

Identifying potential candidates for advanced heart failure therapies using an interpretable machine learning algorithm



Heming Yao, PhD,^{a,1} Jessica R. Golbus, MD, MS,^{b,1} Jonathan Gryak, PhD,^c
Francis D. Pagani, MD, PhD,^d Keith D. Aaronson, MD, MS,^{b,2} and
Kayvan Najarian, PhD^{a,e,f,g,h,2}

From the ^aDepartment of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan; ^bDepartment of Internal Medicine, Division of Cardiovascular Medicine, Michigan Medicine - University of Michigan, Ann Arbor, Michigan; ^cDepartment of Computer Science, Queens College, City University of New York, New York, New York; ^dDepartment of Cardiac Surgery, Michigan Medicine - University of Michigan, Ann Arbor, Michigan; ^eDepartment of Emergency Medicine, University of Michigan, Ann Arbor, Michigan; ^fDepartment of Computer Science and Engineering, University of Michigan, Ann Arbor, Michigan; ^gMichigan Institute for Data Science, University of Michigan, Ann Arbor, Michigan; and the ^hMax Harry Weil Institute for Critical Care Research and Innovation, University of Michigan, Ann Arbor, Michigan.

KEYWORDS:

heart transplantation;
mechanical circulatory
support;
machine learning;
heart failure advanced
therapies;
left ventricular assist
device

BACKGROUND: Systems level barriers to heart failure (HF) care limit access to HF advanced therapies (heart transplantation, left ventricular assist devices). There is a need for automated systems that can help clinicians ensure patients with HF are evaluated for HF advanced therapies at the appropriate time to optimize outcomes.

METHODS: We performed a retrospective study using the REVIVAL (Registry Evaluation of Vital Information for VADs in Ambulatory Life) and INTERMACS (Interagency Registry for Mechanically Assisted Circulatory Support) registries. We developed a novel machine learning model based on principles of tropical geometry and fuzzy logic that can accommodate clinician knowledge and provide recommendations regarding need for advanced therapies evaluations that are accessible to end-users.

RESULTS: The model was trained and validated using data from 4,694 HF patients. When initiated with clinical knowledge from HF and transplant cardiologists, the model achieved an F1 score of 43.8%, recall of 51.1%, and precision of 46.9%. The model achieved comparable performance compared with other commonly used machine learning models. Importantly, our model was 1 of only 3 models providing transparent and parsimonious clinical rules, significantly outperforming the other 2 models. Eleven clinical rules were extracted from the model which can be leveraged in clinical practice.

CONCLUSIONS: A machine learning model capable of accepting clinical knowledge and making accessible recommendations was trained to identify patients with advanced HF. While this model

¹These authors contributed equally to this work and are co-first authors.

²These authors contributed equally to this work and are co-last authors.

Abbreviations: EBM, Explainable boosting machine; FDA, Food and drug administration; HF, Heart failure; INTERMACS, Interagency Registry for Mechanically Assisted Circulatory Support; KCCQ, Kansas City Cardiomyopathy Questionnaire; LVAD, Left ventricular assist device; MCS, Mechanical circulatory support; NYHA, New York Heart Association; REVIVAL, Registry Evaluation of Vital Information for VADs in Ambulatory Life; SVM, Support Vector Machine

Reprint requests: Jessica R. Golbus, MD, MS, Department of Internal Medicine, Division of Cardiovascular Medicine, 2723 Cardiovascular Center, 1500 E. Medical Center Dr., SPC 5853, Ann Arbor, MI 48109-5853. Telephone: 734-243-5045; Fax: 734-615-3326.

E-mail address: jpgolbus@med.umich.edu

was developed for HF care, the methodology has multiple potential uses in other important clinical applications.

J Heart Lung Transplant 2022;41:1781–1789

© 2022 The Author(s). Published by Elsevier Inc. on behalf of International Society for Heart and Lung Transplantation. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Heart failure (HF) is expected to affect greater than 8-million US adults by 2030 with high disease-associated morbidity and mortality.¹⁻⁴ Heart transplantation and durable mechanical circulatory support (MCS) devices such as left ventricular assist devices (LVADs), also called HF advanced therapies, offer selected New York Heart Association (NYHA) class IV patients the best opportunity for long-term survival with improved quality of life.^{5,6} While a supply-demand mismatch does not exist for durable LVAD therapy, for heart transplantation there remains a significant mismatch due to limited donor heart availability. Furthermore, there is a risk of mortal and highly morbid complications with both heart transplantation and LVAD implantation.⁶⁻⁸ Premature heart transplantation and LVAD implantation thus exposes patients to potentially unnecessary adverse outcomes though delayed delivery places patients at risk for clinical deterioration and poor outcomes.⁹ There is thus a tension between premature and delayed delivery of HF advanced therapies.

In the current healthcare environment, the timing of advanced therapies is largely determined by HF and transplant cardiologists, who typically rely on both evidence-based practice and on clinical heuristics to determine the optimal time to deliver such therapies.¹⁰⁻¹² Heuristics, however, are error-prone and the available risk models have limited effectiveness for individual patients.¹³ Furthermore, given the high prevalence of HF in the population, the majority of patients are managed by primary care clinicians or by general cardiologists who lack training in heart transplantation and MCS. Under recognition of illness severity by such clinicians may lead to delayed referral to an HF and transplant cardiologist over which time patients may develop complications precluding advanced therapies. While a number of HF risk models are available, these have limited accuracy for individual patients.¹³⁻¹⁵ This arises, in part, due to their failure to capture multidimensional relationships, a limitation of traditional logistic regression models, and derivation from unrepresentative populations, which are often younger with mild-moderate disease. Additionally, while treatment timing is of critical importance given competing risks, none of the available scores effectively assist clinicians with treatment selection at the bedside. There is thus a critical need for algorithms that can be deployed systematically in a health system's electronic medical record that are capable of identifying patients in need of and potentially eligible for advanced HF therapies. Such a system could be used to prompt general clinicians to refer these patients to an HF and transplant cardiologist to initiate a comprehensive advanced therapies evaluation.

Herein, we describe our process for developing and validating a novel machine learning designed to identify

patients with advanced HF warranting evaluation for heart transplantation and durable LVAD implantation. Unlike other machine-learning models which are often opaque and cannot provide the rationale underlying their recommendations, we report on an interpretable model capable of (1) leveraging clinical knowledge, (2) learning new clinical rules, and (3) providing transparent and accessible recommendations which can be reviewed for validity.¹⁶ We used clinical knowledge from HF and transplant cardiologists to initialize and then train the model using clinical data from the REVIVAL (Registry Evaluation of Vital Information for VADs in Ambulatory Life)¹⁷ and INTERMACS (Interagency Registry for Mechanically Assisted Circulatory Support) registries.^{9,18-20} We demonstrate herein the model's capabilities not only for HF care but also for other similarly sensitive clinical applications in medicine.

Methods

The REVIVAL registry

The REVIVAL registry contains information on 400 patients with advanced systolic HF from 21 US medical centers and has been previously described.¹⁷ As part of the registry, patients were evaluated at up to 6 pre-specified time points over a 2-year period at which time they underwent physical examinations, medication review, functional assessments, and laboratory testing and completed general (EuroQol-5D-5L) and disease-specific (Kansas City Cardiomyopathy Questionnaire [KCCQ]) questionnaires. At each time point, investigators were asked to record whether the participant had been evaluated for heart transplantation or LVAD and the result of that evaluation. Death, heart transplantation, and durable MCS implantation were study end-points with no additional follow-up. For purposes of this analysis, study participants were labeled at each time point as (1) "positive" cases, defined as those who were felt to have advanced HF warranting heart transplantation/LVAD evaluation or (2) "negative" cases, defined as those too well for heart transplantation or LVAD. While a subset of patients labeled as "positive" cases had medical or psychosocial contraindications to advanced therapies, the focus of this model is on identifying patients whose HF is severe enough to warrant a formal advanced therapies evaluation by a HF and transplant cardiologist. As such, these patients were included amongst the positive cases.

The INTERMACS registry

INTERMACS was established as a joint effort of the National Heart, Lung, and Blood Institute, FDA, Centers for Medicare and Medicaid Services, clinicians, scientists, and industry representatives for the purpose of advancing our understanding of durable MCS device therapy.^{9,18-20} The INTERMACS registry is a North American registry of data for adults who received an FDA-approved MCS device for HF at 1 of 170 active INTERMACS

centers. The registry includes information on patient demographics; clinical data including medications, functional status, disease-specific (KCCQ) and general (EQ-5D-3L) quality of life questionnaires, and laboratory values; and clinical outcomes up to 1-year post-MCS implantation or until heart transplantation. We used INTERMACS data from June 27, 2006 through July 26, 2017. For this analysis, data were only extracted at the time of LVAD implantation and all patients were classified as “positive” (i.e., having advanced HF).

Combined dataset

In this study, we combined data from the REVIVAL and INTERMACS registries to train and validate the classification model. We excluded samples with more than 5 missing values in selected variables (discussed in *Variable selection*). The combined dataset includes 4,434 positive cases from 4,406 patients and 1,190 negative cases from 332 patients; of these 86 (1.9%) positive samples from 58 patients and 1,190 (100%) negative samples from 332 patients are from the REVIVAL registry. The training set includes all samples from the INTERMACS registry and 80% of the negative samples from the REVIVAL registry. The remaining negative samples and all positive samples from the REVIVAL registry were equally and randomly split into the validation and test sets. This data split was proposed to better evaluate the model’s generalizability. In the REVIVAL dataset, multiple samples may be from the same patient at different time points. To avoid information leakage in model development, samples from the same patient were in the same validation/test set. In this study, the data split was repeated 10 times for both model development and validation. The average number of samples and patients in the training, validation, and test sets are shown in [Table 1](#).

Variable selection

For this analysis, we chose to focus on select variables with strong associations with advanced HF as described in Tables S1 and S2, many of which are used in routine clinical care. We excluded NYHA classification score and INTERMACS profile scores for model building given the relatively subjective and deterministic nature of these assessments.

Approximate rules from advanced HF and transplant cardiologists

We assembled a panel of 5 HF and transplant cardiologists, none of whom were from Michigan Medicine. We first asked 2 HF and transplant cardiologists to generate a set of clinical rules for identifying patients with advanced HF, using only those variables in the aforementioned datasets. We collated those rules and distributed

them to 3 additional HF and transplant cardiologists, allowing them to add additional rules as appropriate. We then used them to initialize our model. For this study, given its “proof of concept” nature and relatively limited number of variables, we chose to initiate the model with simplified versions of the rules (supplemental methods).

An interpretable machine learning algorithm

We applied our end-to-end tropical geometry-based interpretable machine learning method¹⁶ to identify patients with advanced HF. The structure of our machine learning method is shown in Figure S1(a), where the inputs are values of selected variables from the target patient, and the output is a recommendation regarding advanced therapies eligibility. There are 3 core modules in the proposed method: encoding, rule, and inference modules (Figure S1(a)). The edges connecting nodes represent trainable parameters to be optimized during the training phase. All 3 modules are integral for building the interpretable machine learning algorithm.

With this method, observations of selected clinical variables are encoded into humanly understandable fuzzy concepts in the encoding layer. For example, a heart rate can be encoded into 3 concepts: “low,” “medium,” and “high,” roughly. In the proposed method, the fuzzy models are designed to match the way humans perceive variables and the approximate logical relationships amongst them. Instead of assigning an observation of 1 variable to a single concept through predefined thresholds, the method learns fuzzy membership functions during model training. In the encoding module, membership values, ranging from 0 to 1, are calculated by learned membership functions and then those values are used for the following modules. [Figure 1](#) shows an example of the learned membership functions encoding distance on a 6-minute walk test (DISTWLK).

After observations are encoded into concepts, compositional rules for patient classification are built based on combinations of concepts. These compositional rules are constructed by parameters in the rule module and inference module. The parameters are then optimized during model training. An example of a compositional rule can be written as:

IF the patient has a low level of x_1 *AND* a high level of x_2 , *THEN* the patient is “positive.”

As we demonstrated in prior work, our method can successfully extract hidden rules from a dataset.¹⁶ In addition, the aforementioned rules from HF and transplant cardiologists were used to initialize the network and facilitate model training. As rules from the trained model can capture redundant concepts, we created an algorithm capable of summarizing the most representative rules from single or multiple trained models. The algorithm is summarized in Figure S1(b) and in the supplemental methods. In this study, we repeated the training process 10 times to avoid bias from data splitting and to improve the robustness of rule extraction. The proposed rule summarization algorithm was applied to

Table 1 Average Numbers of Samples in the Training, Validation, and Test Sets From 10 Repetitions

	Training set	Validation set	Test set
Positive samples in REVIVAL	0 ± 0	43 ± 3 (29)	44 ± 3 (29)
Negative samples in REVIVAL	947 ± 8 (234)	115 ± 6 (49)	128 ± 9 (52)
Samples in INTERMACS	4348 ± 0 (4348)	0	0
Total	5295 ± 8 (4582)	158 ± 6 (78)	172 ± 10 (81)

The dataset was split patient-wise so the number of samples in each repetition may vary. Positive samples refer to patients with advanced HF while negative samples refer to patients who were deemed too well for HF advanced therapies. Data is presented as mean samples ± standard deviation (number of patients).

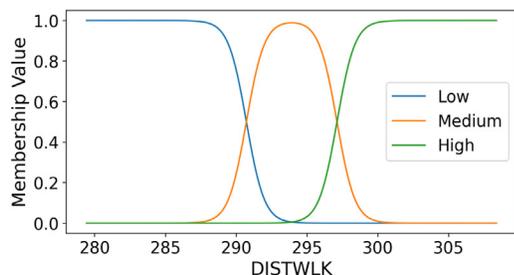


Figure 1 Optimized membership functions to encode distance on a 6-minute walk test (DISTWLK) into 3 concepts.

the 10 trained models. Additional information on the proposed interpretable machine learning model is presented in Supplementary material Section II – A. For comparison, existing machine learning algorithms were implemented and validated on the same dataset (Supplementary material Section II – B, Table S3).

Machine learning model evaluation

We calculated accuracy ((true positive + true negative)/(all positives + all negatives)); precision (true positives/(true positives + false positives)); recall (true positives/(true positives + false negatives)); and F1, the harmonic mean of precision and recall ($2 \cdot \frac{\text{recall} \cdot \text{precision}}{\text{recall} + \text{precision}}$) values. Paired t-tests were performed for bivariate comparisons across models.

Generalization gaps were calculated as the differences between metrics on the validation and test sets. A higher generalization gap indicates greater overfitting. Considering the trade-off between model interpretability and performance, we evaluated whether each trained machine learning model is transparent. In this study, a model is transparent only if the model can explain its predictions in a way understood by humans. We also evaluated each model's ability to generate a parsimonious set of rules, or a set of rules that is limited or succinct. In clinical application, a model that provides a small set of humanly understandable rules is favorable as it can be used to provide guidance and reasoning to decision-makers. Recommendations delivered in the form of interpretable and parsimonious rules are critical for high-stakes decision-making in sensitive applications such as medicine and also key to trust-building.

Results

Classification performance on the dataset

Patient characteristics are shown in Table 2. Table 3 presents our model's classification performance and that of other existing machine learning techniques with all values reflecting the average classification over 10 repetitions on the test set. The proposed method without domain knowledge achieved an F1 score of 40.4%. The integration of clinical rules further improved model performance with an F1 score of 43.8%. When compared to other existing machine learning models our proposed model achieved significantly better performance than other methods capable of providing transparent and parsimonious rules. In addition, our proposed model achieved comparable performance with other commonly used black-box machine learning models.

We also investigated the generalization gaps between the validation and test sets for our model, EBM, random forest, support vector machine (SVM), and XGBoost. Table S4 presents different models' performances on the validation set. Compared to existing machine learning models, our model achieved good performance on this dataset and had significantly smaller generalization gaps (Figure 2).

Finally, we evaluated the models based on their transparency and their ability to provide a parsimonious set of rules that can be interpreted and used by clinicians (Table 3). Our model was 1 of only 3 models providing both transparency and clinical rules, outperforming the other 2 models (Decision Tree, Fuzzy Inference Classifier) with superior F1 and recall values.

Evaluation of the extracted rules

Figure 3 shows eleven representative rules extracted from our trained model. Rules are presented in individual columns with rows corresponding to individual concepts. The first column represents the first representative rule. The color indicates the contribution of individual concepts to each compositional rule with darker colors denoting a greater contribution of that concept to the corresponding rule. For example, Rule 1 can be written as:

IF ALB is low, *AND* EF is low, *AND* pVO2 is low or medium, *AND* KCCQ1HRY is low or medium, *THEN* evaluate for heart transplantation/MCS

In practice, the model would identify patients appropriate for an advanced therapies evaluation based on those 11 rules and the degree to which patients' clinical characteristics match each rule. Specification of the individual rules, however, allows clinicians to understand which rules were used to identify patients and thereby the clinical characteristics that lead to the recommendation for an advanced therapies evaluation. It also allows for knowledge to be gained. For example, had it not been previously known, the aforementioned rule would teach clinicians that a low peak VO2 is an important indicator of advanced HF.

Discussion

Herein we present a transparent machine learning model capable of identifying patients with advanced HF warranting evaluation for LVAD and/or heart transplantation. Firstly, amongst the 3 methods capable of providing transparent and parsimonious rules, our proposed method achieved significantly better performance with the highest F1-score. Thus, our model is more likely to recommend that patients be evaluated for advanced therapies, yet provides a reasonably high probability that referred patients will subsequently be deemed appropriate for an advanced therapy. Arguably in such a scenario, recall, or sensitivity, is most important as failure to recommend an advanced therapies evaluation has significant implications for disease-associated morbidity and mortality. Additionally, as

Table 2 Demographic and Clinical Characteristics of Patients in the INTERMACS and REVIVAL Registries

	REVIVAL "Positive" (n = 86)	REVIVAL "Negative" (n = 1289)	INTERMACS "Positive" (n = 4348)
Age (years), mean	61.9 (9.7)	59.1 (11.9)	57.6 (12.6)
Sex, n (%)	29 (33.7%)	343 (26.9%)	902 (20.7%)
NYHA class, n (%)			
I	0 (0%)	63 (4.9%)	108 (2.5%)
II	4 (4.7%)	506 (39.7%)	26 (0.6%)
IIIA	47 (54.7%)	587 (46.1%)	800 (18.4%)
IIIB	20 (23.3%)	29 (2.3%)	3414 (78.5%)
IV	15 (17.4%)	86 (6.7%)	0 (0%)
Missing	0 (0.0%)	2 (0.2%)	0 (0%)
INTERMACS profile, n (%)			
1	0 (0%)	0 (0%)	269 (6.2%)
2	0 (0%)	1 (0.1%)	1416 (32.6%)
3	14 (16.3%)	3 (0.2%)	1842 (42.4%)
4	23 (26.7%)	64 (5.0%)	699 (16.1%)
5	21 (24.4%)	197 (15.5%)	85 (2.0%)
6	21 (24.4%)	403 (31.7%)	22 (0.5%)
7	7 (8.1%)	604 (47.5%)	15 (0.3%)
Heart rate (bpm), mean	76.3 (13.3)	74.9 (12.3)	87.2 (16.6)
SBP (mmHg), mean	98.8 (10.8)	110.5 (15.9)	106.9 (15.4)

Continuous variables are presented as mean (standard deviation) and categorical variables as count (percentage) for each level. NYHA, New York Heart Association; SBP, systolic blood pressure.

evidenced by its high F1 score, our model has balanced recall and precision, avoiding a scenario in which there would be excess referrals of patients who do not need advanced therapies and thereby overwhelming subspecialist resources. Second, our model achieved comparable performance with other machine learning models with higher complexity but less interpretability, significantly outperforming other transparent models.

Our model has multiple strengths compared to other machine learning models and as such would lend itself to other clinical applications in medicine. First, our model can be initiated with clinical knowledge. In a wide spectrum of

applications, there exist invaluable heuristics and domain expertise. The majority of machine learning models, however, have no mechanism by which to leverage such knowledge for model formation and training. Second, the model captures and then refines the inherent imprecision in clinical decision-making.²¹ For example, patients are often evaluated for advanced therapies when their functional capacity is severely reduced, typically defined by a low 6-minute walk distance or reduced peak VO₂ on a cardiopulmonary exercise test. Rather than requiring specific cutoffs, the model can accommodate the recommendation from clinicians that a "low" peak VO₂ may identify patients eligible

Table 3 Classification Performance of Machine Learning Models on Test Datasets

Model	Accuracy (%)	Recall (%)	Precision (%)	F1 (%)	p-value	Transparent / Parsimonious rules
Our Model	77.5 (75.7-79.3)	51.1 (44.5-57.7)	46.9 (43.8-50.0)	43.8 (40.8-46.8)	N/A	Yes, Yes
Logistic Regression	80.9 (79.3-82.5)	30.9 (26.4-35.4)	53.8 (48.0-59.6)	34.8 (30.4-39.2)	0.005	Yes, No
Naïve Bayes	22.3 (20.6-24.0)	88.6 (86.1-91.1)	18.9 (18.0-19.8)	28.5 (27.2-29.8)	<0.001	Yes, No
Decision Tree	75.1 (72.9-77.3)	44.8 (39.0-50.6)	40.2 (34.9-45.5)	36.1 (33.1-39.1)	0.042	Yes, Yes
Fuzzy Inference Classifier	80.4 (79.2-81.6)	3.3 (-0.7-7.3)	29.9 (2.3-57.5)	5.0 (-1.1-11.1)	<0.001	Yes, Yes
Explainable Boosting Machine	81.0 (79.6-82.4)	48.3 (44.0-52.6)	52.8 (49.3-56.3)	45.7 (42.6-48.8)	0.45	Yes, No
Random Forest	80.4 (78.4-82.4)	58.4 (53.3-63.5)	51.2 (47.2-55.2)	49.7 (46.2-53.2)	0.08	No, No
XGBoost	81.7 (80.2-83.2)	43.0 (38.1-47.9)	55.4 (51.1-59.7)	43.7 (39.9-47.5)	0.32	No, No
Support Vector Machine	76.1 (72.4-79.8)	43.1 (36.8-49.4)	43.8 (36.2-51.4)	37.4 (33.7-41.1)	0.20	No, No

The results are averaged over 10 repetitions. Presented as mean (95% confidence interval). p-value refers to the statistical significance for the F1 score from our proposed model compared to other commonly used methods.

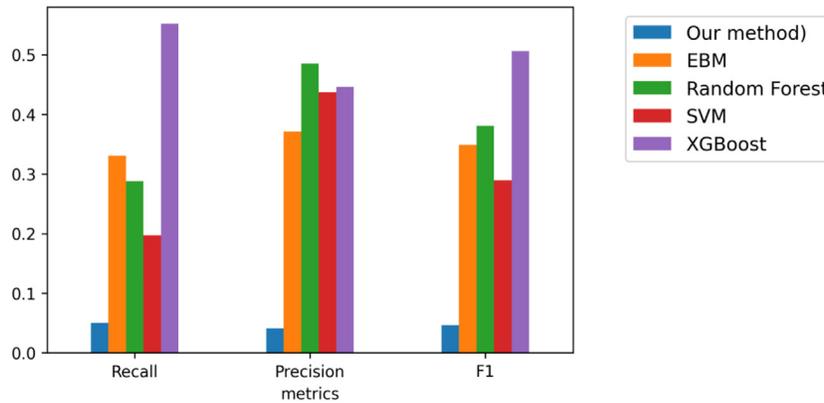


Figure 2 Generalization gaps between the validation sets and test sets.

for advanced therapies. As such, the model does not dichotomize inherently gray variables.

Finally, the model is transparent, meaning that it provides justification for its recommendations. Decision-makers in sensitive applications such as medicine, of which advanced HF is just one example, are less likely to trust recommendations in which no clear justification is provided to support a recommendation. Many commonly used machine learning models, including families of neural networks and SVMs, are amongst the “black box” models whose usages in clinical practice have been limited by a lack of transparency. While more traditional models such as regression trees provide such capabilities, they do not provide a set of rules to explain the logical relations and interactions between variables, rules that can be effectively communicated to clinicians. In addition, some models, such as random forest, list a very large number of rules, ranging from hundreds to thousands, limiting their utility for decision making. Our model overcomes these limitations by providing a clear rationale for recommendations. Thus, clinicians can inspect the rules and identify any potential fallacies in the recommendations, allowing them to update the model in near real-time. This will increase clinicians’ confidence in the model and allow for serial improvement in performance. Furthermore, by representing the rationale for recommendations, the model enables the extraction of new clinical knowledge which can subsequently be applied in practice.

In addition to the strengths of the model itself, our study has multiple strengths. We used data from the INTERMACS and REVIVAL registries with data collected predominantly as part of real-world clinical practice. As such, there is missingness and imprecision in the data similar to real-world settings. The data was also collected from patients at multiple clinical centers with the INTERMACS registry including data on patients from all North American centers implanting FDA-approved LVADs. Our dataset thus captures clinical information on a diversity of patients and reflects varying clinician practice patterns with respect to advanced therapies. Second, the clinical rules used to initiate the model were created by 5 HF and transplant cardiologists, each from a different academic medical center. As such, these rules capture a plurality of perspectives. Finally, we removed subjective variables such as NYHA and

INTERMACS classifications from the model. While this reduced model performance, we viewed these variables as subjective and deterministic as the decision to initiate an advanced therapies evaluation is determined, in part, by clinician perception of illness severity. Thus, the presented model was derived from more objective markers of illness severity.

Our study does have limitations. First, the REVIVAL dataset is relatively small with only 400 patients, only a subset of whom went on to receive advanced therapies. As such there was a relatively limited number of “positive” case examples by which to train the model. To overcome this limitation, we combined cases from the REVIVAL and the INTERMACS datasets. Differences between these datasets included those related to variable measurement as well as to illness severity, with patients in the INTERMACS registry generally having more severe HF. This also led to the majority of patients receiving an advanced therapy, which is not the case for the majority of HF patients in practice. The data split was selected, however, so as to retain the greatest number of positive samples from the REVIVAL registry in the test dataset, which more closely mirrors the real-world setting in terms of advanced therapies delivery, and to then retain only REVIVAL cases in the validation set for algorithm optimization. Second, follow-up in the REVIVAL dataset was terminated at the time of LVAD implantation or heart transplantation. In order to mirror the REVIVAL dataset, post-LVAD and heart transplantation outcomes from the INTERMACS dataset were not included in the model. As such, the model was trained to identify patients warranting evaluation for advanced therapies though does not identify patients most likely to benefit from advanced therapies given the lack of data on post-heart transplantation and LVAD outcomes. Finally, we trained and validated the model using variables already known to be associated with HF severity. While we limited the number of variables in this study, future studies can use an expanded set of variables including those not previously known to be associated with HF, increasing our ability to learn clinical relationships from the model.

In conclusion, while HF advanced therapies have the potential to improve survival and quality of life, our ability to screen a population to identify appropriate candidates and deliver optimally timed therapies is limited. Herein, we

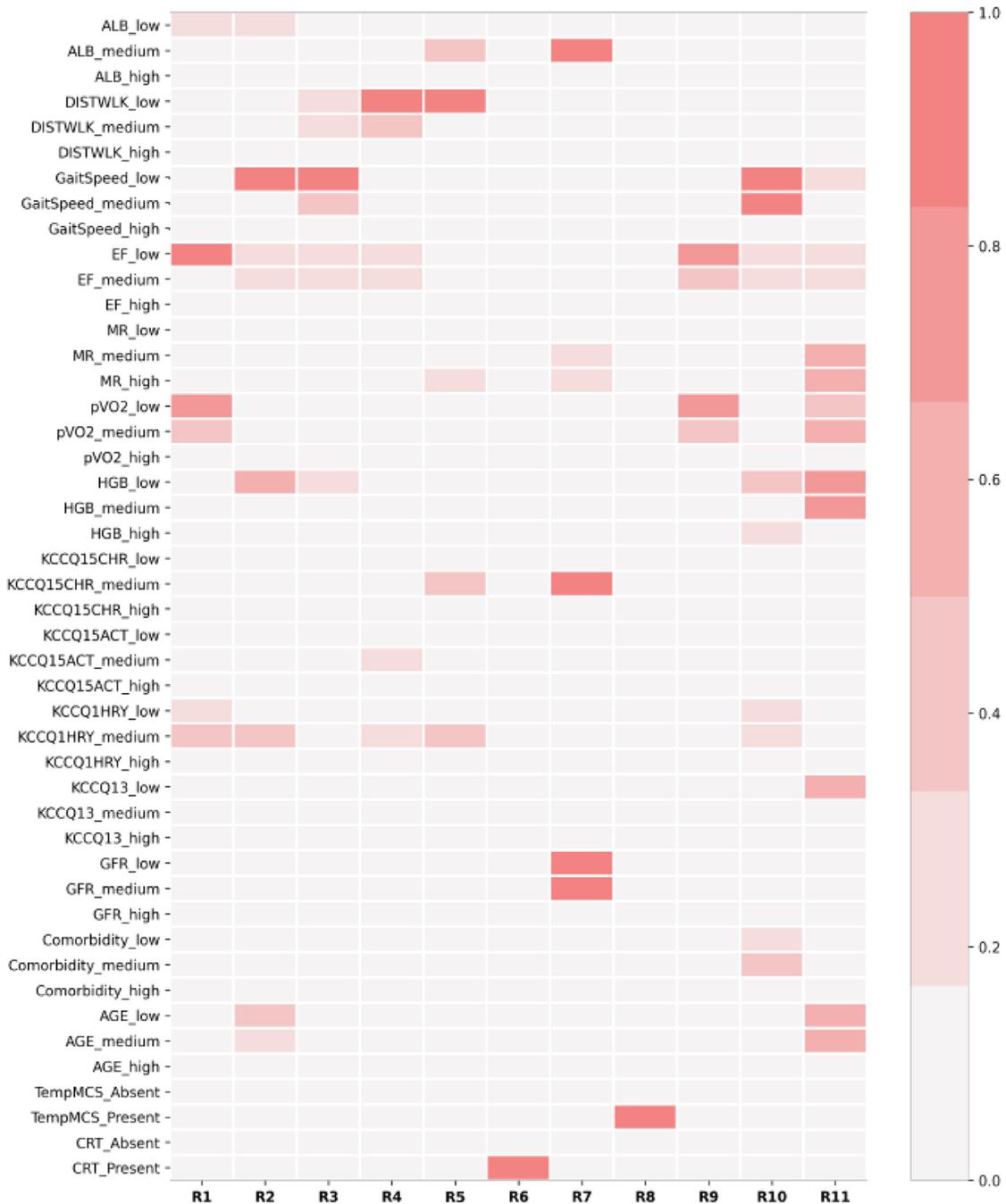


Figure 3 Rule visualization. Each column represents a set of rules with colors representing the contribution of that variable to the composite rule. When completing the KCCQ, patient answers are converted to a numerical value with lower scores indicating greater illness severity. Key: Alb = Albumin; CRT = cardiac resynchronization therapy; DISTWLK = distance walked; EF = ejection fraction; GFE = glomerular filtration rate; HGB = hemoglobin; KCCQ = Kansas City Cardiomyopathy questionnaire; MR = mitral regurgitation; TempMCS = temporary mechanical circulatory support.

present a novel machine learning model capable of identifying patients with advanced HF warranting evaluation for heart transplantation and/or LVAD. We applied the model to the REVIVAL and INTERMACS registries and demonstrated that the model outperforms commonly used machine learning models and provides transparent and accessible recommendations that can inform clinician decision making. Such a model has multiple potential uses in other important and similarly sensitive clinical applications outside of HF care.

Author Contributors

All authors had full access to the raw data sets and take responsibility for the integrity of the data and the accuracy of the analysis. Drs. Golbus, Yao, Aaronson, Gryak, and Najarian contributed to the conceptualization of this work. All authors contributed to funding acquisition. Dr Yao performed the analysis under the supervision of Drs. Gryak and Najarian who verified the data reported in the manuscript.

The first draft of the manuscript was prepared by the co-first authors. It was reviewed and edited by all authors, all of whom made the decision to submit the manuscript.

Disclosures statement

Jonathan Gryak, Kayvan Najarian, and Heming Yao are named inventors on a planned patent application related to the tropical geometry-based interpretable machine learning algorithm described in this manuscript. Keith Aaronson has received institutional funding for clinical trials (Abbott, Amgen, Cytokinetics), honoraria for consulting (Abbott) and for participating on an Independent Physician Quality Panel (Medtronic), and stock options (Procyron). Dr. Golbus receives funding from the NIH (L30HL143700) and receives salary support by an American Heart Association grant (grant number 20SFRN35370008).

Funding

This work was supported by the National Science Foundation (Grant No. 2014003). The Interagency Registry for Mechanically Assisted Circulatory Support (INTERMACS) was carried out as a collaborative study supported by contract HHSN268201100025C from the National Heart, Lung, and Blood Institute. The Registry Evaluation of Vital Information for VADs in Ambulatory Life (REVIVAL) was funded by the National Institutes of Health, National Heart, Lung, and Blood Institute (NHLBI Contract Number: HHSN268201100026C), and the National Center for Advancing Translational Sciences (NCATS Grant Number: UL1TR002240) for the Michigan Institute for Clinical and Health Research (MICHR).

Acknowledgments

Data for this study were provided, in part, by the Registry Evaluation of Vital Information for VADs in Ambulatory Life (REVIVAL). REVIVAL was supported by funding from the National Institutes of Health, National Heart, Lung, and Blood Institute (NHLBI Contract Number HHSN268201100026C) and the National Center for Advancing Translational Sciences (NCATS Grant Number UL1TR002240) for the Michigan Institute for Clinical and Health Research (MICHR). Additional data were provided by the Interagency Registry for Mechanically Assisted Circulatory Support (INTERMACS). INTERMACS was carried out as a collaborative study supported by contract HHSN268201100025C from the National Heart, Lung, and Blood Institute. Opinions expressed in this manuscript do not represent those of REVIVAL, INTERMACS, NHLBI, Centers for Medicare and Medicaid Services (CMS), or U. S. Food and Drug Administration (FDA). The authors would like to acknowledge Dr. Marissa Miller and Dr. Wendy Taddei-Peters for their assistance in obtaining the INTERMACS data. Additionally, the authors would like to thank Drs. Jennifer Cowger, Garrick Stewart, Michelle

Kittleston, Donna Mancini, and Josef Stehlik for their assistance in creating the clinical rules used to initiate the model. Finally, we would like to acknowledge the substantial contributions to the acquisition and to the interpretation of data by the Matthew Konerman MD, MS to be listed as an indexed collaborator.

Supplementary materials

Supplementary material associated with this article can be found in the online version at <https://doi.org/10.1016/j.healun.2022.08.028>.

References

1. Virani SS, Alonso A, Aparicio HJ, et al. Heart disease and stroke statistics-2021 update: a report from the American Heart Association. *Circulation* 2021;143:e254-743.
2. Tsao CW, Lyass A, Enserro D, et al. Temporal Trends in the incidence of and mortality associated with heart failure with preserved and reduced ejection fraction. *JACC: Heart Failure*. 2018;6:678-685.
3. Gerber Y, Weston SA, Redfield MM, et al. A contemporary appraisal of the heart failure epidemic in Olmsted County, Minnesota, 2000 to 2010. *JAMA Internal Med* 2015;175:996. <https://doi.org/10.1001/jamainternmed.2015.0924>.
4. Gustafsson F, Rogers JG. Left ventricular assist device therapy in advanced heart failure: patient selection and outcomes. *Eur J Heart Fail* 2017;19:595-602.
5. Ambardekar AV, Kittleston MM, Palardy M, et al. Outcomes with ambulatory advanced heart failure from the Medical Arm of Mechanically Assisted Circulatory Support (MedaMACS) Registry. *J Heart Lung Transplant* 2019;38:408-17.
6. Guglin M, Zucker MJ, Borlaug BA, et al. Evaluation for heart transplantation and LVAD implantation: JACC council perspectives. *J Am Coll Cardiol* 2020;75:1471-87.
7. Colvin M, Smith JM, Ahn Y, et al. OPTN/SRTR 2019 annual data report: Heart. *Am J Transplant* 2021;21(suppl 2):356-440.
8. Mehra MR, Naka Y, Uriel N, et al. A fully magnetically levitated circulatory pump for advanced heart failure. *N Engl J Med* 2017; 376:440-50.
9. Kormos RL, Cowger J, Pagani FD, et al. The society of thoracic surgeons intermacs database annual report: evolving indications, outcomes, and scientific partnerships. *J Heart Lung Transplant* 2019;38:114-26.
10. Mehra MR, Canter CE, Hannan MM, et al. The 2016 International Society for Heart Lung Transplantation listing criteria for heart transplantation: A 10-year update. *J Heart Lung Transplant* 2016;35:1-23.
11. Feldman D, Pamboukian SV, Teuteberg JJ, et al. The 2013 International Society for Heart and Lung Transplantation Guidelines for mechanical circulatory support: executive summary. *J Heart Lung Transplant* 2013;32:157-87.
12. Morris AA, Khazanie P, Drazner MH, et al. Guidance for timely and appropriate referral of patients with advanced heart failure: a scientific statement from the American Heart Association. *Circulation* 2021;144:e238-50.
13. Canepa M, Fonseca C, Chioncel O, et al. Performance of prognostic risk scores in chronic heart failure patients enrolled in the European Society of Cardiology heart failure long-term registry. *JACC: Heart Failure* 2018;6:452-62. 06/.
14. Alba AC, Agoritsas T, Jankowski M, et al. Risk prediction models for mortality in ambulatory patients with heart failure: a systematic review. *Circ Heart Fail* 2013;6:881-9.
15. Allen LA, Matlock DD, Shetterly SM, et al. Use of risk models to predict death in the next year among individual ambulatory patients with heart failure. *JAMA Cardiol* 2017;2:435-41.
16. Yao H, Derksen H, Golbus JR, et al. A novel tropical geometry-based interpretable machine learning method: Application in prognosis of

- advanced heart failure. *arXiv [csLG]*. 2021. <http://arxiv.org/abs/2112.05071>. Accessed December 9, 2021.
17. Aaronson KD, Stewart GC, Pagani FD, et al. Registry evaluation of vital information for VADs in ambulatory life (REVIVAL): rationale, design, baseline characteristics, and inclusion criteria performance. *J Heart Lung Transplant* 2020;39:7-15.
 18. School of medicine - interagency registry for mechanically assisted circulatory support. Available at: <https://www.uab.edu/medicine/intermacs/>. Accessed January 21, 2022.
 19. Kirklin JK, Naftel DC, Stevenson LW, et al. INTERMACS database for durable devices for circulatory support: first annual report. *J Heart Lung Transplant* 2008;27:1065-72.
 20. Miller MA, Ullisney K, Baldwin JT. INTERMACS (Interagency Registry for Mechanically Assisted Circulatory Support): a new paradigm for translating registry data into clinical practice. *J Am Coll Cardiol* 2010;56:738-40.
 21. Bouchon-Meunier B, Yager RR, Zadeh LA. *Fuzzy Logic and Soft Computing*. Singapore, Singapore: World Scientific Publishing; 1995.