# Decision Path Models for Patient-Specific Modeling of Patient Outcomes

**Antonio Ferreira, PhD[1], Gregory F. Cooper, MD, PhD[1], Shyam Visweswaran, MD, PhD[1]**
**[1]Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA**

**Abstract**

*Patient-specific models are constructed to take advantage of the particular features of the patient case of interest compared to commonly used population-wide models that are constructed to perform well on average on all cases. We introduce two patient-specific algorithms that are based on the decision tree paradigm. These algorithms construct a decision path specific for each patient of interest compared to a single population-wide decision tree with many paths that is applicable to all patients of interest that are constructed by standard algorithms. We applied the patient-specific algorithms to predict five different outcomes in clinical datasets. Compared to the population-wide CART decision tree the patient-specific decision path models had superior performance on area under the ROC curve (AUC) and had comparable performance on balanced accuracy. Our results provide support for patient-specific algorithms being a promising approach for predicting clinical outcomes.*

## 1. Introduction

Clinical care typically entails predicting outcomes under uncertainty [1], such as predicting mortality and serious complications in patients with community acquired pneumonia [2]. Making better predictions is an important healthcare goal since it has the potential to improve clinical decision-making, which in turn can lead to better outcomes in patients. In addition, efficient use of healthcare resources depends on being able to predict accurately when and where a resource is likely to be useful.

In predictive modeling, the typical machine learning paradigm consists of learning a single model from a dataset of patient cases, and then the model is applied to predict outcomes for any future patient case [3]. We call such a model a *population-wide model* because it is intended to be applied to an entire population of future cases. Popular algorithms that are used to learn models from clinical data include decision trees, logistic regression, neural networks, and Bayesian networks. A contrasting paradigm that we describe in this paper consists of learning a model that is specific to the particular features of a given patient case. We call such a model a *patient-specific model* since it is specialized to the particular features of the patient case of interest, and it is optimized to predict especially well for that case. While a population-wide model is optimized to have good predictive performance on average on all future cases, a patient-specific model is optimized to perform well on a specific patient case [3, 4]. This optimization is achieved by specializing the model structure and parameters, as well as the model search, based on the known features of the patient case of interest.

The patient-specific paradigm is motivated by the intuition that in constructing predictive models, all the available information should be utilized including available knowledge of the features of the patient case of interest. This paradigm encompasses a variety of approaches, in all of which different patient cases will potentially result in different models, because the cases contain potentially different values for the features. In one approach, patient-specific models may differ in the variables included in the model, in the interaction among the included variables, and in the strength of the interaction. A second approach is to select models that differ considerably in their predictions for patient case of interest, and combine the differing predictions [5]. A third approach is to identify a subset of the training data that is similar in some way to the patient case of interest and learn a model from the subset. The third approach is the one that is typically used by instance-based machine learning algorithms that are described in Section 2.

Algorithms that learn patient-specific models can provide several advantages over algorithms that learn population-wide models. First, patient-specific models can have better predictive performance because the algorithm takes advantage of any special characteristics of the features in the patient case of interest [6]. Second, a patient-specific model may be simpler than a population-wide model as illustrated by an example given in Section 3, and a simpler model will provide a more succinct explanation. Third, the construction of patient-specific models may be computationally faster, though this advantage is offset by the observation that a patient-specific algorithm has to construct a distinct model for each patient case of interest while a population-wide algorithm has to construct just a

single model. Fourth, patient-specific algorithms often ignore missing features which simplifies the handling of missing features in the patient case of interest.

In this paper, we investigate the performance of two patient-specific algorithms that construct decision path models. For a patient case of interest, these two algorithms construct a decision tree that is optimized for that patient case. In reality, only a path is constructed that contains variables whose values correspond to the values of those features in the patient case. We compare the performance of the patient-specific algorithms with Classification and Regression Trees (CART) which is a commonly used population-wide algorithm for constructing decision trees. We focus on the discriminative performance of the three algorithms and evaluate them using balanced accuracy and the area under the ROC curve (AUC). We use balanced accuracy because it mitigates the inflated performance estimates on imbalanced datasets that is obtained with accuracy.

The remainder of this paper is organized as follows. Section 2 presents background and related work. Section 3 provides details of the patient-specific decision path algorithms that we have developed. Section 4 describes the datasets and experimental methods and Section 5 presents and discusses the results of the patient-specific decision path algorithms on several clinical datasets. Section 6 presents our conclusions.

## 2. Background

In this paper, we are concerned with predictive models and, in particular, we predict a discrete-valued target. Thus, the prediction models are classification models. In the field of machine learning the terms variable (or predictor) and feature are often used interchangeably. We distinguish between the two terms as follows. A *variable* is a quantity that describes an aspect of an object of the world. A *feature* is the specification of a variable and its value. For example, "eye color" is a variable and "eye color = black" is a feature. In this paper, the term feature is used exclusively to refer to a variable-value (or predictor-value) pair. A case (also referred to as instance) is a single object of the world and is described by a list of features. In addition, a case is also labeled with a target value. A dataset is a collection of cases. The test case (or test instance) refers to the patient case of interest for which a patient-specific algorithm constructs a predictive model to predict the target.

Patient-specific models can be constructed using existing instance-based machine learning algorithms. Typically, instance-based algorithms are "lazy", since no model is constructed *a priori* before a test instance becomes available, and a model is learned only when a prediction is needed for a new instance [7]. In contrast, algorithms that learn population-wide models are "eager" since they explicitly build a model from the training data before a test instance becomes available.

In the rest of this section, we briefly describe some of the work described in the literature that is related to instance-based algorithms and patient-specific algorithms, and then we briefly describe population-wide algorithms for constructing decision trees.

### 2.1 Related work

The canonical instance-based algorithm is the nearest-neighbor method. When a test instance is encountered, the training instance that is most similar to the test instance is located and its target value is returned as the prediction [8]. A straight-forward extension to the nearest-neighbor technique is the *k*-Nearest Neighbor (*k*NN) algorithm. For a test instance, this algorithm selects the *k* most similar training instances and either averages or takes a majority vote of their target values. Modified versions of *k*NN have been applied successfully to clinical datasets for diagnosis and knowledge extraction [9].

Other instance-based algorithms take advantage of features in the test instance to learn a model. Friedman et al. [10] described an algorithm called LazyDT that searches for the best CART-like decision tree for a test instance. In contrast to population-wide decision tree models that are constructed by algorithms like CART, LazyDT does not perform pruning and handles only discrete variables. Even with these limitations, when compared to ID3 and C4.5 (standard population-wide algorithms for inducing decision trees) LazyDT had higher accuracies on average.

Zheng et al. [11] developed an instance-based algorithm called the Lazy Bayesian Rules (LBR) learner that constructs a rule tailored to the features of the test instance that is then used to classify it. A LBR rule consists of (1) a conjunction of the features present in the test instance as the antecedent, and (2) a naïve Bayes classifier as the consequent. The naive Bayes classifier consists of the target variable as the parent of all other predictors that do not appear in the antecedent, and the parameters of the classifier are estimated from those training instances that satisfy

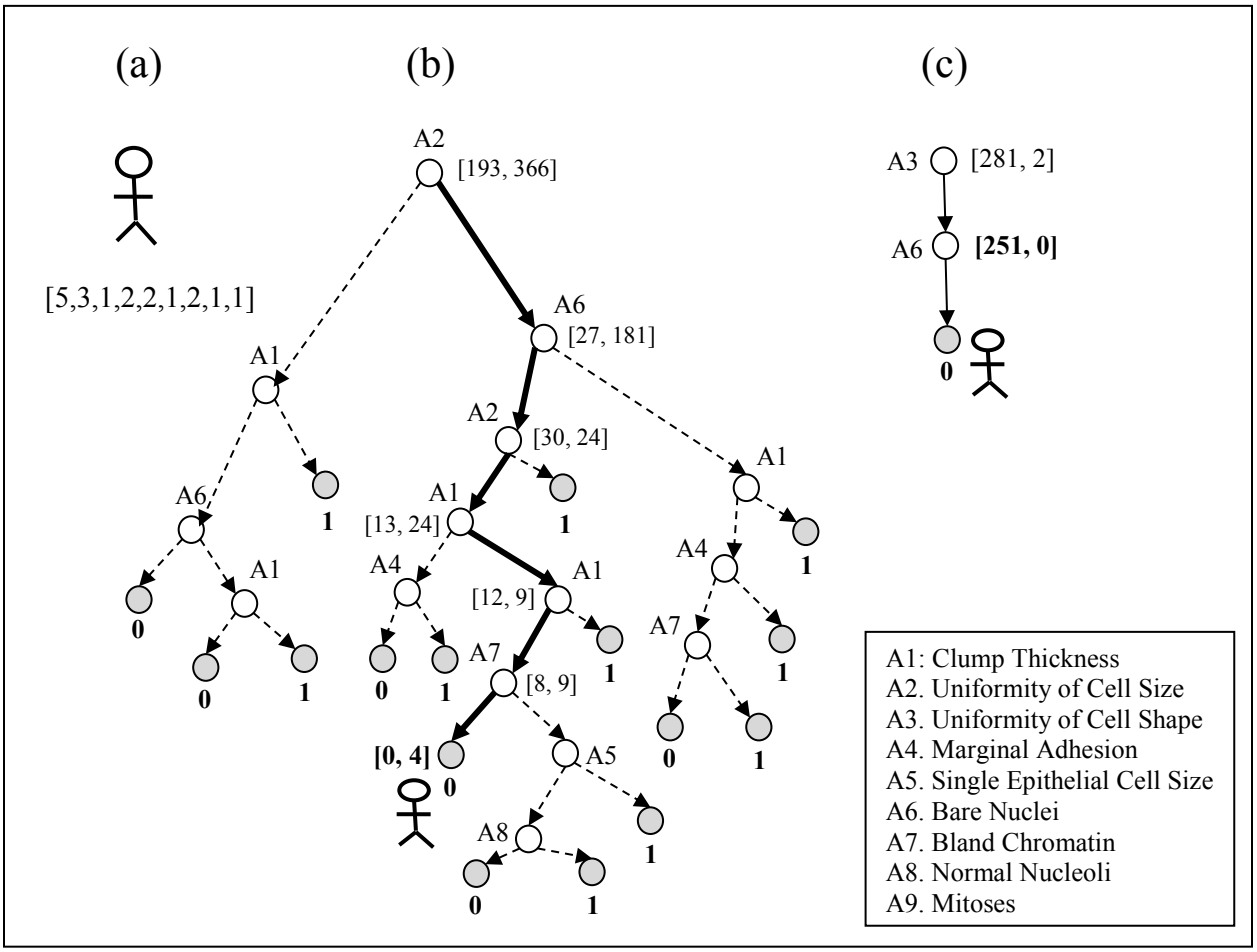


**Figure 1.** Example of a population-wide decision tree and a patient-specific decision path constructed from a breast cancer dataset. Panel (a) represents a patient whose outcome we desire to predict and for whom the values of the nine predictors A1 to A9 are given by the vector as shown. Panel (b) shows the decision tree constructed by the CART algorithm and panel (c) shows the decision path constructed by the PSDP-BA algorithm. For the patient in (a) the decision tree predicts target value 0 via the path given by the solid arrows and the decision path also predicts 0. Note that the decision path is shorter and contains only two predictors while the path in the tree contains six predictors. Also, compared to the path in the decision tree, the patient-specific decision path predicts outcome 0 with higher confidence due to the larger sample size used in estimating the probability parameters.

the antecedent. A greedy step-forward search selects the optimal LBR rule for a test instance to be classified. LBR is an example of an instance-based algorithm that uses values of predictors in the test instance to direct the search for a suitable model in the model space. When compared to range of population-wide algorithms, LBR had higher accuracies on average.

Visweswaran et al. [12, 13] developed and evaluated a patient-specific algorithm that performed Bayesian model averaging over LBR models. This algorithm searched over LBR models using the features of the test case to direct the search. The prediction for the target of the test case was obtained by combining the predictions of the selected models weighted by their posterior probabilities. This algorithm was better than several population-wide algorithms on a pneumonia dataset when evaluated within a clinically relevant range of the ROC curve. Furthermore, patient-specific algorithms that use the features in the test case to direct the search in the model space of Bayesian networks have been developed and applied to predict clinical outcomes with good results [3, 4].

### 2.2 Decision tree algorithms

Population-wide algorithms that construct decision trees recursively partition a dataset to construct a tree model. A model that is learned by these algorithms is called either a classification tree or a regression tree depending on if the

target variable is discrete or continuous. In a tree model, leaves represent target values and branches represent conjunctions of features that lead to those target values. These algorithms usually perform greedy search in a top-down fashion by choosing a predictor at each step that is the next best variable to use in splitting the dataset. The heart of the algorithm is the criterion that is used for selecting the predictors to be included in the tree, and common criteria include the Gini index (used by CART), information gain (used by ID3 and C4.5) and misclassification error [14]. Decision trees are a popular model used in medicine because the tree is easy to interpret and it can be translated to a set of IF-THEN rules either for improved interpretability or for implementation in clinical decision support systems.

## 3. Patient-Specific Decision Path Algorithms

In this section, we describe patient-specific decision path (PSDP) algorithms that are intended to predict well a discrete target of interest, such as a clinical outcome. Given a patient for whom the features are known, these algorithms construct a decision path that contains a subset of the features and apply it to the patient case to predict the target value. The algorithms handle missing features by ignoring those predictors during the construction of the decision path. An illustrative example that contrasts a population-wide decision tree with a patient-specific decision path is shown in Figure 1. In general, a decision path that is constructed for a specific patient may not exist in a population-wide tree constructed from the same training data. Thus, the patient-specific algorithms have the opportunity to construct a simpler model (a shorter path) that may be more accurate for the patient case of interest as shown in Figure 1. We implemented two PSDP algorithms: 1) PSDP-BA that uses *balanced accuracy* as the criterion to select predictors, and 2) PSDP-IG uses *information gain* as the criterion to select predictors. We next describe the PSDP-BA algorithm in detail and then briefly summarize how PSDP-IG differs from PSDP-BA.

### 3.1 PSDP-BA algorithm

The PSDP-BA algorithm uses heuristic search to explore the model space of decision paths since this model space is large. For example, there are $2^m$ decision path models when there are $m$ binary-valued predictors. Though this model space is smaller than the space of population-wide models it is still intractable to perform exhaustive search for typical clinical datasets. Hence heuristic search such as greedy forward-stepping search is appropriate.

A decision path model consists of a conjunction of features that are present in the test case and a probability distribution over the values of the target variable that is estimated from the cases in the training dataset that satisfy the conjunction of features. To control overfitting, we use a Bayesian estimator called the BDeu for estimating the probability parameters of the target. The BDeu estimator has a *prior equivalent sample size* (*pess*) parameter that controls how much smoothing occurs in estimating probability parameters; the higher is *pess*, the greater the smoothing that occurs [15].

The pseudocode for the PSDP-BA algorithm is given in Figure 2. For a given training dataset and test patient case whose target value (e.g., a clinical outcome) we want to predict, PSDP-BA begins with the decision path model that has no predictors. It successively adds predictors to the path and at each step it evaluates each of the remaining predictors using balanced accuracy. Balanced accuracy is defined as the arithmetic mean of sensitivity and specificity, or the average accuracy obtained on each value of the target. A predictor under consideration is temporarily appended to the current path and balanced accuracy is estimated from the training cases that satisfy the temporarily extended path using leave-one-out crossvalidation. More specifically, one of the training cases is left out and the BDeu estimator is used to estimates the probability parameters from the remaining training cases. The target value of the left out training case is predicted using the probability parameters. The process is repeated by leaving out each of the training cases in turn so that a prediction is obtained for each case that satisfies the current path. The balanced accuracy for the predictor under consideration is computed from these predictions using the following expression

$$Bacc = \frac{1}{2}(sensitivity + specificity).$$

Balanced accuracy is superior to accuracy since it takes into account separately the accuracy on each value of the target. Particularly on skewed datasets, very high accuracies can be obtained by the non-discriminatory model that assigns the most prevalent target value in the training dataset to all test patient cases. However, balanced accuracy for such a model will be lower.

```
PSDP-BA (Var, D, Test)
    INPUT:  Var is a set of predictor variables,
            D is a training dataset of patient cases described using Var and target variable Target,
            Test is a test patient case that is not in D and whose Target is to be predicted
    OUTPUT: estimated probability parameters of Target for Test

Path = decision path model with no predictors that is constructed from D
Bacc = balanced accuracy of Path estimated using leave-one-out-crossvalidation (LOOCV) on D
Store Bacc with Path
DBest = D
LOOP till Var is empty or DBest is empty or all cases in DBest have the same value for Target
    BaccBest = 0
    FOR each predictor V in Var whose value is v in Test and is not missing DO
        DTemp = cases in D with V = v
        PathTemp = decision path model that extends Path with V and estimates probability parameters of
          Target from DTemp
        BaccTemp = balanced accuracy of PathTemp estimated using LOOCV on DTemp
        IF BaccTemp > BaccBest
            DBest = DTemp
            VBest = V
            PathBest = PathTemp
            BaccBest = BaccTemp
        END IF
    END FOR
    D = DBest
    Var = remove VBest from Var
    Path = PathBest
    Store BaccBest with VBest in Path
END LOOP
Prune Path bottom-up to the predictor in the path that has the highest balanced accuracy
RETURN the probability parameters of Target that is estimated using Path
```

**Figure 2.** Psuedocode for the patient-specific decision path using balanced accuracy (PSDP-BA) algorithm.

As shown in the pseudocode, the PSDP-BA algorithm extends the path in a greedy forward-stepping fashion until it runs out of predictors, runs out of training cases that satisfy the conjunction of features, or the training cases all have the same target value. Note that in the step for extending the decision path with a new predictor, the algorithm evaluates every predictor that is not already present in the path. Thus, the order in which the predictors are considered does not influence the identification of the next predictor to be added to the path. When the decision path can no longer be extended because one of the termination criteria is satisfied, the path is pruned back to the best-performing predictor in the path. This path is the model that is used to predict the target value of the test patient case.

### 3.2 PSDP-IG algorithm

The PSDP-IG algorithm uses information gain in place of balanced accuracy that is used in the PSDP-BA algorithm. Information gain for a predictor $V$ is given by

$$IG(V) = H(D) - H(D|V),$$

where $D$ is a set of training cases and $H$ denotes entropy and is given by the expression

$$H(X) = -\sum_i P(X = x_i)\, log_2 P(X = x_i).$$

The PSDP-IG algorithm is similar to the Lazy-DT algorithm described by Friedman et al. [10]. However, PSDP-IG differs from Lazy-DT in that it performs pruning and uses BDeu probability parameter estimates when computing the information gain and also to predict the target value of the test patient case.

## 4. Experimental Methods

In this section we describe the clinical datasets on which the algorithms were evaluated, the preprocessing of the datasets, the performance measures used in the evaluation, and the experimental settings used for the algorithms. The clinical datasets included one on pneumonia with one target, one on sepsis with two targets and one on heart failure with two targets. Brief descriptions of the datasets are given in Table 1.

### 4.1 Datasets

**Pneumonia dataset.** The pneumonia dataset that we used contains several hundred variables on patients diagnosed with community acquired pneumonia. The data was collected by the Pneumonia Patient Outcomes Research Team (PORT) using a prospective cohort study of hospitalized and ambulatory care patients in three geographical locations: Pittsburgh, Boston, and Halifax, Nova Scotia. Eligibility criteria were that a patient must (1) be at least 18 years of age, (2) have one or more symptoms suggestive of pneumonia, and (3) have radiographic evidence of pneumonia within 24 hours of presentation [16]. Enrolled patients were followed prospectively to assess their vital status and a variety of outcomes at 30 days after the radiographic diagnosis of pneumonia. One key goal of the PORT project was to develop accurate criteria for prognosis of patients with pneumonia that could provide guidance on which patients should be hospitalized and which patients might be safely treated at home. The data used in our experiments consisted of 2,287 patients diagnosed with community acquired pneumonia and 158 predictors that included demographic information, history and physical examination information, laboratory results, and chest X-ray findings.

The binary target is called dire outcome. A patient was considered to have experienced a dire outcome if any of the following occurred: (1) death within 30 days of presentation, (2) an initial intensive care unit admission for respiratory failure, respiratory or cardiac arrest, or shock, or (3) the presence of one or more specific, severe complications.

**Sepsis dataset.** Sepsis is a syndrome of systemic inflammation in response to infection that can result in multi-system organ dysfunction and failure. The risk factors, causes and prognosis of sepsis are not fully understood. The sepsis data was collected in the GenIMS (Genetic and Inflammatory Markers of Sepsis) project which was coordinated by the Department of Critical Care Medicine in the University of Pittsburgh School of Medicine. GenIMS was a large, multicenter, observational cohort study of patients with community acquired pneumonia (but not necessarily with sepsis) who presented to the emergency departments of 28 hospitals in western Pennsylvania, Connecticut, Michigan, and Tennessee [17]. The data used in our experiments consisted of 1,673 patients who were eventually admitted to a hospital and 21 predictors that included three demographic variables, six clinical variables, two inflammatory markers and 10 genetic variables. The clinical variables are summary variables obtained from data collected at the time of admission and during the first three days of hospital stay.

Two binary targets, which were the focus of investigation in the original study, were selected for prediction: (1) death within 90 days of inclusion in the study, and (2) the development of severe sepsis during the hospitalization.

**Heart failure dataset.** Heart failure affects five million people in the United States leading to about one million hospital admissions each year with a primary discharge diagnosis of heart failure and another approximately two million with a secondary discharge diagnosis of this condition. Accurate evaluation of heart failure patients in the Emergency Department followed by appropriate treatment (including the decision whether to admit a patient to the

**Table 1.** Brief descriptions of the datasets. The # Predictors column gives the number of continuous (cnt) and discrete (dsc) predictors, as well as the total number of predictors (excluding the target variable). The target variables are all binary and denote the presence or absence of a clinical outcome.

| Dataset | # Predictors (cnt + dsc = total) | Target variable | Sample size | Positive outcome count (percent) |
|---|---|---|---|---|
| pneumonia | 38 + 120 = 158 | dire outcome | 2,287 | 261 (11.4%) |
| sepsis-d | 7 + 14 = 21 | death | 1,673 | 189 (11.3%) |
| sepsis-s | 7 + 14 = 21 | severe sepsis | 1,673 | 478 (28.6%) |
| heart failure-d | 11 + 10 = 21 | death | 11,178 | 500 (4.5%) |
| heart failure-c | 11 + 10 = 21 | complications incld. death | 11,178 | 1,255 (11.2%) |

hospital or not) is an important clinical problem. The heart failure data was derived from data collected by 192 general acute care hospitals in Pennsylvania for the year 1999 and consist of heart failure patients who were hospitalized from the Emergency Departments. The data used in our experiments consisted of 11,178 cases and 21 predictors that included demographic, clinical, laboratory, electrocardiographic and radiographic findings. These predictors were identified as prognostic factors in a study that developed a prediction rule to detect patients with heart failure [18].

Two binary targets were selected for prediction: (1) the occurrence of death from any cause during the hospitalization, and (2) the development of one or more serious medical complications (including death) during the hospitalization.

### 4.2 Experiments

The continuous variables were discretized using the entropy-based algorithm developed by Fayyad and Irani [19]. Missing values were imputed using an iterative non-parametric imputation algorithm described by Caruana [20] which has previously been applied to fill in missing predictor values for clinical datasets with good results.

We evaluated the two PSDP algorithms and compared them to the population-wide CART algorithm. We implemented the PSDP algorithms in MATLAB (version R2012a) and for the CART algorithm we used the ClassificationTree.fit implementation in MATLAB (version R2012a). We applied the PSDP-IG, PSDP-BA, and CART algorithms to predict the target in the datasets shown in Table 1. In the PSDP algorithms we set the value of the *pess* parameter to 1. In the CART algorithm we set the impurity criterion to "deviance" which uses information gain to select predictors to include in the decision tree.

We evaluated the algorithms using 20-fold cross-validation. Each dataset was randomly partitioned into 20 approximately equal sets such that each set had a similar proportion of individuals for each target value. For each algorithm, we combined 19 sets and evaluated it on the remaining test set, and we repeated this process once for each possible test set. We thus obtained a prediction for the target for every patient in a dataset. The balanced accuracy and AUC results reported in the next section are computed from these predictions.

We performed all experiments on a PC with two 1.80 GHz Intel Xeon processors and 32 GB of RAM, and running the 64-bit Windows 7 operating system.

### 4.3 Performance measures

The performance of the algorithms was evaluated on two discrimination measures: balanced accuracy and AUC. Both these measures evaluate how well an algorithm differentiates among the values of the target. As mentioned in Section 3.1, balanced accuracy is defined as the arithmetic mean of sensitivity and specificity, or the average accuracy obtained on each value of the target.

## 5. Results

Table 2 gives the balanced accuracies obtained by the three algorithms on five targets. Overall, the two PSDP algorithms perform comparably to CART. Overall, the observed balanced accuracies with the population algorithm (CART) are not statistically significantly different at the 0.05 significance level from those observed with PSDP-BA

**Table 2.** Balanced accuracies for the datasets and targets shown in Table 1. For each algorithm the table gives the mean balanced accuracy obtained from 20-fold cross-validation along with 95% confidence intervals. Highest mean balanced accuracy for each target is in bold.

| Dataset | CART | PSDP-BA | PSDP-IG |
|---|---|---|---|
| pneumonia | **0.6133** [0.5644,0.6623] | 0.6025 [0.5527,0.6523] | 0.5542 [0.5244,0.5841] |
| sepsis-d | **0.6268** [0.5730,0.6805] | 0.6158 [0.5671,0.6645] | 0.6011 [0.5597,0.6424] |
| sepsis-s | 0.6187 [0.5832,0.6542] | 0.6200 [0.5884,0.6516] | **0.6355** [0.6043,0.6667] |
| heart failure-d | 0.5358 [0.5199,0.5518] | **0.5495** [0.5226,0.5763] | 0.5232 [0.5104,0.5361] |
| heart failure-c | 0.5806 [0.5628,0.5984] | **0.5832** [0.5678,0.5987] | 0.5660 [0.5564,0.5767] |

**Table 3.** AUCs for the datasets and targets shown in Table 1. For each algorithm the table gives the mean AUC obtained from 20-fold cross-validation along with 95% confidence intervals. Highest mean AUC for each target is in bold.

| Dataset | CART | PSDP-BA | PSDP-IG |
|---|---|---|---|
| pneumonia | 0.6626 [0.6254,0.6999] | 0.7062 [0.6769,0.7355] | **0.7437** [0.7166,0.7710] |
| sepsis-d | 0.6670 [0.6172,0.7168] | **0.7539** [0.7191,0.7889] | 0.7492 [0.7117,0.7867] |
| sepsis-s | 0.6614 [0.6314,0.6915] | 0.7142 [0.6881,0.7404] | **0.7263** [0.6995,0.7532] |
| heart failure-d | 0.6803 [0.6527,0.7080] | 0.7182 [0.6962,0.7404] | **0.7235** [0.7020,0.7450] |
| heart failure-c | 0.6531 [0.6348,0.6715] | 0.6635 [0.6477,0.6794] | **0.6963** [0.6811,0.7117] |

and PSDP-IG. On comparing the two patient-specific algorithms, though, there is statistically significant difference on the heart failure-c dataset. On this dataset, PSDP-BA with a balanced accuracy of 0.5832 was significantly better than 0.5660 obtained by PSDP-IG.

Table 3 gives the AUCs obtained by the three algorithms on five targets. Overall, the observed AUCs for PSDP-BA and PSDP-IG are statistically significantly higher at the 0.05 significance level compared with the AUCs obtained by CART. The only exception is heart failure-c where the AUCs of CART and PSDP-BA are not statistically significantly different at the 0.05 significance level. On comparing the two patient-specific algorithms, there is no statistically significant difference in performance on any of the datasets.

In summary, the patient-specific algorithms perform comparably to the population-wide algorithm on balanced accuracy but outperform it on AUC.

## 6. Conclusions

We have introduced PSDP, a patient-specific approach for constructing decision path prediction models, and evaluated two algorithms using the PSDP approach on several clinical datasets. The results show that the PSDP algorithms outperform the population-wide algorithm on AUC and maintain comparable performance on balanced accuracy. These results are encouraging and provide support for further investigation into the PSDP approach and more extensive evaluation on a wider range of datasets.

The current PSDP algorithms have several limitations. One limitation is that they handle only discrete variables and continuous variables have to be discretized. A second limitation is that the PSDP algorithms currently use greedy search and better performance may be obtained by using more sophisticated search strategies such as best-first search.

In future work, we plan to examine the complexity of the models generated by the PSDP algorithms and also explore other criteria for selecting predictors. In addition, it will be interesting to explore how patients cluster into subpopulations based on the patient-specific decision paths. Moreover, the presentation of patient-specific decision paths as IF-THEN rules to a domain expert may provide insight into patient subpopulations.

**Acknowledgements**

This research was supported in part by grants NLM HHSN276201000030C from the National Library of Medicine, NIH and IIS-0911032 from the National Science Foundation, and by training grant T15-LM/DE07059 from the National Library of Medicine to the University of Pittsburgh's Biomedical Informatics Training Program.

## References

1. Abu-Hanna, A. and P.J. Lucas, *Prognostic models in medicine. AI and statistical approaches.* Methods Inf Med, 2001. **40**(1): p. 1-5.
2. Cooper, G.F., et al., *Predicting dire outcomes of patients with community acquired pneumonia.* J Biomed Inform, 2005. **38**(5): p. 347-366.

3. Visweswaran, S., et al., *Learning patient-specific predictive models from clinical data.* J Biomed Inform, 2010. **43**(5): p. 669-685.
4. Visweswaran, S. and G.F. Cooper, *Learning instance-specific predictive models.* Journal of Machine Learning Research, 2010. **11**(Dec): p. 3333-3369.
5. Visweswaran, S. *Learning patient-specific models from clinical data*. Department of Biomedical Informatics [PhD dissertation] 2007. Available from: http://etd.library.pitt.edu/ETD/available/etd-11292007-232406/unrestricted/visweswaran_etd__7_Dec_2007.pdf.
6. Gottrup, C., et al., *Applying instance-based techniques to prediction of final outcome in acute stroke.* Artif Intell Med, 2005. **33**(3): p. 223-236.
7. Aha, D.W., D. Kibler, and M.K. Albert, *Instance-based learning algorithms.* Machine Learning, 1991. **6**(1): p. 37-66.
8. Cover, T. and P. Hart, *Nearest neighbor pattern classification.* IEEE Transactions on Information Theory, 1967. **13**(1): p. 21-27.
9. Gagliardi, F., *Instance-based classifiers applied to medical databases: diagnosis and knowledge extraction.* Artif Intell Med, 2011. **52**(3): p. 123-39.
10. Friedman, J.H., R. Kohavi, and Y. Yun. *Lazy decision trees*. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence*. 1996. Portland, Oregon: AAAI Press.
11. Zheng, Z.J. and G.I. Webb, *Lazy learning of Bayesian rules.* Machine Learning, 2000. **41**(1): p. 53-84.
12. Visweswaran, S. and G.F. Cooper. *Instance-specific Bayesian model averaging for classification*. In *Proceedings of the Eighteenth Annual Conference on Neural Information Processing Systems*. 2004. Vancouver, Canada.
13. Visweswaran, S. and G.F. Cooper. *Patient-specific models for predicting the outcomes of patients with community acquired pneumonia*. In *AMIA Annu Symp Proc*. 2005.
14. Duda, R.O., P.E. Hart, and D.G. Stork, *Pattern Classification*. 2000: Wiley Interscience.
15. Heckerman, D., D. Geiger, and D.M. Chickering, *Learning Bayesian networks - the combination of knowledge and statistical data.* Machine Learning, 1995. **20**(3): p. 197-243.
16. Fine, M.J., et al., *Processes and outcomes of care for patients with community-acquired pneumonia: Results from the Pneumonia Patient Outcomes Research Team (PORT) cohort study.* Arch Intern Med, 1999. **159**(9): p. 970-80.
17. Kellum, J.A., et al., *Understanding the inflammatory cytokine response in pneumonia and sepsis: results of the Genetic and Inflammatory Markers of Sepsis (GenIMS) Study.* Arch Intern Med, 2007. **167**(15): p. 1655-63.
18. Auble, T.E., et al., *A prediction rule to identify low-risk patients with heart failure.* Acad Emerg Med, 2005. **12**(6): p. 514-21.
19. Fayyad, U.M. and K.B. Irani. *Multi-interval discretization of continuous-valued attributes for classification learning*. In *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence (IJCAI-93)*. 1993. Chamberry, France.
20. Caruana, R. *A non-parametric EM-style algorithm for imputing missing values*. In *Proceedings of Artificial Intelligence and Statistics*. 2001.