ESC European Society of Cardiology

European Heart Journal - Cardiovascular Imaging (2018) **19**, 730–738
doi:10.1093/ehjci/jey003

Predicting deterioration of ventricular function in patients with repaired tetralogy of Fallot using machine learning

Manar D. Samad[1†], Gregory J. Wehner[2†], Mohammad R. Arbabshirani[1], Linyuan Jing[1], Andrew J. Powell[3], Tal Geva[3], Christopher M. Haggerty[1], and Brandon K. Fornwalt[1,4]*

[1]Department of Imaging Science and Innovation, Center for Health Research, Geisinger Clinic, 100 North Academy Avenue, Danville, 17822-4400 PA, USA; [2]Department of Biomedical Engineering, University of Kentucky, 522 Robotics and Manufacturing Building, Lexington, 40506-0108 KY, USA; [3]Department of Cardiology, Boston Children's Hospital, 300 Longwood Ave, Boston, 02115 MA, USA; and [4]Department of Radiology, Geisinger, 100 North Academy Ave, Danville, 17822 PA, USA

Received 25 October 2017; editorial decision 30 December 2017; accepted 5 January 2018; online publish-ahead-of-print 12 March 2018

Aims Previous studies using regression analyses have failed to identify which patients with repaired tetralogy of Fallot (rTOF) are at risk for deterioration in ventricular size and function despite using common clinical and cardiac function parameters as well as cardiac mechanics (strain and dyssynchrony). This study used a machine learning pipeline to comprehensively investigate the predictive value of the baseline variables derived from cardiac magnetic resonance (CMR) imaging and provide models for identifying patients at risk for deterioration.

Methods and results Longitudinal deterioration for 153 patients with rTOF was categorized as 'none', 'minor', or 'major' based on changes in ventricular size and ejection fraction between two CMR scans at least 6 months apart (median 2.7 years). Baseline variables were measured at the time of the first CMR. An exhaustive variable search with a support vector machine classifier and five-fold cross-validation was used to predict deterioration and identify the most useful variables. For predicting any deterioration (minor or major) vs. no deterioration, the mean area under the curve (AUC) was 0.82 ± 0.06. For predicting major deterioration vs. minor or no deterioration, the AUC was 0.77 ± 0.07. Baseline left ventricular (LV) ejection fraction, LV circumferential strain, and pulmonary regurgitation were most useful for achieving accurate predictions.

Conclusion For the prediction of deterioration in patients with rTOF, a machine learning pipeline uncovered the utility of baseline variables that was previously lost to regression analyses. The predictive models may be useful for planning early interventions in patients with high risk.

Keywords outcomes • cardiac magnetic resonance • congenital heart disease • prediction • cardiac mechanics • ventricular deterioration

Introduction

Tetralogy of Fallot (TOF) is a congenital abnormality of the heart that typically requires surgical repair in early childhood.[1] Advancements in surgical techniques have reduced the early mortality rate to less than 3%.[2] However, the annual mortality rate more than triples, from 0.24%/year to 0.94%/year, 20–30 years after the initial surgical repair largely due to adverse cardiac events, such as sudden cardiac death or worsening heart failure.[2] The increasing mortality risk generally coincides with progressive dilation and dysfunction of both the left and right ventricles (LV and RV, respectively),[3,4] and there is growing evidence of a link between dysfunction and poor outcomes, such as death or sustained ventricular tachycardia.[5,6]

Fortunately, many patients with repaired TOF (rTOF) do not develop progressive ventricular dilation and dysfunction. However, this creates the challenge of predicting which patients are at risk, and the

* Corresponding author. Tel: 570-214-5478; Fax: 570-214-9451. E-mail: bkf@gatech.edu

† These authors contributed equally to this work.

current predictive ability is poor despite several recent efforts. Wald *et al.*[7] investigated whether baseline variables, including clinical parameters and functional measurements derived from routine cardiac magnetic resonance (CMR) imaging, could predict subsequent deterioration in ventricular size and function. They found that no baseline parameters could differentiate between patients whose ventricular function deteriorated over a median 2.2-year follow-up and those whose function remained unchanged. In a subsequent study, Jing *et al.* performed a similar analysis that included baseline measures of cardiac mechanics such as strain and mechanical dyssynchrony, which have been touted as more sensitive measures of cardiac function.[8,9] However, they reported that baseline cardiac mechanics were also not predictive of subsequent ventricular deterioration.[10]

A common limitation of prior studies is the reliance on regression models for assessing the relationship between baseline variables and outcomes. In reality, the interactions mediating physiologic processes may be too complex to be captured using common regression techniques. Fortunately, machine learning, a field of computer science, has emerged with the necessary tools for both identifying complex patterns within data sets and using those patterns to make better predictions.[11] Previously, machine learning has been used in cardiology for the classification of arrhythmias,[12] the classification of constrictive vs. restrictive pericarditis,[13] and the prediction of 1-year mortality in patients with heart failure.[14] Most recently, machine learning has been employed on speckle-tracking echocardiographic data sets to distinguish hypertrophic cardiomyopathy from the physiologic hypertrophy seen in athletes[15] and on coronary angiography data sets to predict 5-year mortality in patients with suspected coronary artery disease.[16] In both cases, machine learning techniques performed better than standard models (e.g. Framingham). Therefore, we propose the use of machine learning to predict deterioration in ventricular size and function in patients with rTOF. Contrary to the results of previous regression techniques, we hypothesize that baseline clinical variables and cardiac mechanics *can* be used to predict deterioration.

Methods

A database search at Boston Children's Hospital identified patients who fulfilled the following criteria: (i) diagnosis of rTOF; (ii) two clinical CMR scans performed at least 6 months apart from May 2005 to March 2012; (iii) no surgical- or catheter-based interventions between CMR scans; and (iv) a 12-lead electrocardiogram (ECG) at the time of first CMR. The search yielded 164 patients, among which 11 were excluded due to incomplete or poor quality imaging. Subsequently, 153 patients (mean age at baseline CMR: 23 ± 14 years, 76 males) were included. The follow-up duration was between 6 months and 5.9 years with a median of 2.7 years (interquartile range 1.9–3.8). Note that this data set is identical to that previously used by Jing *et al.*[10] Using thresholds published by Wald *et al.* that were based on the reproducibility of volumetric and ejection fraction measurements,[7,17,18] patients were categorized into three groups: (i) 'no' deterioration ($n = 38$), (ii) 'minor' deterioration ($n = 78$), or (iii) 'major' deterioration ($n = 37$) according to their change in ventricular size and function between CMR scans (*Table 1*). Eight patients with major deterioration satisfied two of the three criteria while the rest satisfied only one criterion.

We used the same 22 baseline variables as Jing *et al.*,[10] with 2 additional variables: indexed RV mass and age at the first CMR (*Table 2*). A detailed description of the CMR imaging protocol and measurement techniques of CMR variables has been reported.[9,10] A custom feature tracking algorithm

Table 1 Criteria for categorizing patients based on change in ventricular size and function

Deterioration	Increase in RVEDVi (mL/m^2)	Decrease in RVEF (%)	Decrease in LVEF (%)	Selection criterion
None	≤5	≤3	≤3	All
Minor	Not in 'None' or 'Major' group			
Major	≥30	≥10	≥10	Any

Major deterioration was defined as patients who fulfilled any of the three given criteria while no deterioration (none) was defined as patients who fulfilled all three given criteria. The remaining patients were placed into the minor deterioration group. Change in RVEF and LVEF are absolute percent changes.
LVEF, left ventricular ejection fraction; RVEDVi, indexed right ventricular end-diastolic volume; RVEF, right ventricular ejection fraction.

Table 2 Baseline variables used to predict deterioration

	Variables	Mean ± SD or *n* (%)
Demographics	Sex (male)	76 (50%)
	Age at surgical repair (years)	3.5 ± 6.7
	Age at first CMR (years)	23.4 ± 14
Surgical repair procedure[a]	Transannular patch	97 (63%)
	RV-PA conduit	11 (7%)
	RVOT patch	30 (20%)
	Infundibular resection	2 (1%)
	Pulmonary valvotomy	11 (7%)
	Commissurotomy	9 (6%)
Electrocardiogram	QRS duration (ms)	136 ± 27
	Heart rate (HR, beats/min)	80 ± 17
CMR variables	LV dyssynchrony (LV-dyss, ms)	19 ± 11
	RV dyssynchrony (RV-dyss, ms)	57 ± 29
	Interventricular dyssynchrony (Inter-dyss, ms)[b]	-40 ± 20
	LV circumferential strain (LV-Ecc, %)	27 ± 3
	RV circumferential strain (RV-Ecc, %)	18 ± 3
	LV longitudinal strain (LV-Ell, %)	19 ± 3
	RV longitudinal strain (RV-Ell, %)	23 ± 3
	RV end-diastolic volume index (RVEDVi, mL/m^2)	143 ± 36
	RV end-systolic volume index (RVESVi, mL/m^2)	67 ± 23
	LV ejection fraction (LVEF, %)	59 ± 6
	RV ejection fraction (RVEF, %)	53 ± 8
	Pulmonary regurgitation fraction (PR fraction, %)	36 ± 15
	RV mass index (RVMASSi, g/m^2)	33.8 ± 12.5

CMR, cardiac magnetic resonance; LV, left ventricular; RV, right ventricular; RVOT, right ventricular outflow tract; RV-PA, right ventricle-to-pulmonary artery; SD, standard deviation.
[a]Some patients had more than one type of surgical repair.
[b]Negative values indicate delayed contraction of the RV relative to the LV.

was used to extract both LV and RV strains and dyssynchrony from baseline CMR.[9] Briefly, 48–96 segmental strain vs. time curves were generated from 4 to 8 short-axis images covering the ventricles. Peak strains were calculated as the peak of the average strain curve. Cross-correlational delays were calculated for each segmental strain curve when compared to a patient-specific reference curve. The standard deviation of the cross-correlational delays within each ventricle was reported as LV/RV dyssynchrony. The difference between median delays of the two ventricles was defined as inter-ventricular dyssynchrony. Heart rate and QRS duration were obtained from the baseline ECG. Sex and surgical repair procedures were categorical. The remaining variables were continuous.

Machine learning framework

We developed a machine learning pipeline using the Python-based library scikit-learn (*Central illustration*).[19] Specifically, five-fold cross-validation was used with a linear support vector machine (linear-SVM) classifier. SVM is a popular supervised machine learning algorithm used to classify individual samples into two groups. The algorithm uses a set of training examples (e.g. a cohort of volunteers, each represented by a number of variables) with the supervision of known group labels (e.g. patient vs. control) of individual examples. A linear-SVM results in a linear equation of many input variables, which serves as a 'decision boundary' to determine the group label of a test sample. Among many choices of 'decision boundary' separating the two groups, linear-SVM optimally determines the one that ensures maximum separation between the groups in terms of input variables. This maximum separation criterion during training yields improved classification performance compared to other predictive models.

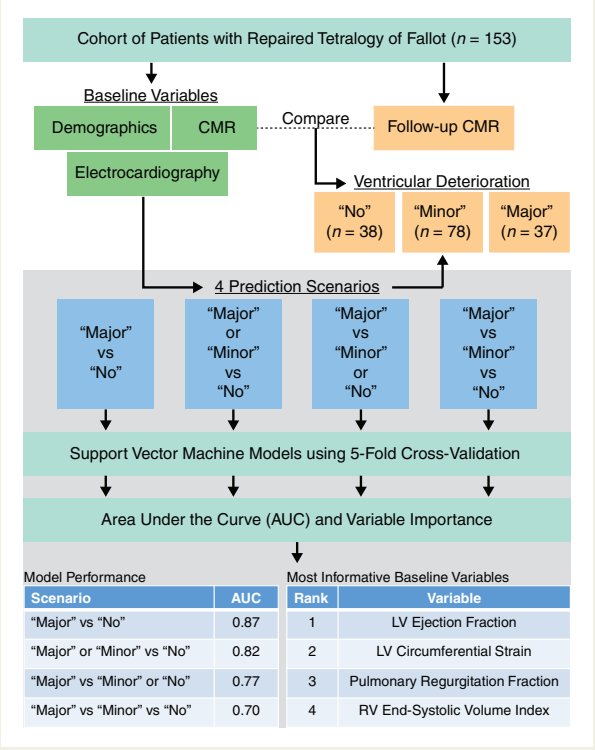
The mean area under the receiver operating characteristic curve (AUC) over the five-fold cross-validations was used as the performance metric. We used an exhaustive search algorithm to find the optimal combination of baseline variables.[20] In an exhaustive variable search, there are ${}^{n}C_{r}$ unique combinations of r variables,

$$ {}^{n}Cr = \frac{n!}{r!(n-r)!} \quad (1) $$

where n is 24 variables in our case. For each experimental scenario below, a total of 16 777 215 variable combinations (ranging from $r = 1$ to $r = 24$) were assessed with the expectation that performance would initially increase with the inclusion of more informative variables and eventually decrease as high numbers of less informative variables would lead to overfitting. The variable combinations were ranked according to their cross-validated mean AUC in order to reveal the optimal variable combination. In addition to reporting the best performing variable combination, the top 100 combinations were inspected to determine the prevalence of the different baseline variables within those high-performing combinations as a measure of variable importance. Finally, Platt scaling was used to convert the discrete predictions of the linear-SVMs into probabilities,[21] which allows a new patient to be linked to the probability of experiencing no, minor, or major deterioration.

Experimental scenarios

Four experimental scenarios were considered for predicting deterioration (*Table 3*). Scenario 1 sought to distinguish those with major deterioration from those without deterioration, which were the two groups analysed in previous subgroup analyses.[10] Scenario 2 sought to distinguish patients with any deterioration (major or minor) from those without deterioration while Scenario 3 attempted to distinguish patients with major deterioration from patients with either minor or no deterioration. Scenario 4 attempted a three-group classification where each group was predicted separately using a *one-vs.-all* binary classification model, and the average AUC over 3 *one-vs.-all* binary classification models was reported.



Scenario	AUC
"Major" vs "No"	0.87
"Major" or "Minor" vs "No"	0.82
"Major" vs "Minor" or "No"	0.77
"Major" vs "Minor" vs "No"	0.70

Rank	Variable
1	LV Ejection Fraction
2	LV Circumferential Strain
3	Pulmonary Regurgitation Fraction
4	RV End-Systolic Volume Index

Central illustration Machine-learning pipeline for predicting deterioration in patients with rTOF. A machine-learning pipeline was designed to evaluate four predictive models and identify the most important baseline variables to predict deterioration in ventricular size and function in patients with repaired tetralogy of Fallot (rTOF). The patients with rTOF were categorized into three deterioration groups using measures from follow-up cardiac magnetic resonance (CMR) scans and then the deterioration was predicted using baseline measurements. The predictive performance was reported using area under the curve (AUC). LV, left ventricular; RV, right ventricular.

RESULTS

Overall predictive performance

All experimental scenarios yielded AUCs above the random chance level of 0.5 (*Figure 1*). Scenario 1 obtained the highest mean AUC (0.87 ± 0.07) was obtained using six variables (*Table 4*). The remaining Scenarios 2, 3, and 4 achieved mean AUCs of 0.82 ± 0.06, 0.77 ± 0.07, and 0.70 ± 0.06 by using an optimum of 6, 5, and 10 variables, respectively.

Scenario 1: major vs. no deterioration

While the best performance was achieved with six variables, results from the best two variables are presented first for visualization. The mean AUC (0.77 ± 0.12) for the best single predictor (LVEF) improved to 0.84 ± 0.08 after adding pulmonary regurgitation (PR) fraction. In PR fraction vs. LVEF variable space, a linear-SVM obtains the decision boundary for the binary classification of patients into either major or no deterioration (*Figure 2A*). The contour lines in *Figure 2B*

Table 3 Experimental scenarios for predicting deterioration

Scenarios	Classification type	Patient grouping	Goal
Scenario 1	Binary	Major ($n = 37$) and no deterioration ($n = 38$)	Predicting major vs. no deterioration
Scenario 2	Binary	Patients with ($n = 115$) and without ($n = 38$) deterioration	Predicting major or minor deterioration vs. no deterioration
Scenario 3	Binary	Patients with ($n = 37$) and without major deterioration ($n = 116$)	Predicting major deterioration vs. minor or no deterioration
Scenario 4	Three-group	Major ($n = 37$), minor ($n = 78$), no deterioration ($n = 38$)	Predicting the three distinct groups: major vs. minor vs. no deterioration

Scenarios 2–4 consider the possible ways to incorporate patients with minor deterioration. In Scenario 2, they are combined with patients with major deterioration; in Scenario 3, they are combined with patients with no deterioration; in Scenario 4 they are treated as a distinct third group.

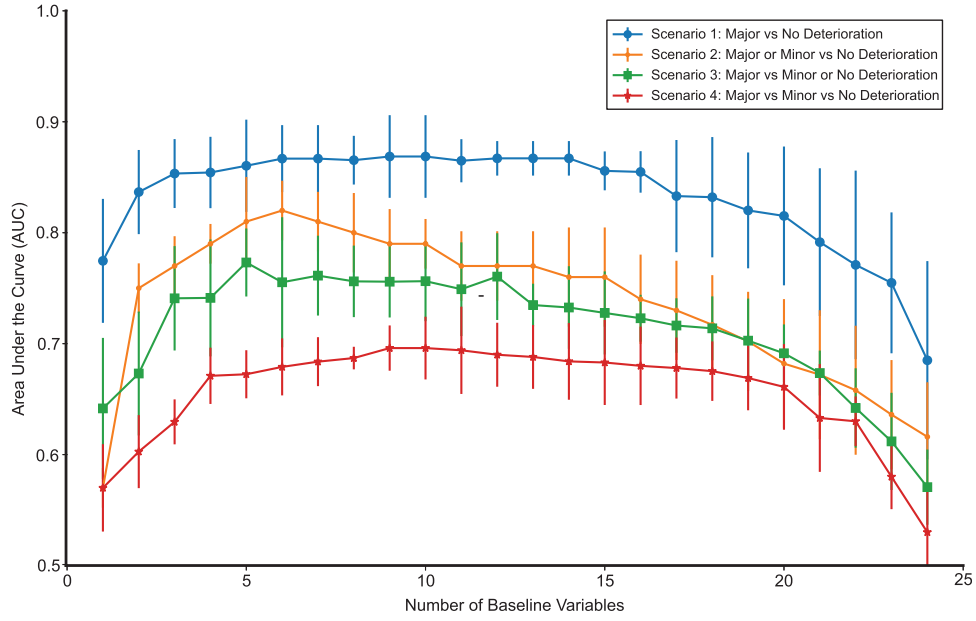


Figure 1 Performance of the predictive models. The best mean cross-validated AUC obtained from exhaustive variable search across the different numbers of baseline predictor variables for each scenario.

indicate the probability that a new patient would be expected to experience major deterioration depending on their baseline LVEF and PR fraction.

The parametric formulation of the decision boundary below can be used to classify a new patient using baseline LVEF and PR fraction.

$$f(\text{LVEF}, \text{PR}) = a * \text{LVEF} + b * \text{PR} + C = 0 \quad (2)$$

where, $a = -0.1936$, $b = -0.0457$, and $c = 13.317$ are the coefficients and LVEF and PR fraction are the values of the baseline variables. Using Eq. (2), the binary decision about a given patient with rTOF can be made using the following criteria:

$f(\text{LVEF}, \text{PR}) > 0$ Patient without deterioration
$f(\text{LVEF}, \text{PR}) < 0$ Patient with major deterioration.

The negative coefficients indicate that lower values for both LVEF and PR fraction are favourable for no deterioration. Equivalent parametric formulations of the binary decision as well as the probabilities shown in *Figure 2B*, which can be used to predict the deterioration of a new patient, are easily calculated from the 6D variable combination that yielded the best AUC; however, the 6D decision boundary is impossible to visualize. In addition to LVEF and PR fraction, the optimal combination of variables included heart rate, interventricular dyssynchrony, RV longitudinal strain, and RVEF.

Scenario 2: major or minor vs. no deterioration

For the prediction of any deterioration vs. no deterioration, the best variable combination (mean AUC of 0.82 ± 0.06) included LVEF, LV

circumferential strain, RVESVi, and three surgical repair variables (transannular patch, infundibular resection, and commisurotomy). The formula for the decision boundary indicated that lower LVEF and RVESVi, higher LV circumferential strain, and the absence of any of the three repair types were favourable for no deterioration (*Table 4*).

Scenario 3: major vs. minor or no deterioration

For the prediction of major deterioration vs. minor or no deterioration, the highest mean AUC was 0.77 ± 0.07 using five baseline variables. Higher longitudinal strains, lower LVEF and PR fraction, and female gender were favourable for only minor or no deterioration (*Table 4*).

Table 4 Best combinations of baseline variables for each experimental scenario

Scenario	Best combination of variables	Mean AUC (standard deviation)
1	LVEF↓, PR↓, HR↑, Inter-dyss↓, RV-Ell↑, RVEF↓	0.87 ± 0.07
2	LVEF↓, LV-Ecc↑, RVESVi↓, transannular patch↓, infundibular resection↓, commissurotomy↓	0.82 ± 0.06
3	LVEF↓, PR↓, LV-Ell↑, RV-Ell↑, Sex[a]	0.77 ± 0.07
4	LVEF, PR, LV-Ecc, LV-Ell, RV-EDV, inter-dyss, QRS, age at surgical repair, age at first CMR, transannular patch	0.70 ± 0.06

Scenario 1 indicates major vs. no deterioration; Scenario 2 indicates major or minor vs. no deterioration; Scenario 3 indicates major vs. minor or no deterioration; Scenario 4 indicates major vs. minor vs. no deterioration. An 'up' arrow indicates that a higher value (or the presence of the variable if categorical) is favourable for predicting no deterioration and vice versa.
CMR, cardiac magnetic resonance; HR, heart rate; Inter-dyss, interventricular dyssynchrony; LV-Ecc, left ventricular circumferential strain; LVEF, left ventricular ejection fraction; LV-Ell, left ventricular longitudinal strain; PR, pulmonary regurgitation; RV-EDV, right ventricular end-diastolic volume; RVEF, right ventricular ejection fraction; RV-Ell, right ventricular longitudinal strain; RVESVi, indexed right ventricular end-systolic volume.
[a]Female is favourable relative to male.

Scenario 4: major vs. minor vs. no deterioration

For the three-group classification, the highest mean AUC (0.70 ± 0.06) was obtained using 10 baseline variables (*Table 4*). Because the three-group classification requires multiple one-vs.-all decision boundaries, and, in particular, one of those decisions is for minor deterioration vs. either no or major deterioration, it is not possible to report a favourable directionality for each variable.

Variable importance

The exhaustive variable search obtained the best performing combination of variables for each experimental scenario. Therefore, to assess the importance of each baseline variable individually, their prevalence within the top 100 variable combinations was measured. For each scenario, the top 100 variable combinations were found to be within 0.05 units from the best AUC.

For Scenario 1, LVEF, PR fraction, HR, and the age at surgical repair had the highest prevalence (*Figure 3*). Scenario 2 demonstrated high prevalence of LVEF, LV circumferential strain, RVESVi, and transannular patch. Scenario 3 frequently selected the right ventricular measurements such as RVEDVi, RVESVi, and RVEF in addition to LVEF, PR fraction, and age at surgical repair. The three-group classification (Scenario 4) identified LVEF, age at surgical repair, age at the first CMR, LV circumferential and longitudinal strains, and transannular patch as the most predictive variables. After averaging the prevalences across the four scenarios, the variable ranking in *Figure 4* shows that LVEF, PR fraction, LV circumferential strain, RVESVi, and age at surgical repair were the top five most informative baseline variables.

Discussion

This study proposed a machine learning pipeline to predict worsening ventricular size and function in patients with rTOF using baseline

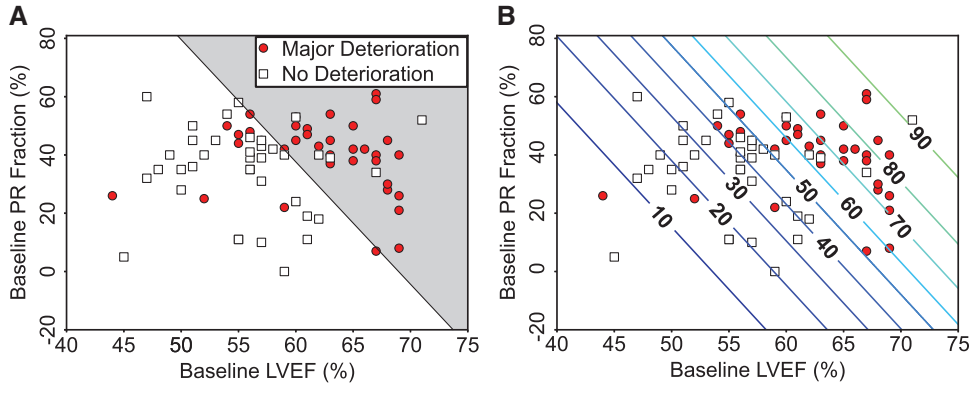


Figure 2 Distinguishing patients with major vs. no deterioration. Visualization of (*A*) the linear decision boundary between major and no deterioration, and (*B*) a contour plot of the probability (%) of experiencing major deterioration as a function of the two most predictive baseline variables (LVEF and PR fraction). LVEF, left ventricular ejection fraction; PR, pulmonary regurgitation.

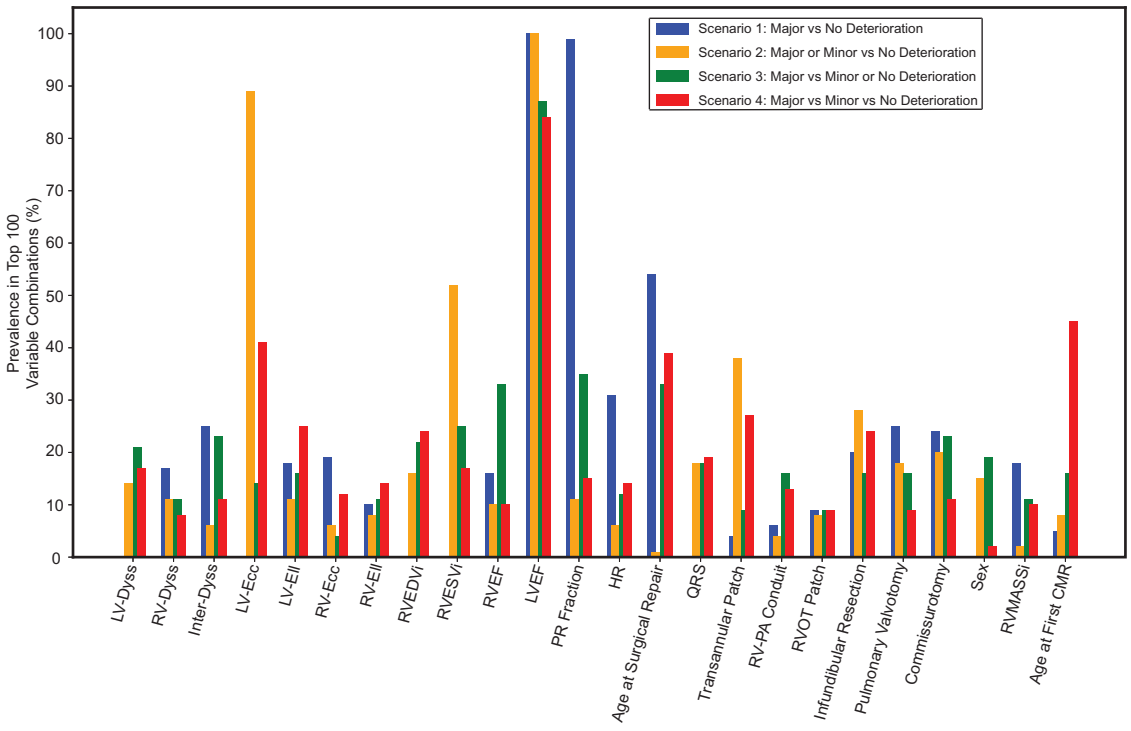


Figure 3 Baseline variable importance for each scenario. Prevalence of individual baseline variables in the top 100 combinations for each scenario. CMR, cardiac magnetic resonance; HR, heart rate; Inter-Dyss, interventricular dyssynchrony; LV-Dyss, left ventricular dyssynchrony; LV-Ecc, left ventricular circumferential strain; LVEF, left ventricular ejection fraction; LV-Ell, left ventricular longitudinal strain; PR, pulmonary regurgitation; RV-Dyss, right ventricular dyssynchrony; RV-Ecc, right ventricular circumferential strain; RVEDVi, indexed right ventricular end-diastolic volume; RVEF, right ventricular ejection fraction; RV-Ell, right ventricular longitudinal strain; RVESVi, indexed right ventricular end-systolic volume; RVMASSi, right ventricular mass index.

clinical and CMR-derived variables. Our major findings included: (i) baseline clinical parameters and measures of cardiac function and mechanics from CMR can be used to predict deterioration in ventricular size and function using machine learning (AUCs of 0.70–0.87 across four categorization scenarios); (ii) LVEF, PR fraction, LV circumferential strain, RVESVi, and age at surgical repair were the five most informative baseline variables; and (iii) the machine learning models provided decision boundaries and probability estimates as a function of baseline variables that can be used to determine the likelihood of deterioration for a new patient. We believe these findings serve as an important proof-of-concept for one of the first applications of machine learning in congenital heart disease and provide new insights with relevance to clinical decision-making.

Deterioration in ventricular size and function in patients with rTOF

Patients with rTOF may experience progressive ventricular dilation and dysfunction, which are independent predictors of adverse events such as sudden cardiac death and sustained ventricular tachycardia.[4,5,22] Therefore, identifying patients at risk for ventricular deterioration may be useful for guiding the frequency of clinical follow-up as well as the timing of interventions. While several studies have explored potential predictors of deterioration, such as clinical history, ECG, and CMR parameters, none have resulted in robust and accurate prediction models.[7,23,24] These studies relied on traditional statistics such as regression and subgroup analyses.

Success of machine learning compared to traditional models

There is mounting evidence that traditional statistical modelling, which is common in cardiology,[25–27] is not as useful as machine learning for patient risk stratification.[11,16] Pertinent examples include the previous multivariate regression models that were unable to predict deterioration in patients with rTOF.[7,10] In addition, while subgroup analyses may identify variables that are significantly different between patient groups (e.g. major vs. no deterioration), these findings do not necessarily help to predict outcomes for individual patients.[28] Finally, statistical analyses that ignore interactions among variables may miss useful variable interactions.

In contrast, our proposed machine learning pipeline predicted deterioration in ventricular size and function by uncovering the predictive ability of baseline variables that was previously lost to traditional techniques. For example, Jing *et al.* reported no difference in PR fraction between the major and no deterioration groups.[10] However, our results demonstrated that combining PR fraction with LVEF substantially increased the model AUC from 0.77 to 0.84 (Scenario 1, *Figure 2*).

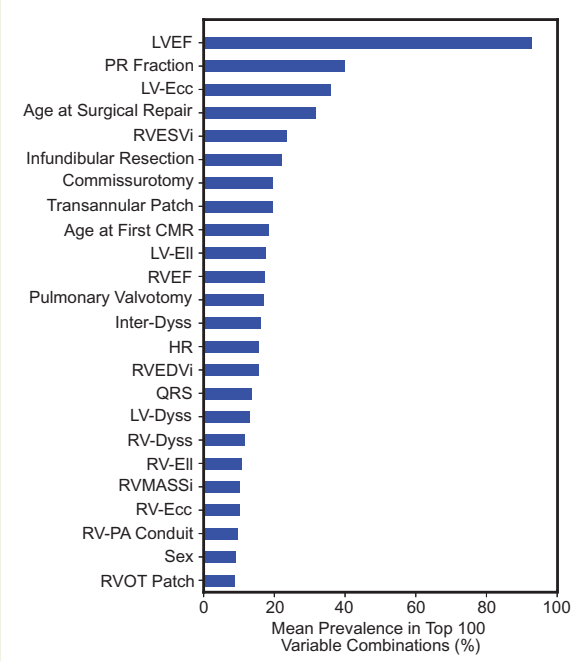


Figure 4 Rank-ordering of the individual baseline variables. Ranking of individual baseline variables based on their mean prevalence in top 100 variable combinations across the four experimental scenarios. CMR, cardiac magnetic resonance; HR, heart rate; Inter-Dyss, interventricular dyssynchrony; LV-Dyss, left ventricular dyssynchrony; LV-Ecc, left ventricular circumferential strain; LVEF, left ventricular ejection fraction; LV-Ell, left ventricular longitudinal strain; PR, pulmonary regurgitation; RV-Dyss, right ventricular dyssynchrony; RV-Ecc, right ventricular circumferential strain; RVEDVi, indexed right ventricular end-diastolic volume; RVEF, right ventricular ejection fraction; RV-Ell, right ventricular longitudinal strain; RVESVi, indexed right ventricular end-systolic volume; RVMASSi, right ventricular mass index.

Furthermore, PR fraction was a component of the six-variable combination that yielded the best AUC (0.87). The surgical repair of TOF often results in PR, which can eventually lead to deterioration in ventricular size and function.[3,29] Therefore, it is significant that a machine learning-based model identified the predictive value of PR fraction where prior statistical models were unsuccessful.

In addition, our study investigated a more comprehensive risk stratification by incorporating the group with minor deterioration (Scenarios 2–4), which is necessary for any future clinical applications. The prediction performance for Scenario 2 (major or minor deterioration vs. none; AUC = 0.82), when compared to Scenario 3 (major deterioration vs. minor or none; AUC = 0.77), suggests that the minor deterioration group is more similar to the major deterioration group. However, the lower AUC achieved for the three-group classification (AUC = 0.70) is indicative of substantial overlap among the three groups. Further research is needed to determine the clinical significance of the minor deterioration group with respect to long-term patient outcomes. Depending on the clinical need for distinguishing the minor group from either or both of the other groups, any of Scenarios 2–4 may be clinically useful.

Identification of the most useful baseline variables

Baseline LVEF was invariably the best individual predictor of deterioration. There was an inverse relationship between LVEF and deterioration such that higher values of LVEF indicated a greater likelihood of deterioration. This same finding was present in two previous studies of patients with rTOF.[7,10] The reason for this inverse relationship is unknown. However, studies in other populations such as asymptomatic volunteers in the MESA study,[30] patients with heart failure,[31,32] and patients admitted to the intensive care unit[33] have demonstrated that increased LVEF (e.g. >65%) can be associated with poor outcomes. RVEF demonstrated a similar inverse relationship with deterioration in Scenario 1; however, this may be due to the presence of PR, which was the second most important variable for predicting deterioration. Higher PR fractions, which may be associated with elevated RVEF in compensated ventricles, were indicative of a greater likelihood for subsequent deterioration.

Other high-ranking variables (*Figure 4*) included LV circumferential strain, RVESVi, and age at surgical repair. Many of these variables were individually identified by prior studies. For example, Orwat *et al.* reported that LV circumferential strain was the strongest predictor of mortality among other cardiac strain measures in patients with rTOF.[27] In addition, RVESVi has been shown to be an important determinant of RV remodelling.[34,35] Gatzoulis *et al.*[36] found that older age at surgical repair was a strong predictor of late mortality. Similar to PR fraction, both the differences in baseline RVESVi and age at surgical repair were previously reported as *insignificant* between patients with major and no deterioration on subgroup analysis.[10] On the other hand, RVEF and RVEDVi were two of the five variables that did reach significance in the previous subgroup analysis.[10] However, these variables were not in the top 10 of variable rank in the present study (RVEF: rank 11, RVEDVi: rank 15), which demonstrates that significant differences between subgroups do not necessarily correlate with a variable's utility to predict outcomes.

Implications

Several features of our machine learning pipeline have implications for future clinical applications and prediction studies. First, linear-SVM provides intuitive decision boundaries for binary classifications of new patients (*Figure 2*). Even with higher numbers of variables, the directionality is readily observed from the coefficients of the decision boundary formula. Furthermore, the categorical decisions are easily converted to probabilities, which could be used to guide management decisions. Similar to predicting deterioration in patients with rTOF, the proposed pipeline could be applied in other clinical scenarios to predict future status by baseline variables. Importantly, our methodology to rank order the importance of clinical variables in predicting outcomes could be applied in many other scenarios to garner new insights into disease pathophysiology.

Limitations

The proposed exhaustive variable search may not be computationally efficient with a large number of baseline variables. Hence, a limited number of variables were selected based on previous knowledge about relevant predictors of adverse outcomes in rTOF to avoid overfitting and impractical computational cost. However, exhaustive

search should be considered when feasible to guarantee identification of optimal combinations of baseline variables, as opposed to existing variable selection algorithms such as forward sequential search and information gain.[37]

The outcome variables used in the current study were based on changes in predefined CMR-derived parameters (RVEDVi, LVEF, and RVEF), which reflect surrogates of progressive remodelling and dysfunction. These outcome variables are different from conventional clinical outcomes such as death; however, multiple previous papers have suggested that these intermediate endpoints of progressive remodelling and dysfunction ultimately relate to adverse clinical outcomes.[4,5,22] Future studies should investigate whether interventions targeted based on machine learning techniques to predict these intermediate endpoints can ultimately improve clinical outcomes.

We performed five-fold cross-validation as an internal validation to avoid bias in the evaluation of the models. The resulting mean AUCs were well above random chance for all four scenarios, which clearly showed previously unknown predictive ability of the baseline variables. However, we used data collected from a single institution, which can miss variations in measurements across institutions. An external validation from a truly independent cohort would be useful to generalize the results. The validation results could be further improved with additional training samples. However, large data sets in patients with repaired TOF linked to longitudinal follow-up data are rare.

The sample size of 153 patients is generally small for machine learning, which motivated some aspects of our pipeline. Linear-SVMs were chosen not only for their ease of interpretation, but also for their resistance to overfitting and paucity of hyperparameters, which typically require large training sets to adequately tune. However, similar sample sizes are common in machine learning studies within the field of cardiology, [$n = 139$ (15); $n = 94$ (13)] and our sample size is relatively large for the field of congenital heart disease.

Due to its lower frame rate compared to echocardiography, MRI may not be optimal for assessing dyssynchrony. However, we addressed this problem by upsampling the segmental strain curves, and using cross-correlational analysis to compare each segmental curve to a patient-specific reference curve, to find a time shift (delay) that best matches the pattern of the segmental strain curve to the reference curve. We have documented good inter-test reproducibility of this method previously, and shown that patients with repaired TOF have higher dyssynchrony compared to healthy controls.[9] Moreover, MRI is able to assess all three types of dyssynchrony: interventricular, right intra-ventricular, and left intra-ventricular, whereas echocardiography cannot reproducibly quantify all of these measures.

Conclusions

For the prediction of ventricular deterioration in patients with rTOF, a machine learning pipeline uncovered the utility of baseline variables that was previously lost to traditional statistical methods. Left ventricular ejection fraction, PR fraction, feature tracking-derived left ventricular circumferential strain, right ventricular end-systolic volume index, and age at surgical repair were the five most important baseline variables for predicting deterioration over a median duration of 2.7 years. The proposed pipeline may be useful for identifying patients with rTOF who are at risk for deterioration and for planning appropriate interventions. While data from a new and larger patient population are needed for external validation, the proposed pipeline could be applied to any clinical scenario where it is desirable to predict future clinical status using baseline variables.

Funding

This work was supported by a National Institutes of Health (NIH) Director's Early Independence Award (DP5 OD-012132), NIH grant number KL2 RR033171 from the National Center for Research Resources and the National Center for Advancing Translational Sciences, and American Heart Association Great Rivers Affiliate via grant 14POST20310025. The content is solely the responsibility of the authors and does not necessarily represent the official views of NIH.

Conflict of interest: None declared.

References

1. Apitz C, Webb GD, Redington AN. Tetralogy of Fallot. *Lancet* 2009;**374**:1462–71.
2. Nollert G, Fischlein T, Bouterwek S, Böhmer C, Klinner W, Reichart B. Long-term survival in patients with repair of tetralogy of Fallot: 36-year follow-up of 490 survivors of the first year after surgical repair. *J Am Coll Cardiol* 1997;**30**:1374–83.
3. Valente AM, Powell AJ. Clinical applications of cardiovascular magnetic resonance in congenital heart disease. *Magn Reson Imaging Clin N Am* 2007;**15**:565–77.
4. Geva T, Sandweiss BM, Gauvreau K, Lock JE, Powell AJ. Factors associated with impaired clinical status in long-term survivors of tetralogy of Fallot repair evaluated by magnetic resonance imaging. *J Am Coll Cardiol* 2004;**43**:1068–74.
5. Knauth AL, Gauvreau K, Powell AJ, Landzberg MJ, Walsh EP, Lock JE *et al.* Ventricular size and function assessed by cardiac MRI predict major adverse clinical outcomes late after tetralogy of Fallot repair. *Heart* 2008;**94**:211–6.
6. Ortega M, Triedman JK, Geva T, Harrild DM. Relation of left ventricular dyssynchrony measured by cardiac magnetic resonance tissue tracking in repaired tetralogy of Fallot to ventricular tachycardia and death. *Am J Cardiol* 2011;**107**:1535–40.
7. Wald RM, Valente AM, Gauvreau K, Babu-Narayan SV, Assenza GE, Schreier J *et al.* Cardiac magnetic resonance markers of progressive RV dilation and dysfunction after tetralogy of Fallot repair. *Heart* 2015;**0**:1–7.
8. Stanton T, Leano R, Marwick TH. Prediction of all-cause mortality from global longitudinal speckle strain: comparison with ejection fraction and wall motion scoring. *Circ Cardiovasc Imaging* 2009;**2**:356–64.
9. Jing L, Haggerty CM, Suever JD, Alhadad S, Prakash A, Cecchin F *et al.* Patients with repaired tetralogy of Fallot suffer from intra- and inter-ventricular cardiac dyssynchrony: a cardiac magnetic resonance study. *Eur Heart J Cardiovasc Imaging* 2014;**15**:1333–43.
10. Jing L, Wehner GJ, Suever JD, Charnigo RJ, Alhadad S, Stearns E *et al.* Left and right ventricular dyssynchrony and strains from cardiovascular magnetic resonance feature tracking do not predict deterioration of ventricular function in patients with repaired tetralogy of Fallot. *J Cardiovasc Magn Reson* 2016;**18**:49–58.
11. Goldstein BA, Navar AM, Carter RE. Moving beyond regression techniques in cardiovascular risk prediction: applying machine learning to address analytic challenges. *Eur Heart J* 2017;**38**:1805–14.
12. Asl BM, Setarehdan SK, Mohebbi M. Support vector machine-based arrhythmia classification using reduced features of heart rate variability signal. *Artif Intell Med* 2008;**44**:51–64.
13. Sengupta PP, Huang Y-M, Bansal M, Ashrafi A, Fisher M, Shameer K *et al.* Cognitive machine-learning algorithm for cardiac imaging. *Circ Cardiovasc Imaging* 2016;**9**:e004330.
14. Ortiz J, Ghefter CG, Silva CE, Sabbatini RM. One-year mortality prognosis in heart failure: a neural network approach based on echocardiographic data. *J Am Coll Cardiol* 1995;**26**:1586–93.
15. Narula S, Shameer K, Salem Omar AM, Dudley JT, Sengupta PP. Machine-learning algorithms to automate morphological and functional assessments in 2D echocardiography. *J Am Coll Cardiol* 2016;**68**:2287–95.
16. Motwani M, Dey D, Berman DS, Germano G, Achenbach S, Al-Mallah MH *et al.* Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: a 5-year multicentre prospective registry analysis. *Eur Heart J* 2016;**52**:468–76.

17. Mooij CF, de Wit CJ, Graham DA, Powell AJ, Geva T. Reproducibility of MRI measurements of right ventricular size and function in patients with normal and dilated ventricles. *J Magn Reson Imaging* 2008;**28**:67–73.
18. Blalock SE, Banka P, Geva T, Powell AJ, Zhou J, Prakash A. Interstudy variability in cardiac magnetic resonance imaging measurements of ventricular volume, mass, and ejection fraction in repaired tetralogy of Fallot: a prospective observational study. *J Magn Reson Imaging* 2013;**38**:829–35.
19. Pedregosa F, Varoquaux G. Scikit-learn: machine learning in Python. *J. Mach Learn Res* 2011;**12**:2825–30.
20. Ferri FJ, Pudil P, Hatef M, Kittler J. Comparative study of techniques for large-scale feature selection. *Pattern Recognit Pract IV* 1994;**16**:403–13.
21. Platt J. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Adv Large Margin Classif* 1999;**10**:61–74.
22. Valente AM, Gauvreau K, Assenza GE, Babu-Narayan SV, Schreier J, Gatzoulis MA *et al*. Contemporary predictors of death and sustained ventricular tachycardia in patients with repaired tetralogy of Fallot enrolled in the INDICATOR cohort. *Heart* 2014;**100**:247–53.
23. Luijnenburg SE, Helbing WA, Moelker A, Kroft LJM, Groenink M, Roos-Hesselink JW *et al*. 5-year serial follow-up of clinical condition and ventricular function in patients after repair of tetralogy of Fallot. *Int J Cardiol* 2013;**169**:439–44.
24. Quail MA, Frigiola A, Giardini A, Muthurangu V, Hughes M, Lurz P *et al*. Impact of pulmonary valve replacement in tetralogy of Fallot with pulmonary regurgitation: a comparison of intervention and nonintervention. *Ann Thorac Surg* 2012;**94**:1619–26.
25. Diller GP, Kempny A, Liodakis E, Alonso-Gonzalez R, Inuzuka R, Uebing A *et al*. Left ventricular longitudinal function predicts life-threatening ventricular arrhythmia and death in adults with repaired tetralogy of Fallot. *Circulation* 2012;**125**:2440–6.
26. Moon TJ, Choueiter N, Geva T, Valente AM, Gauvreau K, Harrild DM. Relation of biventricular strain and dyssynchrony in repaired tetralogy of Fallot measured by cardiac magnetic resonance to death and sustained ventricular tachycardia. *Am J Cardiol* 2015;**115**:676–80.
27. Orwat S, Diller G-P, Kempny A, Radke R, Peters B, Kühne T *et al*. Myocardial deformation parameters predict outcome in patients with repaired tetralogy of Fallot. *Heart* 2016;**102**:209–15.
28. Arbabshirani MR, Plis S, Sui J, Calhoun VD. Single subject prediction of brain disorders in neuroimaging: promises and pitfalls. *NeuroImage* 2017;**145**:137–65.
29. Aboulhosn JA, Lluri G, Gurvitz MZ, Khairy P, Mongeon FP, Kay J *et al*. Left and right ventricular diastolic function in adults with surgically repaired tetralogy of Fallot: a multi-institutional study. *Can J Cardiol* 2013;**29**:866–72.
30. Yeboah J, Rodriguez CJ, Qureshi W, Liu S, Carr JJ, Lima JA *et al*. Prognosis of low normal left ventricular ejection fraction in an asymptomatic population-based adult cohort: the multiethnic study of atherosclerosis. *J Card Fail* 2016;**22**:763–8.
31. Toma M, Ezekowitz JA, Bakal JA, O'connor CM, Hernandez AF, Sardar MR *et al*. The relationship between left ventricular ejection fraction and mortality in patients with acute heart failure: insights from the ASCEND-HF Trial. *Eur J Heart Fail* 2014;**16**:334–41.
32. Curtis JP, Sokol SI, Wang Y, Rathore SS, Ko DT, Jadbabaie F *et al*. The association of left ventricular ejection fraction, mortality, and cause of death in stable outpatients with heart failure. *J Am Coll Cardiol* 2003;**42**:736–42.
33. Paonessa JR, Brennan T, Pimentel M, Steinhaus D, Feng M, Celi LA. Hyperdynamic left ventricular ejection fraction in the intensive care unit. *Crit Care* 2015;**19**:288.
34. Geva T, Gauvreau K, Powell AJ, Cecchin F, Rhodes J, Geva J *et al*. Randomized trial of pulmonary valve replacement with and without right ventricular remodeling surgery. *Circulation* 2010;**122**:S201.
35. Uebing A, Fischer G, Schlangen J, Apitz C, Steendijk P, Kramer H-H. Can we use the end systolic volume index to monitor intrinsic right ventricular function after repair of tetralogy of Fallot? *Int J Cardiol* 2011;**147**:52–7.
36. Gatzoulis M. A, Balaji S, Webber S. A, Siu SC, Hokanson JS, Poile C *et al*. Risk factors for arrhythmia and sudden cardiac death late after repair of tetralogy of Fallot: a multicentre study. *Lancet* 2000;**356**:975–81.
37. Guyon I, Elisseeff A. An introduction to variable and feature selection. *J Mach Learn Res* 2003;**3**:1157–82.