Computers in Biology and Medicine 141 (2022) 105147

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Identification of hemodynamic biomarkers for bicuspid aortic valve induced aortic dilation using machine learning

ARTICLE INFO

Keywords
Bicuspid aortic valve
Hemodynamic biomarker
Machine learning
Pattern recognition
Feature selection

ABSTRACT

Recent advances in medical imaging have confirmed the presence of altered hemodynamics in bicuspid aortic valve (BAV) patients. Therefore, there is a need for new hemodynamic biomarkers to refine disease monitoring and improve patient risk stratification. This research aims to analyze and extract multiple correlation patterns of hemodynamic parameters from 4D Flow MRI data and find which parameters allow an accurate classification between healthy volunteers (HV) and BAV patients with dilated and non-dilated ascending aorta using machine learning. Sixteen hemodynamic parameters were calculated in the ascending aorta (AAo) and aortic arch (AArch) at peak systole from 4D Flow MRI. We used sequential forward selection (SFS) and principal component analysis (PCA) as feature selection algorithms. Then, eleven machine-learning classifiers were implemented to separate HV and BAV patients (non- and dilated ascending aorta). Multiple correlation patterns from hemodynamic parameters were extracted using hierarchical clustering. The linear discriminant analysis and random forest are the best performing classifiers, using five hemodynamic parameters selected with SFS (velocity angle, forward velocity, vorticity, and backward velocity in AAo; and helicity density in AArch) a 96.31 $\pm$ 1.76% and 96.00 $\pm$ 0.83% accuracy, respectively. Hierarchical clustering revealed three groups of correlated features. According to this analysis, we observed that features selected by SFS have a better performance than those selected by PCA because the five selected parameters were distributed according to 3 different clusters. Based on the proposed method, we concluded that the feature selection method found five potentially hemodynamic biomarkers related to this disease.

1. Introduction

Bicuspid aortic valve (BAV) is the most common congenital cardiac defect [1] with a prevalence of 1–2% in the general population [2]. Clinical manifestations of BAV are aortic dilation, aneurysm, and dissection, which typically develop in the ascending aorta (AAo) [3,4] and often extend to the aortic arch (AArch) [5]. Current clinical management of aneurysms in BAV patients relies on quantifying aortic diameter, but its predictive capacity is limited [6]. Therefore, there is a need for new biomarkers to refine disease monitoring and improve patient risk stratification.

The most common BAV leaflet fusion phenotype involves the right-left cusps and right-non-coronary cups, with a prevalence of around 80% and 17%, respectively [3,7]. Recent studies have demonstrated that BAVs cause altered blood flow hemodynamic in the AAo, which implies increased flow asymmetry, helicity, and wall shear stress (WSS) overloads on the aortic wall [8]. The WSS abnormalities are associated with histological and proteolytic of the aortic wall damage demonstrating a role for hemodynamics in the etiology of BAV aortopathy [9–14]. Nevertheless, increased WSS is not a unique feature in BAV disease. Aortic valve stenosis can also subject the aortic wall to high WSS [15]. Furthermore, many other hemodynamic parameters beyond WSS can be used to study aortic blood flow in BAV patients, making it difficult to conclude which has more association to BAV aortopathy, especially with traditional statistical assessment.

Artificial intelligence (AI) methods have gained increased attention in cardiology and cardiovascular imaging [16]. For instance, in cardiac magnetic resonance (CMR), AI methods have been applied to segment left and right ventricles and aorta to enable automatic cardiovascular volume assessment and enhance reproducibility in clinical assessments [17–21]. Pattern recognition, i.e., the automatic discovery of regularities in data through machine learning (ML), has been recently applied to genomic data to stratify BAV patients, identify distinct patterns of aortopathy, and characterize their association with valve morphology [22–24]. The absence of clear associations between several hemodynamics parameters has limited the use of a few uncorrelated parameters to characterize BAV pathology [25].

Assessing the impact of several hemodynamic parameters in BAV dilation is difficult due to extensive data. Different methods of feature selection (FS) can be used to reduce dimensionality and redundancy of data, which allows determining which features discriminate best between two or more classes. Given the increasing capacity to extract a large amount of data from images, FS methods have become essential to achieve effective classifications. FS selection method is mainly used to provide accurate classifier models for classification tasks [26].

This study aimed to identify hemodynamic biomarkers for BAV patients and their relationships with aortic dilation. For that purpose, we analyzed and extracted multiple correlation patterns of hemodynamic parameters, finding which showed high collinearity between them, which allows us to reduce their size to few variables. And finally, we applied machine learning algorithms to discriminate between healthy volunteers (HV) and BAV patients with and without ascending aorta

https://doi.org/10.1016/j.compbiomed.2021.105147
Received 14 October 2021; Received in revised form 13 December 2021; Accepted 13 December 2021
Available online 16 December 2021

dilation.

2. Methods

2.1. Study population

Data obtained from a previous prospective study [5] was used in this work. We included sixty-seven BAV patients with a fusion of right and left coronary cusps (67.16%) or right and noncoronary cusps, AAo diameters were less or equal than 45 mm, and had no severe aortic valve disease (aortic regurgitation $\leq$ III, maximum aortic valve velocity $<$3 m/s by echocardiography). Patients were recruited at the Hospital Universitari Vall d'Hebron (Barcelona, Spain). Other inclusion criteria include age $>$18 years; without any congenital heart disease, including aortic coarctation; no connective tissue disorders; no previous aortic surgery or aortic valve replacement; and no contraindication for CMR. Forty-eight healthy volunteers (HV) matched for age, sex, and body surface area (BSA) were also included. The local ethics committee approved the study, and informed consent was obtained from all participants.

2.2. Cardiovascular Magnetic Resonance protocol

Multi-slice two-dimensional balanced steady-state free precession (b-SSFP) and 4D Flow MRI using Vastly undersampled Isotropic Projection Reconstruction (VIPR) [27,28] were acquired in a clinical GE 1.5T scanner (Signa, General Electric Healthcare, Waukesha, Wisconsin, USA). MRI datasets of the thoracic aorta were acquired with retrospective ECG cardiac gating with free breathing and without administration of an endovenous contrast agent. Acquisitions parameters were: velocity encoding of (VENC) 200 cm/s, a field of view of 400 $\times$ 400 $\times$ 400 mm^3, scan matrix of 160 $\times$ 160 $\times$ 160 (voxel size of 2.5 $\times$ 2.5 $\times$ 2.5 mm^3), flip angle of 8°, repetition time between 4.2 and 6.4 ms, and echo time between 1.9 and 3.7 ms. The data were reconstructed offline with corrections for background phase, eddy currents, and trajectory errors [29] according to each patient's nominal temporal resolution, which was ranged between 21 and 36 ms.

2.3. Aortic diameters and valve morphotype

BAV morphotype and aortic diameters were assessed using cine MR images [30]. The three aortic root cusp-to-commissure diameters were measured using double-oblique cine images at the aortic root level at end-diastole, and the maximum value was retained for the analysis. AAo diameter was measured at the level of the pulmonary artery bifurcation at the end-diastolic phase. To determine the existence of aortic root or ascending dilation, aortic diameters were adjusted with a logarithm transformation to calculate the z-score for both sinuses (zSoV) and AAo (zAAo) accounting for sex, age, and BSA as described by Campens et al. [31]. A z-core cut-off value was used to define the aortic dilation of two standard estimate errors. According to Della Corte's classification, patients were categorized concerning the aorta segment predominantly or exclusively involved in dilation [32]. Therefore, patients were classified as non-dilated (NON-DIL BAV) (zSoV $\leq$2 and zAAo $\leq$2) and dilated (DIL BAV) (zAAo $>$2 and zAAo $>$ zSoV) ascending aorta. Patients with only aortic root dilation were excluded from the study.

2.4. 3D quantification of hemodynamics parameters

A detailed description of methods used to quantify hemodynamics descriptors is described in previous publications [33–36]. We briefly explain the methods next. The quantifications were done through an in-house MATLAB toolbox (The MathWorks Inc., Natick, Massachusetts, USA) [37]. The thoracic aorta was semiautomatically segmented, and a segmentation mask was used to generate a tetrahedral mesh [38]. Afterward, we used cubic interpolation to compute the velocity vector at each mesh node. Thereafter, a finite-element least-squares projection method was used to obtain several continuous 3D maps, including eccentricity, velocity, forward velocity, backward velocity, velocity angle, regurgitation fraction, WSS, WSS axial, WSS circumferential, oscillatory shear index (OSI), vorticity, axial vorticity, axial circulation, helicity density, viscous dissipation, energy loss, and kinetic energy were generated. Finally, eight different regions were analyzed in the thoracic aorta, four for each segment in the AAo and AArch. In each region, we analyzed the mean value for each hemodynamic parameter at an averaged peak systolic data, corresponding to the average at one time-frame before peak systole, one at peak systole, and one time-frame after to reduce noise in the data, except for regurgitation fraction and oscillatory shear index (OSI). These parameters were calculated using information along the entire cardiac cycle [30].

2.5. Hemodynamic parameters analysis: machine learning algorithm

A machine learning model was designed to select hemodynamic parameters that differentiate among three classes: HV, NON-DIL BAV, and DIL BAV. The imaging process pipelines extract seventeen hemodynamic features in each of the two segments of the aorta. Then, the classifiers assigned the extracted features from 4D-flow CMR to one of these classes.

To build the classifiers, we firstly reduced the dimensionality of the data. For this purpose, hemodynamic parameters were selected using sequential forward selection (SFS) and principal component analysis (PCA). We chose five features using SFS with Fisher objective function and exhaustive search, as shown in Supplementary Table S1. We used singular value decomposition to perform PCA. PCA generates a set of new features; each being a linear transformation of the original elements. We decompose the data matrix X of $n \times p$ size, where n is the number of subjects and p the number of features, using singular value decomposition, i.e.

$$\mathbf{X} = \mathbf{USV}^{\mathrm{T}}$$

where $\mathbf{U}$ is a unitary matrix, $\mathbf{S}$ is the diagonal matrix of singular values s_i and $\mathbf{V}$ the right singular vector Therefore, we can write the covariance matrix as

$$\mathbf{C} = \frac{1}{n}\mathbf{XX}^{\mathrm{T}} = \frac{1}{(n-1)}\mathbf{VSU}^{\mathrm{T}}\mathbf{USV}^{\mathrm{T}} = \mathbf{V}\frac{\mathbf{S}^2}{(n-1)}\mathbf{V}^{T}$$

Meaning that the right singular vector $\mathbf{V}$ are principal directions and that singular values are related to the eigenvalues of covariance matrix $\mathbf{C}$, via $\lambda_i = s_i^2/(n-1)$. The principal components are defined by $\mathbf{V} = \mathbf{USV}^{\mathrm{T}}\mathbf{V} = \mathbf{US}$. We selected only the dominant eigenvectors, representing 95% of the data. Then, we took each vector's norm in the new space and leveraged scores. Finally, we obtained the indices of the vectors with the largest leverage scores (see Supplementary Table S2) [39].

After features selection, we tested different classifiers. We used the following classifiers: k-nearest neighbors (KNN) with 5, 7, 9 neighbors, linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), minimum distance, Mahalanobis distance, support vector machine (SVM) using both the linear and radial basis function kernel (RBF), neural network, and random forest. We used a neural network with multilayer perceptron architecture with one hidden layer that contains 15 nodes. In our network, the activation function in the output layer was a "softmax" unit. Of note, in our experiments, this configuration obtained the highest accuracy. We did not use individual decision trees to develop the random forest because they tend to overfit. To reduce the effects of overfitting and improve generalization, we used bootstrap-aggregated decision trees to combine the results of many decision trees. Therefore, to maximize the variance explanation of the dependent variable, a variable is selected at each spit/node. In each round of training, 1000 decision trees were generated with a maximum allowed tree depth of five.

We used stratified cross-validation to evaluate the performance of

the classification. The holdout method was used to divide the data into ten folds (90% of the data were used for training and 10% for testing) because it has become the standard method in practical terms [37]. To evaluate the stability of the classifier, we repeated this experiment ten times, interchanging training and testing data. For each time, the performance defined as the rate of samples correctly classified was computed as η_i, for $i = 1, \ldots, 10$. A confusion matrix was constructed based on prediction results in each training and validation sample, and the corresponding accuracy, precision, sensitivity, and specificity were calculated as the mean of the ten percentages of the true classifications that are tabulated in each case: $\eta = (\eta_1 + \ldots + \eta_{10})/10$. In addition, the mean ROC area under the curve (AUC) and 95% confidence interval were computed [40].

To validate both the algorithms and hemodynamic features, the performance of different pattern recognition classifiers was measured to quantify the amount of variance between the three classes of subjects. High classification accuracy shows that the proposed system of algorithms and features can be used to differentiate among the groups.

On the other hand, we used t-Distributed Stochastic Neighbor Embedding (t-SNE) to visualize high-dimensional datasets (classification with all features, SFS, and PCA features selected) [41].

The algorithms were implemented in MATLAB using Balu [42] and the Statistics and Machine Learning MATLAB Toolbox.

2.6. Hemodynamic parameters analysis: hierarchical clustering

In order to extract multiple correlation patterns from hemodynamic parameters and identify highly correlated data, we used the correlation matrix-based hierarchical clustering method. This method grants a dataset into a multilevel cluster tree or dendrogram. The distance (dissimilarity) is defined as one minus Pearson's correlation coefficient. It expressed that if two U-fiber maps had a shorter distance, they are similar, i.e., the distance is 0, the correlation coefficient is 1. The variable that quantifies an effective representation of the pattern dissimilarities in the dendrogram is the cophenetic correlation (the correlation between original and cophenetic distances). We used the average linkage method to calculate the cophenetic distance to get an average intercluster distance that allows higher cophenetic correlations. It provided the dendrogram illustrating the correlation matrix's hierarchical structure and derived the final cluster. At last, based on inconsistency coefficients, we broke down the dendrogram obtaining clusters visually differentiated by colors [43,44].

2.7. Statistical analysis

The software GraphPad Prism version 6.0.1 (GraphPad Software Inc., San Diego, California, USA) was used for statistical analysis. In population demographics, normal distribution was evaluated using the Shapiro-Wilk test. Student *t-test* and Mann-Whitney *U test* were applied to find differences between groups for continuous parameters with normal and non-normal distributions, respectively. For categorical variables we applied $\chi 2$ test. A p-value < 0.05 was considered statistically significant.

3. Results

3.1. Demographics

Demographical and clinical data are described in Table 1. HV and BAV patients were matched in terms of age, sex, and body surface area. The BAV patients presented higher diameter and Z-score than HV at both the aortic root and AAo. Seventy-three percent of BAV patients had AAo dilation. We did not find any differences other than aortic diameters, and Z scores between patients with and without aortic dilation.

Table 1
Demographical and clinical data for the healthy volunteers (HV) and BAV patients. Quantitative data are expressed as the mean ± SD. BSA, body surface area; DBP, diastolic blood pressure; EAo, aortic stenosis; SBP, systolic blood pressure; IAo, aortic insufficiency; and SoV, sinus of Valsalva. * indicates statistically significant differences ($p < 0.05$).

	HV	BAV	p-value	BAV Dilation Group		
				NON-DIL	DIL	p-value
N	48	67		18	49	
Age (year)	48.71 ± 12.57	47.74 ± 15.06	0.998	46.68 ± 14.35	48.28 ± 15.44	0.698
Sex (female: male)	23:25	31:44	0.514	5:13	24:25	0.121
Weight (kg)	70.81 ± 10.53	72.18 ± 13.25	0.587	74.72 ± 13.05	72.42 ± 13.59	0.534
Height (cm)	171.23 ± 7.82	169.45 ± 10.85	0.364	172.33 ± 8.60	168.69 ± 11.33	0.221
BSA (m^2)	1.83 ± 0.16	1.83 ± 0.21	0.995	1.88 ± 0.20	1.82 ± 0.22	0.349
IAo (%)			…			0.246
None	…	73.33		88.89	75.51	
Mild	…	10.67		5.56	12.25	
Moderate	…	16		5.56	12.24	
EAo (%)						0.359
None	…	89.55		94.44	87.76	
Mild	…	5.97		5.56	6.12	
Moderate	…	4.48			6.12	
SBP (mmHg)	130.15 ± 18.85	134.88 ± 17.55	0.114	133.78 ± 18.12	135.17 ± 17.05	0.773
DBP (mmHg)	73.41 ± 10.09	76.43 ± 8.73	0.128	77.39 ± 6.99	77.33 ± 9.03	0.981
Diameter SoV (mm)	30.32 ± 3.9	35.91 ± 4.69	<0.001*	33.28 ± 3.74	36.08 ± 4.41	0.019*
Diameter AAo (mm)	27.89 ± 3.71	39.37 ± 6.74	<0.001*	32.72 ± 4.17	42.82 ± 5.19	<0.001*
Z score SoV	−0.25 ± 1.18	1.30 ± 1.30	<0.001*	0.27 ± 0.82	1.33 ± 1.05	<0.001*
Z score AAo	−0.14 ± 0.91	2.89 ± 1.52	<0.001*	0.98 ± 1.07	3.71 ± 0.98	<0.001*

3.2. Hemodynamic biomarkers selection with SFS and PCA

From the seventeen hemodynamic features obtained, the five features that best differentiated the three classes (HV, NON-DIL BAV, and DIL BAV) were selected using both SFS and PCA by eliminating highly correlated or constant features that maximized accuracy (see Supplementary Table S1).

Regarding SFS, Fig. 1 a. shows the maximum separability obtained by each of the seventeen hemodynamic features as assessed in both aortic regions. The five variables presenting with the lowest separability and resulting in the higher accuracy once used in the different classifiers (Supplementary Fig. 1) were retained. They correspond to velocity angle, forward velocity, vorticity, backward velocity in AAo, and helicity density in AArch. Fig. 1 b. shows the 3D feature space obtained using three features (velocity angle in AAo, forward velocity in AAo, and helicity density in AArch) for visualization purposes, showing good separability among classes. Fig. 2 a. shows the two principal components of PCA. Dimension 1 explains 45.21% variation in the data while Dimension 2 explains 23.73% variation. Together they explain 68.94% of the variation.

Regarding PCA, Fig. 2 b. shows PCA correlation circle that represented the quality of representation of the features on a factor map. The better its representation on the factor map, is the variable closer to the circle's center. This means that variables located closer than to the center of the plot are less important. Finally, Fig. 2 c. shows a bar graph of the quality of representation of the variables on factor maps on all the dimensions. The five top-performing features were forward velocity,

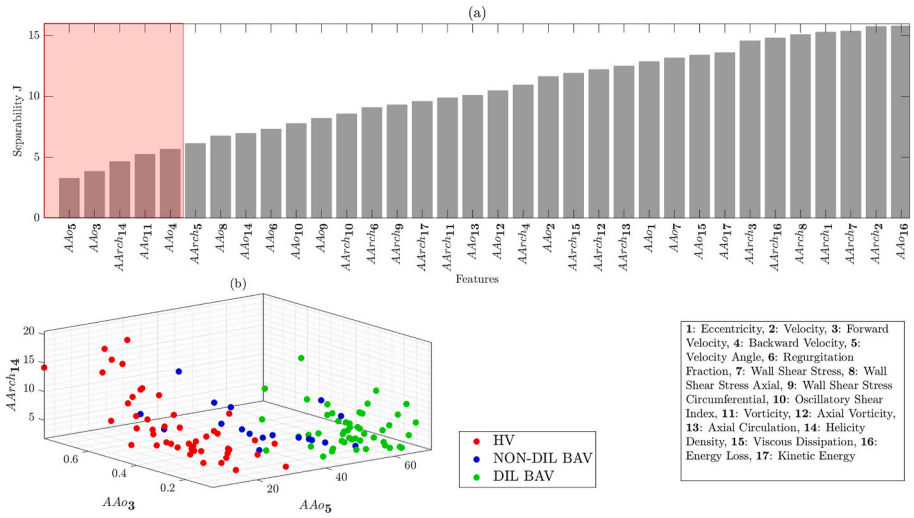

Fig. 1. (a): Feature selection using sequential forward selection (SFS). There are five selected features, and they correspond to velocity angle in AAo, forward velocity in AAo, helicity density in AArch, vorticity in AAo, and backward velocity in AAo (red rectangle). (b): Feature Space in 3D.

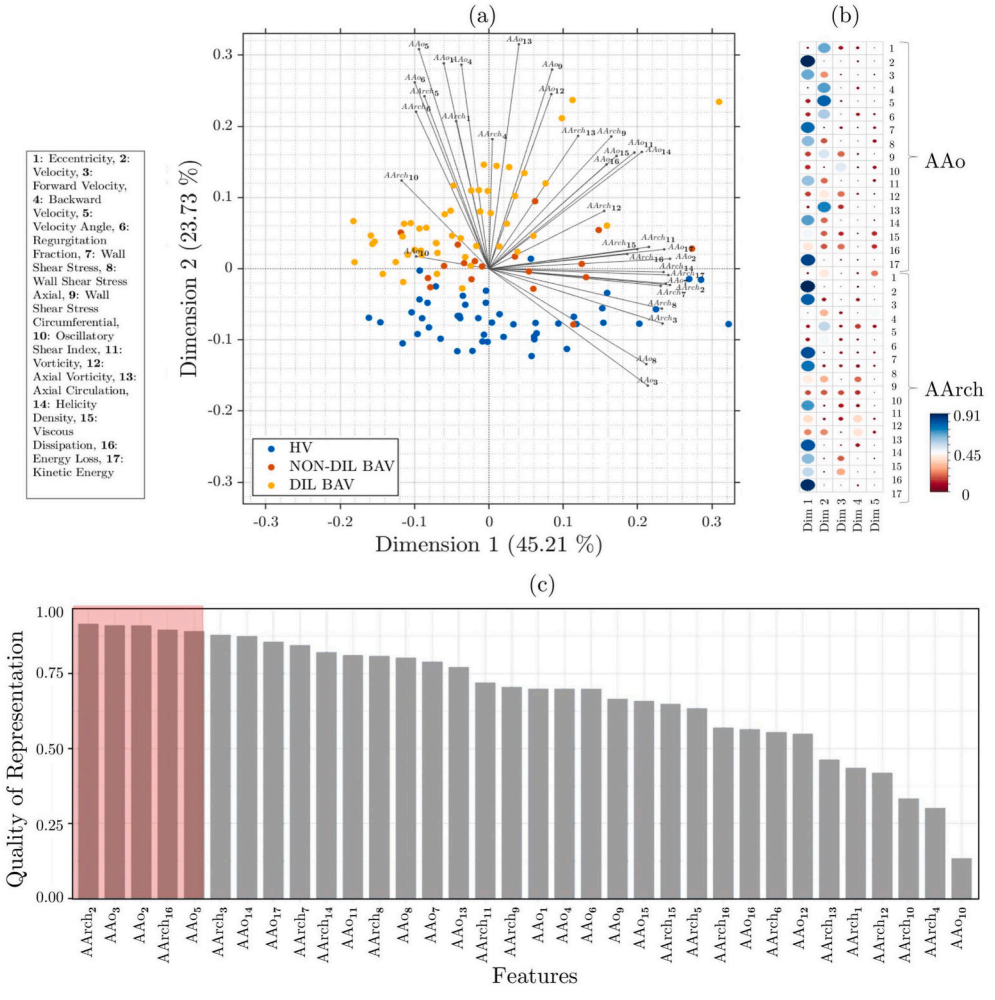

Fig. 2. (a) The two principal components PCA. The distance between variables and the origin measures the quality of the variables on the factor map. Variables that are away from the origin are well represented on the factor map. (b) PCA correlation circle that represented the quality of representation of the features on factor map. The better its representation on the factor map, is the variable closer to the circle's center. This means that variables located closer than to the center of the plot are less important (c) Bar graph of quality of representation. There are five selected features, and they correspond to velocity in AArch, forward velocity in AAo, velocity in AAo, energy loss in AArch, and velocity angle in AAo (red rectangle).

velocity, velocity angle in AAo, and velocity and energy loss in AArch.

The computational time of the feature selection is short (approximate 0.05s) because we are dealing with a small number of features.

3.3. Classification results

Both simple (e.g., minimum distance and linear discriminant analysis) and more complex (e.g., SVM and neural networks) classifiers were tested using as input either all features or the five selected by SFS or

PCA.

First, we used t-SNE as a tool to visualize high-dimensional data, as showing in Fig. 3. Fig. 3 a. shows t-SNE with 34 features (17 parameters in each of the two segments). HV (red) and DIL BAV (blue) groups are separated, but few HV are located close to the DIL BAV group. Nevertheless, the NON-DIL BAV group is not clearly separated from the DIL BAV group, and two NON-DIL BAV subjects are grouped in the HV class. Fig. 3 b. shows t-SNE with five selected features from SFS. Three groups are visualized; still, a few NON-DIL BAV subjects are grouped into HV and DIL BAV groups, respectively. Finally, Fig. 3 c. shows t-SNE with five selected features from PCA. Their behavior is similar to SFS's figure, but the distance in the three groups is lower, and more NON-DIL BAV subjects can be classified as DIL BAV compared to SFS results.

The best result was obtained by combining the five features selected by SFS in LDA, getting a 96.31 ± 1.76% classification accuracy on HV, NON-DIL, and DIL BAV datasets. Other classifiers, as KNN and SVM-Linear, resulted in an accuracy of over 86% and 91.34% using SFS (Supplementary Table 2). Using PCA as a feature selection, almost all classifiers were close to 90% accuracy.

The second-best result was obtained by combining the five hemodynamic features selected by SFS and random forest, with a 96.00 ± 0.83% accuracy. Actually, there were not statistical differences between random forest and the LDA in precision, sensitivity, and specificity in DIL BAV and HV classes, but in NON-DIL BAV class (Table 2). The LDA had a better overall performance in the NON-DIL BAV class. However, it showed a larger variance than random forest for precision, sensitivity, and specificity in the cross-validation experiment. A diagram is showing in Supplementary Fig. 1. Each node contains the feature ID and threshold used for splitting. The position of some features, e.g., the relative distance from the root, in the random forest reflects the strength of association between diameter and hemodynamic parameters and BAV disease. For example, backward velocity in AAo is the optimal splitting feature. The optimal splitting feature found for the subsets is forward and backward velocity in AAo in the second layer. We can measure the association between aortic dilation and hemodynamic parameters in BAV disease by summarizing each feature's overall random forest based on these ranks. Moreover, we computed the predictor importance estimates from the random forest that grows trees using all variables extracted, as showed in Fig. 4. Bar graph stores the increase in mean square error (MSE) averaged over all trees in ensemble and divided by the standard deviation taken over the trees for each feature. The bars with the highest values contain the information of the most important features. This suggests that velocity angle in AAo is the most important predictor, followed by backward velocity, eccentricity, axial circulation, and regurgitation in AAo. The average processing time for feature selection and classification was 42s in a 2.3 GHz Intel i7 processor equipped with 8GB of RAM.

Fig. 5 shows ROC curves for both combinations with the best performance (LDA and random forest) using five features selected by SFS. We noted that the ROC curves for HV and DIL BAV classes are high similar, indicating that the methods can distinguish these classes. Nevertheless, the NON-DIL BAV class has lower results, and this class's imbalance may jeopardize the results. These classification methods achieved the best ROC AUC, sensitivity, specificity, precision across training and validation samples, and stratified cross-validations (Table 2).

Additionally, we performed another experiment to classify only two classes (NON-DIL BAV and DIL BAV groups). We applied the methodology previously described and used hemodynamic features in AAo and AArch. Feature selection algorithm SFS found their five-top performing features were: velocity angle in AAo, regurgitation fraction in AArch, eccentricity in AAo, backward velocity in AAo, and oscillatory shear index in AAo. PCA's five best-performing features were: velocity in AArch, forward velocity in AArch, velocity in AAo, kinetic energy in AArch, and forward velocity in AAo. Using features selected by PCA, almost all classifiers get close to 86% accuracy. The best results were obtained by combining SFS-selected features using an LDA classifier with 96.18 ± 2.34% (see Supplementary Table S3).

3.4. Hemodynamic parameters correlation

Fig. 6 a. shows the Pearson correlation matrix among all hemodynamic parameters for all regions and subjects. Supplementary Fig. 2 shows p-values obtained by the linear regression between all hemodynamic parameters. Several parameters show good correlations (e.g., eccentricity and WSS in AAo), which indicate that some hemodynamic parameters are highly correlated and can be divided into clusters.

A total of three clusters were identified as shown in Fig. 6 b. All three clusters combine at a much higher dendrogram distance and can be treated as individual groups for analysis. Cluster 1 (green): OSI, regurgitation fraction, velocity angle, and eccentricity in all regions; and backward velocity in AAo. Cluster 2 (red): axial circulation, WSS

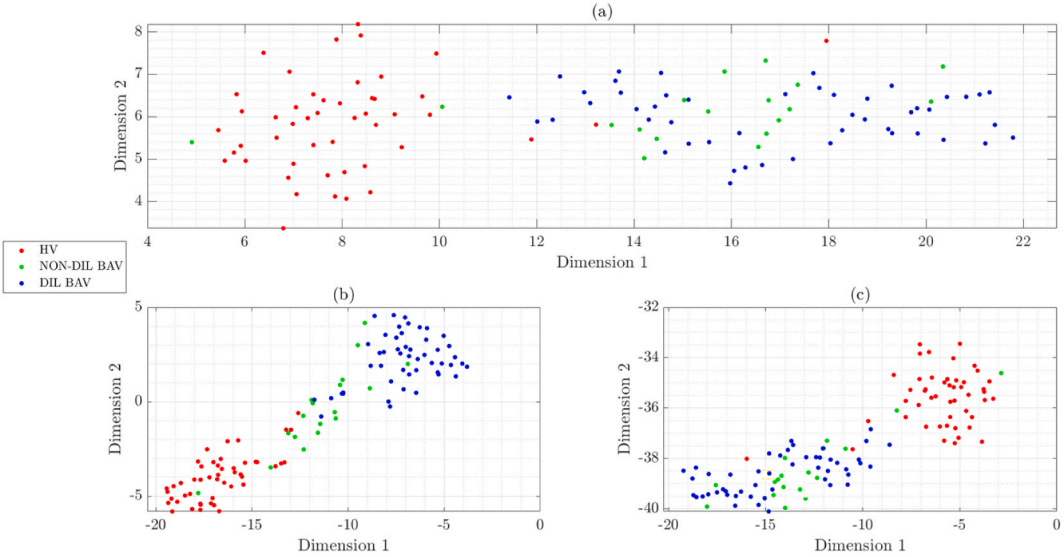

Fig. 3. t-SNE: t-Distributed Stochastic Neighbor Embedding. (a) All features, five selected features by: (b) SFS, and (c) PCA. SFS and PCA results show a good separation of the groups. But PCA results show a lower distance between groups, and more NON-DIL BAV subjects can be classified as DIL BAV compared to SFS results.

Table 2
Accuracy, precision, specificity, and sensitivity of different combinations of classifiers and all features, and five features selected by SFS and PCA. Each experiment was done using 10-fold cross-validation and repeated 10 times with confidence interval 95%. Bold type means statistically significant between the LDA and random forest for all features, SFS, and PCA, respectively (p-value < 0.05).

		LDA			Random Forest		
		All features	SFS	PCA	All features	SFS	PCA
HV class	**Precision (%)**	100.00 ± 0.00	100.00 ± 0.00	100 .00 ± 0.00	98.01 ± 1.19	99.49 ± 1.12	99.09 ± 1.30
	Specificity (%)	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	99.05 ± 0.81	99.45 ± 0.84	100.00 ± 0.93
	Sensitivity (%)	94.17 ± 2.01	97.49 ± 6.51	97.48 ± 7.44	95.39 ± 1.51	94.22 ± 1.30	97.13 ± 1.62
NON-DIL BAV class	**Precision**	**69.38 ± 48.42**	**88.42 ± 32.40**	**63.49 ± 4.80**	**67.44 ± 9.45**	**78.02 ± 4.16**	**79.83 ± 9.80**
	Specificity (%)	**96.49 ± 5.51**	**98.41 ± 4.15**	**95.33 ± 7.51**	**94.32 ± 1.53**	**96.39 ± 0.65**	**96.43 ± 1.70**
	Sensitivity (%)	**80.36 ± 37.39**	**86.03 ± 22.05**	**78.42 ± 4.29**	**80.39 ± 6.41**	**99.40 ± 2.71**	**95.02 ± 9.57**
DIL BAV class	**Precision**	92.48 ± 7.02	93.01 ± 1.38	94.01 ± 1.00	95.50 ± 3.22	100.00 ± 0.75	99.03 ± 2.84
	Specificity (%)	95.32 ± 9.80	96.44 ± 7.12	96.43 ± 7.71	97.09 ± 2.19	100.00 ± 0.55	99.44 ± 2.09
	Sensitivity (%)	98.05 ± 6.33	99.59 ± 2.60	93.29 ± 1.10	93.44 ± 2.70	97.04 ± 1.32	95.32 ± 2.70
Accuracy (%)		93.86 ± 2.24	96.31 ± 1.76	91.05 ± 2.29	92.00 ± 1.80	96.00 ± 0.83	96.00 ± 2.70

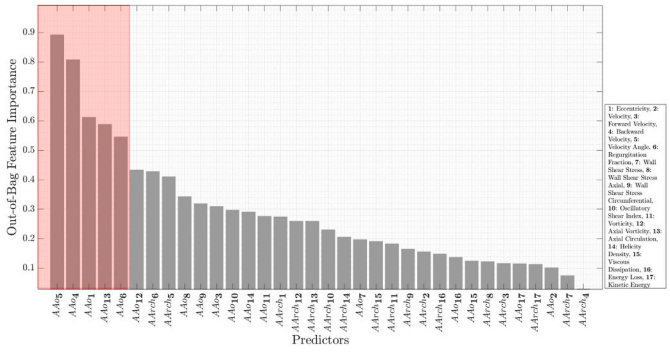

Fig. 4. Predictors' importance estimation from random forest. The five top-performing features were: velocity angle, backward velocity, eccentricity, axial circulation, and regurgitation fraction in AAo (red rectangle).

circumferential, and axial vorticity in all regions; backward velocity in AArch; and helicity density, vorticity, viscous dissipation, and energy loss in AAo. Finally, cluster 3 (blue): kinetic energy, velocity, WSS, WSS axial, and forward velocity in all regions; and energy loss, viscous dissipation, vorticity, and helicity density in AArch. This analysis shows high collinearity between the variables, which would allow us to reduce their size to a few variables.

According to hierarchical cluster analysis, we determined the clustering corresponding to each feature selected by SFS and PCA. SFS: cluster 1(green): velocity angle and backward velocity in AAo, cluster 2 (red): vorticity in AAo; and cluster 3 (blue): forward velocity and helicity density in AAo and Aarch, respectively. PCA: cluster 1(green): velocity angle in AAo, cluster 2 (red): velocity in AAo, and cluster 3 (blue): forward velocity in AAo, energy loss, and velocity in AArch. Hence, we can assume that features by SFS have a better performance than those selected by PCA because it has a wide representation of each clustering.

Since most parameters were selected from the AAo, we repeated the methodology previously described, by selecting only three features from SFS and PCA in the AAo, by eliminating highly correlated or constant features that maximized accuracy. Each classification experiment is shown in Supplementary Tables S4 and S5. The best result was obtained with the QDA classifier using three features selected by SFS: cluster 1 (green): velocity angle, cluster 2 (red): vorticity, and cluster 3 (blue): forward velocity, achieving an average of 94.90 ± 2.05% classification accuracy.

4. Discussion

Using ML, we have devised a differentiation algorithm for BAV with aortic dilation based on hemodynamic parameters derived from 4D flow CMR. After comparing multiple ML methods, the results showed that the accuracy gained with feature selection vs. all features in the final classifiers used is not that considerable [9–15]. Nevertheless, considering the large number of flow descriptors proposed to classify BAV patients with aortic dilation., the use of feature selection algorithms allows for the reduction of the number of input variables used to develop a predictive model without losing accuracy. Therefore, we found that combining five hemodynamic features selected by SFS and applying them to the LDA classification algorithm achieves the best performance

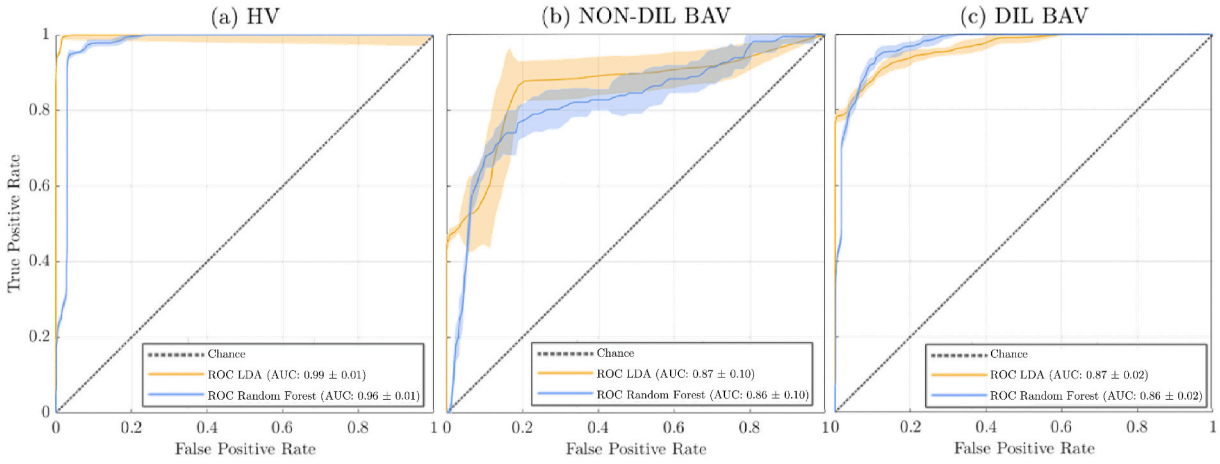

Fig. 5. The ROC – curve for LDA and random forest with five features selected by SFS, displaying for the three classes in the mean ± SD. The NON-DIL BAV class has lower results in both classifiers, and this class's imbalance may jeopardize the results.

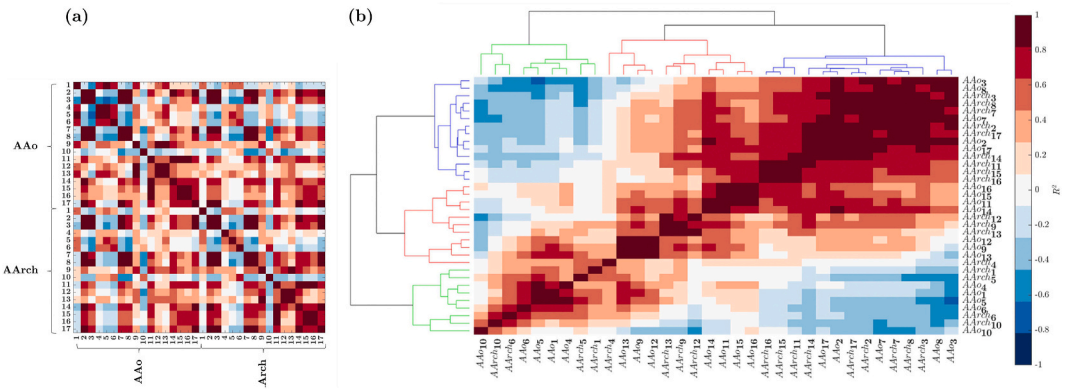

Fig. 6. (a) Correlation matrix obtained by the linear regression, (b) dendrogram and hierarchical clustering result based on average linkage method, for all hemodynamic parameters of HV and BAV patients, in AAo and AArch regions.

with an accuracy of 96.31 ± 1.76%, which is higher than the accuracy of random forest (96.00 ± 2.70%).

Both classification tasks with LDA and random forest showed better performance when including AAo and AArch than only features in AAo (SFS reaching 93.27 ± 2.34% accuracy and random forest resulting in 94.00 ± 2.00% accuracy). This result suggests that for the classification it is important to include parameters in AAo and AArch. Further, both classifiers did not show significant differences, but in the NON-DIL BAV class. For this class, the LDA was sensitive to the selected test data as showed large variance for precision, sensitivity, and specificity in the cross-validation experiment.

Feature selection algorithm SFS found five-top performing features including: velocity angle, forward velocity, vorticity, backward velocity in the AAo, and helicity density in the AArch. PCA's five best-performing features were: velocity angle, forward velocity and velocity, in the AAo, and velocity and energy loss in the AArch. Interestingly, the most important parameters found by Random Forest were velocity angle, backward velocity, eccentricity, axial circulation, and regurgitation all of them localized in the AAo. Thus, algorithms consistently identify velocity angle as key descriptors of BAV hemodynamics, a result in line with previous research, and most of them highlighted the importance of forward and backward velocity components and the role of rotational flow descriptors, such as helicity, circulation and vorticity [11,45].

Notably, the algorithms did not select WSS or its components, all previously related cross-sectional and longitudinal data with dilation in BAV. This may have resulted from averaging these biomarkers over aortic wall regions or reflect their relatively lower reproducibility than bulk flow measures. Furthermore, the present WSS assessment may be limited in evaluating the spatiotemporal complexity of this biomarker [46]. Alternatively, the Eulerian method to analyze WSS topological skeleton by identifying and classifying WSS fixed points and manifolds in complex vascular geometries can increase the chance of finding mechanistic explanations to clinical observations as presented by Mazzi et al. [47], analysis that may be added in the ML classification algorithms in future works.

The structure and information of bootstrap-aggregated decision trees were extracted to count and analyze the extent of the influence of various hemodynamic parameters on BAV dilation to determine the parameters most closely related to the dilation of the aorta in this disease. This suggests that velocity angle in AAo is the most important predictor, followed by backward velocity, eccentricity, axial circulation, and regurgitation in AAo. These variables align with those identified in previous studies, which related high asymmetrical shear stresses with aortic dilation in BAV disease [48]. However, a decision tree further allows for the identification of the relative importance of each flow descriptor in the classification task, showing how velocity angle and flow eccentricity, two descriptors of asymmetric flow, backward velocity and axial circulation, and regurgitation bringing information of flow rotation are the dominant factors. The proposed decision tree model could differentiate the three classes with 96.00 ± 0.83 accuracy using five features selected by SFS. Nevertheless, this decision tree was our second-best result and appeared to be most helpful in determining HV and DIL BAV classes. Instead of the LDA that ascertains better the three classes, including the NON-DIL BAV class.

Aortic stenosis in BAV patients has been reported that increased with the patient's age [49,50]. Therefore, the patient's age can be a possible confounding factor in the classifiers. We executed another experiment, including age as an input parameter. However, features selection algorithms, SFS and PCA, did not find age as their top-best performing features (see Supplementary Table S6).

In this study, the hierarchical clustering method provided an alternative for reliable correlation between hemodynamic parameters from 4D-flow CMR. By classifying them into three different clusters according to their similarities, the resulting dendrogram provides a good representation of the relationship of various parameters in two aorta regions. According to hierarchical cluster analysis, we can assume that features by SFS have a better performance than those selected by other feature selection algorithms because it has a wide representation of each clustering: cluster 1(green) velocity angle and backward velocity in AAo, cluster 2 (red) vorticity in AAo; and cluster 3 (blue): forward velocity and helicity density in AAo and AArch, respectively. When statistical modeling is used to pursue a predictive aim, Gregorich et al. showed that two highly correlated independent variables will lead to high variance in the predictions, even if both variables are relevant for prediction. In small samples, it may then be beneficial to omit one of the pair to decrease that variance, even if this incurs some new bias in the predictions [51]. Further, O'Brien shows that multicollinearity is not a sufficient reason to eliminate variables from a model. A more important criterion to consider when contemplating dropping a variable from model is 'model influence' [52]. Although, we studied the correlation and clustering of the features, this information was not used to intervene in the ML model since the latter selects the features automatically. Instead, we used the cluster information to explain the relation of the features selected by ML and the localization of these parameter across the different clusters.

One of the strengths of our study is that it provides a comprehensive overview of the relative performance of different ML algorithms for BAV aortopathy classification. These results can be used to guide researchers in the selection of an appropriate ML algorithm for their studies. Hence, non-linear interactions can be associated with the selected features that better identify HV and BAV patients.

4.1. Limitations

Considering the small number of subject data, we did not explore the use of advanced deep learning algorithms. Instead, we used classical ML

algorithms such as random forest and SVM. However, with both methods, we achieved a high classification accuracy. Deep networks require extremely large datasets to achieve high performance. In future studies, we will include more data in our dataset to perform advanced deep learning methods and compared them with classical ML algorithms results. Another limitation of the current study is the small number of NON-DIL BAV types, which unbalanced the analyzed classes. However, the cross-validation assessment aimed to reduce the effect of this issue in the classification output.

Additionally, from the acquisition point of view, the movement of the aorta along the cardiac cycle was not considered in this study since technical limitations in 4D-flow CMR acquisitions, as poor contrast and low signal-to-noise ratio, make it challenging to obtain a time-resolved segmentation of the aorta.

Further, in this study, we averaged parameters along the circumference of each region, which can induce a sub-estimation of local values. Nevertheless, assessing the local impact of all hemodynamic parameters for classification would require more patients from a statistical point of view, as there would be more parameters than subjects.

Finally, another limitation is the absence of longitudinal data. Performing a similar study would elucidate if the parameters that best classify BAV patients with and without aortic dilation would also be the best predictor for aortic dilation in those patients. A paper with longitudinal outcomes was recently published; however, only WSS was assessed as a predictor for dilation in that study [15].

5. Conclusions

The main contributions of the paper are twofold. On the one hand, we analyzed and extracted multiple correlation patterns of hemodynamic parameters, finding which parameters showed high collinearity between them, which allows us to diminish their size to a few variables. Also, we defined five hemodynamic features that best classify HV and BAV with and without aortic dilation using SFS: velocity angle, forward velocity, vorticity, and backward velocity in AAo, and helicity density in AArch. The best-performing methods were with features selected by SFS in LDA and random forest classifiers with 96.31 $\pm$ 1.76% and 96.00 $\pm$ 0.83%, respectively. Moreover, we found five features by SFS: velocity angle, eccentricity, backward velocity, and oscillatory shear index in AAo, and regurgitation fraction in AArch, that best classified BAV patients' groups (NON-DIL BAV and DIL BAV classes) using LDA classifier with 96.18 $\pm$ 2.34% accuracy.

Funding

This work has been funded by projects PIA-ACT192064, the Millennium Nucleus on Cardiovascular Magnetic Resonance NCN17_129, and ICN2021_004 of the Millennium Science Initiative of the National Agency for Research and Development, ANID. The authors also thanks to Fondecyt project 1181057 also by ANID. Franco P. thanks to ANID – PCHA/Doctorado-Nacional/2018–21180391. Sotelo J. thanks to CONICYT - FONDECYT Postdoctorado 2017 #3170737 and ANID - FONDECYT de Iniciación en Investigación #11200481. Guala A. has received funding from Spanish Ministry of Science, Innovation and Universities (IJC2018-037349-I).

Author contributions

All authors were actively involved in reviewing and drafting the manuscript. All authors have approved the final version of this manuscript. Mainly, P.F. processed the MRI dataset, developed the study's design, performed the statistical analysis, and wrote the manuscript. J.S. developed the finite element method to evaluate the hemodynamic parameters and edited the manuscript. A.G. managed the acquisition of the CMR studies and edited the manuscript. L.DS. participated in the design of the study to recruit patients and edited the manuscript. A.E. participated in the design of the study to recruit patients and edited the manuscript. J.RP. participated in the design of the study to recruit patients and edited the manuscript. D.M. participated in the study's design, developed the Balu Toolbox, and edited the manuscript. R.S. participated in the study's design and edited the manuscript. S.U. participated in the study's design and edited the manuscript.

Summary

Background: Recent advances in medical imaging have demonstrated the existence of altered hemodynamics in bicuspid aortic valve (BAV) patients. Therefore, there is a need for new hemodynamic biomarkers to refine disease monitoring and improve patient risk stratification.

Objective: This research aims to analyze and extract multiple correlation patterns of hemodynamic parameters from 4D Flow MRI data and find which parameters allow an accurate classification between healthy volunteers (HV) and BAV patients with dilated and non-dilated ascending aorta using machine learning.

Methods: 4D flow MRI data of 48 healthy volunteers (HV) and 67 BAV (73% with AAo dilation) patients were acquired in a 1.5T GE-MR Signa Scanner and using the VIPR technique. We obtained the velocity gradient from 4D flow data using a finite-element least-squares projection. Sixteen hemodynamic parameters (e.g., velocity, wall shear stress, vorticity) were calculated in the ascending aorta (AAo) and aortic arch (AArch) at peak systole. Then, a machine learning model was designed to select hemodynamic parameters that adequately separate HV and BAV patients (non- and dilated ascending aorta). A sequential forward selection (SFS) and principal component analysis (PCA) were used to reduce the dimensionality of the data as input for the classifiers. We used the following classifiers: minimum distance, linear discriminant analysis (LDA), k-nearest neighbors (KNN) with 5, 7, 9 neighbors, quadratic discriminant analysis, Mahalanobis distance, support vector machine in both its linear and radial basis function kernel, neural network, and random forest. The performance of the classifiers was evaluated using cross-validation with ten folds and repeated ten times for each classification. Additionally, the Pearson correlation method was used to calculate the correlation matrix between all the hemodynamic parameters. Hierarchical clustering was then applied to classify its rows/columns into different groups.

Results: The LDA and random forest are the best performing classifiers, using five hemodynamic parameters selected with SFS (velocity angle, forward velocity, vorticity, and backward velocity in AAo; and helicity density in AArch) a 96.31 $\pm$ 1.76% and 96.00 $\pm$ 0.83% accuracy, respectively. Hierarchical clustering revealed three groups of correlated features. On the other hand, hierarchical clustering revealed three groups of correlated features. According to this analysis, we observed that features selected by SFS have a better performance than those selected by PCA because the five selected parameters were distributed according to 3 different clusters: cluster 1: velocity angle and backward velocity in AAo, cluster 2: vorticity in AAo, and cluster 3: forward velocity and helicity density in AAo and AArch, respectively.

Conclusions: We present an algorithm based on machine learning that characterizes HV from BAV patients with and without aortic dilation using hemodynamic parameters from 4D-flow MRI. Based on the proposed method's performance, it can be concluded that the feature selection method found five potentially hemodynamic biomarkers related to this disease.

Declaration of competing interest

The authors declare that they have no conflict of interest.

Acknowledgements

Millennium Science Initiative of the Ministry of Economy,

Development and Tourism, grant Nucleus for Cardiovascular Magnetic Resonance and ICN2021_004. We are also grateful to Biomedical Imaging Center at Pontificia Universidad Católica de Chile and Hospital Universitari Vall d'Hebron for support this research.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2021.105147.

References

[1] J.W. Kang, H.G. Song, D.H. Yang, S. Baek, D.H. Kim, J.M. Song, et al., Association between bicuspid aortic valve phenotype and patterns of valvular dysfunction and bicuspid aortopathy: comprehensive evaluation using MDCT and echocardiography, JACC Cardiovasc Imaging 6 (2) (2013) 150–161.

[2] S.C. Siu, C.K. Silversides, Bicuspid aortic valve disease, J. Am. Coll. Cardiol. 22;55 (25) (2010) 2789–2800, https://doi.org/10.1016/j.jacc.2009.12.068.

[3] A. Evangelista, P. Gallego, Calvo-Iglesias, et al., Anatomical and clinical predictors of valve dysfunction and aortic dilation in bicuspid aortic valve disease, Heart 104 (7) (2018) 566–573, https://doi.org/10.1136/heartjnl-2017-311560.

[4] E. Girdauskas, J. Petersen, N. Neumann, et al., Novel approaches for BAV aortopathy predition – is there a need for cohort studies and biomrakers? Biomolecules 8 (3) (2018) 58, https://doi.org/10.3390/biom8030058.

[5] L. Dux-Santoy, A. Guala, G. Teixido-Tura, A. Ruiz-Muñoz, G. Maldonado, N. Villalva, et al., Increased rotational flow in the proximal aortic arch is associated with its dilation in bicuspid aortic valve disease, Eur Heatr J - Cardiovasc Imaging 20 (12) (2019) 1407–1417, https://doi.org/10.1093/ehjci/jez046.

[6] L.A. Pape, T.T. Tsai, E.M. Isselbacher, J.K. Oh, P.T.O. Gara, A. Evangelista, et al., Aortic diameter > 5.5 cm is not a good predictor of type A aortic dissection observations from the international registry of acute aortic dissection (IRAD), Circulation 116 (2007) 1120–1127, https://doi.org/10.1161/CIRCULATIONAHA.107.702720.

[7] B.M. Schaefer, M.B. Lewin, K.K. Stout, E. Gill, et al., The bicuspid aortic valve: an integrated phenotypic classification of leaflet morphology and aortic root shape, Heat 94 (2008) 1634–1638.

[8] J. Liu, J.A. Shar, P. Sucosky, Wall shear stress directional abnormalities in BAV aortas: toward a new hemodynamic predictor of aortopathy? Front. Physiol. 14 (9) (2018) 993.

[9] A. Pasipoularides, Clinical-pathological correlations of BAV and the attendant thoracic aortopathies. Part 1: pluridisciplinary perspective on their hemodynamics and morphomechanics, J. Mol. Cell. Cardiol. 133 (2019) 223–232, https://doi.org/10.1016/j.yjmcc.2019.05.017.

[10] B.D. Allen, P. van Ooij, A.J. Barker, et al., Thoracic aorta 3D hemodynamics in pediatric and young adult patients with bicuspid aortic valve, J Magn Reason Imaging 42 (4) (2015) 954–963, https://doi.org/10.1002/jmri.24847.

[11] M.M. Bissell, A.T. Hess, L. Biasiolli, et al., Aortic dilation in bicuspid aortic valve disease: flow pattern is a major contributor and differs with valve fusion type, *Circ Cardiovasc* Imaging 6 (2013) 499–507, https://doi.org/10.1161/CIRCIMAGING.113.000528.

[12] S. Atkins, P. Sucosky, Etiology of bicuspid aortic valve disease: focus on hemodynamics, World J. Cardiol. 6 (12) (2014 Dec 26) 1227–1233, https://doi.org/10.4330/wjc.v6.i12.1227.

[13] G. Soulat, M.B. Scott, B.D. Allen, et al., Association of regional wall shear stress and progressive ascending aorta dilation in bicuspid aortic valve, S1936-878X, JACC Cardiovasc Imaging 21 (2021 Aug 11), https://doi.org/10.1016/j.jcmg.2021.06.020, 00510-6.

[14] Guala A, Dux-Santoy L, Teixido-Tura G, et al., Wall Shear Stress Predicts Aortic Dilation in Patients with Bicuspid Aortic Valve. *In press* JACC cardiovascular imaging.

[15] K. Boudoulas, B. Wolfe, S. Lilly, et al., The aortic stenosis complex: aortic valve, atherosclerosis, aortopathy, J. Cardiol. 65 (2015) 377–382, https://doi.org/10.1016/j.jjcc.2014.12.021.

[16] D. Dey, P. Slomka, P. Leeson, et al., Artificial intelligence in cardiovascular imaging: JACC state-of-the-art review, J. Am. Coll. Cardiol. 73 (11) (2019) 1317–1335, https://doi.org/10.1016/j.jacc.2018.12.054.

[17] M. Avendi, A. Kheradvar, Jafarkhani, Automatic segmentation of the right ventricle from cardiac MRI using a learning-based approacg, Magn. Reson. Med. 78 (6) (2017 Dec) 2439–2448, https://doi.org/10.1002/mrm.26631.

[18] L. Tan, R. McLaughlin, E. Lim, et al., Fully automated segmentation of the left ventricle in cine cardiac MRI using neural network regression, J. Magn. Reson. Imag. 48 (1) (2018 Jul) 140–152, https://doi.org/10.1002/jmri.25932.

[19] Q. Tao, W. Yan, Y. Wang, et al., Deep learning-based method for fully automatic quantification of left ventricle function from cine MR images: a multivendor, multicenter study, Radiology 290 (1) (2019) 81–88.

[20] M. Campello, P. Gkontra, C. Izquierdo, C. Mart, A. Sojoudi, P.M. Full, et al., Multicentre, multi-vendor and multi-disease cardiac segmentation: the M&ms challenge, IEEE Trans. Med. Imag. (2021) 1–12, https://doi.org/10.1109/TMI.2021.3090082.

[21] J. Aviles, G.D. Maso Talou, O. Camara, M.C. Marcos, X. Morales Ferez, D. Romero, et al., Domain adaptation for automatic aorta segmentation of 4D flow magnetic resonance imaging data from multiple, Functional Imaging and Modeling of the Heart. FIMH 2021. Lecture Notes in Computer Science 12738 (2021) 112–121.

[22] C. Wojnarski, E. Roselli, J. Idrees, et al., Machine-learning phenotypic classification of bicuspid aortopathy, J. Thorac. Cardiovasc. Surg. 155 (2) (2018 Feb) 461–469, https://doi.org/10.1016/j.jtcvs.2017.08.123, e4.

[23] B. Ambale-Venkatesh, X. Yang, C. O Wu, et al., Cardiovascular event prediction by machine learning: the multi-ethnic study of atherosclerosis, Circ. Res. 121 (9) (2017 Oct 13) 1092–1101, https://doi.org/10.1161/CIRCRESAHA.117.311312.

[24] E. Cantor, R. Salas, H. Rosas, et al., Biological knowledge-slanted random forest approach for the classification of calcified aortic valve stenosis, BioData Min. 14 (1) (2021 Jul 23) 35, https://doi.org/10.1186/s13040-021-00269-4.

[25] E. Girdauskas, M. Borger, M. Secknus, et al., Is aortopathy in bicuspid aortic valve disease a congenital defect or a result of abnormal hemodynamics? A critical reappraisal of a one-sided argument, Eur. J. Cardio. Thorac. Surg. 39 (2011) 809–814.

[26] D. Mery, Computer Vision for X-Ray Testing: Imaging, Systems, Image Databases, and Algorithms, first ed., Springer, 2015.

[27] T. Gu, F. Korosec, W. Bloch, et al., PC VIPR: a high-speed 3D phase-contrast method for flow quantification and high-resolution angiography-, AJNR Am J Neuroradiol 26 (4) (2005) 743–749.

[28] K. Johnson, M. Markl, Improved SNR in phase contrast velocimetry with five-point balanced flow encoding, Mgn Reson Med 63 (2010) 349–355, https://doi.org/10.1002/mrm.22202.

[29] K. Johnson, D. Lum, P. Turski, et al., Improved 3D phase contrast MRI with off-resonance corrected dual Echo VIPR, Magn. Reson. Imaging 60 (6) (2012) 1329–1336, https://doi.org/10.1002/mrm.21763.Improved.

[30] J. Rodríguez-Palomares, L. Dux-Santoy, A. Guala, et al., Aortic Flow patterns and wall shear stress maps by 4D-flow cardiovascular magnetic resonance in the assessment of aortic dilatation in bicuspid aortic valve disease, J. Cardiovasc. Magn. Reson. 20 (1) (2018 Apr 26) 28, https://doi.org/10.1186/s12968-018-0451-1.

[31] L. Campens, L. Demulier, K. De Groote, et al., Reference values for echocardiographic assessment of the diameter of the aortic root and ascending aorta spanning all age categories, Am. J. Cardiol. 114 (6) (2014) 914–920.

[32] A. Della Corte, C. Bancone, G. Dialetto, et al., The ascending aorta with bicuspid aortic valve: a phenotypic classification with potential prognostic significance, Eur. J. Cardio. Thorac. Surg. 46 (2) (2014) 240–247.

[33] J. Sotelo, J. Urbina, I. Valverde, et al., 3D quantification of wall shear stress and oscillatory shear index using a finite-element method in 3D CINE PC-MRI data of the thoracic aorta, IEEE Trans. Med. Imag. 35 (6) (2016 Jun) 1475–1487, https://doi.org/10.1109/TMI.2016.2517406. Epub 2016 Jan 14. PMID: 26780787.

[34] J. Sotelo, L. Dux-Santoy, A. Guala, et al., 3D axial and circumferential wall shear stress from 4D flow MRI data using a finite element method and a laplacian approach, Magn. Reson. Med. 79 (5) (2018 May) 2816–2823, https://doi.org/10.1002/mrm.26927.

[35] J. Sotelo, L. Dux-Santoy, A. Guala, et al., Comprehensive Analysis of Hemodynamic Parameters in Patients with Bicuspid Aortic Valve Using 4D Flow Data and a Finite Element Method, ISMRM annual meeting, Jun 16-21th, Paris, France, 2018.

[36] L. Dux-Santoy, A. Guala, J. Sotelo, et al., Low and oscillatory wall shear stress is not related to aortic dilation in patients with bicuspid aortic valve: a time-resolved 3-dimensional phase-contrast magnetic resonance imaging study, Arterioscler. Thromb. Vasc. Biol. 40 (1) (2020 Jan) e-10–e20.

[37] J. Sotelo, 4D-Flow-Matlab-Toolbox. https://github.com/JulioSoteloParraguez/4D-Flow-Matlab-Toolbox, 2021.

[38] Q. Fang, D. Boas, Tetrahedral mesh generation from volumetric binary and gray-scale images, Proceedings of IEEE International Symposium on Biomedical Imaging: From Nano to Macro (2009) 1142–1145.

[39] M.E. Wall, A. Rechtsteiner, L.M. Rocha, Singular value decomposition and principal component analysis, in: D.P. Berrar, Granzow M. Dubitzky (Eds.), A Practical Approach to Microarray Data Analysis, Springer, Boston, MA, 2003, pp. 91–109.

[40] I. Witten, E. Frank, Data Mining: Practical Machine Learning Tolls and Techniques, second ed., 2005. San Mateo.

[41] L. van der Maaten, G. Hinton, Visualizing Data using t-SNE, J. Mach. Learn. Res. 9 (86) (2008) 2579–2605.

[42] D. Mery, BALU: A Matlab Toolbox for Computer Vision, Pattern Recognition and Image Processing, 2011. http://dmery.ing.puc.cl/index.php/balu.

[43] Y. Gu, C. Wang, A study of hierarchical correlation clustering for scientific volume data, Advances in Visual Computing 6455 (2010), 978-3-642-17276-2.

[44] G. Ciaburro, MATLAB for Machine Learning: Practical Examples of Regression, Clustering and Neural Networks, Publisher: Packt Publishing, 2017, ISBN 978-1788398435.

[45] R. Lorenz, J. Bock, A.J. Barket, et al., 4D flow magnetic resonance imaging in bicuspid aortic valve disease demonstrates altered distribution of aortic blood flow helicity, Magn. Reson. Med. 71 (4) (2014 Apr) 1542–1553, https://doi.org/10.1002/mrm.24802.

[46] K. Calò, D. Gallo, A. Guala, J. Rodriguez-Palomares, S. Scarsoglio, L. Ridolfi, U. Morbiducci, Combining 4D flow MRI and complex networks theory to characterize the hemodynamic heterogeneity in dilated and non-dilated human ascending aortas, Ann. Biomed. Eng. 49 (9) (2021 Sep) 2441–2453.

[47] V. Mazzi, D. Gallo, K. Calò, M. Njafi, M.O. Khan, G. De Nisco, D.A. Steinman, U. Morbiducci, A Eulerian method to analyze wall shear stress fixed points and manifolds in cardiovascular flows, Biomech. Model. Mechanobiol. 19 (5) (2020 Oct) 1403–1423.

[48] R. Mahadevia, A.J. Barker, S. Schnell, et al., Bicuspid aortic cusp fusion morphology alters aortic three-dimensional outflow patterns, all shear stress, and

expression of aortopathy, Circulation 29 (2014) 673–682, https://doi.org/10.1161/CIRCULATIONAHA.113.003026.
[49] M. Lewin, C. Otto, The bicuspid aortic valve: advense outcomes from infancy to old age, Circulation 111 (7) (2005 Feb 22) 832–834, https://doi.org/10.1161/01.CIR.0000157137.59691.0B.
[50] M. Ferencik, L.A. Pape, Changes in size of ascending aorta and aortic valve function with time in patients with congenitally bicuspid aortic valves, Am. J. Cardiol. 92 (2003) 43–46.
[51] M. Gregorich, S. Strohmaier, D. Dunker, G. Heinze, Regression with highly correlated predictors: variable omission is not the solution, Int. J. Environ. Res. Publ. Health 18 (8) (2021 Apr 17) 4259.
[52] O'Brien, Dropping highly collinear variables from a model: why it typically is not a good idea, Soc. Sci. Q. 98 (2017) 360–375.

Pamela Franco[a,b,c], Julio Sotelo[a,c,d], Andrea Guala[e], Lydia Dux-Santoy[e], Arturo Evangelista[e], José Rodríguez-Palomares[e], Domingo Mery[f,h], Rodrigo Salas[d,h], Sergio Uribe[a,c,g,h,*]

[a] *Biomedical Imaging Center, School of Engineering, Pontificia Universidad Católica de Chile, Santiago, Chile*

[b] *Electrical Engineering Department, School of Engineering, Pontificia Universidad Católica de Chile, Santiago, Chile*

[c] *Millennium Nucleus for Cardiovascular Magnetic Resonance, Cardio, MR, Chile*

[d] *School of Biomedical Engineering, Universidad de Valparaíso, Valparaíso, Chile*

[e] *Department of Cardiology, Hospital Universitari Vall d'Hebron, Vall d'Hebron Institut de Recerca (VHIR), Universitat Autònoma de Barcelona, Barcelona, Spain*

[f] *Department of Computer Science, Pontificia Universidad Católica de Chile, Santiago, Chile*

[g] *Radiology Department, School of Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile*

[h] *Instituto Milenio Intelligent Healthcare Engineering, Chile*

[*] Corresponding author. Department of Radiology and Biomedical Imaging Center, School of Medicine. Pontificia Universidad Católica de Chile. Marcoleta 367, Santiago, Chile.
E-mail address: suribe@uc.cl (S. Uribe).