Original Article

MDM Medical Decision Making

Medical Decision Making
2019, Vol. 39(8) 1032–1044

DOI: 10.1177/0272989X19879095
journals.sagepub.com/home/mdm
SAGE

Probabilistic Graphical Modeling for Estimating Risk of Coronary Artery Disease: Applications of a Flexible Machine-Learning Method

Alind Gupta, Justin J. Slater, Devon Boyne, Nicholas Mitsakakis, Audrey Béliveau, Marek J. Druzdzel, Darren R. Brenner, Selena Hussain, and Paul Arora

Objectives. Coronary artery disease (CAD) is the leading cause of death and disease burden worldwide, causing 1 in 7 deaths in the United States alone. Risk prediction models that can learn the complex causal relationships that give rise to CAD from data, instead of merely predicting the risk of disease, have the potential to improve transparency and efficacy of personalized CAD diagnosis and therapy selection for physicians, patients, and other decision makers. **Methods.** We use Bayesian networks (BNs) to model the risk of CAD using the Z-Alizadehsani data set—a published real-world observational data set of 303 Iranian patients at risk for CAD. We also describe how BNs can be used for incorporation of background knowledge, individual risk prediction, handling missing observations, and adaptive decision making under uncertainty. **Results.** BNs performed on par with machine-learning classifiers at predicting CAD and showed better probability calibration. They achieved a mean 10-fold area under the receiver-operating characteristic curve (AUC) of 0.93 ± 0.04, which was comparable with the performance of logistic regression with L1 or L2 regularization (AUC: 0.92 ± 0.06), support vector machine (AUC: 0.92 ± 0.06), and artificial neural network (AUC: 0.91 ± 0.05). We describe the use of BNs to predict with missing data and to adaptively calculate prognostic values of individual variables under uncertainty. **Conclusion.** BNs are powerful and versatile tools for risk prediction and health outcomes research that can complement traditional statistical techniques and are particularly useful in domains in which information is uncertain or incomplete and in which interpretability is important, such as medicine.

Keywords
artificial intelligence, Bayesian networks, Bayesian statistics, cardiology, coronary artery disease, graphical models, health economics and outcomes research (HEOR), machine learning, risk modeling, risk prediction, statistical models

Date received: February 4, 2019; accepted: August 28, 2019

With the ever-increasing availability of electronic health records populating real-world evidence databases and the simultaneous increase in the number of variables available for modeling, incorporation of machine-learning techniques may be able to improve traditional methods of disease risk prediction. One such area is cardiovascular diseases (CVDs), which are the leading causes of mortality and disease burden worldwide.[1] Coronary artery disease (CAD) is the most common CVD. It is caused by reduced blood flow to the heart due to the buildup of plaque in coronary arteries that can lead to chest pain and myocardial infarction.[2] In 2016, CAD was the cause of 1

Corresponding Author
Paul Arora, Lighthouse Outcomes. 1 University Avenue, 3rd Floor, Toronto, ON, M5J 2P1, Canada (parora@lighthouseoutcomes.com).

in 7 deaths and an economic burden of $199.6 billion in the United States alone.[3] Several decades of research have identified risk factors associated with CAD such as hypertension, high total cholesterol, dietary sodium, and smoking[4] that have been assimilated into clinical guidelines to guide risk stratification, interventions, and management of disease for patients with suspected or diagnosed CAD.[5]

According to a 2012 study, the American College of Cardiology (ACC) and its European counterparts recommend using either the Diamond and Forrester model[6] or the Duke clinical score[7] to estimate the probability of CAD in patients with chest pain without a history of CAD, despite these models overestimating CAD risk.[8] Decision analyses have also reported that the choice of diagnostic investigation in patients with chest pain is guided mainly by the estimated risk of CAD, and therefore, improved CAD risk estimation may greatly enhance clinical care.[9,10] A 2010 report based on a US national cardiovascular data registry of the ACC found that only 41% of patients without history of CAD who underwent elective catheter-based coronary angiography procedures were diagnosed with obstructive CAD, prompting calls for improvements in CAD risk stratification that could increase the diagnostic yield of the test and reduce associated costs and risk from the invasive procedure and incidental findings.[11] A large retrospective pooled analysis of existing cohorts concluded that updated prediction models including age, sex, symptoms, and CAD risk factors can allow for more accurate estimation of the pretest probability of CAD in low-prevalence populations and that the addition of coronary calcium scores to the prediction models improves estimation.[8] With the unprecedented amount of data available about patient demographics, biomarkers, genetics, and medical history, machine-learning methods that can learn from data should be explored as supplements to traditional rule-based risk prediction models.

Bayesian networks (BNs) are probabilistic graphical models that are well suited for reasoning under uncertainty inherent in medicine and biology.[12,13] Developed in the late 1980s,[14] BNs are widely used in the fields of finance, engineering, and computer science but are relatively new tools in the health and outcomes research space. BNs offer several advantages over regression-based methods, such as the ability to relax assumptions of a rigid relationship between independent covariates and a target variable by identifying these relationships from data.[13] BNs explicitly model conditional dependence relationships between all variables of interest as a joint probability distribution (JPD), and data can be used to learn these relationships. This can yield more intuitive probabilistic models that are useful for causal reasoning,[15] risk prediction, and evidence-based decision analysis under uncertainty[16] and that can handle missing data, correlated predictors, and multiple correlated outcomes. BNs consist of qualitative and quantitative components. The qualitative component is a directed acyclic graph (DAG) made of nodes representing random variables and edges representing direct probabilistic or causal relationships between nodes. The quantitative component represents the set of parameters associated with each node in the network that specifies its probability distribution conditional on the nodes that have an edge directed to it (i.e., its parents). These components can be learned from data computationally, constructed from existing knowledge manually (e.g., using published research or expert opinion), or a combination of both. A detailed statistical description of BNs is outside the scope of this article and can be found elsewhere.[13,14,17] Here, we focus on applying BN methodology for risk prediction using CAD data as an example.

We completed a scoping review of the literature applying BNs for risk estimation in CAD demonstrating their use for prediction for various data sets (Supplementary Table S1). For example, Chen et al.[18] found BNs to perform on par with support vector machines (SVMs), artificial neural networks (ANNs), and decision trees for predicting CAD risk with the added advantage of revealing relationships between variables that captured known causal relationships from literature. Huan et al.[19] and Mäkinen et al.[20] used BNs to discover gene regulatory networks that contribute to CAD risk. Others have used BNs for modeling both longitudinal and cross-sectional data containing information on demographics, clinical profiles, and biomarkers to identify key drivers of CAD risk and disease progression and the relationships between them.[21–23] Therefore, BNs show promise at being able to discover meaningful dependencies between variables of interest and in analyzing complex data for disease risk.

Our objective in this article is to demonstrate how BNs can be used to develop a model for predicting CAD risk

Lighthouse Outcomes, Toronto, ON, Canada (AG, JJS, DB, DRB, PA); Cumming School of Medicine, University of Calgary, Calgary, AB, Canada (DB, DRB); Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada (NM, SH, PA); Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada (NM); Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, ON, Canada (AB); Faculty of Computer Science, Bialystok University of Technology, Bialystok, Poland (MJD). The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article. The authors received no financial support for the research, authorship, and/or publication of this article.

from a real-world data set of visitors to a cardiovascular research center in Iran.[24,25] We aim to describe the process of using BNs for discovering interactions between variables, belief updating with partial observations, risk prediction in comparison, and adaptive decision making based on the prognostic value PV of variables for predicting CAD.

Methods

Description of the Extended Z-Alizadehsani Data Set and Preprocessing

The extended Z-Alizadehsani data set is a publicly available real-world data set from the UCI Machine Learning Repository.[26] It contains records of 303 visitors to Rajaei Cardiovascular, Medical and Research Center in Tehran, Iran, who were tested for CAD using angiography. It contains 59 total variables, including 4 target variables. Variables present in the data set can be categorized into demographic information, symptoms, electrocardiography results, biomarkers, and indicators of CAD (Table 1; Supplementary Table S2). Of the 303 visitors, 216 (71%) were diagnosed with CAD. The Z-Alizadehsani data set has previously been used to evaluate machine-learning methods because of its completeness and clinical relevance.[27–29]

We discarded variables containing a single unique value, namely, exertional CP. The extended data set contains 4 target variables: LAD, LCX, and RCA, which are indicators of stenosis for 3 different carotid arteries, and Cath, which is the result of angiography and can be either CAD or normal. A patient was diagnosed with CAD if angiography revealed that his or her artery diameter narrowing was $\geq$50% (i.e., at least 1 of LAD, LCX, or RCA was stenotic). We retained only the single target variable "Cath" as an aggregate indicator of stenosis for this article.

BN Modeling

For a set of N categorical variables $\{X_1, X_2, \ldots, X_N\}$, the JPD $P(X_1, X_2, \ldots, X_N) = \prod_{i=1}^{N} P(X_i|X_{i+1}, \ldots, X_N)$. Given a DAG over the variables, we define for any variable X_i its parent set $Pa(X_i)$ containing all nodes that have a directed edge ($\rightarrow$), which represents a probabilistic or causal relationship, to X_i. For a BN induced by a DAG, the JPD factorizes as follows: $P(X_1, X_2, \ldots, X_N) = \prod_{i=1}^{N} P(X_i|X_{i+1}, \ldots, X_N) = \prod_{i=1}^{N} P(X_i|Pa(X_i))$. For example, given the 2-node BN with the configuration $X_1 \rightarrow X_2$, the JPD $P(X_1, X_2) = P(X_2|X_1)P(X_1)$. In a BN, any variable X_i is independent of all other nodes conditioned on its Markov blanket $MB(X_i)$, which is the set containing $Pa(X_i)$, the children of X_i (i.e., have a directed edge from X_i), and parents of children of X_i. $MB(X_i)$, if fully observed, provides all the information about X_i in a BN.

BN modeling is performed in 2 steps: step 1 is called "structure learning" and involves learning a DAG from patterns in the data according to some user-specified metric, and step 2 involves estimation of the numerical parameters associated with each node in the network. There are several algorithms for learning BNs from data,[30–33] common implementations of most of which assume either multinomial or Gaussian factorizations of JPD for reasons of computational tractability. We used discrete BNs for modeling because they lack assumptions of normality, linear relationships, and because their parameters can be represented as probability tables that are easy to understand for nonexperts.[34] For use in discrete BNs, we discretized continuous variables using clinically relevant cutoffs where available (for body mass index [BMI] and biomarkers) or else by discretizing them by sample quantiles into 3 bins (for age, height, and weight) to minimize the number of parameters to estimate.

For BN structure learning, we used the hill-climbing algorithm implemented in the R package *bnlearn*[35] for minimizing the Bayesian information criterion), a penalized maximum likelihood metric. The hill-climbing algorithm is a greedy local search technique for optimization that performs sequential edge modifications (edge additions, reversals, and deletions) to find a DAG that maximizes a given network score.[13] To recover a robust structure for Figure 1, we calculated significant edges from 10,000 bootstrapped replicates of data and averaged them.[36,37] For the BN in Figure 1, incoming edges to age and sex were blacklisted. Numerical parameters were learned from data using Bayesian estimation with a Dirichlet prior distribution over parameters and equivalent sample size of 1.[35] In our DAG learned algorithmically from observational data, edges represent probabilistic but not necessarily causal relationships, and edge orientations do not imply cause-and-effect relationships but the orientation that was most commonly identified in high-scoring networks by hill climbing in the bootstrap set.

For comparison, we also tested an edgeless BN, in which all variables are assumed to be independent as a control, and the commonly used naïve Bayes classifier, which assumes a constrained model that assumes that all covariates are independent and the target variable is dependent on all covariates.

Table 1

Feature Type	Feature Name	Feature Description	Range
Demographic	Age	Age (years)	30–86
	Weight	Weight (kg)	48–120
	Length	Height (cm)	140–188
	Sex	Gender	Male, Female
	BMI	Body mass index (kg/m^2)	18–41
	DM	Has diabetes mellitus	Yes, No
	HTN	History of hypertension	Yes, No
	Current smoker	Is a current smoker	Yes, No
	Ex-smoker	History of cigarette consumption	Yes, No
	FH	History of heart disease in first-degree relatives	Yes, No
	Obesity	Yes if BMI >25, else no	Yes, No
	CRF	Chronic renal failure	Yes, No
	CVA	Cerebrovascular accident	Yes, No
	Airway disease	Has airway disease	Yes, No
	Thyroid disease	Has thyroid disease	Yes, No
	CHF	Congestive heart failure	Yes, No
	DLP	Dyslipidemia	Yes, No
Symptoms	BP	Blood pressure (mm Hg)	90–190
	PR	Pulse rate (ppm)	50–110
	Edema	Has edema	Yes, No
	Weak peripheral pulse	Has weak pulse in extremities	Yes, No
	Lung rales	Has lung rales (crackles)	Yes, No
	Systolic murmur	Has heart murmurs during systole	Yes, No
	Diastolic murmur	Has heart murmurs during diastole	Yes, No
	Typical chest pain	Has typical chest pain	Yes, No
	Dyspnea	Has dyspnea	Yes, No
	Function class	NYHA functional classification	1, 2, 3, 4
	Atypical	Has atypical chest pain	Yes, No
	Nonanginal CP	Nonanginal chest pain	Yes, No
	Exertional CP	Exertional chest pain	Yes, No
	Low Th Ang	Low-threshold angina	Yes, No
Electrocardiography	Q wave	Q wave	Yes, No
	ST elevation	Elevated ST segment	Yes, No
	ST depression	Depressed ST segment	Yes, No
	T inversion	T wave inversion	Yes, No
	LVH	Left ventricular hypertrophy	Yes, No
	Poor R progression	Poor R wave progression	Yes, No
	BBB	Bundle branch block	Left, Right, None
Biomarkers	FBS	Fasting blood sugar (mg/dL)	62–400
	Cr	Creatinine (mg/dL)	0.5–2.2
	TG	Triglyceride (mg/dL)	37–1050
	LDL	Low-density lipoprotein (mg/dL)	18–232
	HDL	High-density lipoprotein (mg/dL)	15–111
	BUN	Blood urea nitrogen (mg/dL)	6–52
	ESR	Erythrocyte segmentation rate (mm/h)	1–90
	HB	Hemoglobin (g/dL)	8.9–17.6
	K	Potassium (mEq/L)	3.0–6.6
	Na	Sodium (mEq/L)	128–156
	WBC	White blood cell (cells/mL)	3700–18,000
	Lymph	Lymphocyte (%)	7–60
	Neut	Neutrophil (%)	32–89
	PLT	Platelet (1000/mL)	25–742
	EF	Ejection fraction (%)	15–60
	Region RWMA	Regional wall motion abnormality	0, 1, 2, 3, 4
	VHD	Valvular heart disease	Normal, Mild, Moderate, Severe

(continued)

Table 1 (continued)

Feature Type	Feature Name	Feature Description	Range
Outcome	Cath	Cardiovascular disease, result of angiography	CAD, Normal
	LAD	Stenosis: left anterior descending	Stenotic, Normal
	LCX	Stenosis: left circumflex artery	Stenotic, Normal
	RCA	Stenosis: right coronary artery	Stenotic, Normal

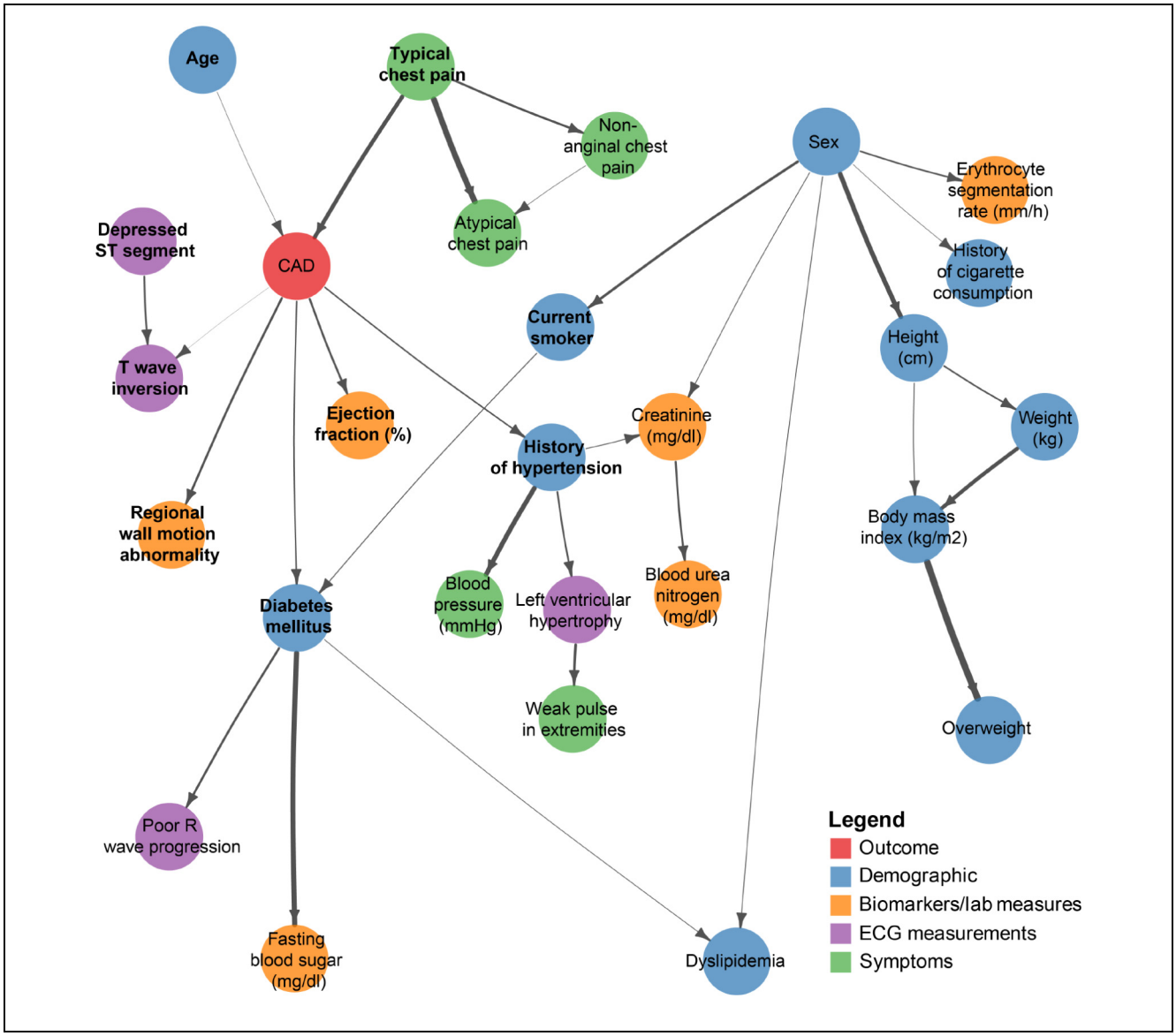

Figure 1 Bayesian network structure of the extended Z-Alizadehsani data set. Nodes are colored according to their grouping from Table 1. Variables in the Markov blanket of coronary artery disease (red node) are shown in boldface. Edge line widths correspond to log-transformed mutual information between adjacent nodes, with a thicker edge representing stronger statistical dependence. For simplicity, unconnected nodes are not shown.

Logistic Regression Analysis

For comparing the performance of BNs for CAD risk prediction, we used logistic regression with L2 penalty (also known as Ridge regression) or L1 penalty (also known as Lasso). Regularization involves specification of a penalty term that helps with learning less complex models that generalize better to unseen data.[38] For L2 and L1 regularization, the penalty parameter that controls the strength of regularization was tuned using cross-validation (CV) over the training set using the *glmnet*[39] package. Continuous variables were not discretized prior to regression analysis because it is unnecessary.

SVM

The implementation of SVM with a radial kernel (*svmRadial*) from the *caret* package[40] was used. The radial kernel yielded a slightly higher area under the receiver-operating characteristic curve (AUC) compared with a linear kernel. Data were centered and scaled for use with SVMs.

ANN

The *nnet* implementation of ANNs in *caret* was used. The decay (regularization) rate was chosen to be 0.5 based on CV as it yielded the highest test set accuracy. The optimal number of units in the hidden layer (2 to 6) was chosen using a grid search over the training set. ANNs were also tested using the *keras* package with varying depth, size of hidden layers, and regularization (dropout and L2 penalty); however, no combination of hyperparameters that we tested yielded a higher AUC than the *caret* implementation. Data were centered and scaled for use with ANNs.

Measuring Predictive Performance: Classification and Calibration

We aimed to comply with TRIPOD guidelines for transparent reporting of risk prediction models.[41]

We used 10 times iterated 10-fold CV to measure the performance of BNs and regression models for predicting CAD. For 10-fold CV, the data set is divided into 10 equally sized subsets. A single subset (called "test set") is used to assess the performance metrics of a model trained on the other 9 subsets (called "training set"). This step is repeated 10 times, once for each subset of data, and the resulting metrics are averaged. Tenfold CV is a common method for model validation and testing and is useful for measuring the capacity of a model to generalize to unseen data. In all cases, models were fitted only on training data and never on the complete data set to prevent optimism in calculated classification metrics.

Using 10-fold CV, we calculated the following performance metrics: true-positive rate (TPR), false-positive rate (FPR), and AUC (mean AUC ± standard deviation from 10 independent iterations of 10-fold CV). TPR is the proportion of true-positive instances (Cath = CAD) that are classified correctly, while FPR is the proportion of negative instances (Cath = Normal) that are classified correctly by the model of interest. These were used for plotting the receiver-operating characteristic (ROC) curve, which maps TPR versus FPR at different probability thresholds used for binary classification. The AUC is a widely used measure of classification performance, with an AUC of 1 representing perfect classification and an AUC of 0.5 representing classification at random.

To calculate the probability calibration, we binned predicted probabilities by deciles and plotted the mean values of the empirical frequency of CAD and the predicted probabilities of CAD for each bin. A perfectly calibrated model outputs classification probability scores that mirror actual class probabilities and is suitable for interpretation as a confidence level for calculating risk. To quantify the divergence in tested model calibration from a perfectly calibrated model, we calculated the sum of Euclidean distances between the mean predicted probability and empirical frequency of CAD for each bin. A divergence value of zero indicates perfect calibration.

Mutual Information

Mutual information was used to calculate the strength of probabilistic dependence between adjacent nodes (i.e., nodes connected by an edge) in the BN from their estimated JPD. Unlike correlation, mutual information is calculated from probability distributions over random variables that is not restricted to a linear dependence. Mutual information $I(X;Y)$ between 2 random variables X and Y quantifies the information shared between these variables (i.e., their level of statistical dependence).[42] If X and Y are independent or uncorrelated random variables, that is, $P(X,Y) = P(X)P(Y)$, then $I(X;Y) = 0$. Otherwise, $I(X;Y)$ is positive: $I(X,Y) = \sum_{x\in X}\sum_{y\in Y} p(x,y) \cdot log_2\left(\frac{p(x,y)}{p(x)p(y)}\right)$.

Prognostic Values

To quantify the prognostic utility of a variable X in being able to predict outcome Y, we used conditional entropy $H(Y|X)$,[42] where $H(Y|X) = -\sum_{x\in X}\sum_{y\in Y} p(x,y)\cdot log_2\left(\frac{p(x,y)}{p(x)}\right) = -\sum_{x\in X} p(x)\sum_{y\in Y} p(y|x)\cdot log_2 p(y|x) H(Y|X)$ quantifies the average reduction in uncertainty about Y conditioned on observing the variable X. In formal terms, it is the expected value of the uncertainty in the conditional distribution $p(Y|X)$ weighted by the probability of observing each value of $x \in X$. If X is perfectly predictive of Y $\forall x \in X$, then $H(Y|X) = 0$ (zero entropy). Conversely, if X and Y are independent, then $H(Y|X) = H(Y)$; that is, observing X provides no additional information for predicting Y. Therefore, $0 \le H(Y|X) \le H(Y)$. Here,

$$H(Y) = -\sum_{x\in X} p(y)\cdot \log_2 p(y)$$

PVs reported herein were calculated by normalizing $H(Y|X)$ by $H(Y)$ and subtracting the value from 1 such that a PV of 1 represents a perfectly (deterministically) prognostic and 0 represents no prognostic utility. Formally, for the outcome variable Y,

$$PV(X) = 1 - \frac{H(Y|X)}{H(Y)}, 0 \ge PV(X) \ge 1$$

Conditional entropy provides a succinct scalar quantification of the information gained about a random variable having observed another random variable, accounting for the uncertainty by the use of expectations over probability distributions estimated by the BN.

Software

We used R and the *bnlearn* package (version 4.3) for BN learning and inference. *caret* (version 6.0.81) and *glmnet* (version 2.0.16) were used for implementations of other machine-learning algorithms.

Funding

This study had no external funding source.

Results

Description of BN Structure Learned from Data

To learn a robust network structure from the relatively small number of records ($\sim$303 records compared with 55 variables) in this data set, we averaged network structures learned from 10,000 bootstrap replicates of the data using a confidence level based on empirical frequency of edges, as described previously.[37] Figure 1 shows the BN containing Cath and 26 other connected variables with which it shares dependencies. The 28 variables that are not a part of this network were assumed to be statistically independent of the nodes in Figure 1, and vice versa, at least under the modeling assumptions. Thus, BNs can be used for variable selection to reduce the number of variables in a data set to the most informative or relevant ones for predicting disease risk. For learning, we considered age and sex to be "primary" influences and prohibited incoming edges to these 2 nodes. It is also possible to force an edge between 2 nodes if a relationship between them is known from prior knowledge. Although not necessary, incorporation of expert knowledge can help guide purely computational data-driven approaches in finding network structures that are sensible to the subjective opinions of users and possibly enhance their confidence in the reasonability or appropriateness of the model based on visual inspection alone, increasing their uptake.

The graphical nature of BNs lends itself well to reasoning and model criticism by decision makers without a high level of technical expertise because it is possible to visually inspect if it has discovered known or expected relationships, which may form the basis for model updating. The BN in Figure 1 was able to learn relationships between creatinine and blood urea nitrogen, which are both by-products of nitrogen metabolism, and height, weight, BMI, and obesity, as well as fasting blood sugar and diabetes mellitus, and hypertension and blood pressure. It is important to note that the orientation of edges in this network does not represent actual cause-and-effect relationships since these are impossible to derive from purely observational data; the presence of an edge simply implies that the nodes that it connects are correlated and may be causally related. The generation of a causal graph often requires interventional data and/or domain knowledge.[14,15] Edges in BNs correspond to interactions between covariates in regression analysis, which are explicitly represented and naturally modeled by a BN at the cost of a generally higher model complexity in terms of the number of parameters.

Graphical Reasoning

We describe some patterns of graphical reasoning with BNs using the subgraph representing nodes in the Markov blanket of Cath (red nodes in Figure 1). The Markov blanket of Cath, which consists of variables that are parents, children, or parents of children of Cath,

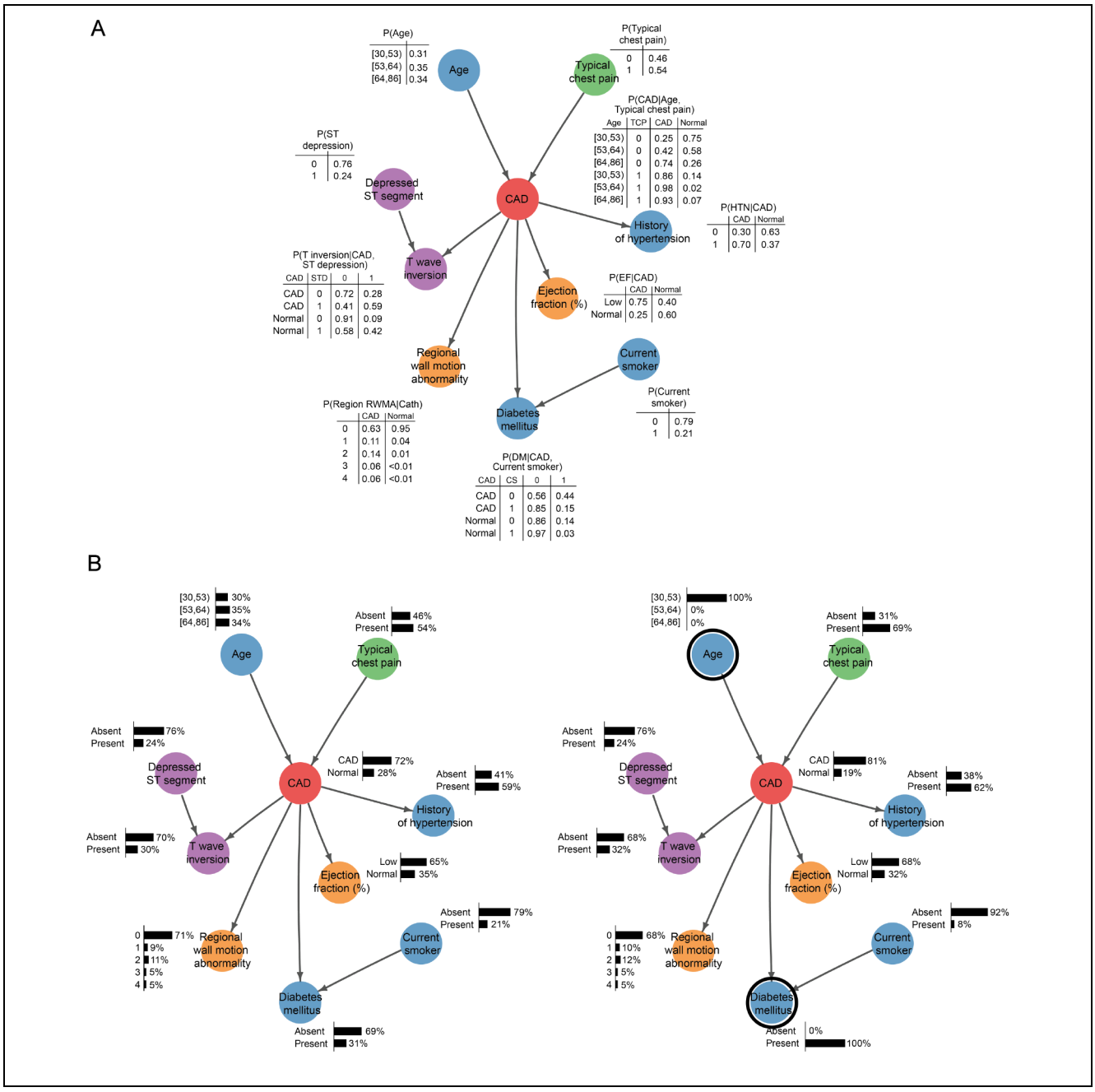

Figure 2 Quantitative components of the subgraph from Figure 1 corresponding to Cath (indicator of coronary artery disease) and variables in its Markov blanket. (A) Joint probability tables. (B, left panel) Marginal probabilities of nodes given no evidence. (B, right panel) Marginal probabilities given a person's age (between 30 and 52 years) and diabetes mellitus status (diabetic). Nodes that were observed as evidence are marked by thick black circles.

summarizes all the information about CAD risk; Cath is independent of all other nodes if the values for variables in its Markov blanket are known. These are shown in Figure 2A along with the fitted conditional probability tables that quantify their relationships. The variables in the Markov blanket of Cath consist of known CAD risk factors such as age, diabetes, hypertension, and smoking status (current smoker), and disease symptoms such as

Table 2

Model Type	Model	AUC from 10-fold CV (Mean ± Standard Deviation)
Linear regression	Ridge regression	0.92 ± 0.06
	LASSO regression	0.92 ± 0.05
Bayesian network	Naïve Bayes	0.93 ± 0.04
	Hill climbing	0.93 ± 0.04
Other	Support vector machine	0.92 ± 0.06
	Artificial neural network	0.91 ± 0.05

typical chest pain, T inversion, and lower ejection fraction. Obesity, another known CAD risk factor, is not directly connected to CAD and, according to the data, informs CAD risk through sex, current smoker, and diabetes mellitus in order of graphical proximity to CAD. Essentially, knowing that a person is obese is superfluous for predicting their CAD risk if their diabetes status is known in this population. This is due to the independence assumptions induced by local factorization of the JPD by BNs (see the Methods section). This property of BNs can be used to order the importance of variables visually, for example, when deciding the most important measurements or tests to perform. More sophisticated techniques for quantifying variable importance that use the conditional probability tables may be used as well (see the Prognostic Values section) and are implemented in BN software such as GeNIe (BayesFusion).[43] Note that this graphical reasoning applies to any variable of interest, not just the traditional outcome variable Cath in this case. We can, for example, determine visually that sex, DM, and Cath, which form the Markov blanket of current smoker, are most informative for predicting whether a person, in this population, currently smokes.

Updating Beliefs from Observations

In medicine, information is often incomplete or obtained sequentially. For instance, a physician might not be able to measure all relevant risk factors of a person to predict his or her risk of CAD on the spot. BNs provide an inferential framework of probabilistic "beliefs" that can be updated as new evidence becomes available. The left panel in Figure 2B shows these probabilistic beliefs, which are the marginal probability distributions of the nodes in the network without any observed evidence calculated from the fitted parameters shown in Figure 2A. Intuitively, these represent the population-level frequencies of variables. For a hypothetical person who is a 50-year-old diabetic, BNs can be used to calculate their individualized CAD risk given this newly observed evidence, increasing it from 72% to 81%. Marginal probabilities of other nodes in the Markov blanket of Cath are recalculated to adjust to the new evidence, leading to, for example, an increase in the probability of having hypertension from 59% to 62%. In this way, a trained BN model can handle missing data and sequential observations by inferring their values from observed data and update their probabilities as new evidence is incorporated. A user can query arbitrary hypothetical scenarios to help guide reasoning about outcomes, such as through value-of-information analysis in decision frameworks based on BNs.[16] These properties of BNs have been used previously for clinical decision support systems for diagnosis of cognitive disorders,[44] cancer,[45] and cardiovascular event prediction,[46] particularly in combination with some expert-defined rules or information from literature and meta-analysis.[47,48]

Risk Prediction: Classification Performance and Probability Calibration

Next, we evaluated the classification performance of BNs learned using the hill-climbing algorithm in comparison with logistic regression with ridge (L2) and LASSO (L1) penalties, categorical naïve Bayes classifier, and the commonly used nonlinear classifiers SVM and ANN to predict CAD in the Z-Alizadehsani data set. Our results showed comparable ROC curves and AUCs for all models using iterated 10-fold CV: ridge regression (AUC: 0.92 ± 0.06), LASSO regression (AUCs: 0.92 ± 0.05), naïve Bayes (AUC: 0.93 ± 0.04), hill climbing (AUC: 0.93 ± 0.04), SVM (AUC: 0.92 ± 0.06), and ANN (AUC: 0.91 ± 0.05; Figure 3A; Table 2). Our results suggest that BNs can perform on par with classifiers for disease classification even though they are not trained to directly minimize classification error, unlike other predictive models, which is consistent with prior work on real-world clinical data sets.[49,50]

In addition to classification performance, we tested the probability calibration of these models, which is a measure

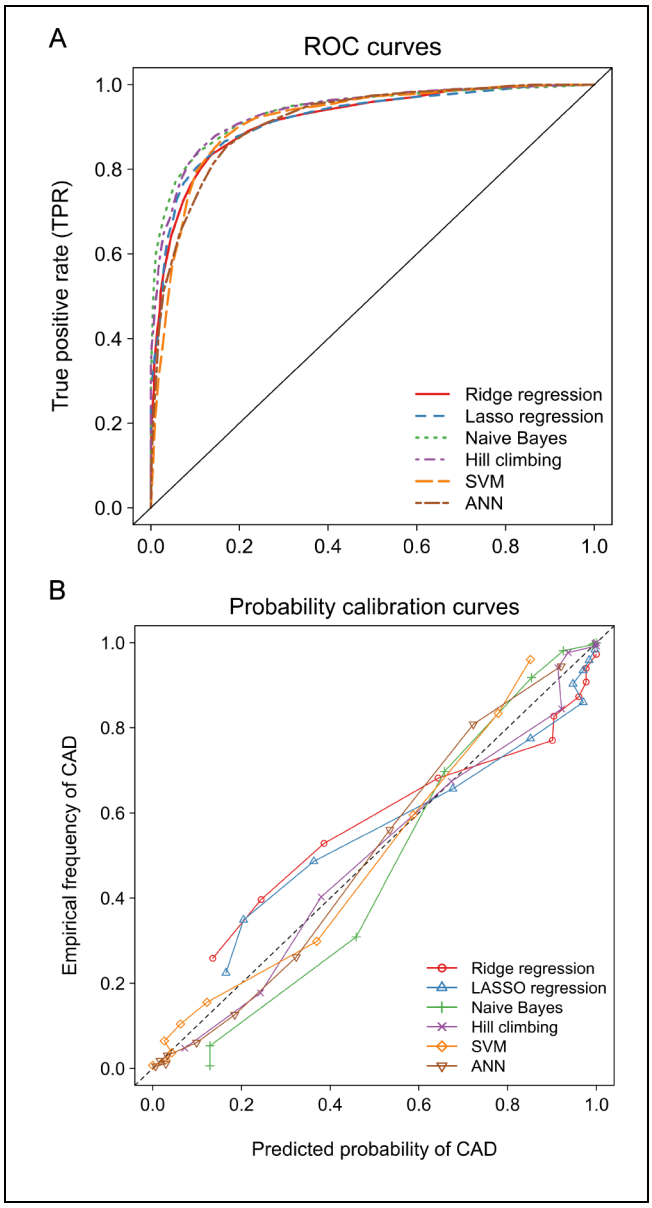

Figure 3 Predictive performance of Bayesian networks compared with logistic regression, support vector machine (SVM), and artificial neural network (ANN) for classification of coronary artery disease (CAD). (A) Mean receiver-operating characteristic (ROC) curves from 10 iterations of 10-fold cross-validation. ROC values were interpolated where necessary. The diagonal line represents classification at random. (B) Mean values of the predicted probabilities of CAD binned by deciles plotted against the empirical frequency of CAD in each bin. The dashed diagonal line represents perfect calibration.

of how well the probabilities output by models correspond to the observed risk of disease. Probability scores from well-calibrated models can be interpreted as confidence about disease risk, whereas poorly calibrated models over- or underestimate disease risk even though they may be able to classify well when tuned. As shown in Figure 3B and quantified in Table 3, networks learned with hill climbing produced the most well-calibrated probability scores of all the models that we tested. BNs showed slightly overconfident risk estimates—overpredicting lower risk and underpredicting higher risk—which is undesirable. The source of such bias in a BN-based risk prediction model may need to be addressed during model validation.

Adaptive Individualized Decision Making for Prognosis of CAD

In clinical settings, decisions often need to made in the presence of variables with uncertain outcomes. For example, a physician might wish to recommend a test for a patient from a range of available tests based on the patient's specific characteristics based on which one would be most informative of CAD, even though the actual outcome of the test is not known at the time of recommendation for that specific patient. Here, we describe the use of BNs for adaptively ranking variables by how prognostic they are for CAD. Figure 4 shows variables ranked by their PV, a measure based on conditional entropy that quantifies how much observing a variable reduces the uncertainty in predicting CAD. For a patient without known characteristics, typical chest pain is the most prognostic variable and should be tested first (Figure 4A). Now, if we find that the patient does not suffer from typical chest pain, we can recalculate PVs by conditioning on this new evidence using the BN. Figure 4B shows that for this person, region RWMA is the most prognostic variable. Atypical chest pain drops from being the third most prognostic to the 40th, and age rises from the seventh most prognostic to the second. This individualized ranking of PVs can be used as an adaptive decision tool to rank the order of tests to perform in a clinic based on most informative of CAD. In practice, such decisions need to be balanced with other constraints, such as cost and risks of a potentially invasive test. BNs can be extended to support more complex decision models such as influence diagrams, which are described elsewhere.[16]

Conclusion

In this article, we described the process of modeling CAD risk from a real-world data set using BNs. We showed that BNs provide an intuitive graphical representation of

Table 3

Model Type	Model	Divergence of Predicted Probability from Empirical Frequency of CAD (Mean ± Standard Deviation)
Linear regression	Ridge regression	0.63 ± 0.02
	LASSO regression	0.46 ± 0.02
Bayesian network	Naïve Bayes	1.03 ± 0.03
	Hill climbing	0.19 ± 0.02
Other	Support vector machine	0.27 ± 0.02
	Artificial neural network	0.23 ± 0.02

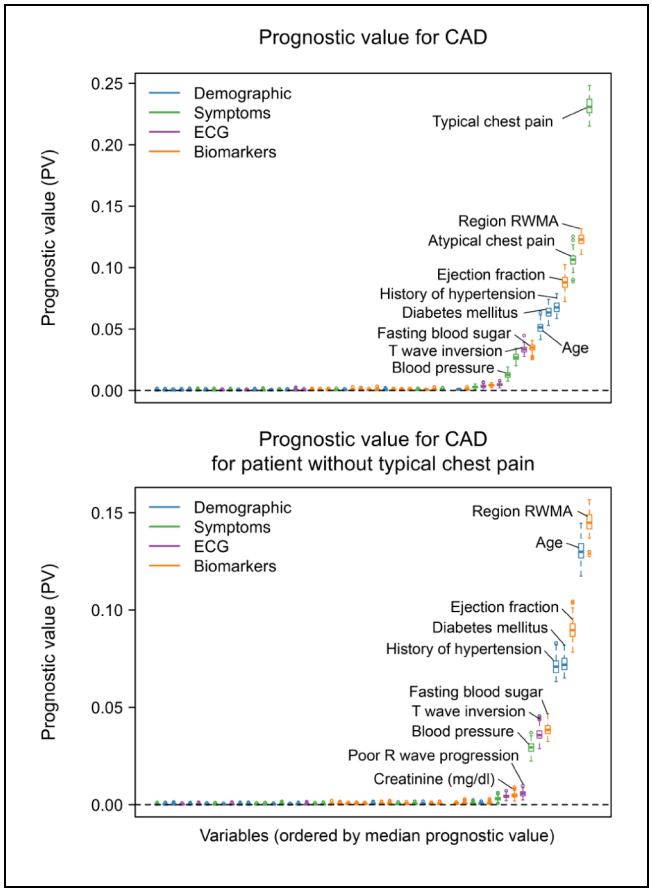

Figure 4 Prognostic values (PVs) for variables for predicting coronary artery disease (CAD). The top 10 most prognostic variables are labeled. (Top) PVs for all variables without any observations. (Bottom) PVs for variables after observing that a patient does not present with typical chest pain. Note how age becomes much more prognostic for a person without typical chest pain than in the general population, and atypical chest pain drops out as one of the top 10 prognostic variables.

dependencies between variables in the data that is useful for visualization and reasoning in human-facing front-end tools and that they can function on par with traditional methods for disease risk prediction such as logistic regression while benefitting from their capacity as generative models to handle missing data. If the only goal of analysis is prediction and a large amount of data is available, then models such as BNs that learn a JPD may be ill-suited compared with discriminative models such as logistic regression because of the stronger assumptions about the data-generating process.[51] BNs may perform at par with or better than traditional regression-based approaches in domains with little to modest amounts of data because of their ability to incorporate prior knowledge and to handle uncertainty. It is important to note that our model is not clinically useful for CAD estimation at this stage because it may suffer from effects of spectrum bias. We have not tested its generalizability to independent external data taken from different settings and populations in comparison with the Z-Alizadehsani data set.

Various software packages and graphical user interfaces for working with BNs exist, including GeNIe, Netica, and BayesiaLab, which can facilitate analysis and interactive visualizations. Because of their visual nature, BNs are suitable for generating clinical decision support system tools,[45,46,48] which are useful for physicians and patients who must make difficult decisions about treatment based on conflicting evidence, subjective preferences, and hypothetical scenarios. Clinical decision support systems are also beneficial from a payer perspective for health technology assessments to evaluate value of therapy. Finally, there is a larger systems application for BNs that draws on their Bayesian properties of being updateable systems. BNs can be embedded in large real-world data sets, such as large-volume hospital databases or administrative health records systems, to continually learn with each new record that is added, forming the foundation for dynamic and living risk-modeling systems. Such endeavors can improve the value of real-world data sets and update the quality of risk modeling and clinical knowledge by using all available clinical data

to guide medical decision making and incorporate uncertainty inherent in the field of medical outcomes research.

Acknowledgments

We thank Neha Sati and Kathryn Zhang for assistance with the scoping review of BNs in CAD risk prediction and disease modeling.

ORCID iDs

Nicholas Mitsakakis https://orcid.org/0000-0001-9041-9152
Paul Arora https://orcid.org/0000-0001-6617-3631

Supplemental Material

Supplementary material for this article is available on the *Medical Decision Making* Web site at http://journals.sagepub.com/home/mdm.

References

1. Roth GA, Johnson C, Abajobir A, et al. Global, regional, and national burden of cardiovascular diseases for 10 causes, 1990 to 2015. *J Am Coll Cardiol*. 2017;70(1):1–25.
2. Nabel EG, Braunwald E. A tale of coronary artery disease and myocardial infarction. *N Engl J Med*. 2012;366(1):54–63.
3. Benjamin EJ, Muntner P, Bittencourt MS. Heart disease and stroke statistics—2019 update: a report from the American Heart Association. *Circulation*. 2019;139(10):e56–e528.
4. Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. *Circulation*. 1998;97(18):1837–47.
5. Task Force Members, Montalescot G, Sechtem U, et al. 2013 ESC guidelines on the management of stable coronary artery disease: The Task Force on the management of stable coronary artery disease of the European Society of Cardiology. *Eur Heart J*. 2013;34(38):2949–3003.
6. Diamond GA, Forrester JS. Analysis of probability as an aid in the clinical diagnosis of coronary-artery disease. *N Engl J Med*. 1979;300(24):1350–8.
7. Pryor DB, Harrell FE Jr, Lee KL, Califf RM, Rosati RA. Estimating the likelihood of significant coronary artery disease. *Am J Med*. 1983;75(5):771–80.
8. Genders TS, Steyerberg EW, Hunink MG, et al. Prediction model to estimate presence of coronary artery disease: retrospective pooled analysis of existing cohorts. *BMJ*. 2012;344:e3485.
9. Min JK, Gilmore A, Budoff MJ, Berman DS, O'Day K. Cost-effectiveness of coronary CT angiography versus myocardial perfusion SPECT for evaluation of patients with chest pain and no known coronary artery disease. *Radiology*. 2010;254(3):801–8.
10. Genders TS, Meijboom WB, Meijs MF, et al. CT coronary angiography in patients suspected of having coronary artery disease: decision making from various perspectives in the face of uncertainty. *Radiology*. 2009;253(3):734–44.
11. Patel MR, Peterson ED, Dai D, et al. Low diagnostic yield of elective coronary angiography. *N Engl J Med*. 2010;362(10):886–95.
12. Pearl J. Reasoning under uncertainty. *Annu Rev Comp Sci*. 1990;4(1):37–72.
13. Koller D, Friedman N, Bach F. *Probabilistic Graphical Models: Principles and Techniques*. Cambridge, MA: MIT Press; 2009.
14. Pearl J. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Burlington, MA: Morgan Kaufmann; 1988.
15. Pearl J. *Causality: Models, Reasoning and Inference*. Cambridge, UK: Cambridge University Press; 2009.
16. Fenton N, Neil M. *Risk Assessment and Decision Analysis with Bayesian Networks*. Boca Raton, FL: CRC Press; 2012.
17. Bishop CM. *Pattern Recognition and Machine Learning*. New York: Springer; 2006.
18. Chen Q, Li G, Leong T-Y. Predicting coronary artery disease with medical profile and gene polymorphisms data. In: *Medinfo 2007: Proceedings of the 12th World Congress on Health (Medical) Informatics; Building Sustainable Health Systems*. Amsterdam, the Netherlands: IOS Press; 2007.
19. Huan T, Zhang B, Wang Z, et al. A systems biology framework identifies molecular underpinnings of coronary heart disease. *Arterioscler Thromb Vasc Biol*. 2013;33(6):1427–34.
20. Mäkinen V-P, Civelek M, Meng Q, et al. Integrative genomics reveals novel molecular pathways and gene networks for coronary artery disease. *PLoS Genet*. 2014;10(7):e1004502.
21. Wei Z, Zhang XL, Rao HX, Wang HF, Wang X, Qiu LX. Using the Tabu-search-algorithm-based Bayesian network to analyze the risk factors of coronary heart diseases. *Zhonghua Liu Xing Bing Xue Za Zhi*. 2016;37(6):895–9.
22. McGeachie M, Ramoni RL, Mychaleckyj JC, et al. An integrative predictive model of coronary artery calcification in arteriosclerosis. *Circulation*. 2009;120(24):2448.
23. Exarchos KP, et al. Prediction of coronary atherosclerosis progression using dynamic Bayesian networks. In: *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Piscataway, NJ: IEEE; 2013.
24. Alizadehsani R, Habibi J, Hosseini MJ, et al. A data mining approach for diagnosis of coronary artery disease. *Comput Methods Programs Biomed*. 2013;111(1):52–61.
25. Alizadehsani R, Zangooei MA, Hosseini MJ, et al. Coronary artery disease detection using computational intelligence methods. *Knowledge-Based Systems*, 2016;109:187–97.
26. Dheeru D, Taniskidou EK. *UCI Machine Learning Repository*. Irvine, CA: University of California, School of Information and Computer Science; 2017.

27. Arabasadi Z, Alizadehsani R, Roshanzamir M, Moosaei H, Yarifar AA. Computer aided decision making for heart disease detection using hybrid neural network-Genetic algorithm. *Comput Methods Programs Biomed*. 2017;141:19–26.
28. Babič F, Olejár J, Vantová Z, Paraličet J. Predictive and descriptive analysis for heart disease diagnosis. In: *2017 Federated Conference on Computer Science and Information Systems (FedCSIS)*. Piscataway, NJ: IEEE; 2017.
29. Yadav C, Lade S, Suman MK. Predictive analysis for the diagnosis of coronary artery disease using association rule mining. *International Journal of Computer Applications*. 2014;87(4).
30. Colombo D, Maathuis MH. Order-independent constraint-based causal structure learning. *Journal of Machine Learning Research*, 2014;15(1):3741–82.
31. Tsamardinos I, Aliferis CF, Statnikov A. Algorithms for large scale Markov blanket discovery. In: *Proceedings of the Sixteenth International Florida Artificial Intelligence Research Society Conference*, May 12–14, 2003, St. Augustine, FL
32. Russell SJ, Norvig P. *Artificial Intelligence: A Modern Approach*. Kuala Lumpur, Malaysia: Pearson Education Limited; 2016.
33. Tsamardinos I, Brown LE, Aliferis CF. The max-min hill-climbing Bayesian network structure learning algorithm. *Machine Learn*. 2006;65(1):31–78.
34. Scutari M, Denis J-B. *Bayesian Networks: With Examples in R*. London: Chapman and Hall/CRC; 2014.
35. Scutari M. Learning Bayesian networks with the bnlearn R package. *arXiv*. 2009;0908:3817.
36. Friedman N, Goldszmidt M, Wyner A. Data analysis with Bayesian networks: a bootstrap approach. In: *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*. Burlington, MA: Morgan Kaufmann; 1999.
37. Scutari M, Nagarajan R. Identifying significant edges in graphical models of molecular networks. *Artif Intell Med*. 2013;57(3):207–17.
38. Hastie T, Tibrhirani R, Friedman J. The elements of statistical learning: data mining, inference and prediction. *Mathematical Intelligencer*. 2005;27(2):83–5.
39. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*. 2010;33(1):1.
40. Kuhn M. Building predictive models in R using the caret package. *J Stat Softw*. 2008;28(5):1–26.
41. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMC Med*. 2015;13(1):1.
42. Cover TM, Thomas JA. *Elements of Information Theory*. New York: John Wiley & Sons; 2012.
43. Druzdzel MJ. SMILE: structural modeling, inference, and learning engine and GeNIe: a development environment for graphical decision-theoretic models. In: AAAI-99 Proceedings. Menlo Park, CA: American Association for Artificial Intelligence; 1999.
44. Seixas FL, Zadrozny B, Laks J, Conci A, Muchaluat Saade DC. A Bayesian network decision model for supporting the diagnosis of dementia, Alzheimer's disease and mild cognitive impairment. *Comput Biol Med*. 2014;51:140–58.
45. Sesen MB, Nicholson AE, Banares-Alcantara R, Kadir T, Brady M. Bayesian networks for clinical decision support in lung cancer care. *PloS One*. 2013;8(12):e82349.
46. Tylman W, Waszyrowski T, Napieralski A, et al. Real-time prediction of acute cardiovascular events using hardware-implemented Bayesian networks. *Comput Biol Med*. 2016;69:245–53.
47. Sesen MB, Peake MD, Banares-Alcantara R, et al. Lung cancer assistant: a hybrid clinical decision support application for lung cancer care. *J R Soc Interface*. 2014;11(98):20140534.
48. Yet B, Perkins ZB, Rasmussen TE, Tai NR, Marsh DW. Combining data and meta-analysis to build Bayesian networks for clinical decision support. *J Biomed Inform*. 2014;52:373–85.
49. Gao P, Zhou X, Wang ZN. Which is a more accurate predictor in colorectal survival analysis? Nine data mining algorithms vs. the TNM staging system. *PLoS One*. 2012;7(7):e42015.
50. 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 Making*. 2018;38(7):822–33.
51. Ng AY, Jordan MI. On discriminative vs. generative classifiers: a comparison of logistic regression and naive Bayes. In: *NIPS'01 Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic*. Cambridge, MA: MIT Press; 2002. p 841–8.