Intelligently Characterizing Patient Hemodynamic Phenotypes for Advanced Heart Failure in the ESCAPE Trial Using Learned Multi Valued Decision Diagrams

Josephine Lamp[1], Yuxin Wu[1], Steven Lamp[2], Lu Feng[1], Sula Mazimba[3]
[1]Department of Computer Science, University of Virginia, Charlottesville, VA, USA
[2]Department of Computer Science, Colorado State University, Fort Collins, CO, USA
[3]Department of Cardiovascular Medicine, University of Virginia, Charlottesville, VA, USA

Abstract

Identifying patients at risk of advanced heart failure (HF) is critical in order to provision the correct treatment decisions and implement advanced, life-saving therapies at the correct time. For high risk heart failure patients, hemodynamic assessments can allow clinicians to better understand the patient state and their risk levels, providing a greater window of time for treatment decisions.[3] To this end, we present the development of an advanced heart failure hemodynamic risk score that takes in single point of care measures as input, and uses a diverse set of features (including invasive and composite hemodynamics and other clinical measures) to return a score of 1 to 5, indicating the probability of a specified outcome, such as mortality, rehospitalization, or readmission in 30 days. Our risk scores are learned using a machine learning methodology that takes advantage of the explainability and expressivity of Multi Valued Decision Diagrams (MVDDs), trained on the ESCAPE dataset,[4] which includes a rich and inclusive feature set of invasive hemodynamics, demographics, lab values, medications, quality metrics and other health measures. In addition to providing a concrete score metric, our solution provides and visualizes the learned patient *phenotypes*, which characterize the set of features and thresholds that were used to determine the resulting risk score. Our open source tool implementation is developed completely in Python 3, interfaced with a web server that allows live risk score prediction, allowing other researchers to quickly explore, extend, or prototype with our tool. All code is publicly available from this Github repository: https://github.com/jozieLamp/HemoPheno4HF, and the web server may be accessed here: http://hemopheno.pythonanywhere.com/. Through the use of our snapshot risk scores and explainable phenotypes, it is our intention that our solution will facilitate clinical decision making to improve the effectiveness and timing of treatment decisions, ultimately resulting in better patient outcomes. Moreover, using our MVDDs, we are able to make predictive decisions using incomplete data, which is particularly advantageous in clinical scenarios where clinicians must make quick decisions, even when they are not presented with all patient information.

Team Expertise

Our team is well suited to make advancements in this area due to our combined clinical and technical expertise in advanced heart failure, hemodynamic phenotyping, and risk assessment and machine learning methodologies for clinical diseases. Mazimba has a proven history of clinical innovation in hemodynamic research and has worked with the ESCAPE dataset extensively, including publishing multiple clinical research articles.[2,3,10] Additionally, Feng and Lamp have proven experience in the development of machine learning and predictive risk analysis models in other safety-critical and clinical disease areas such as Type I Diabetes.[6,8,9,14]

Introduction

Identifying high risk advanced HF patients early on in the care continuum is critical in order to implement advanced, life-saving therapies such as mechanical support, device implantation or transplant allocation.[3] Determining patient risk states (i.e. identifying high vs low risk patients) in the context of outcomes including rehospitalization and mortality is particularly useful, especially since additional rehospitalizations are associated with increased mortality rates.[1] For high risk HF patients, hemodynamic assessments can allow clinicians to better understand disease trajectory, thereby facilitating stratification of patients for timely treatment decisions.[3] Established HF risk scores such as the Seattle Heart Failure Risk score do not incorporate invasive hemodynamics for acutely decompensated advanced HF patients.[7]

To this end, we present the development of an advanced HF hemodynamic risk score that takes in single point of care measures as input, and uses an inclusive and diverse set of features (including invasive/composite hemodynamics and diagnostic data) to return a score of 1 to 5, indicating the probability of a specified outcome, such as mortality, rehospitalization, or readmission in 30 days. For example, a risk score of 5 (high) may indicate the patient has a $> 40\%$ probability of the specified outcome, whereas a score of 1 (low) may indicate a $< 10\%$ probability of the outcome. Our risk scores are learned using a machine learning methodology, trained on a rich feature set from the ESCAPE dataset.[4] In addition to providing a concrete score metric, our solution provides the learned patient *phenotypes*, which characterize the set of features and thresholds that were used to determine the resulting risk score. For example, a returned risk score of 5 may be characterized by the phenotype *(Arrhythmia = True or MI = True) and Age >= 70 and (PCWP > 29.5 or MPAP > 77) and Creatinine <= 1.55.* In this case, the use of the logical "or" operator indicates that the risk score can be satisfied by either feature (i.e., the patient may have Arrhythmia or MI, and PCWP or MPAP to still be in this risk category).

Our machine learning methodology takes advantage of the explainability and expressivity of Multi Valued Decision Diagrams (MVDDs), which are directed, acyclic graph structures that represent logical functions. In the MVDDs, nodes represent features, edges represent logical operators ("and", "or") with parameter threshold values, and leaf nodes represent the final score classification. As such, the "path" through the graph may be returned to provide a descriptive patient phenotype, such as the one described previously. We argue that MVDDs are well suited to classification tasks and the representation of heart failure phenotypes over other black box models because of their flexibility and interpretability in characterizing the relationships between features used to return a specific score. Moreover, MVDDs are resilient to missing data, due to their use of logical operators. For example, in the above phenotype, PCWP or MPAP may be used for rule prediction, and as such, when a feature is missing from the provided data, alternative features may be used in their place to still allow for score prediction. Our solution is completely open source and is developed entirely in Python 3 using only publicly available code packages. Our tool also includes an interface with a front-end web server that provides live risk score calculation, for ease of use. In summary, our solution presents the following contributions: (1) Development of a novel hemodynamic advanced HF risk score that uses a diverse feature set including invasive hemodynamics, calculated composite hemodynamics (e.g., mean arterial pressure (MAP),

mean pulmonary artery pressure (MPAP), cardiac power index (CPI), pulse pressure (PP), percent plasma volume (PV)), demographics, lab values, medications, quality metrics and other health metrics; (2) A MVDD machine learning methodology that returns and visualizes the diverse patient phenotypes used to categorize patient risk scores; (3) An easy to use, extensible, open source tool developed in Python 3 that other researchers may quickly explore or prototype with.

Our solution can facilitate clinician decision making, not only through the use of a risk stratification score to gain a snapshot understanding of the patient state and their likelihood of mortality and rehospitalization outcomes, but also through the explainability of the produced phenotypes. Such a methodology will improve the impact and effect of clinician treatment decisions and maximize and ensure the correct timing of the given treatments, ultimately resulting in better patient outcomes including reduced mortality and rehospitalization rates.

Background

Decision Trees. Decision trees are a well-known classification methodology used in supervised machine learning tasks to develop models to predict specific target classes.[5] The decision trees are generated by recursively splitting the input data into successively smaller and increasingly homogenized subsets using rules derived from the feature set. A so-called splitting criterion is used to determine how pure the resulting nodes are after a split and when to stop the splitting. Decision trees are advantageous because they are easy to train and interpretable. However, they are prone to overfitting issues and do not scale to large datasets. In our project, we first learn an initial predictive tree model, before transforming it into a MVDD.

MVDD. Decision Diagrams are discrete structures which represent boolean functions in directed, acyclic graphs, and are advantageous due to their ability to efficiently store and manipulate large sets of logical data.[13] In particular, Multi Valued Decision Diagrams (MVDDs) extend the basic syntax of decision diagrams to express more varieties of logical functions and allow parameter thresholds on edges. For our purposes, MVDDs are well suited to classification tasks and the representation of HF phenotypes because they allow increased flexibility in characterizing feature relationships and are highly interpretable. Unlike decision trees, MVDDs can still predict outcomes from missing data because multiple substitutable features may contribute to the same prediction score. For example, a high risk score of 5 may be characterized by the following phenotype: *(PCWP > 18.5 or HR > 25.5) and MPAP <= 35.17*. In this case, the feature PCWP *or* HR may be used for rule prediction. In addition, MVDDs use efficient learning and manipulation techniques using parallel processing to scale to large datasets.

Methodology

Data. We used the ESCAPE trial[4] which contains 433 patients (mean age 56.1 +/- 13.9 years, 25.9% female), of which 65.1% experienced a negative outcome of rehospitalization, mortality or both, during six months of follow-up. Though on the smaller side, the ESCAPE dataset contains a rich feature set that includes demographics (age, race, sex, prior medical conditions), diagnostics (HR, BP, ejection fraction), quality metrics (NYHA, MLHFS), lab values and medications. Of the total, invasive hemodynamics including right atrial pressure

(RAP), pulmonary artery pressure (PAP), pulmonary capillary wedge pressure (PCWP), and cardiac index (CI) were recorded for 209 patients at both baseline and prior to the removal of a heart catheter.

Data Preprocessing. We developed two subsets of the overall ESCAPE dataset, in order to explore phenotypes for invasive hemodynamics as well as phenotypes between *composite* hemodynamics and other factors, henceforth referred to as the Hemo and All Data datasets, respectively. For each dataset, we first combined all data variables into a single file and preprocessed the data, including removing outliers and normalizing the values (necessary to reduce bias in machine learning training). The patient values recorded at baseline and discharge were treated as two separate data records, to increase the total number of training records. We also calculated additional composite hemodynamic factors from other recorded features (e.g. HR, BP, etc.) including mean arterial pressure (MAP), mean pulmonary artery pressure (MPAP), cardiac power index (CPI), pulse pressure (PP), percent plasma volume (PV), ratio of pulse pressure to heart rate, systemic vascular resistance (SVR) and mixed venous O_2. These metrics were specifically chosen because of their predictive quality found in other studies.[2,3,10]

Supervised Label Generation. Next, we generated the risk score label sets for our models for each of the four outcomes: Mortality, Rehospitalization, Readmission in 30 days or All Outcomes. We devised a risk score from 1 to 5 which characterized the probability with which the outcome would occur, with 1 indicating low risk and 5 indicating high risk, as shown in Figure 1. For each dataset and outcome, we combined the features using PCA, clustered the patients into 5 groups corresponding to each of the risk probabilities using K-Means, and then assigned score labels to each data record based on the cluster it belonged to.

Risk Score	Outcome Chance	Risk Category
1	< 10%	Low
2	10 - 20%	Low-Intermediate
3	20 - 30%	Intermediate
4	30 - 40%	Intermediate-High
5	> 40%	High

Figure 1: Risk Score Meaning

Learning Multi Valued Decision Diagrams. To learn the MVDDs, we first took advantage of the efficient training capabilities of decision trees and learned a multi-class decision tree for our risk labels, using the CART algorithm and separating the nodes using the splitting criterion of gini index or entropy. We then used the base structure of the decision tree to learn a MVDD. Explicitly, we performed an exhaustive search in which we replaced the boolean edges (True/False) from the decision tree with logical operators ("and", "or") and calculated the accuracy of the diagram. From there, we selected the MVDDs with the best accuracies. It was important to transform the decision trees into MVDDs so that missing data can still be learned on, and also to increase the expressivity with which our trees can describe patient risk phenotypes. HF patients can present varied combinations of many different features and the strict structure of decision trees may not be conducive to representation of such diverse risk phenotypes. For example, one high risk HF patient may present out of range values for a combination of hemodynamics including RAP, PCWP and CI, whereas another may present out of range values for a different combination of features such as systemic hypotension/hypertension and SVR. A decision tree may learn a strict structure that only encompasses nodes with RAP, PCWP, and CI, thereby missing or misclassifying patients

who may present other symptoms for the same risk category. However, with an MVDD, we can learn rules that better encompass the wide variety of patient variability in risk types. For example, an MVDD may learn a rule that uses (*RAP and PCWP and CI) or SVR*.

In the generation of our MVDDs, we performed feature selection techniques including backward elimination, recursive, and embedded feature selection, and also used clinical expertise to select the most informative features. We generated MVDDs for the Hemo dataset, and for two different feature subsets of the All Data datasets (v1 and v2), for each of the four outcome risk stratifications (Mortality, Rehospitalization, Readmission in 30 days, All Outcomes). The Hemo, All Data v1 and All Data v2 datasets contain 418, 866 and 866 data records and 27, 122, and 44 features, respectively. In addition, in order to maximize the training capabilities of our small dataset, we used 5-fold cross validation, in which 80% of the data in the split was used for training and the other 20% was held out for validation purposes.

Results

MVDD Generation. From our supervised risk label generation, we were able to effectively stratify our patient cohorts into the five label classes. For sake of space, we only report results for the All Outcomes label, but all experiments and additional graphs are available in our open source code repository. An example graph of the K-Means clusters for the score provision in the Hemo dataset is shown in Figure 2a. We were successfully able to learn MVDDs that clearly delineate our risk stratification groups from 1 (low) to 5 (high). An example MVDD for the Hemo dataset is shown in Figure 3.

(a) Clustering for Hemo Dataset All Outcome Label

Dataset	Accuracy	Precision	Recall	F1 Score
Hemo	0.801(+/- 0.110)	**0.807(+/- 0.115)**	0.801(+/- 0.110)	0.802(+/- 0.112)
All Data v1	0.906(+/- 0.056)	**0.911(+/- 0.056)**	0.906(+/- 0.056)	0.904(+/- 0.059)
All Data v2	**0.785(+/- 0.046)**	0.777(+/- 0.078)	**0.785(+/- 0.046)**	0.774(+/- 0.062)

(b) Accuracy Metrics for the All Outcome Label

Score Name	HF Focus Area	Outcome Prediction	Number Features	Includes Hemo	Allows Missing Data	Open Source	Accuracy
Hemo	**HF, Advanced**	**Mortality, Rehospitalization & Readmission in 30 days**	27	**Yes**	**Yes**	**Yes**	0.801
All Data v1			122				**0.906**
All Data v2			44				0.785
SHFM	HF, General	1, 2, 3 year Mortality Curves	40	No	**Yes**	**Yes**	**0.971-1.0**
MAGGIC	HF, General	3 year Mortality	14	No	No	**Yes**	**0.88-0.95**
EUROMACS	HF, Early Right Sided	Mortality, Length of ICU Stay	5	**Yes**	No	No	0.67

(c) Comparison of HF Risk Score Approaches

Figure 2: Results

Accuracy. Our accuracy metrics are reported in Figure 2b. Averaged across our 5-fold cross validation, we were able to obtain accuracy rates of 0.801 for the Hemo dataset, 0.906 for the All Data v1 dataset and 0.785 for the All Data v2 dataset. We also plotted the ROC curves for each risk class in the Hemo and All Data v1 datasets in Figure 4. It is worth noting that our MVDD models perform the worst at classifying patients in intermediate risk levels (i.e. risk score of 3.) This may be because the majority of the ESCAPE trial patient records are at the extreme ends of the spectrum (high or low risk). As such, there are less training examples representing the intermediate risk class for the MVDDs to train on. We were unable to perform independent validation using another dataset because there are currently no other datasets available that include invasive hemodynamics. As a result, we are currently working to validate our results in a prospective manner among patients who receive hemodynamic assessments within the

UVA Medical Center.

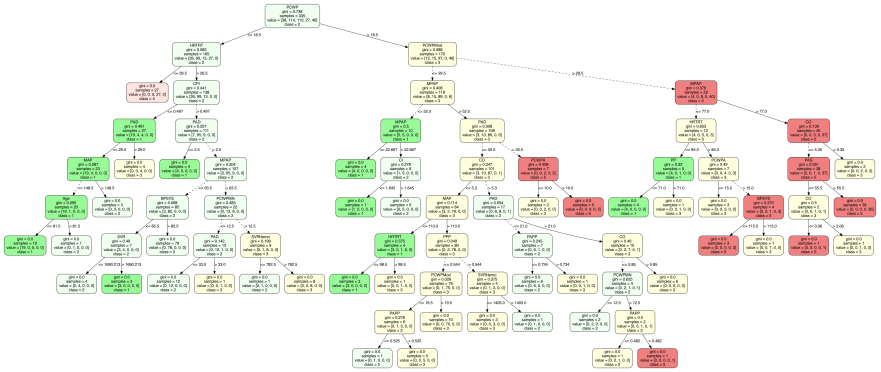
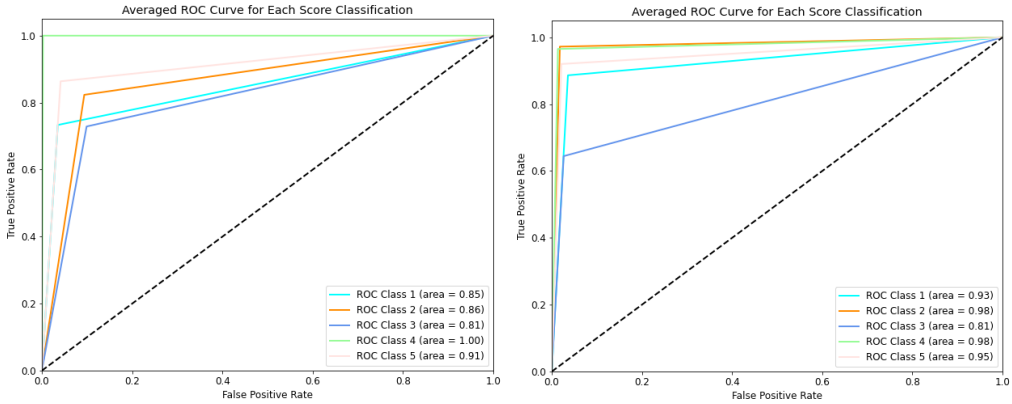


Figure 3: Example MVDD Learned on the Hemo Dataset for the All Outcome Label. Green nodes represent low risk and red nodes represent high risk.



(a) Hemo Dataset (b) All Data v1 Dataset

Figure 4: Averaged ROC Curves for the All Outcome Label Classes Across Cross Validation.

Comparison to Previous Heart Failure Risk Scores. Previous heart failure risk scores have been developed including the Seattle Heart Failure Risk score (SHFM),[7] Meta-analysis Global Group in Chronic Heart Failure score (MAGGIC),[11] and the European Registry for Patients with Mechanical Circulatory Support Right-Sided Heart Failure Risk score (EUROMACS).[12] It is our intention that our risk score would be complementary to previous risk methodologies, in which our score is used to provision risk stratification for *advanced* HF patients requiring invasive hemodynamic monitoring, and others may be used to gain an understanding of risk for more general HF patients. A comparison of each of the risk scores is shown in Figure 2c.

Tool Availability and Deployment. Our implementation is completely open source and developed in Python 3. The tool package is clearly commented and includes a jupyter notebook runner file such that others can quickly and easily explore, extend, or prototype on top of the tool. In addition, our implementation includes a deployed web server which provides a live risk score prediction for the Hemo dataset for ease of clinical use. We plan to include an implementation for the All Data dataset in the future. All code is publicly available from the Github repository: https://github.com/jozieLamp/HemoPheno4HF, and the live web server may be accessed here: http://hemopheno.pythonanywhere.com/.

Discussion, Conclusion & Future Work

Impact. The risk score can provide incremental information in categorizing HF risk among patients with invasive hemodynamics. Such risk stratification provides a granular characterization of advanced HF patients with acute decompensation states and may help identify patients at risk of death or rehospitalization. This may provide information on the timely allocation of advanced HF interventions as well as aggressive escalation of HF therapies to prevent readmissions. The generated phenotypes from the MVDDs can be used to inform clinicians and better understand relationships between features, particularly hemodynamics and other factors. Additionally, the MVDDs can be used to elucidate new phenotypes for advanced HF research that can inform future clinical research studies. For example, they can increase and better the understanding of the effect of patient physiology and hemodynamics for advanced HF symptoms. Moreover, generation of MVDDs using different combinations of feature sets (i.e. new combinations of hemodynamics and other clinical metrics) can allow for initial exploration and knowledge discovery of advanced HF patient phenotypes which may be further investigated in future clinical research studies.

Innovation. Our solution is able to identify phenotypes that combine diverse features together, in order to better stratify the resulting risk scores. Our phenotypes include a more diverse and inclusive set of feature combinations, beyond what is currently used in other clinical risk scores, such as the Seattle Heart Failure risk score. Our implementation is novel due to its use of invasive hemodynamic factors, as well as the use of additional predictive composite hemodynamic metrics that can be used even for patients who don't have invasive hemodynamic assessments, (e.g. MPAP, CPI, pulmonary artery wedge pressure, etc.). The phenotypes developed from the MVDDs combine the hemodynamic features in new ways and with other features, including demographics, medications, labs and other clinical indicators. In addition, with our MVDDs we are able to make predictive decisions of the risk score using incomplete data, which is particularly advantageous in clinical scenarios where clinicians must make quick decisions, even when they are not presented with all of the patient information (i.e. patient features).

Conclusion & Future Work. Our solution develops a novel hemodynamic advanced HF risk score that uses a diverse feature set including invasive hemodynamics, calculated composite hemodynamics, demographics, lab values, medications, quality metrics and other health metrics, using a MVDD machine learning methodology that returns and visualizes the diverse patient phenotypes used to categorize patient risk scores. We also provide an extensible open source tool implementation that other researchers may quickly explore, extend, or prototype with. Ultimately, our solution can facilitate clinician decision making, not only through the use of a quick risk stratification score to gain a snapshot understanding of the patient state, but also through the explainability of the produced phenotypes that are used in determination of the patient risk stratification. In future work, we plan to complete the prospective validation of our tool and finish the web server deployment. We would also like to investigate the use of invasive hemodynamics recorded temporally across longer time periods in our risk phenotypes especially among patients with remote implantable wireless pulmonary sensors that transmit hemodynamic data for patient monitoring.

References

[1] Ali Ahmed, Wilbert S Aronow, and Jerome L Fleg. Higher new york heart association classes and increased mortality and hospitalization in patients with heart failure and preserved left ventricular function. *American heart journal*, 151(2):444–450, 2006.

[2] Kenneth C Bilchick, Nathaniel Chishinga, Alex M Parker, David X Zhuo, Mitchell H Rosner, LaVone A Smith, Hunter Mwansa, Jacob N Blackwell, Peter A McCullough, and Sula Mazimba. Plasma volume and renal function predict six-month survival after hospitalization for acute decompensated heart failure. *Cardiorenal Medicine*, 8(1):61–70, 2018.

[3] Kenneth C Bilchick, Eliany Mejia-Lopez, Peter McCullough, Khadijah Breathett, Jamie L Kennedy, Jose Tallaj, James Bergin, Salpy Pamboukian, Mohammad Abuannadi, and Sula Mazimba. Clinical impact of changes in hemodynamic indices of contractile function during treatment of acute decompensated heart failure. *Journal of Cardiac Failure*, 24(1):43–50, 2018.

[4] Cynthia Binanay, Robert M Califf, Vic Hasselblad, CM O'connor, MR Shah, G Sopko, LW Stevenson, GS Francis, CV Leier, LW Miller, et al. Evaluation study of congestive heart failure and pulmonary artery catheterization effectiveness: the escape trial. *Jama*, 294(13):1625–1633, 2005.

[5] Leo Breiman, Jerome Friedman, Charles J Stone, and Richard A Olshen. *Classification and regression trees*. CRC press, 1984.

[6] Sanjian Chen, Lu Feng, Michael R Rickels, Amy Peleckis, Oleg Sokolsky, and Insup Lee. A data-driven behavior modeling and analysis framework for diabetic patients on insulin pumps. In *2015 International Conference on Healthcare Informatics*, pages 213–222. IEEE, 2015.

[7] Andreas P Kalogeropoulos, Vasiliki V Georgiopoulou, Grigorios Giamouzis, Andrew L Smith, Syed A Agha, Sana Waheed, Sonjoy Laskar, John Puskas, Sandra Dunbar, David Vega, et al. Utility of the seattle heart failure model in patients with advanced heart failure. *Journal of the American College of Cardiology*, 53(4):334–342, 2009.

[8] Josephine Lamp, Carlos E Rubio-Medrano, Ziming Zhao, and Gail-Joon Ahn. Exsol: Collaboratively assessing cybersecurity risks for protecting energy delivery systems. In *2019 7th Workshop on Modeling and Simulation of Cyber-Physical Energy Systems (MSCPES)*, pages 1–6. IEEE, 2019.

[9] Josephine Lamp, Simone Silvetti, Marc Breton, Laura Nenzi, and Lu Feng. A logic-based learning approach to explore diabetes patient behaviors. In *International Conference on Computational Methods in Systems Biology*, pages 188–206. Springer, 2019.

[10] Sula Mazimba, Jamie LW Kennedy, David Zhuo, James Bergin, Mohammad Abuannadi, Jose Tallaj, and Kenneth C Bilchick. Decreased pulmonary arterial proportional pulse pressure after pulmonary artery catheter optimization for advanced heart failure is associated with adverse clinical outcomes. *Journal of cardiac failure*, 22(12):954–961, 2016.

[11] Ulrik Sartipy, Ulf Dahlström, Magnus Edner, and Lars H Lund. Predicting survival in heart failure: validation of the maggic heart failure risk score in 51 043 patients from the s wedish h eart f ailure r egistry. *European journal of heart failure*, 16(2):173–179, 2014.

[12] Osama II Soliman, Sakir Akin, Rahatullah Muslem, Eric Boersma, Olivier C Manintveld, Thomas Krabatsch, Jan F Gummert, Theo MMH De By, Ad JJC Bogers, Felix Zijlstra, et al. Derivation and validation of a novel right-sided heart failure model after implantation of continuous flow left ventricular assist devices: the euromacs (european registry for patients with mechanical circulatory support) right-sided heart failure risk score. *Circulation*, 137(9):891–906, 2018.

[13] Arvind Srinivasan, Timothy Ham, Sharad Malik, and Robert K Brayton. Algorithms for discrete function manipulation. In *1990 IEEE international conference on computer-aided design*, pages 92–93. IEEE Computer Society, 1990.

[14] William Young, John Corbett, Matthew S Gerber, Stephen Patek, and Lu Feng. Damon: a data authenticity monitoring system for diabetes management. In *2018 IEEE/ACM Third International Conference on Internet-of-Things Design and Implementation (IoTDI)*, pages 25–36. IEEE, 2018.