Lung Cancer Survival Prediction Using Instance-Specific Bayesian Networks

Fattaneh Jabbari[1]([✉]), Liza C. Villaruz[2], Mike Davis[3],
and Gregory F. Cooper[1,3]

[1] Intelligent Systems Program, University of Pittsburgh, Pittsburgh, USA
{faj5,gfc}@pitt.edu
[2] UPMC Hillman Cancer Center, Pittsburgh, USA
villaruzl@upmc.edu
[3] Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, USA
midavis@pitt.edu

Abstract. Lung cancer is the most common cause of cancer-related death in men worldwide and the second most common cause in women. Accurate prediction of lung cancer outcomes can help guide patient care and decision making. The amount and variety of available data on lung cancer cases continues to increase, which provides an opportunity to apply machine learning methods to predict lung cancer outcomes. Traditional population-wide machine learning methods for predicting clinical outcomes involve constructing a single model from training data and applying it to predict the outcomes for each future patient case. In contrast, instance-specific methods construct a model that is optimized to predict well for a given patient case. In this paper, we first describe an instance-specific method for learning Bayesian networks that we developed. We then use the Markov blanket of the outcome variable to predict 1-year survival in a cohort of 261 lung cancer patient cases containing clinical and omics variables. We report the results using AUROC as the evaluation measure. In leave-one-out testing, the instance-specific Bayesian network method achieved higher AUROC on average, compared to the population-wide Bayesian network method.

Keywords: Lung cancer survival prediction · Instance-specific modeling · Population-wide modeling · Bayesian network classifiers

1 Introduction

Lung cancer is the most frequent cause of cancer-related death in men worldwide and the second most common cause in women [4], despite significant improvements in diagnosis and treatment during the past decade. Lung cancer is typically divided into two major subtypes: small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC), where the latter is more prevalent. The overall 5-year survival rate for lung cancer is 19% but it can be increased to 57% if diagnosis occurs at a localized stage of the disease [2], which is not often the case. Accurate

prediction of lung cancer outcomes is important, because it can facilitate patient care and clinical decision making. With rapid advancements in technology, large amounts of lung cancer data with clinical and molecular information are being collected and made available. This trend provides an opportunity for researchers to apply machine learning techniques to predict outcomes of lung cancer.

Most machine learning methods for predicting clinical outcomes construct a single model M from training data. M is then applied to predict outcomes in future instances. We refer to such a model as a *population-wide model* because it predicts outcomes for a future population of instances. It may be difficult for population-wide models to perform well in domains in which instances are highly heterogeneous. Studies have revealed that heterogeneity exists both within individual lung cancer tumors and between patients [20]. In such domains, a reasonable approach is to learn a model that is tailored to a particular instance (e.g., a patient), which we refer to as an *instance-specific model*. An instance-specific approach builds a model M_T for a given instance T from the features that we know about T (e.g., clinical and molecular features) and from a training set of data on many other instances. It then uses M_T to predict the outcome for T. This procedure repeats for each instance that is to be predicted.

In this paper, we use Bayesian network (BN) classifiers to predict patient survival in a cohort of 261 lung cancer patients with clinical and omics information. More specifically, we apply a score-based instance-specific BN learning method, called IGES, which we introduced in [14]. The IGES algorithm searches the space of BNs to learn a model that is specific to an instance T by guiding the search based on T's features (i.e., variable-value pairs). We also apply a state-of-the-art score-based population-wide BN learning method, called GES [14], as a control method. These methods are summarized in Sect. 4. The main goal of this paper is to investigate the effectiveness of instance-specific modeling in predicting survival outcomes for lung cancer patients.

2   Related Work

Various population-wide methods such as neural networks, support vector machines, random forests, and Bayesian models have been applied in cancer research to predict cancer survival and reported AUROCs of ∼0.80 [1,8,19]. A review of such methods is provided in [16]. None of these methods learn instance-specific models. Additionally, they use different sets of predictors and different patient cohorts, so a direct comparison to the results in the current paper is not possible.

Several machine learning methods have been developed to learn instance-specific models. Zheng and Webb [22] introduced a lazy Bayesian rule learning (LBR) method to construct a model that is specific to a test instance. In an LBR rule, the antecedent is a conjunction of the variable-value pairs that are present in the test instance and the consequent is a local naïve Bayes classifier in which the target variable is the parent of the variables that do not appear in the antecedent. Visweswaran and Cooper [21] developed an instance-specific Markov Blanket

(ISMB) algorithm that searches over the space of Markov blankets (MBs) of the target variable by utilizing the features of a given test instance. ISMB first uses a greedy hill-climbing search to find a set of MBs that best fit the training data. Then, it greedily adds single edges to the MB structures from the previous step, if doing so improves the prediction of a given test instance. Ferreira et al. [10] implemented two patient-specific decision path (PSDP) algorithms using two variable selection criteria: balanced accuracy and information gain. A decision path is a conjunction of features that are present in a given test instance and a leaf node that contains the probability distribution of the target variable.

Recently, Lengerich et al. [17] introduced a personalized linear regression model that learns a specific set of parameters for each test case based on the idea that the similarity between personalized parameters is related to the similarity between the features. Accordingly, a regularizer is learned to match the pairwise distance between the features (i.e., variable-value pairs) and the personalized regression parameters. In other related work, Cai et al. [5] developed a method to learn tumor-specific BN models from data; this method uses bipartite BNs and makes other assumptions that restrict its generality.

The IGES method [14] that we use in this paper is different from the methods mentioned above. IGES differs from a population-wide method in a principled way; it learns a model for each instance T that is optimized for T, while a population-wide model is designed to learn a single model that is optimized for a population of instances. Also, the IGES method learns a more general model than does LBR [22], ISMB [21], PSDP [10], and tumor-specific BN models [5], because it learns Bayesian network models, which can be used for both prediction and causal discovery [13–15].

3   Preliminaries

A Bayesian network (BN) encodes probabilistic relationships among a set of variables. A BN consists of a pair $(\mathcal{G}, \boldsymbol{\Theta})$, where $\mathcal{G}$ is a directed acyclic graph (DAG) and $\boldsymbol{\Theta}$ is a set of parameters for $\mathcal{G}$. A DAG $\mathcal{G}$ is given as a pair $(\boldsymbol{X}, \boldsymbol{E})$, where $\boldsymbol{X} = \{X_1, X_2, ..., X_n\}$ denotes a set of nodes that correspond to domain variables and $\boldsymbol{E}$ is a set of directed edges between variables. The presence of an edge $X_i \rightarrow X_j$ (X_i is called the parent and X_j is called the child) denotes that these variables are probabilistically dependent. The absence of an edge denotes that X_i and X_j are conditionally probabilistically independent. A set of DAGs that encode the same independence and dependence relationships are statistically indistinguishable from observational data; such DAGs are called Markov equivalent. The second component, $\boldsymbol{\Theta}$, is a set of parameters that encodes the joint probability distribution over $\boldsymbol{X}$, which can be efficiently factored based on the parent-child relationships in $\mathcal{G}$.

As mentioned above, the edges present and absent in a DAG represent conditional dependence and independence relationships between variables, respectively. Any such relationship should hold for all combinations of values of the variables in the BN. There is a more refined form of conditional independence

that holds only in a specific context, which is known as *context-specific independence* (CSI) [3]. Figure 1 shows a BN that includes two CSI relationships: $X_4 \perp\!\!\!\perp \{X_2, X_3\}|X_1 = 0$ and $X_4 \perp\!\!\!\perp X_3|\{X_1 = 1, X_2 = 1\}$. The IGES method models such local independence structures for each instance T, which results in a more expressive BN structure and a more efficient BN parameterization.

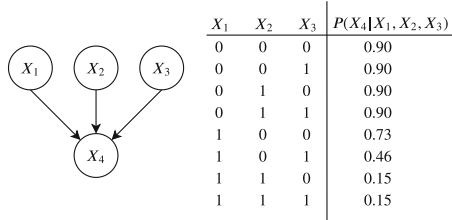| X_1	X_2	X_3	$P(X_4|X_1, X_2, X_3)$
0	0	0	0.90
0	0	1	0.90
0	1	0	0.90
0	1	1	0.90
1	0	0	0.73
1	0	1	0.46
1	1	0	0.15
1	1	1	0.15

Fig. 1. An example Bayesian network with two context-specific independencies (CSIs): $X_4 \perp\!\!\!\perp \{X_2, X_3\}|X_1 = 0$ and $X_4 \perp\!\!\!\perp X_3|\{X_1 = 1, X_2 = 1\}$.

4   Methodology

4.1   Population-Wide Greedy Equivalence Search (GES)

Score-based methods are one of the main approaches to learn BN structure from data that involve (1) a scoring function to measure how well the data (and optional background knowledge) support a given DAG and (2) a search strategy to explore the space of possible DAGs. Since recovering the data-generating BN is an NP-hard problem [6], these methods often utilize a greedy search strategy. GES [7] is a state-of-the-art score-based BN learning algorithm that searches over the Markov equivalence classes of DAGs using two local graph operations: single edge addition and single edge removal. First, it greedily adds single edges to the current graph as long as doing so leads to score improvement. It then greedily removes single edges as long as doing so results in a higher score. Under reasonable assumptions, the GES algorithm converges to the data-generating BN or one Markov equivalent to it [7].

The Bayesian Dirichlet (BD) scoring function [12] is used to score a BN with discrete variables and is calculated as follows:

$$\text{BD}(\mathcal{G}, D) = P(\mathcal{G}) \cdot \prod_{i=1}^{n} \prod_{j=1}^{q_i} \frac{\Gamma(\alpha_{ij})}{\Gamma(\alpha_{ij} + N_{ij})} \cdot \prod_{k=1}^{r_i} \frac{\Gamma(\alpha_{ijk} + N_{ijk})}{\Gamma(\alpha_{ijk})}, \qquad (1)$$

where $P(\mathcal{G})$ is the prior structure probability of $\mathcal{G}$. The first product term is over all n variables, the second product is over the q_i parent instantiations of variable X_i, and the third product is over all r_i values of variable X_i. The term N_{ijk} is the number of cases in the data in which variable $X_i = k$ and its parents

$\boldsymbol{Pa}(X_i) = j$; also, $N_{ij} = \sum_{k=1}^{r_i} N_{ijk}$. The term α_{ijk} is a Dirichlet prior parameter that may be interpreted as representing "pseudo-counts" and $\alpha_{ij} = \sum_{k=1}^{r_i} \alpha_{ijk}$. The pseudo-counts can be defined to be evenly distributed, in which case Eq. (1) represents the so-called *BDeu* score [11]:

$$\alpha_{ijk} = \frac{\alpha}{r_i \cdot q_i}, \tag{2}$$

where α is a positive constant called the prior equivalent sample size (PESS).

4.2   Instance-Specific Greedy Equivalence Search (IGES)

We introduced an instance-specific version of GES called IGES in [14]. Similar to GES, IGES is a two-stage greedy BN learning algorithm that uses an instance-specific score (IS-Score). First, it runs GES using the training data D and the BDeu score to learn a population-wide BN model $\mathcal{G}_{PW}$; this is the population-wide BN for all test instances. In the second stage, IGES uses the data D, the population-wide model $\mathcal{G}_{PW}$, and a single test instance T to learn an instance-specific model for T, which is denoted as $\mathcal{G}_{IS}$. To do so, IGES starts with $\mathcal{G}_{PW}$ and runs an adapted version of GES with a specialized IS-Score to maximize the marginal likelihood of the data given both BN models $\mathcal{G}_{PW}$ and $\mathcal{G}_{IS}$.

   For each variable X_i, the IS-Score is calculated at the parent-child level and is composed of two components: (1) the instance-specific structure that includes X_i's parents in $\mathcal{G}_{IS}$, which we denote by $\boldsymbol{Pa}_{IS}(X_i)$ and (2) the population-wide structure that includes X_i's parents in $\mathcal{G}_{PW}$, which we denote by $\boldsymbol{Pa}_{PW}(X_i)$. In order to score the instance-specific structure $\boldsymbol{Pa}_{IS}(X_i) \rightarrow X_i$, we use the instances in data that are similar to the current test instance T in terms of the values of the variables in $\boldsymbol{Pa}_{IS}(X_i)$. These instances are selected based on the values of $\boldsymbol{Pa}_{IS}(X_i)$ in T; let $D_{Pa_{IS}(X_i)=j}$ be the instances that are similar to T assuming $\boldsymbol{Pa}_{IS}(X_i) = j$. Then the instance-specific score for $\boldsymbol{Pa}_{IS}(X_i) \rightarrow X_i$ given data $D_{Pa_{IS}(X_i)=j}$ is as follows:

$$P(D_{Pa_{IS}(X_i)=j}|\boldsymbol{Pa}_{IS}(X_i) \rightarrow X_i) = \frac{\Gamma(\alpha_{ij})}{\Gamma(\alpha_{ij} + N_{ij})} \cdot \prod_{k=1}^{r_i} \frac{\Gamma(\alpha_{ijk} + N_{ijk})}{\Gamma(\alpha_{ijk})}, \tag{3}$$

where r_i denotes all values of X_i, N_{ijk} is the number of instances in $D_{Pa_{IS}(X_i)=j}$ in which $X_i = k$, and $N_{ij} = \sum_{k=1}^{r_i} N_{ijk}$. The terms α_{ijk} and $\alpha_{ij} = \sum_{k=1}^{r_i} \alpha_{ijk}$ are the corresponding Dirichlet priors.

   Since the instances in $D_{Pa_{IS}(X_i)=j}$ are being used to score the instance-specific structure, they should no longer be used to also score the population-wide structure; therefore, the score for $\boldsymbol{Pa}_{PW}(X_i) \rightarrow X_i$ must be adjusted accordingly. We re-score $\boldsymbol{Pa}_{PW}(X_i) \rightarrow X_i$ using the remaining instances in $D_{Pa_{IS}(X_i)\neq j}$ as follows:

$$P(D_{Pa_{IS}(X_i)\neq j}|\boldsymbol{Pa}_{PW}(X_i) \rightarrow X_i) = \prod_{j'=1}^{q_i} \frac{\Gamma(\alpha_{ij'})}{\Gamma(\alpha_{ij'} + N_{ij'})} \cdot \prod_{k=1}^{r_i} \frac{\Gamma(\alpha_{ij'k} + N_{ij'k})}{\Gamma(\alpha_{ij'k})}, \tag{4}$$

where r_i and q_i are the number of possible values of X_i and $\boldsymbol{Pa}_{PW}(X_i)$, respectively. $N_{ij'k}$ is the number of instances in $D_{Pa_{IS}(X_i) \neq j}$ for which $X_i = k$ and $\boldsymbol{Pa}_{PW}(X_i) = j'$, and $N_{ij'} = \sum_{k=1}^{r_i} N_{ij'k}$. The terms $\alpha_{ij'k}$ and $\alpha_{ij'} = \sum_{k=1}^{r_i} \alpha_{ij'k}$ are the corresponding Dirichlet priors.

Finally, to obtain the overall score for variable X_i using all instances in D, we multiply these two scores. In essence, this method searches for the most probable context-specific BN in light of how well (1) the instance-specific model predicts data for instances like the current test instance, and (2) the population-wide model predicts data for the remaining instances. See [14] for a detailed description of the IS-Score and the IGES method.

5   Experimental Results

This section describes a comparison of the performance of an instance-specific machine learning method to its population-wide counterpart when predicting 1-year survival in lung cancer patients, using clinical and molecular data, which are described in Sect. 5.1. We used Bayesian network classifiers as machine learning models to predict the target variable. More specifically, we applied IGES and GES methods to learn instance-specific and population-wide BN classifiers.

To predict the target variable, we first ran the IGES and GES methods to construct a BN structure over all variables (i.e., the predictors and target). Then, we obtained the Markov blanket (MB) of the target variable that includes the variable's parents, children, and its children's parents. Finally, we calculated the probability distribution of the target variable given its MB and output the most probable outcome as the prediction. We report evaluation criteria to measure the effectiveness of the instance-specific BN model versus the population-wide BN model. In particular, as a measure of discrimination, we report the area under the ROC curve (AUROC) when predicting 1-year survival. We also report the differences between the variables in the MB of the target variable found by the instance-specific models compared to the population-wide model.

5.1   Data Description

This was a retrospective analysis of banked tumor specimens that were collected from patients with lung cancer at the University of Pittsburgh Medical Center (UPMC) in 2016. Baseline demographics, smoking history, staging, treatment, and survival data were collected through the UPMC Network Cancer Registry. We replaced the missing values of the predictor variables with a new category called "missing" and removed the cases for which the value of the outcome variable was not known. Demographic and clinical characteristics of the 261 patients are summarized in Table 1. DNA sequencing was performed using the Ion AmpliSeq$^{\text{TM}}$ Cancer Panel (Ion Torrent, Life Technologies, Fisher Scientific). Gene rearrangements of ALK, ROS1, and RET, and MET amplification were detected using FISH. PD-L1 SP263 and PD-L1 22C2 assays were performed on lung cancer samples to determine the PD-L1 tumor proportion score

Table 1. One-year survival given demographic and clinical characteristics. A 95% confidence interval is included for each sub-group of patients.

Variable name	Variable value	Greater than 1 year	
		# Patients (Total)	% Patients (Confidence Interval)
Age	22–62	54 (84)	64.29 (54.04, 74.54)
	63–72	41 (88)	46.59 (36.17, 57.01)
	73–88	45 (89)	50.56 (40.17, 60.95)
Sex	Female	80 (135)	59.26 (50.97, 67.55)
	Male	60 (126)	47.62 (38.9, 56.34)
Race	White	119 (224)	53.13 (46.58, 59.66)
	Black	16 (31)	51.61 ((34.02, 69.2)
	Other	5 (6)	83.33 (53.33, 113.33)
Tobacco history	Cigar/pipe smoker	0 (1)	0
	Cigarette smoker	42 (85)	49.41 (38.78, 60.04)
	Never used	22 (32)	68.75 (52.69, 84.81)
	Previous tobacco use	76 (142)	53.52 (45.32, 61.72)
	Snuff/chew/smokeless	0 (1)	0
Diagnosis	Adenocarcinoma	53 (89)	59.55 (49.35, 69.75)
	Squamous	3 (7)	42.86 (45.3, 62.06)
	Other	11 (29)	37.93 (6.20, 79.52)
	NA	73 (136)	53.68 (20.27, 55.59)

(TPS). Table 2 provides information about the type, name, and description of the variables that are included in the lung cancer dataset. The outcome-prediction research reported here was performed under the auspices of study protocol number PRO15070164 from the University of Pittsburgh Institutional Review Board (IRB).

5.2   Results

We performed leave-one-out cross-validation on the lung cancer dataset. For each instance T, we used T as a test instance and all the remaining instances as the training set D. We applied IGES search using T, D, and the IS-Score to learn an instance-specific BN model for T, which is called $\mathcal{G}_T$ and is used to predict the outcome for T. We also applied GES search using D and the BDeu score to learn a population-wide BN model for T; this BN model is denoted by $\mathcal{G}_{PW}$ and is used to predict the outcome for T. We repeated this procedure for every instance in the dataset. Note that this leave-one-out cross-validation does not involve any hyperparameter tuning; therefore, we do not expect it to

Table 2. Type, name, and description of the variables of the lung cancer dataset.

Variable type	Variable name	Variable description
Demographics	Sex, Race, Age, Tobacco History	
Clinical	Site	Location of tumor
	Surgical Procedure	Type of surgical resection or biopsy
	Diagnosis	Lung cancer type (Adenocarcinoma, Squamous, Other, NA)
	Mets at Dx-Brain, Mets at Dx-Bone, Mets at Dx-Distant Lymph Nodes, Mets at Dx-Lung, Mets at Dx-Liver, Mets at Dx-Other	Location of metastasis at diagnosis (Dx), if any
	Histo Behavior ICD-O-3	Histological classification
	cT, cN, cM, cStage Group	Clinical staging
	pT, pN, pM, pStage Group, Pathologic Stage Descriptor	Pathologic staging
Molecular	PD-L1 IHC, PD-L1 Comment	PD-L1 immunohistochemistry measures the amount of PD-L1 staining on tumor cells
	MET, KRAS, EGFR-summary, EGFR-Exon-18, EGFR-Exon-19, EGFR-Exon-20, EGFR-Exon-21, BRAF, PIK3CA, ALK Mutation	Status of gene mutations
	ALK IHC	ALK gene immunohistochemistry
	ALK Trans	ALK gene translocation
	ROS Trans	ROS gene translocation
	RET Trans	RET gene translocation
	cMET Ratio	Measurement of cMET gene amplification

cause overfitting. We used an efficient implementation of GES, called FGES [18], which is available in the Tetrad system[1]. We used prior equivalence sample sizes of PESS = $\{0.1, 1.0, 10.0\}$ for both IGES and GES methods. IGES also has a parameter that penalizes the structural difference between the population-wide and instance-specific BN models, called κ ($0.0 < \kappa \leq 1.0$), where a lower value

---

[1] https://github.com/cmu-phil/tetrad.

indicates more penalty for differences (see [14] for details). We report the results of using multiple values of κ.

Table 3 shows the AUROC results on the lung cancer dataset, using both GES and IGES searches; boldface indicates that the results are statistically significantly better, based on DeLong's non-parametric test [9] at a 0.001 significance level. The results indicate that with PESS = 1.0 and κ = 1.0, the instance-specific search resulted in the highest AUROC; also, for almost all values of κ, IGES performs better. Table 3 also suggests that it is important to define PESS properly when applying a Bayesian method on a dataset with small to moderate sample size, which is the case in this paper.

Table 3. AUROC of the GES and IGES methods on the lung cancer dataset. Boldface indicates statistically significantly better results.

Method	GES	IGES									
PESS	–	$\kappa = 0.1$	$\kappa = 0.2$	$\kappa = 0.3$	$\kappa = 0.4$	$\kappa = 0.5$	$\kappa = 0.6$	$\kappa = 0.7$	$\kappa = 0.8$	$\kappa = 0.9$	$\kappa = 1.0$
0.1	0.68	0.67	0.70	0.70	0.70	0.70	0.71	0.71	0.70	0.71	0.72
1.0	0.68	0.68	**0.75**	**0.75**	**0.75**	**0.75**	**0.75**	**0.74**	**0.75**	**0.75**	**0.81**
10.0	0.65	**0.75**	**0.75**	**0.77**	**0.76**	**0.75**	**0.72**	**0.73**	**0.73**	0.69	0.70

Table 4. Comparison of the target variable's Markov blanket (MB) in the instance-specific BNs vs. the population-wide BN for PESS = 1.0 and κ = 1.0.

(a) Structural differences of the variables in the MBs in instance-specific BNs vs. the population-wide BN.

# Added	# Deleted	# Reoriented	% Patients
0	0	0	16.9
5	0	0	10.7
4	0	0	7.7
1	0	0	6.9
3	0	2	4.2
0	0	2	4.2
6	0	0	3.8
other	other	other	45.6

(b) Percentage of top-7 variables of the MBs of instance-specific BNs. The MB of the population-wide BN includes the first two variables denoted by *.

Variable name	% Occurrence in patients
EGFR-Exon-19*	98.1
Mets at Dx-Other*	92.8
Race	37.9
EGFR-Exon-18	35.6
EGFR-Exon-20	31.8
cM	26.8
cStage Group	24.5

Table 4a shows the results of comparing the target variable's MB in the instance-specific models versus the population-wide models with PESS = 1.0 and κ = 1.0. It indicates that in 16.9% of the patient cases, the MB of the target variable was exactly the same in instance-specific and population-wide BNs. Also, in 10.7% of the cases, the MB of the target variable had 5 additional variables in instance-specific models compared to the population-wide model. Table 4b also shows the percentage of the 7 variables that occurred the most in the instance-specific MBs. Table 4 supports that instance-specific structures exist for the lung cancer cases in the dataset we used.

6   Discussion and Conclusions

In this paper, we evaluated the performance of an instance-specific BN classifier, which uses the IGES algorithm, to predict 1-year survival for 261 lung cancer patient cases. We compared IGES results to its population-wide counterpart, GES. We compared IGES to GES method for two reasons. First, the goal of this study was to evaluate the effectiveness of instance-specific modeling in predicting lung-cancer survival; therefore, we wanted the only difference between the two methods to be instance-specific versus population-wide modeling, while keeping the type of machine learning classifier the same (i.e., Bayesian networks). Since to date we have only implemented instance-specific and population-wide algorithms for learning BNs, we compared these two methods. Additionally, BNs continue to be an important machine learning method for clinical outcome prediction because they generally perform well and provide interpretable models that clinicians can understand relatively well. We compared the predictive performance using AUROC and the structural differences between the instance-specific and population-wide BNs. The results provide support that the instance-specific models are often different and have better predictive performance than the population-wide ones. Future extensions include (1) tuning hyperparameters of the methods such as PESS and κ, and (2) implementing instance-specific versions of other machine learning classification methods and comparing them to their population-wide counterparts.

References

1. Agrawal, A., Misra, S., Narayanan, R., et al.: Lung cancer survival prediction using ensemble data mining on SEER data. Sci. Program. **20**(1), 29–42 (2012)
2. American Cancer Society: Cancer Facts & Figures 2020. American Cancer Society, Atlanta (2020)
3. Boutilier, C., Friedman, N., Goldszmidt, M., Koller, D.: Context-specific independence in Bayesian networks. In: Proceedings of Uncertainty in Artificial Intelligence (UAI), pp. 115–123 (1996)
4. Bray, F., Ferlay, J., Soerjomataram, I., et al.: Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J. Clin. **68**(6), 394–424 (2018)
5. Cai, C., Cooper, G.F., Lu, K.N., et al.: Systematic discovery of the functional impact of somatic genome alterations in individual tumors through tumor-specific causal inference. PLoS Comput. Biol. **15**(7), e1007088 (2019)

6. Chickering, D.M.: Learning Bayesian networks is NP-complete. In: Learning from Data, pp. 121–130. Springer (1996). https://doi.org/10.1007/978-1-4612-2404-4_12

7. Chickering, D.M.: Optimal structure identification with greedy search. J. Mach. Learn. Res. **3**, 507–554 (2003)

8. Deist, T.M., Dankers, F.J., Valdes, G., et al.: Machine learning algorithms for outcome prediction in (chemo) radiotherapy: An empirical comparison of classifiers. Med. Phys. **45**(7), 3449–3459 (2018)

9. DeLong, E.R., DeLong, D.M., Clarke-Pearson, D.L.: Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. Biometrics **44**(3), 837–845 (1988)

10. Ferreira, A., Cooper, G.F., Visweswaran, S.: Decision path models for patient-specific modeling of patient outcomes. In: AMIA Annual Symposium Proceedings, pp. 413–421 (2013)

11. Heckerman, D.: A tutorial on learning with Bayesian networks. In: Learning in Graphical Models, pp. 301–354. Springer (1998)

12. Heckerman, D., Geiger, D., Chickering, D.M.: Learning Bayesian networks: The combination of knowledge and statistical data. Mach. Learn. **20**(3), 197–243 (1995)

13. Jabbari, F., Cooper, G.F.: An instance-specific algorithm for learning the structure of causal Bayesian networks containing latent variables. In: Proceedings of the SIAM International Conference on Data Mining (SDM), pp. 433–441 (2020)

14. Jabbari, F., Visweswaran, S., Cooper, G.F.: Instance-specific Bayesian network structure learning. Proc. Mach. Learn. Res. **72**, 169–180 (2018)

15. Jabbari, F., Visweswaran, S. and Cooper, G.F.: An empirical investigation of instance-specific causal Bayesian network learning. In: IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 2582–2585. IEEE (2019)

16. Kourou, K., Exarchos, T.P., Exarchos, K.P., et al.: Machine learning applications in cancer prognosis and prediction. Comput. Struct. Biotechnol. J. **13**, 8–17 (2015)

17. Lengerich, B.J., Aragam, B., Xing, E.P.: Personalized regression enables sample-specific pan-cancer analysis. Bioinformatics **34**(13), i178–i186 (2018)

18. Ramsey, J.D.: Scaling up greedy equivalence search for continuous variables. arXiv preprint arXiv:1507.07749 (2015)

19. Sesen, M.B., Kadir, T., Alcantara, R.B., et al.: Survival prediction and treatment recommendation with Bayesian techniques in lung cancer. In: AMIA Annual Symposium Proceedings, pp. 838–847 (2012)

20. Travis, W.D., Brambilla, E., Noguchi, M., et al.: International association for the study of lung cancer/American Thoracic Society/European Respiratory Society international multidisciplinary classification of lung adenocarcinoma. J. Thorac. Oncol. **6**(2), 244–285 (2011)

21. Visweswaran, S., Cooper, G.F., Chickering, M.: Learning instance-specific predictive models. J. Mach. Learn. Res. **11**(Dec), 3333–3369 (2010)

22. Zheng, Z., Webb, G.I.: Lazy learning of Bayesian rules. Mach. Learn. **41**(1), 53–84 (2000)