

Development of Predictive Models for Continuous Flow Left Ventricular Assist Device Patients using Bayesian Networks

Natasha A. Loghmanpour¹, Manreet K. Kanwar², Raymond L. Benza², Srinivas Murali² and James F. Antaki^{1,*}

¹Department of Biomedical Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA

²Division of Cardiovascular Diseases, Allegheny General Hospital, Pittsburgh, Pennsylvania, USA

Abstract: *Background:* Existing prognostic tools for patient selection for ventricular assist devices (VADs) such as the Destination Therapy Risk Score (DTRS) and newly published HeartMate II Risk Score (HMRS) have limited predictive ability, especially with the current generation of continuous flow VADs (cfVADs). This study aims to use a modern machine learning approach, employing Bayesian Networks (BNs), which overcomes some of the limitations of traditional statistical methods.

Methods: Retrospective data from 144 patients at Allegheny General Hospital and Integris Health System from 2007 to 2011 were analyzed. 43 data elements were grouped into four sets: demographics, laboratory tests, hemodynamics, and medications. Patients were stratified by survival at 90 days post LVAD.

Results: The independent variables were ranked based on their predictive power and reduced to an optimal set of 10: hematocrit, aspartate aminotransferase, age, heart rate, transpulmonary gradient, mean pulmonary artery pressure, use of diuretics, platelet count, blood urea nitrogen and hemoglobin. Two BNs, Naïve Bayes (NB) and Tree-Augmented Naïve Bayes (TAN) outperformed the DTRS in identifying low risk patients (specificity: 91% and 93% vs. 78%) and outperformed HMRS predictions of high risk patients (sensitivity: 80% and 60% vs. 25%). Both models were more accurate than DTRS and HMRS (90% vs. 73% and 84%), Kappa (NB: 0.56 TAN: 0.48, DTRS: 0.14, HMRS: 0.22), and AUC (NB: 80%, TAN: 84%, DTRS: 59%, HMRS: 59%).

Conclusion: The Bayesian Network models developed in this study consistently outperformed the DTRS and HMRS on all metrics. An added advantage is their intuitive graphical structure that closely mimics natural reasoning patterns. This warrants further investigation with an expanded patient cohort, and inclusion of adverse event outcomes.

Keywords: Risk Stratification, Heart Failure, Bayesian, Decision Support, Prognosis, VAD, Risk Score.

INTRODUCTION

Ventricular Assist Devices (VADs) have made great progress over the past 25 years in demonstrating clinical benefit and durability, and are steadily gaining traction in the treatment of advanced HF [1]. They provide alternatives for patients awaiting transplant (*Bridge to Transplant* – BTT), for those ineligible for transplant (*Destination Therapy* – DT) or in patients where the eventual transplant candidacy is unknown (*Bridge to Decision* – BTd). Current estimates of potential VAD candidates in the US ranges from 80,000 to 200,000 per year [1]. However, high incidence of adverse events diminishes the enthusiasm for this therapy, and hence stifles the rate of growth of its acceptance. Likewise the resulting hospital readmissions of VAD patients –currently exceeding 50% within the first six months of initial discharge – negates potential cost-savings [2]. As VAD use steadily increases [3-5], this has become an increasingly apparent obstacle to widespread application of this therapy.

This problem can be ameliorated by more judicious selection of patients and by implanting patients earlier in the progression of disease [6, 7]. This has motivated the development of risk scores to stratify patients based on the probability of clinical outcome. The most commonly cited of such scores for left ventricular assist device (LVAD) therapy is the Destination Therapy Risk Score (DTRS). However, it has demonstrated mediocre performance [8] in estimating mortality when applied to modern continuous flow LVAD (cfLVAD) populations. This is in part due to its derivation from a patient cohort with first generation pulsatile flow pumps [9]. Accordingly, a more recent risk score was derived exclusively from patients receiving the HeartMate II, the most widely used cfVAD. This HeartMate II Risk Score (HMRS), however, demonstrated only marginal improvement over the DTRS. This score was able to predict 90-day survival based on five variables with an AUC of 70%, but only when applied to the same data from which it was derived. Furthermore, the predictors of long term (1 year) survival was limited to only two variables, age and implant center experience [10]. We hypothesize that this is due to the shortcomings of traditional multivariate analysis, which cannot take into account nonlinear inter-relationships between significant clinical variables.

*Address correspondence to this author at the Department of Biomedical Engineering, 700 Technology Drive, Carnegie Mellon University, Pittsburgh, PA, 15219, USA; Tel: 412-268-9857; Fax: 412-268-9807; E-mail: antaki@cmu.edu

The purpose of this study was therefore to explore the use of modern Machine Learning methods, specifically Bayesian Networks (BNs) to provide a superior prognostic model for survival of LVAD recipients. The advantages of this algorithm include: its intuitive structure, efficient statistical pattern recognition, and tolerance of missing or erroneous data elements.

METHODS

Patient Cohort

The dataset was derived retrospectively from a cohort of 144 patients who received an LVAD at Allegheny General Hospital (AGH) (n=88) and Integris Health System (IHS) (n=56) from January 2007 to December 2011. The inclusion criteria were: known device implant date, a minimum of 90 day post-op follow up information, at least 50% completeness of medical history and the use of a cfLVAD.

Pre-Operative Variable Selection and Data Preprocessing

The machine learning methods used for the present study were built upon techniques used previously by our group for decision support of optimal VAD weaning [11] and the need for right ventricular support due to right ventricular failure in LVAD recipients [12-14]. Retrospective data elements were grouped into four sets (total of 43 elements): demographics (n=9), laboratory tests (n=12), hemodynamics (n=15) and medications (n=7). The predictive class value was 90-day outcome and the patients were divided into two risk groups: *low risk*, defined as those who survived at least 90 days post LVAD implant; and *high risk*, those who died within 90 days. A subsequent pre-processing step was performed to fill in any missing data elements and discretize continuous variables. Data were extracted from preoperative day 14 to 1. For variables with multiple values, the value closest to the time of surgery was used. In circumstances where data elements were missing, data imputation techniques were implemented: mean for the continuous variables, mode for the category variables.

Bayesian Networks

One technique within the broad field of data mining is the Bayesian Network (BN) [15] classifier, which is an efficient algorithm that can recognize causal relationships or correlations among complex clinical variables, as well as relationships between each

variable and the outcome (i.e. survival). BNs are acyclic directed graphs representing joint probability distributions over sets of variables. Every node in the graph represents a random variable. Lack of an arc between two nodes represents conditional independence between the variables that these nodes represent. Nodes are quantified by means of conditional probability tables (CPTs) associated with each node, describing the likelihood of the variable's values conditioned upon the values of adjacent nodes. For each node in the graph network, mutually exclusive and cumulatively exhausted states were defined and the CPT embedded. The joint probability is expressed as:

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P(X_i | \text{parents}(X_i))$$

Where $P(X_i | \text{parents}(X_i))$ represents the conditional probability of variable X_i , given the occurrence of its parent nodes.

The Naïve Bayes (NB) [16] classifier seen in Figure 1 is a simplified Bayesian classifier that uses Bayes theorem to predict which class a patient belongs to. Let C denote the class (mortality in this case) of the instance (patient), where c_1 =alive and c_2 =dead, and let $X = \{x_1, x_2, \dots, x_n\}$ be a vector of pre-operative variables (demographics, laboratory tests etc.) denoting the observed variable values (e.g. female or blood pressured of 90). For a patient with pre-operative variables X , NB uses the following equation to calculate the probability of the patient belonging to c_1 or c_2 :

$$\Pr^*(c) = \arg \max_{c \in \{c_1, c_2\}} p(C = c | x_1, x_2, \dots, x_n) = \arg \max_{c \in \{c_1, c_2\}} p(C = c) \prod_{i=1}^n p(X_i = x_i | C = c)$$

Where $p(X_i = x_i | C = c)$ represents the conditional probability of the pre-operative variable X_i in specific states x_i in the different classes c_1 and c_2 .

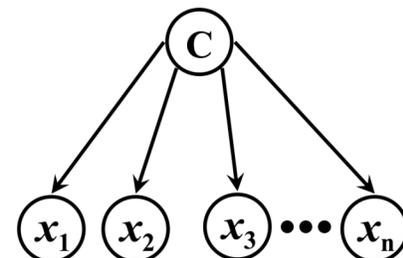


Figure 1: Representation of a Naïve Bayes classifier.

In this study, we investigated both the NB and Tree-Augmented Naïve Bayes (TAN) [17, 18] (Figure 2), where the former allows for only a single parent node and the latter allows two or more arrows to any single node. For example, NB could have INR connected to (associated with) survival and TAN could have INR connected to (associated with) both survival and age. TAN computes the conditional mutual information between pre-operative variables:

$$I(X_j, X_k | C) = \sum_{x_j, x_k, c_i} P(X_j, X_k, c_i) \log_2 \frac{P(x_j, x_k | c_i)}{P(x_j | x_i)P(x_k | c_i)}$$

Where X_j and X_k are pre-operative variables with the values x_j and x_k , respectively; C is the class node; $P(x_j, x_k, c_i)$ is the joint probability of $X_j = x_j$ and $X_k = x_k$; $C = c_i$ in the dataset; $P(x_j, x_k | c_i)$, $P(x_j | x_i)$ and $P(x_k | c_i)$ are the corresponding conditional probabilities. This measures the mutual dependence between two pre-operative variables given the class variable. Building the maximum weighted spanning tree using this index, TAN represents the dependences among the pre-operative variables, as well as the class variable.

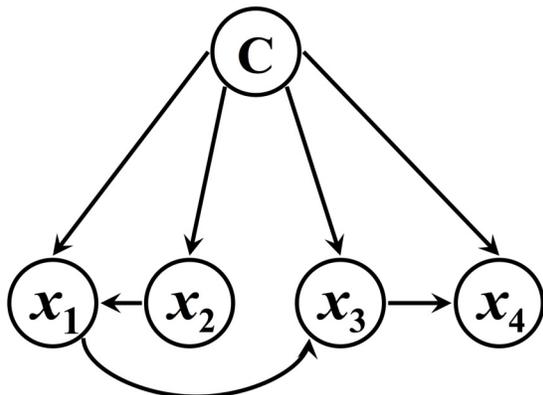


Figure 2: Representation of a TAN classifier with four pre-operative variables.

The example provided in Figure 2 would be represented as:

$$p(C|x_1, x_2, x_3, x_4) \propto p(C)p(x_1|C, x_2)p(x_2|C)p(x_3|C, x_1)p(x_4|C, x_3)$$

The addition of edges between nodes allows for dependence between the variables, which is often the case in medical datasets. For additional details of the underlying BN theory and derivation, please refer to [19]. The final NB and TAN models were based on a subset of clinical variables (called features) out of the original 43. This was done by a process known as *feature selection*, which reduces the total number of clinical variables to avoid over-fitting the model to the

dataset. Using this method, the set of pre-operative variables was first ranked by Chi-Square (χ^2) analysis:

$$\chi^2 = \sum_{ij} (O_{ij} - E_{ij})^2 / E_{ij}$$

Where O =observed outcome and E =expected outcome. Once the variables are ranked, the top 10 most predictive variables were included in the model.

A two-step process was used to train and test the model to avoid overfitting to the data. First, a ten-fold cross validation was used for optimization of the algorithm parameter and model comparison. To perform this step, the data is subdivided into 10 sets of size $n/10$, where the model is trained on 9 of the sets and tested on 1. This is repeated 10 times and the mean accuracy is used to assess the final model's performance. In the second step, the resulting classifier/model was then evaluated on an independent testing/holdout dataset that has not been observed by the model. A training set was comprised of 66% of the data records while the testing set used the remaining 33%. The models were derived, built and implemented using an open-source machine learning software library (WEKA, Waikato Environment for Knowledge Analysis).

Calculation of HMRS and DTRS

The DTRS uses a sum-of-points to calculate the weighted risk score and identify patients as low, medium or high risk. The risk factors are: platelet count $\leq 148 \times 10^3 / \mu\text{L}$ (7 points) + serum albumin $\leq 3.3 \text{ g/dL}$ (5 points) + international normalized ratio > 1.1 (4 points) + vasodilator therapy (4 points) + mean pulmonary artery pressure $\leq 25 \text{ mmHg}$ (3 points) + aspartate aminotransferase $> 45 \text{ U/mL}$ (2 points) + hematocrit $\leq 34\%$ (2 points) + blood urea nitrogen $> 51 \text{ U/dL}$ (2 points) + no intravenous inotropes (2 points). Presence of a risk factor is coded as 1 and absence as 0, which is then multiplied by the associated points and summed to give the final score. Low = 0 to 8 points, medium = 9 to 16, high = 17 to 19 and very high ≥ 19 where the predicted 90-day mortality rate was $<10\%$, 10-50%, 50-70% and $>70\%$, respectively.

The HMRS is calculated by the following simple equation: center experience (0 if ≥ 15 VADs/year, 1 if ≤ 15 VADs/year), age (per 10 years), Albumin, Creatinine, and INR. The HMRS equation = $(0.0274 \times [\text{age in years}]) - (0.723 \times [\text{albumin}]) + (0.74 \times [\text{creatinine}]) + (1.136 \times [\text{INR}]) + (0.807 \times [\text{center LVAD volume} < 15])$. Patients are then categorized into low

risk (HMRS < 1.58), medium risk (HMRS: 1.58-2.48) and high risk (HMRS > 2.48), where the 90-day mortality rate is predicted as 8%, 11% and 25%, respectively.

Statistical Analysis

The differences between groups (high vs. low risk) were evaluated by chi-square test. The comparison of continuous variables between the high and low risk patients was performed using either the Student *t*-test when normally distributed or the Wilcoxon rank sum test when distributed otherwise using SAS (version 9.3, Cary North Carolina). Categorical variables were compared using the Fisher exact test [20] with a commercial software package (MedCalc, version 9.5.0.0, Mariakerke, Belgium). Odds ratios (OR) and confidence intervals (CI) were also computed for both the high and low risk patients.

Sensitivity was defined as the ratio of true positives to the sum of true positives and false negatives (i.e. high risk patients correctly predicted as high risk). *Specificity* was defined as the ratio of true negatives to the sum of true negatives and false positives (i.e. low risk patients correctly predicted as low risk). Standard receiver operator characteristic (ROC) curves were constructed to illustrate overall sensitivity and specificity. The area under the ROC curve (AUC %) was calculated as an index of overall performance. These three metrics in addition to the Kappa statistic was used for comparison to the DTRS and HMRS.

Survival Analysis

Patient survival estimates were plotted using Kaplan-Meier curves. Differences between stratified groups was compared using log-rank statistics [21]. The plots were created using JMP (Version 10. SAS Institute Inc., Cary, NC). Survival was calculated from the day of implant. Patients were censored if they received a heart transplant or device explant, which excluded 16 patients from the 144 cohort.

RESULTS

A summary of the pre-implant patient characteristic statistics can be seen in Tables 1 and 2. The overall accuracy, Kappa statistic and AUC % of both the NB and TAN models (Figure 3) exceeded the DTRS and HMRS in all three parameters (Table 3). The associated ROC curves (Figure 4) reveal the dramatic difference in performance between the Bayesian models (Figure 3) and traditional scores (NB: 79.5%

and TAN: 83.6% vs. the DTRS: 59% and HMRS: 59%) when each was applied to our patient cohort. Likewise, the confusion matrix (Table 4) reveals the clear superiority of the NB and TAN (80% and 60%, respectively) at correctly identifying high risk patients as compared to the DTRS and HMRS (40% and 25%, respectively). Similarly, the false negatives (high risk patients inaccurately predicted as low risk) were much lower for the NB and TAN (20% and 40%) than both the DTRS and HMRS (60% and 75%). Although the NB and TAN (90.9% and 93.2%) performed similar to the HMRS (93.5%) in terms of specificity, they all outperformed the DTRS (78.2%). The NB, TAN and HMRS had similar frequency of false positives (9.1%, 6.8% and 6.4%), while the DTRS (21.4%) had a much higher percentage of incorrectly classified low risk patients.

Both the NB and TAN BN models were built from datasets using 10 most predictive variables, which emerged from feature selection based on their predictive power. The features listed in order from highest to lowest predictive power: hematocrit (%), aspartate aminotransferase (IU/liter), age, heart rate (beats/min), transpulmonary gradient (mm Hg), mean pulmonary artery pressure (mm Hg), use of diuretics, platelet count (10^9 /liter), blood urea nitrogen (mg/dl) and hemoglobin (g/dl). Many of the predictive features identified overlap with the DTRS (platelet count, mean pulmonary artery pressure, aspartate aminotransferase, hematocrit %, and BUN), yet only one overlapped with the HMRS (Age). There was no common variable that was included in all four predictive models.

Survival Outcome

The Kaplan Meier survival curves are provided in Figures 5-8 and summary in Table 5 for the DTRS, HMRS, NB, and TAN, with the patient population stratified into high and low risk. (The Kaplan Meier curve for the actual survival is provided in Figure 9 for reference). Both the DTRS and HMRS curves were derived from 136 patients (16 of the 144 were censored due to heart transplant within 90 days of VAD implant) and the NB and TAN were derived from 43 patients since the model performance was evaluated on the test set (48 patients, 5 of which were censored due to heart transplant within 90 days of VAD implant). The difference between the two groups was statistically significant for all models (log ranks: DTRS = 0.008, HMRS = 0.006, NB = <0.0001, TAN = 0.001).

Table 1: Baseline Characteristics

Patient Characteristics	All Data mean \pm SD (n=144)	High Risk (n=20)	Low Risk (n=124)	p-value OR (95% CI)
Demographics				
Age	58.4 \pm 12	65 \pm 7.4	57.4 \pm 12.3	0.01 1.08 (1.02-1.14)
Female, %	28 (20)	7 (35)	21 (17)	(female) 0.06 2.6 (0.94-7.41)
Ischemic etiology, %	81 (56)	12 (60)	69 (55)	(ischemic) 0.72 1.20 (0.46-3.13)
Weight (kg)	94.5 \pm 21.2	94.2 \pm 21.6	94.5 \pm 21.2	0.96 0.99 (0.98-1.02)
BMI (kg/m ²)	30.7 \pm 6.2	31.7 \pm 6.8	30.5 \pm 6.1	0.46 1.03 (0.95-1.11)
Implant Urgency*, %	Elective 60 (42) Emergent 33 (23) Urgent 51 (35)	Elective 9 (45) Emergent 5 (25) Urgent 6 (30)	Elective 51 (41) Emergent 28 (22) Urgent 45 (36)	0.86†
LVAD indication, %	BTT 84 (58) DT 43 (30) BTD 17 (12)	BTT 14 (70) DT 4 (20) BTD 2 (10)	BTT 70 (56) DT 39 (31) BTD 15 (12)	0.51‡
Laboratory values				
HGB (g/dl)	11 \pm 1.9	10.2 \pm 1.8	11.1 \pm 1.9	0.04 0.74 (0.56-0.99)
WBC (*10E+09/L)	9.6 \pm 3.7	9.6 \pm 3.9	9.6 \pm 3.6	0.75 1.02 (0.90-1.16)
HCT (%)	33.9 \pm 5.4	31.4 \pm 5.9	34.2 \pm 5.2	0.03 0.90 (0.81-0.99)
Albumin (g/dl)	3.2 \pm 0.5	3.1 \pm 0.4	3.2 \pm 0.6	0.31 0.63 (0.26-1.52)
ALT (IU/L)	108.9 \pm 350.1	71.4 \pm 106.8	114.9 \pm 374.9	0.72 1.00 (0.99-1.00)
Total Bilirubin (mg/dl)	1.3 \pm 1.1	1.7 \pm 1.9	1.3 \pm 1	0.09 1.34 (0.9-1.9)
INR	1.5 \pm 0.9	1.5 \pm 0.3	1.5 \pm 1	0.83 0.93 (0.51-1.71)
BUN (mg/dl)	29.9 \pm 18.3	37.8 \pm 24.8	28.6 \pm 16.8	0.04 1.03 (1.00-1.05)
Creatinine (mg/dl)	1.4 \pm 0.7	1.6 \pm 1	1.4 \pm 0.6	0.13 1.54 (0.87-2.71)
AST (IU/L)	154.1 \pm 799.8	69.7 \pm 75.4	167.7 \pm 861.1	0.66 1.00 (0.99-1.00)
Platelet Count (*10E+09/L)	189.8 \pm 78.7	164.6 \pm 95.8	193.8 \pm 75.2	0.12 0.99 (0.99-1.00)
Sodium (MEQ/L)	134.9 \pm 5.1	133.6 \pm 5.7	135.1 \pm 5	0.22 0.94 (0.86-1.04)
IABP, %	55 (38)	8 (40)	47 (38)	0.72 1.20 (0.46-3.13)

(Table 1). Continued.

Patient Characteristics	All Data mean \pm SD (n=144)	High Risk (n=20)	Low Risk (n=124)	p-value OR (95% CI)
Hemodynamics				
PA systolic (mmHg)	55.7 \pm 15.9	54.6 \pm 12.8	55.8 \pm 16.4	0.76 0.99 (0.96-1.03)
PA diastolic (mmHg)	26 \pm 8.2	23.5 \pm 6.5	26.4 \pm 8.4	0.34 0.97 (0.90-1.03)
PA mean (mmHg)	38.1 \pm 10.6	36.1 \pm 7.2	38.4 \pm 11	0.41 0.98 (0.93-1.03)
RV systolic (mmHg)	56.3 \pm 14.1	54 \pm 11.8	56.6 \pm 14.5	0.47 0.99 (0.95-1.02)
RV diastolic (mmHg)	10.2 \pm 6.2	8.9 \pm 6	10.5 \pm 6.2	0.39 0.92 (0.76-1.11)
RAP (mmHg)	13.9 \pm 6.8	14.2 \pm 4.9	13.9 \pm 6.9	0.55 0.96 (0.84-1.09)
PCWP (mmHg)	25 \pm 9.4	24.1 \pm 8.5	25.1 \pm 9.6	0.80 0.99 (0.93-1.05)
TPG (mmHg)	13.4 \pm 7	11.7 \pm 6.1	13.6 \pm 7.1	0.41 0.95 (0.84-1.07)
PVR (WU)	155.5 \pm 214.2	155.1 \pm 204.9	155.5 \pm 216.2	0.90 1.00 (0.99-1.00)
PA sat (%)	52 \pm 11.6	53.3 \pm 13.6	51.8 \pm 11.4	0.62 1.01 (0.97-1.06)
CO (liters/min)	4.4 \pm 1.6	4.6 \pm 1.5	4.4 \pm 1.6	0.68 1.06 (0.79-1.42)
CI (liters/min/m ²)	2.06 \pm 0.7	2.3 \pm 0.6	2 \pm 0.7	0.17 1.59 (0.82-3.11)
HR (bpm)	89.2 \pm 17.7	85.5 \pm 13.7	89.8 \pm 18.2	0.32 0.99 (0.96-1.01)
Medication, %				
Digoxin	76 (53)	9 (45)	67 (54)	0.45 0.70 (0.27-1.80)
Diuretic	84 (58)	6 (30)	78 (63)	0.008 0.25 (0.09-0.70)
ACE inhibitors or A-II antagonists	55 (38)	5 (25)	50 (40)	0.20 0.49 (0.17-1.44)
β -blockers	74 (51)	10 (50)	64 (52)	0.89 0.94 (0.36-2.41)
Vasodilators	31 (21)	4 (20)	27 (28)	0.86 0.90 (0.28-2.91)
Antiarrhythmic	48 (33)	7 (35)	41 (33)	0.86 1.09 (0.40-2.94)
Inotrope	99 (69)	12 (60)	87 (70)	0.36 0.64 (0.24-1.70)

BMI: body mass index, BTT: bridge to transplant, BTD: bridge to decision, DT: destination therapy, HGB: hemoglobin, WBC: white blood cell count, HCT: hematocrit, ALT: Alanine Aminotransaminase, INR: international normalized ratio, BUN: blood urea nitrogen, AST: Aspartate Aminotransferase, IABP: intra-aortic balloon pump, PA: pulmonary artery, RV: right ventricular, RAP: right atrial pressure, PCWP: pulmonary capillary wedge pressure, TPG: transpulmonary pressure gradient, PVR: pulmonary vascular resistance, CO: cardiac output, CI: cardiac index, HR: heart rate.

*Elective – LVAD scheduled as an outpatient, Urgent – LVAD scheduled same admission as heart failure decompensation, Emergent – following extracorporeal membrane oxygenation, IABP, acute coronary syndrome or ventricular tachycardia.

†Compared elective vs. urgent and emergent.

‡Compared BTT vs. DT.

Table 2: LVAD Devices

Device	N (%)
HeartMate II	128 (88)
Levacor	5 (3.4)
Duraheart	8 (5.5)
Jarvik	1 (0.7)
Ventrassist	2 (1.4)

The DTRS (Figure 5) predicted 32 high risk and 96 low risk patients. Survival for high and low risk groups was 94% and 93%, respectively, at 30 days post-LVAD, 72% and 84% at 90 days, 50% and 71% at 180 days, and 22% and 40% at 1 year.

The HMRS (Figure 6) predicted 11 high risk and 117 low risk patients. Survival for high and low risk groups was 73% and 92%, respectively, at 30 days

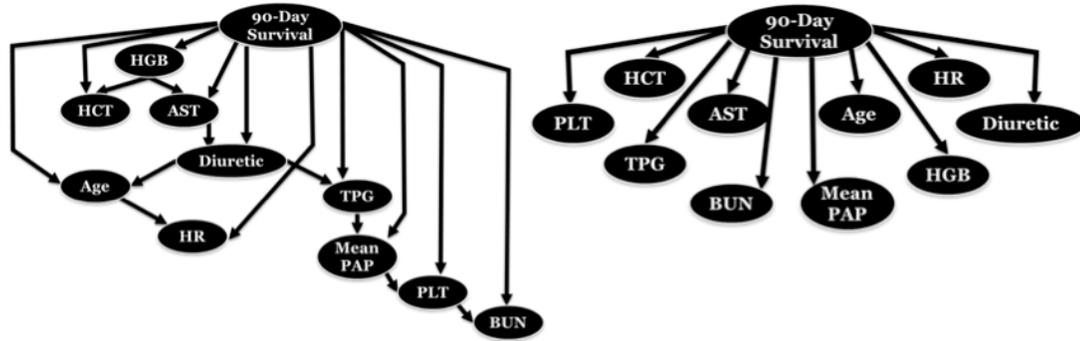


Figure 3: (right) Naïve Bayes model for 90-day survival using the top 10 most predictive features (left) Tree-Augmented Naïve Bayes model. Clinical variables: HGB: hemoglobin, HCT: hematocrit, BUN: blood urea nitrogen, AST: Aspartate Aminotransferase, PAP: pulmonary artery pressure, TPG: transpulmonary pressure gradient, HR: heart rate, PLT: platelet count.

Table 3: Performance on AGH and Integris Dataset

	NB	TAN	DTRS	HMRS
Accuracy (%)	89.8	89.8	72.9	84.0
Kappa	0.56	0.48	0.14	0.22
AUC (%)	79.5	83.6	59.0	59.0

NB: Naïve Bayes, TAN: Tree-Augmented Naïve Bayes, DTRS: Destination Therapy Risk Score, HMRS: HeartMate II Risk Score.

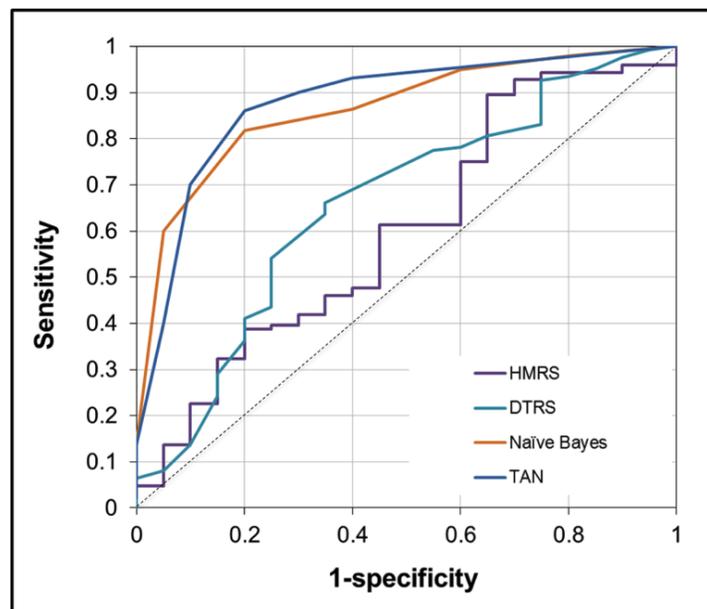


Figure 4: Receiver operating characteristic curve of the HeartMate Risk Score (AUC 59%), Destination Therapy Risk Score (AUC 59%), Naïve Bayes model (AUC 79.5%), TAN model (AUC 83.6%) and line of unity (50%).

Table 4: Confusion Matrix

	Predicted High Risk	Predicted Low Risk
True High Risk	True Positive (Sensitivity) NB (80 %) TAN (60) DTRS (40) HMRS (25)	False Negative NB (20 %) TAN (40) DTRS (60) HMRS (75)
True Low Risk	False Positive NB (9.1 %) TAN (6.8) DTRS (21.8) HMRS (6.4)	True Negative (Specificity) NB (90.9 %) TAN (93.2) DTRS (78.2) HMRS (93.5)

NB: Naïve Bayes, TAN: Tree-Augmented Naïve Bayes, DTRS: Destination Therapy Risk Score, HMRS: HeartMate II Risk Score.

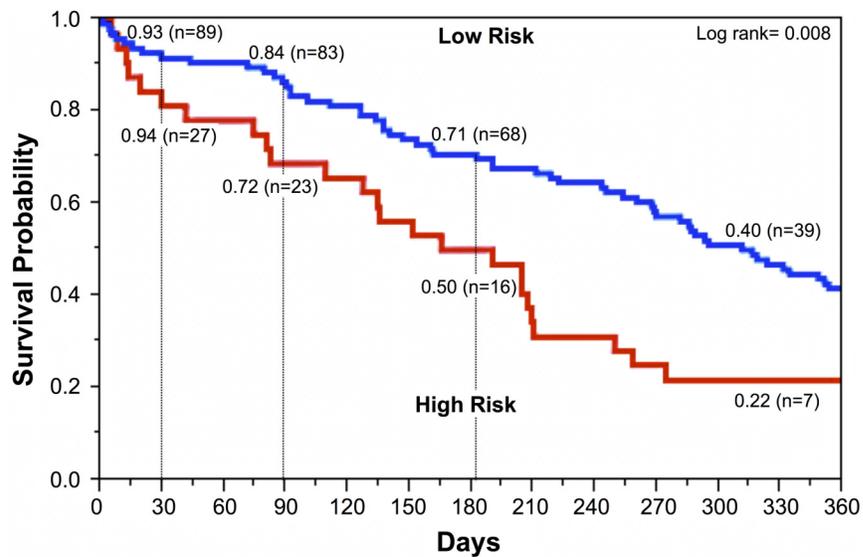


Figure 5: Kaplan Meier survival curve for the Destination Therapy Risk Score.

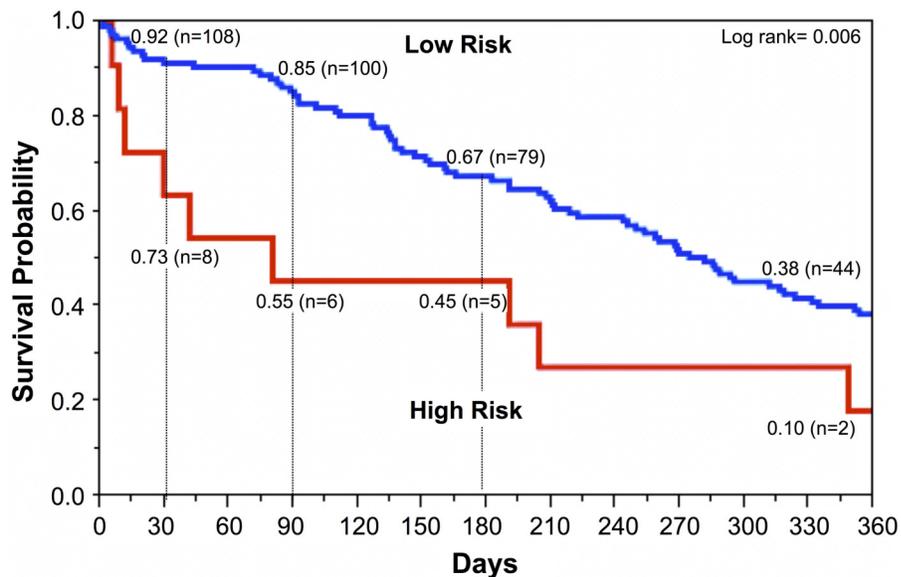


Figure 6: Kaplan Meier survival curve for the HeartMate II Risk Score.

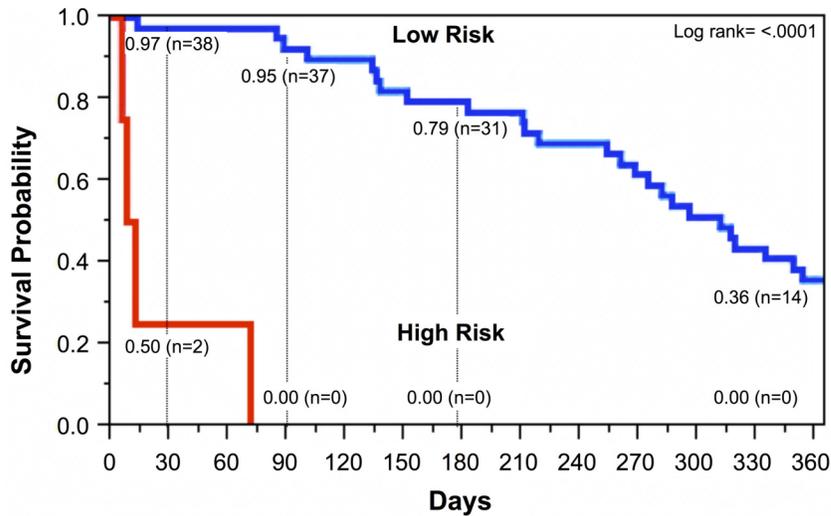


Figure 7: Kaplan Meier survival curve for the Naïve Bayes model.

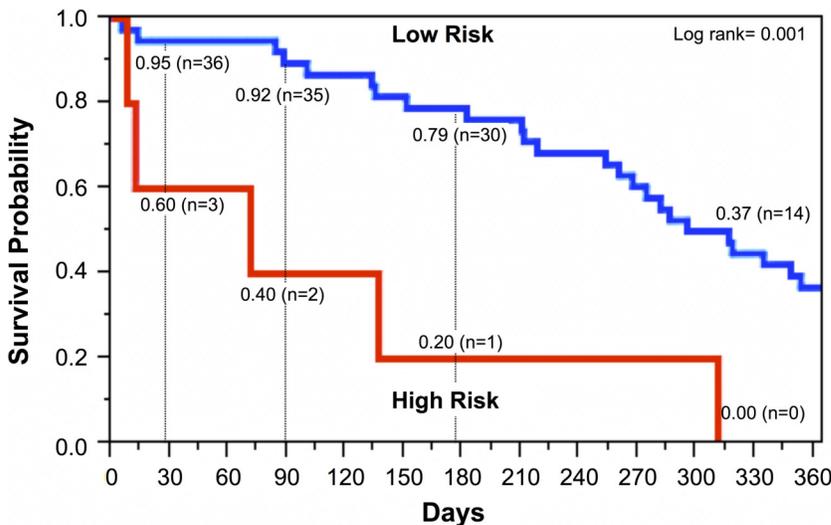


Figure 8: Kaplan Meier survival curve for the Tree-Augmented Naïve Bayes model.

Table 5: Kaplan Meier Survival Curve Summary

Risk Score	Risk Group (N)	Mean Time to Event	30-Day Survival N (%)	90-Day Survival N (%)	1-Year Survival N (%)
DTRS	High (32)	177 days	27 (94%)	23 (72%)	7 (22%)
	Low (96)	313 days	89 (93%)	83 (84%)	39 (40%)
HMRS	High (11)	80 days	8 (73%)	6 (55%)	2 (10%)
	Low (117)	281 days	108 (92%)	100 (85%)	44 (38%)
Naïve Bayes	High (4)	10 days	1 (25%)	0 (0%)	0 (0%)
	Low (39)	311 days	38 (97%)	37 (95%)	14 (36%)
TAN	High (5)	71 days	3 (60%)	2 (40%)	0 (0%)
	Low (38)	305 days	36 (95%)	35 (92%)	14 (37%)

TAN: Tree-Augmented Naïve Bayes, DTRS: Destination Therapy Risk Score, HMRS: HeartMate II Risk Score.

post-LVAD, 55% and 85% at 90 days, 45% and 67% at 180 days, and 10% and 38% at 1 year.

The NB model (Figure 7) predicted 4 high risk and 39 low risk patients. Because of the precipitous death

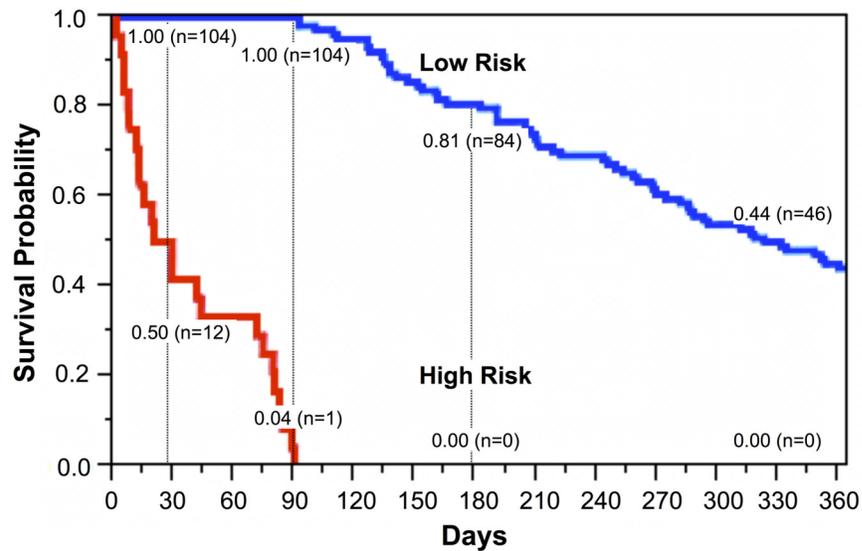


Figure 9: Kaplan Meier survival curve for the actual clinical outcomes.

rate of the NB high risk group, survival comparisons were made at 8 and 12 days, which were 75% and 50%, respectively. No patient identified as high risk survived beyond 71 days, consistent with the Kaplan Meier for the actual outcomes (Figure 9). Survival for the low risk group was 97% at 30 days post-LVAD, 95% at 90 days, 79% at 180 days, and 36% at 1 year.

Finally, the TAN model (Figure 8) predicted 5 high risk and 38 low risk patients. Survival for the high risk group was 60% at 30 days post-LVAD, 40% at 90 days, 20% at 180 days and 0% at 1 year. Survival for the low risk group was 95% at 30 days post-LVAD, 92% at 90 days, 79% at 180 days, and 37% at 1 year.

DISCUSSION

Although Bayesian statistics were first introduced (by Bayes) over 300 years ago, their utility in the medical workflow was only recognized within the past 25 years [22]. Over the past ten years decision support models based on BNs have been published for a wide variety of medical disciplines [23-28]. This study is however the first application of such models on a cohort of LVAD recipients.

Nearly all the variables in the models have been previously cited as being predictive for mortality and morbidity of LVAD recipients, including hematocrit [9], aspartate aminotransferase [9], age [10, 29], mean pulmonary artery pressure [9], platelet count [9] and blood urea nitrogen [9]. The two classifiers investigated here exhibited different strengths and weaknesses. Although the NB classifier is less sophisticated than the TAN, it performed the best of all four scores that were

evaluated. However, as the number patients and variables increase, the performance of TAN is likely to improve, while the NB would likely plateau and then diminish. This is due to the unique characteristic of the TAN classifier that accounts for both the influence of each independent variable on other variables. For example, hemoglobin < 9.25 g/dL was found to decrease the chance of survival twofold irrespective of other variables; whereas the influence of platelet count on survival was dependent upon the mean pulmonary artery pressure. (Data not shown.) With mean PAP less than 45.5 mmHg, platelet count < $123 \times 10^3/\mu\text{L}$ decreased survival twofold; but with PAP greater than this cutoff, the platelet count decreased as much as four-fold. Hematocrit < 29% increased chance of death by 3.5 times, which was further amplified to nearly six-fold when hemoglobin was also between 9.25 and 10.35 g/dL. Blood urea nitrogen > 44.5 mg/dL increased risk of death twofold, but when platelet count > $258 \times 10^3/\mu\text{L}$, mortality was reduced by nearly the same amount.

The suboptimal performance of the DTRS when applied to a cohort of continuous flow devices has already been published [8]. The HMRS was published within the past year, and therefore has yet to be compared to other patient cohorts or against other risk scores. In contrast to the traditional Frequentist statistical methods comprised of weighted combinations of independent variables, BNs provide: 1) a rigorous probabilistic framework in which to perform inference on unknown or predictive variables, 2) a means of learning probabilistic information from data, 3) the ability to capture the nonlinear inter-relationships

amongst independent variables 4) the ability to accommodate missing data elements or sparse data records (common to medical datasets) and 5) a visual representation that is easy to interpret by human experts. Overall, these qualities provide more an accurate depiction of human decision making process and improved performance. Consequently they are more likely to be adopted into the clinical workflow than an anonymous “black box” type risk score.

The primary limitation in this study was, the moderate size of patient cohort used to create the models. Another potential limitation was institutional bias by including only two clinical programs. Missing data elements is another major obstacle that is often unavoidable when performing retrospective studies using medical data. Although we were able to use WEKA to substitute the missing variables with suitable surrogates, there is no true replacement for actual measured/observed data. We will address these limitations in future planned studies that expand the number of patient, broaden the endpoints to include adverse events such as stroke and infection, and extend the timeline to 6-months and 1-year. This will also provide an opportunity to update our existing prognostic model for RV failure [12, 13].

Another important step towards translation of this model to clinical practice will be to incorporate it into a computer application that may be consulted at key decision points. This will also offer an opportunity to introduce expert knowledge and user customization – both features hopefully enhancing the ultimate utility of the model.

BNs are able to more closely model the natural clinical decision-making process as compared to traditional risk scores and therefore offer a valuable tool for medical decisions such as identifying candidates most likely to benefit from VAD therapy. This study demonstrated a remarkable improvement over the existing risk scores (DTRS and HMRS) with respect to accuracy, Kappa, sensitivity, and specificity, also illustrated by predicted Kaplan-Meier survival curve that most closely resembles the actual outcomes. The BNs consistently outperformed the DTRS and HMRS due to (1) their abilities to learn from prior probability, (2) account for relationships between variables and (3) tolerate missing data elements. These results encourage continued validation and expansion of the models with a prospective, multicenter study.

ACKNOWLEDGEMENT

The authors would like to thank Dr. Trevor Snyder, Dr. James Long and Dr. Douglas Horstmanhof at Integris Health System for the contribution of retrospective data.

FUNDING SOURCE

The research was supported in part by the NIH grant R01HL086918-01 and we would like to thank the head Principal Investigator Dr. John Gorcsan for his support.

DISCLOSURES

None of the authors have any financial relationship with any commercial entities that have an interest in the subject of the presented manuscript or other conflicts of interest to disclose.

REFERENCES

- [1] Lietz K, Miller LW. Destination therapy: Current results and future promise. *Seminars in thoracic and cardiovascular surgery* 2008; 20(3): 225-33. <http://dx.doi.org/10.1053/j.semthcvs.2008.08.004>
- [2] Bonde P, Dew MA, Meyer D, Tallaj JJ, Martin T, Hollifield KA, *et al.* 4 national trends in readmission(rea) rates following left ventricular assist device (lvad) therapy. *The Journal of heart and lung transplantation: the official publication of the International Society for Heart Transplantation* 2011; 30(4): S9.
- [3] Rose EA, Frazier OH. Resurrection after mechanical circulatory support. *Circulation* 1997; 96(2): 393-5.
- [4] Schocken DD, Arrieta MI, Leaverton PE, Ross EA. Prevalence and mortality rate of congestive heart failure in the united states. *J Am Coll Cardiol* 1992; 20(2): 301-6. [http://dx.doi.org/10.1016/0735-1097\(92\)90094-4](http://dx.doi.org/10.1016/0735-1097(92)90094-4)
- [5] Murali S. Mechanical circulatory support with the novacor lvas: World-wide clinical results. *The Thoracic and cardiovascular surgeon* 1999; 47 Suppl 2: 321-5. <http://dx.doi.org/10.1055/s-2007-1013191>
- [6] Miller LW. Left ventricular assist devices are underutilized. *Circulation* 2011; 123(14): 1552-8. <http://dx.doi.org/10.1161/CIRCULATIONAHA.110.958991>
- [7] Stewart GC, Stevenson LW. Keeping left ventricular assist device acceleration on track. *Circulation* 2011; 123(14): 1559-68. <http://dx.doi.org/10.1161/CIRCULATIONAHA.110.982512>
- [8] Teuteberg JJ, Ewald GA, Adamson RM, Lietz K, Miller LW, Tatroles AJ, *et al.* Risk assessment for continuous flow left ventricular assist devices: Does the destination therapy risk score work? *J Am Coll Cardiol* 2012; 60(1): 44-51. <http://dx.doi.org/10.1016/j.jacc.2012.02.032>
- [9] Lietz K, Long JW, Kfoury AG, Slaughter MS, Silver MA, Milano CA, *et al.* Outcomes of left ventricular assist device implantation as destination therapy in the post-rematch era - implications for patient selection. *Circulation* 2007; 116(5): 497-505. <http://dx.doi.org/10.1161/CIRCULATIONAHA.107.691972>
- [10] Cowger J, Sundareswaran K, Rogers JG, Park SJ, Pagani FD, Bhat G, *et al.* Predicting survival in patients receiving

- continuous flow left ventricular assist devices: the heartmate II risk score. *J Am Coll Cardiol* 2013; 61(3): 313-21.
<http://dx.doi.org/10.1016/j.jacc.2012.09.055>
- [11] Santelices LC, Wang Y, Severyn D, Druzdzel MJ, Kormos RL, Antaki JF. Development of a hybrid decision support model for optimal ventricular assist device weaning. *The Annals of thoracic surgery* 2010; 90(3): 713-20.
<http://dx.doi.org/10.1016/j.athoracsur.2010.03.073>
- [12] Wang Y, Simon MA, Bonde P, Harris BU, Teuteberg JJ, Kormos RL, *et al.* Decision tree for adjuvant right ventricular support in patients receiving a left ventricular assist device. *J Heart Lung Transplant* 2012; 31(2): 140-9.
<http://dx.doi.org/10.1016/j.healun.2011.11.003>
- [13] Wang Y, Simon M, Bonde P, Harris BU, Teuteberg JJ, Kormos RL, *et al.* Prognosis of right ventricular failure in patients with left ventricular assist device based on decision tree with smote. *IEEE transactions on information technology in Medicine and Biology Society* 2012; 16(3): 383-90.
<http://dx.doi.org/10.1109/TITB.2012.2187458>
- [14] Wang Y, Rose CP, Ferreira A, McNamara DM, Kormos RL, Antaki JF. A classification approach for risk prognosis of patients on mechanical ventricular assistance. *Proceedings of the International Conference on Machine Learning and Applications International Conference on Machine Learning and Applications* 2010: 293-8.
- [15] Pearl J. *Probabilistic reasoning in intelligent systems: Networks of plausible inference*: Morgan Kaufmann; 1988.
- [16] John GH, Langley P, editors. *Estimating continuous distributions in bayesian classifiers*. *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*; 1995: Morgan Kaufmann Publishers Inc.
- [17] Friedman N, Geiger D, Goldszmidt M. Bayesian network classifiers. *Machine Learning* 1997; 29(2-3): 131-63.
<http://dx.doi.org/10.1023/A:1007465528199>
- [18] Chow C, Liu C. Approximating discrete probability distributions with dependence trees. *Information Theory, IEEE Transactions on* 1968; 14(3): 462-7.
<http://dx.doi.org/10.1109/TIT.1968.1054142>
- [19] Heckerman D. *A tutorial on learning with bayesian networks*: Springer; 1998.
- [20] Fisher RA. *Statistical methods for research workers* 1934.
- [21] Kaplan EL, Meier P. Nonparametric-estimation from incomplete observations. *J Am Stat Assoc* 1958; 53(282): 457-81.
<http://dx.doi.org/10.1080/01621459.1958.10501452>
- [22] Ashby D. Bayesian statistics in medicine: A 25 year review. *Statistics in medicine* 2006; 25(21): 3589-631.
<http://dx.doi.org/10.1002/sim.2672>
- [23] Hoot N, Aronsky D, editors. *Using bayesian networks to predict survival of liver transplant patients*. *AMIA annual symposium proceedings*; 2005: American Medical Informatics Association.
- [24] Lucas P, editor. *Bayesian networks in medicine: A model-based approach to medical decision making*. *Proceedings of the EUNITE workshop on Intelligent Systems in patient Care*; 2001.
- [25] Eom J-H, Kim S-C, Zhang B-T. Aptacds-e: A classifier ensemble-based clinical decision support system for cardiovascular disease level prediction. *Expert Systems with Applications* 2008; 34(4): 2465-79.
<http://dx.doi.org/10.1016/j.eswa.2007.04.015>
- [26] Sadeghi S, Barzi A, Sadeghi N, King B. A bayesian model for triage decision support. *International Journal of Medical Informatics* 2006; 75(5): 403-11.
<http://dx.doi.org/10.1016/j.ijmedinf.2005.07.028>
- [27] Gevaert O, De Smet F, Timmerman D, Moreau Y, De Moor B. Predicting the prognosis of breast cancer by integrating clinical and microarray data with bayesian networks. *Bioinformatics* 2006; 22(14): e184-e90.
- [28] Charitos T, van der Gaag LC, Visscher S, Schurink KA, Lucas PJ. A dynamic bayesian network for diagnosing ventilator-associated pneumonia in icu patients. *Expert Systems with Applications* 2009; 36(2): 1249-58.
<http://dx.doi.org/10.1016/j.eswa.2007.11.065>
- [29] Holman WL, Kormos RL, Naftel DC, Miller MA, Pagani FD, Blume E, *et al.* Predictors of death and transplant in patients with a mechanical circulatory support device: A multi-institutional study. *The Journal of Heart and Lung Transplantation* 2009; 28(1): 44-50.
<http://dx.doi.org/10.1016/j.healun.2008.10.011>

Received on 21-07-2014

Accepted on 30-10-2014

Published on 06-11-2014

<http://dx.doi.org/10.6000/1929-6029.2014.03.04.11>© 2014 Loghmanpour *et al.*; Licensee Lifescience Global.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.