and further help improve practitioner performance (Garg, Adhikari, & Devereaux, 2005). In addition, the medical practitioner can use accurate survivability prediction for various treatment options as a reference tool to analyze decisions. As the understanding of cancer therapy progresses, the type of treatment available varies; as a result, the complexity of optimal treatment planning for a particular patient also increases. Thus, the proposed model can be included as a part of decision-making support for medical professionals. Once the model is integrated into the hospital information system, physicians can evaluate treatments versus survivability odds and make an informed decision.

When prioritizing multiple treatment options, patient preference is often incorporated. For example, a study that integrates patient preference and treatment decisions for prostate cancer patients found that patients felt more involved and mutually responsible with oncologists when additional treatment was needed (Johnson *et al.*, 2016). Therefore, patients could also benefit from the model to weigh the outcome of different treatments when listing their preferences.

In addition, this model may not only positively affect the patient outcome but also the efficiency of the healthcare provider. Knowing what survival outcomes might happen will bring knowledge to the caregiver to utilize their resources better and help them improve their patients' health. Moreover, linking EHR with the decision support tool will provide a recommendation at the exact time it is needed and enhance the model performance as more information enters the system.

Our approach starts with applying various ML algorithms that vary in accuracy and interpretability. For instance, ANN is generally considered highly accurate but difficult to interpret, whereas logistic regression is the opposite. Nevertheless, the first level's primary goal was to predict a patient's survival to select the appropriate BN model. Then, the second level provides the probabilistic inference. This approach integrates an accurate classification algorithm with the interpretable nature of BNs and offers multiple benefits for using BN models with different survival targets.

First, we avoided incorporating bounded survival periods while including different survival rates. Especially for the mid-survival (5–10_{yr}), it is desirable to expect the survivability of a treatment to be an unbounded period. For instance, it is more optimistic to inform patients of an expectation of living more than five years than between five and ten years. Additionally, this approach enables us to involve more patients with higher survivability periods when training the BN.

Second, the network structure may differ for different survival targets when learning partially or entirely from data. Wen *et al.* (2017) and Franco, Steyerberg, Hu, Mackenbach and Nusselder (2007) link diabetes and other medical conditions with shorter life expectancy. When higher survivability targets are used, such as 15-year or 20-year, survival may be heavily determined by medical conditions and age rather than stage or treatments. However, we did not find a direct relationship between the health conditions we evaluated and survival in our network structures with the exception of personal history of breast cancer. Another reason why longer survival should be considered is that several cancer types have a more than 90% chance of surviving for more than five years, such as breast and skin cancers.

Third, and most importantly, treatment recommendations differ for five-year and ten-year survival targets when the same evidence is observed. As a result, creating a BN with a single target does not discover the whole picture for treatment outcomes. In the two inference cases presented here, patients who received all three types of treatment were more likely to be associated with survival of more than five years but less than 10.

Although BNs are able to perform both classification and inference tasks, using a single BN model may limit the ability to discover the differences in dependency between variables

and may not be as effective when there are more classes. This is because a single BN model will rely on explicit assumptions about the relationships between the variables, which may not accurately reflect the true dependencies between the variables in all cases. Using multiple BN models and appropriate model selection techniques can allow for a more comprehensive analysis of the data and can help to identify any discrepancies or inconsistencies in the dependencies between variables, particularly when there are more classes.

Nonetheless, we acknowledge that the experiments described here have several limitations. Foremost, our study includes only diseased patients, as it is the only method to compute the survivability period. As a result, few patients fall into the long survivability class, and therefore there is not enough information to discover accurate factors for this group. Moreover, surgery information was not included in treatment plans. In addition, this approach lacked the sequence of treatments and the time span between them.

7. Conclusions

This study presents a novel approach to providing clinical decision support for treatment selection for breast cancer patients considering survival expectations. It starts with utilizing ML algorithms to predict the survivability class. Subsequently, it employs the appropriate BN model for probabilistic inference. It was prominent that the treatment preference relies on the BN's survival targets, as shown in our inference results. In the two cases presented, five-year and ten-year targets were associated with different treatments that yielded a high probability of survival. In the end, this decision support tool does not replace the intuition or judgment of the medical practitioner, but it could serve as a reference for physicians and as a resource to educate and involve patients in their treatment decisions to ultimately enhance the health of the patient. The results of this study provide sufficient motivation to pursue the proposed approach, potentially with more survival classes and for different cancer types, in order to provide a more comprehensive analysis of the data and gain insight into the dependencies between variables for different cancer types and survival classes.

Incorporating temporal data is critical to providing a sequence of treatments. This also opens another channel to improve the performance of survival models and treatment plan selection with appropriate intervention time. Thus, the plan to include temporal information in the future work to explore questions like “How does the sequence of treatments change the probability of survival?” and “When is the best intervention time?”

References

- Berry, D. A., Cronin, K. A., Plevritis, S. K., Fryback, D. G., Clarke, L., Zelen, M., . . . Feuer, E. J. (2005). Effect of screening and adjuvant therapy on mortality from breast cancer. *The New England Journal of Medicine*, *353*, 1784–1792. doi: [10.1056/NEJMoa050518](https://doi.org/10.1056/NEJMoa050518).
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). Smote: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. doi: [10.1613/jair.953](https://doi.org/10.1613/jair.953).
- Choi, J. P., Han, T. H., & Park, R. W. (2009). A hybrid Bayesian network model for predicting breast cancer prognosis. *Journal of Korean Society of Medical Informatics*, *15*, 49. doi: [10.4258/jksmi.2009.15.1.49](https://doi.org/10.4258/jksmi.2009.15.1.49).
- Cruz-Ramírez, N., Acosta-Mesa, H. G., Carrillo-Calvet, H., Alonso Nava-Fernández, L., & Barrientos-Martínez, R. E. (2007). Diagnosis of breast cancer using Bayesian networks: A case study. *Computers in Biology and Medicine*, *37*, 1553–1564. doi: [10.1016/j.compbiomed.2007.02.003](https://doi.org/10.1016/j.compbiomed.2007.02.003).
- Delen, D., Walker, G., & Kadam, A. (2005). Predicting breast cancer survivability: A comparison of three data mining methods. *Artificial Intelligence in Medicine*, *34*, 113–127. doi: [10.1016/j.artmed.2004.07.002](https://doi.org/10.1016/j.artmed.2004.07.002).

- Endo, A., Shibata, T., & Tanaka, H. (2008). Comparison of Seven Algorithms to Predict breast cancer survival. *Journal of Biomedical Fuzzy Systems Association*, 13, 6.
- Evans, R. S. (2016). Electronic health records: Then, now, and in the future. *Yearbook of Medical Informatics*, (Supp. 1), S48–S61. doi:[10.15265/YYS-2016-s006](https://doi.org/10.15265/YYS-2016-s006).
- Forsberg, J. A., Eberhardt, J., Boland, P. J., Wedin, R., & Healey, J. H. (2011). Estimating survival in patients with operable skeletal metastases: An application of a Bayesian belief network. *PLoS One*, 6, e19956. doi: [10.1371/journal.pone.0019956](https://doi.org/10.1371/journal.pone.0019956).
- Franco, O. H., Steyerberg, E. W., Hu, F. B., Mackenbach, J., & Nusselder, W. (2007). Associations of diabetes mellitus with total life expectancy and life expectancy with and without cardiovascular disease. *Archives of Internal Medicine*, 167, 1145–1151. doi: [10.1001/archinte.167.11.1145](https://doi.org/10.1001/archinte.167.11.1145).
- Friedman, N., Geiger, D., & Goldszmidt, M. (1997). Bayesian network classifiers. *Machine Learning*, 29, 131–163. doi: [10.1023/A:1007465528199](https://doi.org/10.1023/A:1007465528199).
- Gámez, J. A., Mateo, J. L., & Puerta, J. M. (2011). Learning Bayesian networks by hill climbing: Efficient methods based on progressive restriction of the neighborhood. *Data Mining and Knowledge Discovery*, 22, 106–148. doi: [10.1007/s10618-010-0178-6](https://doi.org/10.1007/s10618-010-0178-6).
- Garg, A. X., Adhikari, N. K., & Devereaux, P. J. (2005). Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: A systematic review. *Journal of the American Medical Association*, 293, 16.
- Gevaert, O., Smet, F. D., Timmerman, D., Moreau, Y., & Moor, B. D. (2006). Predicting the prognosis of breast cancer by integrating clinical and microarray data with Bayesian networks. *Bioinformatics*, 22, e184–e190. doi: [10.1093/bioinformatics/btl230](https://doi.org/10.1093/bioinformatics/btl230).
- Hunt, D. L., Haynes, R. B., Hanna, S. E., & Smith, K. (1998). Effects of computer-based clinical decision support systems on physician performance and patient outcomes: A systematic review. *Journal of the American Medical Association*, 280, 1339. doi: [10.1001/jama.280.15.1339](https://doi.org/10.1001/jama.280.15.1339).
- Imani, F., Chen, R., Tucker, C., & Yang, H. (2019). Random forest modeling for survival analysis of cancer recurrences. In *2019 IEEE 15th International Conference on Automation Science and Engineering (CASE). Presented at the 2019 IEEE 15th International Conference on Automation Science and Engineering (CASE)* (pp. 399–404). doi: [10.1109/COASE.2019.8843271](https://doi.org/10.1109/COASE.2019.8843271).
- Islami, F., Guerra, C. E., Minihan, A., Yabroff, K. R., Fedewa, S. A., Sloan, K., . . . Jemal, A. (2022). American Cancer Society's report on the status of cancer disparities in the United States, 2021. *Cancer Journal for Clinicians*, 72, 112–143. doi: [10.3322/caac.21703](https://doi.org/10.3322/caac.21703).
- Jayasurya, K., Fung, G., Yu, S., Dehing-Oberije, C., De Ruyscher, D., Hope, A., . . . Dekker, A. L. A. J. (2010). Comparison of Bayesian network and support vector machine models for two-year survival prediction in lung cancer patients treated with radiotherapy. *Medical Physics*, 37, 1401–1407. doi: [10.1118/1.3352709](https://doi.org/10.1118/1.3352709).
- Johnson, D. C., Mueller, D. E., Deal, A. M., Dunn, M. W., Smith, A. B., Woods, M. E., . . . Nielsen, M. E. (2016). Integrating patient preference into treatment decisions for men with prostate cancer at the point of care. *Journal of Urology*, 196, 1640–1644. doi: [10.1016/j.juro.2016.06.082](https://doi.org/10.1016/j.juro.2016.06.082).
- Kahn, C. E., Roberts, L. M., Shaffer, K. A., & Haddawy, P. (1997). Construction of a Bayesian network for mammographic diagnosis of breast cancer. *Computers in Biology and Medicine*, 27, 19–29. doi: [10.1016/S0010-4825\(96\)00039-X](https://doi.org/10.1016/S0010-4825(96)00039-X).
- Kotsiantis, S. B., Kanellopoulos, D., & Pintelas, P. E. (2006). Data preprocessing for supervised learning. *International Journal of Computer Science*, 1, 111–117.
- Li, J., Zhou, Z., Dong, J., Fu, Y., Li, Y., Luan, Z., & Peng, X. (2021). Predicting breast cancer 5-year survival using machine learning: A systematic review. *PLOS One*, 16, e0250370. doi: [10.1371/journal.pone.0250370](https://doi.org/10.1371/journal.pone.0250370).
- Lotfnezhad Afshar, H., Ahmadi, M., Roudbari, M., & Sadoughi, F. (2015). Prediction of breast cancer survival through knowledge discovery in databases. *Global Journal of Health Science*, 7, 392. doi: [10.5539/gjhs.v7n4p392](https://doi.org/10.5539/gjhs.v7n4p392).

- Murphy, E. V. (2014). Clinical decision support: Effectiveness in improving quality processes and clinical outcomes and factors that may influence success. *Yale Journal of Biology and Medicine*, 87, 187–197.
- Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Morgan Kaufmann.
- Pedersen, R. N., Esen, B. Ö., Mellemkjær, L., Christiansen, P., Ejlersen, B., Lash, T. L., . . . Cronin-Fenton, D. (2022). The incidence of breast cancer recurrence 10-32 Years after primary diagnosis. *JNCI Journal of the National Cancer Institute*, 114, 391–399. doi: [10.1093/jnci/djab202](https://doi.org/10.1093/jnci/djab202).
- Sesen, M. B., Nicholson, A. E., Banares-Alcantara, R., Kadir, T., & Brady, M. (2013). Bayesian networks for clinical decision support in lung cancer care. *PLoS One*, 8, e82349. doi: [10.1371/journal.pone.0082349](https://doi.org/10.1371/journal.pone.0082349).
- Shin, H., & Nam, Y. (2014). A coupling approach of a predictor and a descriptor for breast cancer prognosis. *BMC Medical Genomics*, 7, S4. doi: [10.1186/1755-8794-7-S1-S4](https://doi.org/10.1186/1755-8794-7-S1-S4).
- Siegel, R. L., Miller, K. D., Fuchs, H. E., & Jemal, A. (2022). Cancer statistics, 2022. *Cancer Journal for Clinicians*, 72, 7–33. doi: [10.3322/caac.21708](https://doi.org/10.3322/caac.21708).
- Stojadinovic, A., Bilchik, A., Smith, D., Eberhardt, J. S., Ward, E. B., Nissan, A., . . . Steele, S. R. (2013). Clinical decision support and individualized prediction of survival in colon cancer: Bayesian belief network model. *Annals of Surgical Oncology*, 20, 161–174. doi: [10.1245/s10434-012-2555-4](https://doi.org/10.1245/s10434-012-2555-4).
- Taskesen, E. (2020). *Learning Bayesian Networks with the bnlearn Python Package*. (Version 0.3.22) [Computer software]. Available from: <https://erdogant.github.io/bnlearn>
- Wallace, C. (2014). *Causal discovery via MML*. (Version 1.4.2) [Computer software], Monash University.
- Wallace, C. S., & Korb, K. B. (1999). Learning linear causal models by MML sampling. In *Causal Models and Intelligent Data Management*, 89–111. doi: [10.1007/978-3-642-58648-4_7](https://doi.org/10.1007/978-3-642-58648-4_7).
- Wen, C. P., Chang, C. H., Tsai, M. K., Lee, J. H., Lu, P. J., Tsai, S. P., . . . Wu, X. (2017). Diabetes with early kidney involvement may shorten life expectancy by 16 years. *Kidney International*, 92, 388–396. doi: [10.1016/j.kint.2017.01.030](https://doi.org/10.1016/j.kint.2017.01.030).
- Witteveen, A., Nane, G. F., Vliegen, I. M. H., Siesling, S., & IJzerman, M. J. (2018). Comparison of logistic regression and Bayesian networks for risk prediction of breast cancer recurrence. *Medical Decision Making*, 38, 822–833. doi: [10.1177/0272989X18790963](https://doi.org/10.1177/0272989X18790963).

Corresponding author

Robin Qiu can be contacted at: robinqiu@psu.edu