Abstract—Right ventricular failure following implantation of a left ventricular assist device (LVAD) is a significant complication, which increases morbidity and mortality. Consequently, researchers have sought predictors that may identify patients at risk. However they have lacked sensitivity and/or specificity. This study investigated the use of a decision tree technology to explore the preoperative data space for combinatorial relationships that may be more accurate and precise. We retrospectively analyzed the records of 183 patients with initial LVAD implantation at the Artificial Heart Program of the University of Pittsburgh Medical Center between May 1996 and Oct. 2009. Among those patients, 27 later required a right ventricular assist device (RVAD+) and 156 remained on LVAD (RVAD-) until the time of transplantation or death. A synthetic minority oversampling technique (SMOTE) was applied to the RVAD+ group to compensate for the disparity of sample size. Twenty-one resampling levels were evaluated, with decision tree model built for each. Among these models, the top 6 predictors of the need for an RVAD were transpulmonary gradient (TPG), age, international normalized ratio (INR), heart rate (HR), aspartate aminotransferase (AST), prothrombin time and RV systolic pressure. TPG was identified to be the most predictive variable in 15 out of 21 models, and constituted the first splitting node with 7 mmHg as the breakpoint. The sensitivity of the models was found to improve monotonically, although asymptotically with oversampling; while the specificity was diminished to a lesser degree. The model built upon 5X synthetic RVAD+ oversampling, found to provide the best compromise between sensitivity and specificity, included TPG (layer 1), age (layer 2), right atrial pressure (layer 3), HR (layer 4, 7), INR (layer 4, 9), alanine aminotransferase (layer 5), white blood cell count (layer 5, 6, 7), the number of inotrope agents (layer 6), creatinine (layer 8), AST (layer 9, 10) and cardiac output (layer 9). It exhibited 85% sensitivity, 83% specificity and 0.87 area under the ROC curve, which was found to be greatly improved compared to previous published studies.

Index Terms—right ventricular assist device, heart failure, modeling (decision tree), SMOTE

I. INTRODUCTION

Heart failure affects approximately 5.8 million people in the United States each year, the majority of which suffer from left ventricular systolic dysfunction [1]. Mechanical circulatory support, particularly with left ventricular assist devices (LVADs) is an established option for treatment of patients with end-stage heart failure. Multi-center randomized clinical trials (REMATCH [2], INTRiEPiD [3] and HeartMate II DT [4]) found that long-term support by ventricular assist device (VAD) affords superior survival when compared with optimal medical therapy alone. However, right ventricular (RV) failure is a major cause of morbidity and mortality in patients after implantation of LVAD, with a reported incidence of 10% to 30% [5-9]. Severe RV failure results in renal and hepatic dysfunction and under-filling of the LVAD [10]. Peri-operative RV failure also adversely affects outcomes of patients who are ultimately bridged to transplant [7, 11, 12]. In cases where RV failure does not respond to medical therapy, a right ventricular assist device (RVAD) may be the only recourse. Approximately one third of those diagnosed with RV failure receive an RVAD within two days of placement of the LVAD [13]. Therefore it would be advantageous to identify patients who are likely to need both left and right ventricular assist devices (BiVAD) to reduce the need for a delayed RVAD insertion, associated second operation and inferior outcomes [14].

It is equally important to identify patients who are not likely to encounter RV failure so as to avoid the unnecessary complications of RVAD implantation. RVAD increases (effectively doubles) many of the risk factors associated with a single VAD such as thrombosis, infection, and mechanical failure. The surgery for BiVAD implantation is longer, with prolonged cardiopulmonary bypass and anesthesia, which leads
to increased morbidity and delayed discharge [15]. Therefore, a preoperative decision tool to predict the need of RV support would be a welcome asset, particularly for the borderline cases of RV failure.

Development of RV failure is a result of a complex interaction between pre-operative conditions, intra-operative factors and immediate post-operative hemodynamic status. This in turn makes it difficult to predict the incidence of RV failure in an individual patient [16-18]. Previous predictors of RV failure identified by uni/multivariate traditional statistical analysis have been prognostically inconsistent when evaluated in an independent sample [5, 7, 9, 18-23]. The right ventricular failure risk score (RVFRS) proposed by the University of Michigan demonstrated a high positive predictive value (80%) of RV failure in LVAD candidates (based on a threshold value of 5.5), yet the overall sensitivity was 35% [18].

This study investigated the use of data mining techniques to explore the preoperative data space for combinatorial relationships to develop an improved decision tool for predicting the need of RVAD support. These techniques have recently enjoyed growing popularity for improving statistical tools to predict future trends and discover unknown patterns. Decision tree is one such technique that has been used extensively in medicine [24-30]. It has proven to be reliable and effective, providing high classification accuracy with a simple representation of gathered knowledge [31]. Because of its tree structure, it can be easily interpreted by clinicians and therefore more likely to be used to support decision-making than, say, a “black box” type method. Our study also adopted data over-sampling and feature selection techniques to better represent the patient cohort, discover predictive clinical variables and improve the model performance.

II. MATERIALS AND METHODS

A. Patient Cohorts

This study retrospectively analyzed 183 patients enrolled in the Artificial Heart Program at the University of Pittsburgh Medical Center from May 1996 to Oct. 2009. These patients initially received an LVAD, among whom 27 later required a right ventricular assist device (RVAD+) and 156 remained on LVAD until the time of transplantation or death (RVAD-). Devices used for RV support were: Thoratec PVAD (n=15; Thoratec, Pleasanton, CA, US), Abiomed BVS 5000 (n=3; Abiomed, Danvers, MA, US), CentriMag blood pump (n=8; Thoratec, Pleasanton, CA, US), and Biomedicus RVAD (n=1; Medtronic Biomedical, Minneapolis, MN, US). These devices were extracorporeal blood pumps connected in series with native right ventricle to augment pulmonary flow. The Thoratec PVAD and Abiomed BVS 5000 are so-called “first generation” devices that provide pulsatile flow by intermittent pneumonic pressurization of a blood sac [32]. The CentriMag and Biomedicus pump are “second generation” devices which provide continuous flow from a rotating impeller [33].

B. Pre-operative Variable Selection

A total of 39 pre-operative variables were selected based on a survey of the literature and their availability, categorized in four groups: 1) patient demographics (n=6), 2) laboratory tests (n=13), 3) hemodynamics (n=14), and 4) medications (n=6). (See Fig. 1). Data from the University of Pittsburgh Medical Center Cardiothoracic Transplantation (CT) Program’s Transplant Patient Management System (TPMS) were reviewed retrospectively. The TPMS is a password-protected, HIPAA compliant web-based prospective data collection for all patients who receive mechanical circulatory support and is approved by the University of Pittsburgh Institutional Review Board. Data were extracted from preoperative day 14 to 1. For variables having multiple values, the value closest to the time of surgery was used. Missing data elements were imputed using one of three common methods: mean, median, and k-nearest neighbor (km) described elsewhere [34, 35]. For each data element, the imputation method was chosen that produced the least prediction square error by 10-fold cross validation. These are indicated in Fig. 1, where the k-nearest neighbor method is designated according to the number of nearest instances (1nn, 3nn, 5nn, 7nn, 10nn, 15nn, 20nn). A rank-based scale, Chi-Square statistics, was used to optimize the set of pre-operative variables (features) so as to avoid overfitting [36].

C. Data Sampling

In the present cohort, the ratio of RVAD+ to RVAD- patients was approximately 1:6, a highly imbalanced dataset. Misclassifying the RVAD+ minority patients to RVAD-majority patients was felt to lead to greater risk than the reverse clinical scenario (e.g. misclassifying an RVAD- patient as RVAD+). To emphasize the correct classification of RVAD+ patients, we employed Synthetic Minority Over-sampling Technique (SMOTE), which maximizes the use of available data [37, 38]. SMOTE was used to generate new synthetic cases for this study. The new synthetic sample variables were computed from each original RVAD+ clinical record as well as the k nearest neighbors in Euclidian distance. Since only 27 RVAD+ minority cases exist in this study, k=5 was chosen to balance the possibility of random selection against similarity. Continuous variable values were created by taking the difference between original RVAD+ sample variable and one of its nearest neighbors and then multiplying that difference by a random number between 0 and 1. The resulting number was added to the corresponding original sample variable to attain the new feature of the synthetic sample. Binary variable values were selected from one of the nearest neighbors. As a result, the synthetic cases had the attributes with similar values to the existing RVAD+ patients, as contrasted with mere replications provided with over-sampling. The objective was to increase the representation of the RVAD+ patients in the dataset, to amplify the decision region and to improve the correct classification of RVAD+ patients, while representing the structure of original cohorts. The SMOTE algorithm was applied to create new synthetic RVAD+ samples in increments of 100% (1X). To verify that patients, while representing the structure of SMOTE
did not change the original data structure, we applied principal component analysis (PCA) to compare the variance of the original dataset and that of synthetic datasets through the comparison of the Eigenvalues [37].

D. Decision Tree Analysis

We employed a well-known decision tree algorithm C4.5, implemented in an open-source software library (WEKA, J48, University of Waikato, New Zealand) [36]. This algorithm employs a criterion of highest information gain (IG) to select features. It is a top-down algorithm, seeking at each level an attribute to split the cohort so that the RVAD+ and RVAD-patients are best separated; and then recursively applies this criterion in the sub-problem domain to build a complete tree. The definition of IG used here was based on Claude Shannon’s work on information theory:

\[
IG(X) = H(Y) - H(Y|X) 
\]

\[
H(Y) = -\sum_{i=1}^{2} P(y_i) \log_2 P(y_i) 
\]

\[
H(Y|X) = -\sum_{j=1}^{2} \sum_{i=1}^{2} P(x_j|y_i) P(y_i) \log_2 P(y_i|X) 
\]

where \(H(Y)\) is the entropy of the decision variable \(y_i\); \(H(Y|X)\) is the entropy of decision variable \(Y\) given \(X\); \(H(Y|X)\) is the conditional probability of \(y_i\) for a given value of feature \(x_j\). Information gain \(IG(X)\) is therefore the difference between the entropy before and after observing a given feature. By recursively applying the highest IG criterion to select predictive features, C4.5 produces a tree of classificatory “branches” terminating in “leaves” with respect to a previously chosen target classification [39]. The Sub-tree raising operation was employed for post pruning to minimize the error rates estimated using 0.25 Confidence Factor [36]. In this study, we also varied the termination criterion, namely the minimum number of patients in each leaf \((m=1, 2, 3, 4, 5, \ldots, 20)\) to build a generalized tree and avoid overfitting. A sensitivity analysis of the effect of Confidence Factor \((0.1 - 0.5)\) upon termination criterion did not reveal a great dependency on the final results.

E. Performance Measures

To evaluate the performance of the models, we used a 10-fold cross-validation technique. It divided the data into 10 mutually

Fig. 1. Initial pool of pre-operative data elements (prior to feature selection) comprised of four categories: demographics, hemodynamics, laboratory values and medications. The associated interpolation methods are indicated as k nearest neighbor (knn), median, and mean, along with the frequency of missing elements. (Diagnosis: ischemic/non-ischemia; DeviceGeneration: pulsatile/continuous flow; HGB: hemoglobin; WBC: white blood cell count; HCT: hematocrit; ALT: alanine aminotransferase; TBIL: total bilirubin; INR: international normalized ratio; PTT: prothrombin time; BUN: blood urea nitrogen; AST: aspartate aminotransferase; PLT: platelet count; PA_Sys/Dia/Mean: systolic/diastolic/mean pulmonary arterial pressure; RV_Sys/Dia/Mean: systolic/diastolic right ventricular pressure; RAP: right atrial pressure; PCWP: pulmonary capillary wedge pressure; TPG: transpulmonary gradient; PVR: pulmonary vascular resistance; PA sat: pulmonary arterial oxygen saturation; CI: cardiac index; HR: heart rate; IABP: intra-aortic balloon pump; ACE: angiotensin-converting enzyme; A-II: angiotensin II receptor.)
exclusive subsets and then combined nine of those as a training set and then evaluating the 10th left-out subset. Thus, the decision tree model is identified on ten different, but overlapping training sets, and evaluated on 10 completely unique testing sets. The associated performance measures were as follows:

1) True positive (TP): the number of patients in which the algorithm agrees with the historical clinical decision to implant an RVAD, RVAD+/+

2) True negative (TN): the number of patients that both algorithm and expert agree to forgo an RVAD, RVAD-/-

3) False positive (FP): the number of patients that the model predicts RVAD implantation while historical decision was to forgo RVAD, RVAD-/+

4) False negative (FN): the number of patients that the model prediction disagrees with the historical decision of RVAD implantation, RVAD+/+

5) Sensitivity: the TP rate; $= TP/(TP+FN)$
6) Specificity: the TN rate; $= TN/(TN+FP)$
7) Accuracy: the percentage of the correctly classified records $= (TP+TN)/(TP+FN+TN+FP)$

8) Area under the curve (AUC): the area under the receiver operating characteristic (ROC) curve [40], a commonly used metric for the classifier performance.

9) Kappa statistics: the degree of agreement between the classifier’s predictions and reality by considering the proportion of predictions that might occur by chance [41].

F. Calculation of Right Ventricular Failure Risk Score

For each patient, we also calculated a previously reported RVFRS, which is a weighted score introduced by Matthew, et al. [18]. This particular risk assessment was chosen from the literature because it demonstrated competitive AUC of 0.73. It is based on multivariate regression, and is computed as a weighted sum of 4 pre-operative variables: vasopressor requirement (assigned 4 points), AST $> 80$ IU/L (assigned 2 points), bilirubin $> 2.0$ mg/dL (assigned 2.5 points) and creatinine $> 2.3$ mg/dL (assigned 3 points.) Accordingly, the RVFRS was calculated for each patient as the sum of points awarded for the presence of each of the above 4 pre-operative variables, and in turn classified the patients into three categories: high risk (total points $>= 5.5$); medium risk (4.0 - 5.0); and low risk ($= 3.0$) based on the published breakpoints.

III. RESULTS

A. Right Ventricular Support Decision Tree (RVSDT) Model

The data mining process consisted of four stages, as summarized in Fig. 2. In the first stage, 20 oversampled datasets were generated with incremental 1X synthetic RVAD+ samples appended to the original dataset, for a total of 21 datasets. For each of these datasets, attribute selections were performed in the second stage to rank the pre-operative features according to Chi-square statistics. This lead to a ranked list of 39 attributes.

In the third stage, 20 different RVSDT models were constructed using incrementally larger termination criteria (m=1,2, … , 20). In the forth stage, the performance of 780 (39 x 20) resulting RVSDT models for each of the 21 datasets were evaluated by 10-fold cross validation with regard to sensitivity, specificity, AUC, and kappa statistics. By selecting the model with highest AUC value per dataset, twenty-one RVSDT models were analyzed.

Fig. 3 demonstrates the effects of oversampling on performance. There was a dramatic improvement in sensitivity and kappa statistics between 1X and 5X, and a moderate improvement with respect to AUC. The original dataset showed a RVAD+/+ sensitivity of 22%, an AUC of 0.678 and a kappa statistics of 0.169. By visual inspection, a point of diminishing returns was identified at the 5X level, for which sensitivity was improved to 85%, the AUC to 0.879 and the kappa to 0.679 while specificity was slightly reduced to 83%. Beyond this initial 4-fold improvement of the sensitivity and kappa, additional concatenation of synthetic RVAD+ samples to the original cohorts did not substantially benefit the sensitivity, AUC and kappa, yet negatively affected the specificity (RVAD-/+) rate to 72%.

The corresponding RVSDT built upon the dataset with 5X synthetic RVAD+ samples is shown in Fig. 4. The first split was based on transpulmonary gradient (TPG), with the threshold of 7 mm Hg. For patients with TPG below this breakpoint, this model directly assigns them to RVAD- group. Above this threshold leads to further branches in the tree, wherein the second split was age having a breakpoint of 59. The third level split for each branch was right atrial pressure (RAP), however having different thresholds depending on age: 18 mm Hg for those younger than 59 years versus at 10 mm Hg for those older than 59. This demonstrates the apparent nonlinear relationship among these pre-operative features. With the increasing depth, the RVSDT model educed more complicated patterns among the RVAD+ and RVAD- patient samples, with elevated international normalized ratio (INR) and the white blood cells count (WBC) consistently associated with greater occurrence of RVAD insertion.

B. Validity of Results
The consistency of the twenty synthetic datasets was evaluated by PCA. It was found that the Eigenvalues of the original dataset and these twenty datasets were similar resulting in an averaged distribution variation within 1%. From this, we conclude that SMOTE effectively compensated for the sparse nature of RVAD+ samples without introducing significant artifact.

Likewise, the structure of RVSDT models with the best AUC values for 21 datasets were found to contain similar high-ranked variables, the most common of which being: TPG, age (in 15 models); INR (in 13 models); heart rate (HR), aspartate aminotransferase (AST; in 12 models); prothrombin time (PTT) and right ventricular systolic pressure (RV_Sys; in 10 models.)

C. Comparison with RVFRS

For each patient in this cohort of 183 patients, we calculated the corresponding RVFRS score and separated the cohort into low (n=116), medium (n=36) and high risk (n=31) groups for the implantation of RVAD. When compared to the historical clinical decision, it was found that only 18.5% of RVAD+ patients were identified as high risk according to RVFRS and 64.1% of RVAD- patients were identified as low risk. The RVFRS score also mislabeled 59.3% of these RVAD+ patients as low risk and 16.7% of RVAD- patients as high risk. In comparison, the current RVSDT model exhibited 85% sensitivity (RVAD+/+ rate), 83% specificity (RVAD-/ rate), 0.679 kappa statistics, 0.87 AUC value as well as overall 84% accuracy. (See Fig. 3)

IV. DISCUSSION

RV failure is one of the leading complications associated with LVAD implantation. However, the complex pathophysiology of post-operative RV failure [18] and care [42] makes its prediction difficult, therefore hindering the optimal a priori device selection for an individual candidate [12, 17, 42, 43]. By utilizing modern machine learning techniques, the decision tree model presented herein exhibited superior prognostic ability compared to univariate analysis or linear traditional multivariate analysis. One explanation for the improved performance is the ability of the algorithms to capture the synergistic and non-linear relationships among pre-operative features; whereas the traditional statistical methods implicitly assume orthogonality of features. This study implies that the mutual information among pre-operative features is important for RVAD prognosis in LVAD patients, which, to some extent, reflects the complexity of the physiology and pathology. In addition to differences in accuracy, the models differ in transparency. The decision tree model for example is easily interpreted by physicians following the tree branches. By contrast, there is more “blind faith” in accepting the results of a purely mathematically derived risk analysis.

The RVSDT model presented here includes predictive pre-operative variables that are supported by previous studies of right ventricular failure and further offers intriguing new observations. For example, high TPG has been shown to be a significant predictor of right ventricular dysfunction after LVAD implantation [44] and was consistently identified by the model as the first splitting variable. However, the RVSDT revealed that for patients with TPG greater than 7 mm Hg, additional factors should be taken into consideration. Age, identified previously by Fukamachi et al. to be an important predictor of RVAD support [19], was also identified to be a discriminating variable by the RVSDT, however as a secondary splitting variable. Similar to recent findings [15, 18, 23], elevated INR and WBC also were identified as indicators of the need for RVAD. However their relative importance was found

![Performance of the evaluation metrics in RVSDT Models with highest AUC value for each dataset.](image-url)
to be interdependent. For example, for those with TPG > 7 mm Hg, age < 59, and RAP > 18 mm Hg, an INR > 2.6 is indicative of the need for RVAD implantation. However, for patients with RAP < 18 mm Hg, the RVSDT indicates that the concomitant presence resting HR > 110 bpm and WBC > 15.6 \times 10^9/L further predicts the need for RVAD support. A final example is the relative importance of the use of inotropes. Matthew et al. reported that this variable decreases the risk of RV failure overall. However the Decision Tree found that the subset of these patients with WBC < 7.9 \times 10^9/L did not require RVAD support. Neither did those patients requiring RVAD having the combination of factors WBC > 7.9 \times 10^9/L, age > 41, CO > 4.4 L/min and AST > 80 IU/L.

A limitation of the current study is the sparse nature of the patient data used for training and evaluation as well as the imbalanced distribution of patients between two groups. The application of SMOTE data sampling technique greatly compensated for this deficit. SMOTE operates in the “feature space” rather than “data space.” The synthetic cases will both increase the data space and amplify the characteristics of the minority class (RVAD+ decision boundary) without introducing significant artifacts. This was illustrated by the similar PCA among original/augmented datasets. The efficacy of SMOTE to improve performance of classifiers has also been demonstrated in several previous studies in both the medical domain and other areas [38, 45-47]. However, the single-center experience would likely impact the generality of this model. The model would clearly benefit from a larger and preferably multi-center data set to eliminate the single-center bias and further corroborate its statistical confidence.

Since the classification for this study was minimizing the “error” between the model prediction and historical clinical decision, the model as it stands serves essentially to replicate the expert judgment. Therefore, its clinical utility in its present form would be twofold: (1) to transfer expertise from experienced, successful medical centers to those less experienced; and (2) to prompt the clinician to reconsider his/her initial evaluation prior to LVAD insertion with respect the prognosis of RV failure.

There are many circumstances in which a bi-ventricular assistance is plainly obvious, such as the presentation of ventricular tachycardia, cardiogenic shock with multi-organ failure or confounding issues such as a ventricular septal defect. Likewise there are circumstances in which an RVAD is clearly unnecessary and does not require a computer to make a decision. Therefore it may be argued that the utility of a decision support model is limited to discriminating the marginal cases: to determine who could get by without an RVAD despite initial marginal RV function, who will recover with a temporary RVAD and who will require chronic support.

Cases wherein the model did not correctly predict the historical clinical decision may be explained either by inconsistencies in the clinical decision, or by deficits of the model. Clinical judgment is far more comprehensive than pure mathematics [48]. Accordingly, there may exist additional, perhaps subconscious, factors [48] that were overlooked by the model (i.e. excluded from the initial set of candidate features.)
In the absence of a random clinical trial, it is currently impossible to definitively corroborate whether the historical decisions were necessarily “correct.” For example it cannot be determined if a particular RVAD+ patients would have survived without an RVAD; or conversely if an RVAD- could have been salvaged with an RVAD. This is complicated by the definition of “optimal outcome.” Although a mathematical model may be predictive of mortality and morbidity, the ultimate clinical decision must consider other factors, such as long-term outcome and quality of life. If and when such data become available, it would be possible to optimize the decision model with respect to outcome, rather than replicating clinical judgment – as is being accomplished in other fields of medicine [49, 50].

In summary, the RVSDT described here has exhibited the ability to replicate expert judgment with 85% sensitivity and 83% specificity. It therefore has the potential to facilitate decision making, particularly by those with less experience, and in ambiguous circumstances where there is disagreement amongst experts. However, the assessment of its perceived utility by clinicians was beyond the scope of the present study. The retrospective nature of this study precludes a definitive assessment of the accuracy of either expert or model decision inasmuch as it is impossible to foretell the outcome of the converse decision: whether a patient receiving an RVAD would or would not have fared equally well without an RVAD, and vice versa. Future development of this decision model will therefore benefit from additional prospective data beyond the single center used in this study. Improvements to both the quality and quantity of the data should improve its performance, and hopefully the breadth of its utility.

ACKNOWLEDGMENT

Special thanks to Michelle D. Navoney who was the honest broker for the TPMS database. Special thanks to Professor Lukasz Kurgan and Professor Marek Druzdzel who offered constructive suggestions to the methodology of this study. We are also grateful for reviewers’ valuable comments to corroborate this study.

REFERENCES


