A machine learning cardiac magnetic resonance approach to extract disease features and automate pulmonary arterial hypertension diagnosis

Andrew J. Swift 1,2†, Haiping Lu2,3†, Johanna Uthoff3, Pankaj Garg1, Marcella Cogliano1, Jonathan Taylor4, Peter Metherall4, Shuo Zhou3, Christopher S. Johns4, Samer Alabed1,4, Robin A. Condliffe5, Allan Lawrie1, Jim M. Wild1, and David G. Kiely5

1Department of Infection, Immunity and Cardiovascular Disease, University of Sheffield, Western Bank, Sheffield S10 2TN, UK; 2INSIGNEO, Institute for In Silico Medicine, The University of Sheffield, The Pam Liversidge Building, Sir Frederick Mappin Building, F Floor, Mappin Street, Sheffield, S1 3JD, UK; 3Department of Computer Science, The University of Sheffield, 211 Portobello, Sheffield, S1 4DP, UK; 4Radiology Department, Sheffield Teaching Hospitals NHS Foundation Trust, Glossop Rd, Sheffield S10 2JF, UK; and 5Sheffield Pulmonary Vascular Disease Unit, Royal Hallamshire Hospital, Sheffield Teaching Hospitals NHS Foundation Trust, Glossop Rd, Sheffield S10 2JF, UK

Received 25 September 2019; editorial decision 2 January 2020; accepted 3 January 2020; online publish-ahead-of-print 30 January 2020

Aims

Pulmonary arterial hypertension (PAH) is a progressive condition with high mortality. Quantitative cardiovascular magnetic resonance (CMR) imaging metrics in PAH target individual cardiac structures and have diagnostic and prognostic utility but are challenging to acquire. The primary aim of this study was to develop and test a tensor-based machine learning approach to holistically identify diagnostic features in PAH using CMR, and secondarily, visualize and interpret key discriminative features associated with PAH.

Methods and results

Consecutive treatment naive patients with PAH or no evidence of pulmonary hypertension (PH), undergoing CMR and right heart catheterization within 48 h, were identified from the ASPIRE registry. A tensor-based machine learning approach, multilinear subspace learning, was developed and the diagnostic accuracy of this approach was compared with standard CMR metrics. Two hundred and twenty patients were identified: 150 with PAH and 70 with no PH. The diagnostic accuracy of the approach was high as assessed by area under the curve at receiver operating characteristic analysis (P < 0.001): 0.92 for PAH, slightly higher than standard CMR metrics. Moreover, establishing the diagnosis using the approach was less time-consuming, being achieved within 10 s. Learnt features were visualized in feature maps with correspondence to cardiac phases, confirming known and also identifying potentially new diagnostic features in PAH.

Conclusion

A tensor-based machine learning approach has been developed and applied to CMR. High diagnostic accuracy has been shown for PAH diagnosis and new learnt features were visualized with diagnostic potential.

Keywords

machine learning • tensor • pulmonary arterial hypertension • diagnosis • right ventricle

Introduction

Pulmonary arterial hypertension (PAH) is a life-shortening condition, and if untreated, it leads to right heart failure and death with a median survival of less than 3 years. The condition has an insidious onset with disease usually advanced at diagnosis.1 In response to distal vascular remodelling, the proximal pulmonary vasculature becomes dilated and stiffened.2–5 As the disease progresses the right ventricular (RV)
dilates, hypertrophies, and deteriorates. In addition, other features including changes in interventricular septal configuration, atrial size and function, and pericardial and pleural effusions can occur.

Echocardiography is recommended in suspected pulmonary hypertension (PH) via assessing systolic pulmonary arterial pressure.

Cardiovascular magnetic resonance (CMR) imaging protocols can assess a large number of characteristics in a single examination. In contrast to the 10’s of millions of voxels within an imaging dataset, derived data used for diagnostic and prognostic purposes is limited. Image analysis is time-consuming, particularly in PAH where the automatic segmentation algorithms typically fail due to ventricular deformation and reproducibility is lower for the right ventricle than the left.

Volumetric assessment of the right ventricle with CMR is superior to that achieved with echocardiography, though remains challenging. Machine learning can rapidly and consistently identify relevant information from vast 3D/4D data usually inaccessible to the reporting physician. Multilinear subspace learning is an emerging tensor-based machine learning approach that reduces the dimensionality of multidimensional data by directly mapping their tensor representations to a low-dimensional space, with recent application to automatically identify features in registered brain magnetic resonance imaging images to discriminate between different brain conditions. However, this multilinear subspace learning approach to extract, visualize, and interpret diagnostic features have not previously been applied to cardiovascular imaging. Success of such a machine learning diagnostic approach could improve standardization and diagnostic rates, particularly in less experienced centres. Even in experienced centres, machine learning may allow a more focused, enhanced, and standardized assessment by considering computer learnt diagnostic features alongside expert interpretation of the images.

To our knowledge, no studies have used multilinear subspace learning to extract, interpret, and visualize the range of diagnostic features in suspected PH, not to mention without the necessity for manual segmentation.

The primary aim of this study was to develop and test multilinear subspace learning for CMR imaging to identify and learn diagnostic features in patients with suspected PAH without the need for manual segmentation. The secondary aim was to visualize and interpret key discriminative features identified by the machine learning method.

Methods

Patients

Consecutive treatment naïve patients with PAH or no evidence of PH (no PH) referred for CMR between December 2014 and February 2017 at a PH referral centre were identified. Diagnosis was made following multidisciplinary team assessment. Inclusion required CMR and right heart catheterization to be performed within 48 h. Ethical approval was granted from the local ethics committee and institutional review board for this retrospective study and written consent was waived (ref c06/Q2308/8).

MR acquisition

CMR was performed on a GE HDx (GE Healthcare, Milwaukee, USA) whole-body scanner at 1.5 T using an eight-channel cardiac coil. Short-axis cine images were acquired using a cardiac gated multislice balanced SSFP sequence (20 frames per cardiac cycle, slice thickness 8 mm, FOV 48, matrix 512 × 512, BW 125 KHz/pixel, TR/TE 3.7/1.6 ms). Mid-chamber cine images in addition to a stack of images in the short-axis plane with slice thickness of 8 mm (2 mm interslice gap) were acquired fully covering both ventricles from base to apex, end-systole was considered as the smallest cavity area, and end-diastole was defined as the first cine cardiac phase of the R-wave triggered acquisition or largest volume. Patients were scanned in the supine position with a surface coil and retrospective electrocardiogram gating.

Image preprocessing

Registration

CMR images from different patients/sequences were registered using paired landmark points. Three landmarks were manually labelled on the first frame of each image sequence and reviewed (AJS). The short-axis image landmarks are inferior and superior hinge points and the inferolateral inflection point of the RV free wall. The four-chamber images landmarks are left ventricular (LV) apex, mitral, and tricuspid annuli. One image sequence in a dataset was considered the reference sequence to register the rest of the sequences against. A geometric transformation was fit to these three landmark point pairs, including translation, rotation, scaling, and reflection (flipping). This transformation was then used to warp a particular sequence against the reference sequence to accomplish registration. The distances between the warped landmarks and the reference landmarks were computed and the maximum among the three distance values was examined for quality assurance.

Masking

We studied two elliptical masks (small and large) to focus the machine learning approach on the relevant anatomy, i.e. a region including the whole heart. The two masks were automatically generated from five boundary landmarks manually drawn (AJS) on the first frame of the reference sequence, which needs to be done only once for short axis or four chambers.

Scaling

To reduce the negative effect (e.g. overfitting) of the small training sample size (~200) on the machine learning model, we rescaled the original spatial resolution of 512 × 512 (0.94 mm) to smaller scales 32 × 32 (15 mm), 64 × 64 (7.50 mm), 128 × 128 (3.75 mm), and 256 × 256 (1.89 mm) to find a resolution best for prediction.

Machine learning methodology

Machine learning builds models to automatically learn and improve from data examples rather than being explicitly programmed. The machine learning system in this article consists of (i) a feature extraction step to learn a low-dimensional representation from high-dimensional input, (ii) a feature selection step to focus on a small portion of extracted features, and (iii) a classifier to learn a decision function from selected features that distinguish which candidate category a new observation belongs. Due to the small training sample size (~200) relative to the high input dimensionality (32 × 32 × 20 = 20 480 even at the smallest scale 32 × 32), complex machine learning models such as random forests and neural networks are susceptible to overfitting. Therefore, we choose simple linear methods in the following, which are also more transparent and interpretable than complex, non-linear methods. Alternative choices at each step are possible but not the focus of this article.

A sequence of K CMR images of a standard size $I \times J$ is viewed as a single sample (example) of data, with size $I \times J \times K$, which is a multidimensional array called a tensor in mathematics. There were N training sequences available for training. We first used multilinear principal component analysis (MPCA), a fundamental multilinear subspace learning
algorithm that extends principal component analysis (PCA) to tensor input, to reduce the dimensionality of each (tensor) sample from $I \times J \times K$ to MPCA features of size $P \times Q \times R$ typically with $P < I$, $Q < J$, and $R < K$ using three projection matrices of size $I \times P$, $J \times Q$, and $K \times R$. During training, these three matrices are optimized to maximize the variation captured in the $N$ mapped $P \times Q \times R$ tensors and these optimized matrices are the output of the MPCA learning algorithm. During testing, the learnt three matrices map a new $I \times J \times K$ tensor input of $I \times J \times K$ into a $P \times Q \times R$ tensor as its low-dimensional representation.

**Right heart catheterization and clinical assessment**

Diagnostic classification of PAH was made using standard criteria following multi-professional assessment. Right heart catheterization was performed using a balloon-tipped 7.5-Fr thermodilution catheter (Becton-Dickinson, USA). Right heart catheterization was performed via the internal jugular vein using a Swan-Ganz catheter. Right heart catheterization indices required to define PAH were mean pulmonary artery pressure (mPAP) $\geq 25$ mmHg at rest with a pulmonary arterial wedge pressure (PAWP) of $\leq 15$ mmHg. Pulmonary vascular resistance (PVR) was determined as follows: $\text{PVR} = (\text{mPAP} - \text{PAWP})/\text{cardiac output (CO)}$. CO was measured by thermodilution technique. Cardiac index is calculated by CO/body surface area. No PH was defined as suspected PH with mean pulmonary arterial pressure at right heart catheterization less than $25$ mmHg.

**Statistical analysis**

Comparisons of CMR between patients with PAH and patients without PH were analysed using an independent t-test for continuous data, and $\chi^2$ and Fisher’s exact test for categorical data. Receiver operating characteristic (ROC) analysis was used to test the diagnostic strength of established CMR indices and multilinear subspace learning for the detection of the presence or absence of PAH. ROC curve analysis results are presented as area under the curve (AUC). Equal sensitivity and specificity were chosen from the ROC to report the accuracy ($\text{sensitivity}=\text{specificity}$), positive predictive value (PPV), and negative predictive values (NPVs).

Ten-fold cross-validation was used to evaluate the proposed machine learning approach in terms of AUC. The patients were divided into 10 disjoint subsets (folds). We used nine subsets for training (learning the MPCA projection, determine $S$, and learning the classifier) and the remaining one subset for prediction (evaluation). This process was repeated 10 times so that each subset was used for prediction exactly once, with the average over 10 repeats reported. In addition, a subgroup analysis was performed assessing the accuracy for differentiation of patients with and without idiopathic PAH (IPAH).

A $P$-value $<0.05$ was determined to be statistically significant. The CMR indices were analysed using SPSS 22 (SPSS, Chicago, IL, USA) to obtain the statistics. The proposed machine learning approach was implemented and studied using MATLAB (version R2017b, MathWorks). The core MATLAB code is available at http://www.mathworks.com/matlabcentral/fileexchange/26168.

**Results**

**Patients**

Of 1122 patients with suspected PH at their diagnostic visit, 220 patients with PAH or no PH were identified having CMR and
RHC within 48 h (see Figure 3). One hundred and fifty patients were identified with PAH and 70 patients were found not to have PH. Among 150 patients with PAH, 69 patients were diagnosed with IPAH, 58 with PAH associated with connective tissue disease, 11 with PAH associated with portal hypertension, 10 with congenital heart disease, and 2 patients with PAH associated with drugs and toxins. Table 1 presents the demographics, right heart catheter, and CMR indices for patients with and without PAH, and IPAH alone. Patients with PAH had higher mRAP, mPAP, PVR, and lower cardiac index and mixed venous oxygen saturations, all with $P < 0.0001$ compared to patients with no PH. In addition, higher RV end-diastolic volume and end-systolic volume and lower RV ejection fraction (all $P < 0.001$) were recorded for patients with PAH vs. patients without PH.
Learnt features, interpretation, and discovery

The average number of learnt features identified following MPCA and feature selection with 10-fold cross-validation on four-chamber images were 22 (out of 1215 total features on average, after MPCA) and 49 (out of 1242 total features on average) for no PH vs. PAH and IPAH, respectively. More features were identified on short-axis images though the total numbers of features were smaller; 66 (out of 212 total features on average) and 92 (out of 252 total features on average) for differentiating no PH vs. PAH and IPAH, respectively. These numbers were determined in a data-driven approach on the training data. Because this machine learning model is fully linear, these learnt features can be mapped back to the original images to determine voxel-wise weights obtained from linear classifiers for diagnosis and interpretation.

For short-axis imaging, features identified were most frequently located in the region of the interventricular septum (Figure 4A). The learnt temporal features were most frequently identified and discriminative at end-systole. Four-chamber analysis also was highly diagnostic for PAH, where features associated with PAH (highlighted in red) were identified adjacent to the interventricular septum (Figure 4B).

New features were discovered on short-axis imaging, located at the interventricular septum at end-diastole/early systole which were indicative of no PH (highlighted in green) (Figure 4C). New features were also discovered on four-chamber images, green features are shown at the basal LV lateral wall at end-diastole indicating normality (Figure 4D), and red features are noted in the same location at end-systole indicating PAH.

Diagnostic utility

PAH diagnosis

Table 2 presents the ROC analysis results for established indices, as well as the proposed machine learning approach with scaling factor of $64 \times 64$ (7.50 mm) for short axis and $128 \times 128$ (3.75 mm) for four chambers, and with the best performing classifier (SVM/LR). The table also reports the respective confidence intervals. From 10-fold cross-validation of the machine learning approach, the average AUC for differentiation of patients with PAH from patients without PH was 0.90 for the short-axis cine images and 0.86 for the four-chamber cine images (Figure 5). This was achieved using the small elliptical mask for both short-axis and four-chamber images, rescaled to $64 \times 64$ and $128 \times 128$, respectively. Table 3 shows diagnostic accuracy using different image scaling factors for short-axis and four-chamber images.
cine images using the small ellipse mask. On short-axis and four-chamber imaging, the large elliptical mask and using no mask had weaker diagnostic accuracy (Table 2). SVM and LR both had good accuracy, with small differences only. For differentiation of PAH from without PAH, SVM had an AUC of 0.90, while LR had 0.87 for the short-axis cine images, and SVM had an AUC of 0.86, while LR had 0.78 for the four-chamber cine images. Table 4 presents the sensitivity, specificity, PPV, and NPV of the machine learning approach. The results from existing manually drawn CMR diagnostic measurements were similarly high, where ventricular mass index and septal angle have AUC of 0.92 and 0.97, respectively.

**Preprocessing results**

Registration was successful in all cases as judged by visual inspection. Error distance measured in unit pixel (in an image of $512 \times 512$) was 7.1 (SD 3.7, range 0.32–20.1) for four-chamber images and 4.1 (SD 2.3, range 0.3–10.6) for short axis. The small elliptical mask consistently gave the best classification results for both short-axis and four-chamber imaging (outperforming large mask and no mask).

**Discussion**

This study shows for the first time that a machine learning workflow based on multilinear subspace learning applied to CMR cine images can provide a rapid diagnostic assessment without manual
segmentation of images and can differentiate patients with and without PAH with high accuracy. The diagnostic utility of the machine learnt image and temporal features was marginally higher than that of time-consuming, manually drawn diagnostic CMR measurements used in PAH, with comparable variations in confidence intervals. However, the proposed machine learning workflow has much lower cost and much less variability than the manual segmentation-based indices produced by PH imaging experts at a specialist centre. Learnt features were visualized in feature maps corresponding to cardiac phases, confirming known and also identifying potentially new diagnostic features in PAH.

Identified features were most frequently located in the region of the interventricular septum for short-axis imaging. We postulate that the diagnostic value of interventricular septal deviation has been learnt automatically by the machine. We know that septal deviation is a feature with added diagnostic value, particularly when measured at end-systole.\(^9,25\) In keeping with higher accuracy of septal position at end-systole, the learnt temporal features were most frequently identified at end-systole as providing the highest discriminative power. New features were discovered on short-axis imaging, e.g. green features located at the interventricular septum at end-diastole/early systole are indicative of normality (Figure 4C).

Four-chamber analysis also was highly diagnostic for PAH, with red features identified adjacent to the interventricular septum as expected, Figure 4B but also in the RV free wall, particularly the tricuspid annulus at end-diastole likely relating to abnormal free wall contraction. New features were also discovered on four-chamber images. Green features were shown at the basal LV lateral wall at end-diastole indicating normality (Figure 4C and D), and PAH features were noted in the same location at end-systole indicating PAH. These features may reflect impact of PAH on LV function.

Blood pool disease features in the right ventricle were identified. RV blood pool features indicating PAH diagnosis (red, Figure 4A and B) have been found during systole and those indicating normality...
Furthermore, it is known that the reproducibility of RV measurements is inferior to those of the left ventricle. The manual landmarking of only three points in our machine learning approach is easier and faster to perform (a few seconds) and would require little training to undertake.

A major advantage of the proposed tensor-based machine learning approach is the removal of the requirement for manual segmentation. Typically regions of interest, lines, or angles are drawn by experienced observers directly on the images to derive cardiac and pulmonary measurements. Such measurements are affected by inter-observer variability even in expert hands, particularly for RV measurements. It is known that the reproducibility of RV measurements is inferior to those of the left ventricle. The manual landmarking of only three points in our machine learning approach is easier and faster to perform (a few seconds) and would require little training to undertake.

Automatic detection and segmentation of the ventricles using machine learning is also showing promise. The data derived from these approaches may reduce the need for manual adjustments that are currently labour intensive, especially for the right ventricle. Such approaches are however limited by the accuracy of manual segmentation used to train the algorithms. The approach proposed in the present paper has the great advantage of assessing all pixels within a mask in each image of a sequence and has the potential to pick up diagnostic features from within or outside of the ventricles, e.g., atrial size and function, and pericardial and pleural effusions. This allows the computer to identify discriminative features in an unbiased fashion. A small number of diagnostic features were identified outside of the heart (e.g., Figure 4C), which indicates the need for further work to investigate their pathophysiological significance and potential diagnostic value. Other applications include treatment response assessment, subtyping, severity assessment, and pressure/PVR prediction. In particular, the range of prognostic features from the right and left ventricles, the atria, and features of decompensation such as pericardial effusion may be identified using the proposed approach.

It is known that PH is underdiagnosed and machine learning may help to improve diagnostic rates and ultimately improve outcomes. A prior study using machine learning to measure RV tissue motion has shown good prognostic value in PH, and association with metabolic profiles. In the development of the proposed machine learning approach, we aim to gain insights into the features identified for further analyses (e.g., biomarker identification/treatment). Therefore, we favour relatively simple, yet interpretable machine learning methods, such as linear methods, over highly complex, but opaque black-box methods, such as the popular deep neural networks. In particular, linear methods allow direct visualization of the importance of individual features to assist interpretation.

The length of time to manually analyse a CMR study is typically 10–15 min, depending on the complexity of the analysis. There can be variability between observers in their approach to the analysis and there are also subtle differences in the software provided by vendors. The machine learning approach proposed here, if used to test an individual short-axis sequence, would take less than 10 s, including manual landmarking (dominating the time cost), and automated preprocessing and diagnostic output. The automated process after landmarking would take less than 1 s.

Masks of different sizes were studied with best results obtained using smaller more focused elliptical masks around the heart. This finding was expected given the effect of PH on the heart and the characteristic imaging features of high pulmonary arterial pressure. For both IPAH and PAH diagnosis, high accuracy was found even when short-axis images were rescaled to 64 × 64 and four-chamber images (scale: 128 × 128).
were rescaled to 128 × 128, compared with the original scale 512 × 512 (see Table 3). At these lower resolutions, the structural information may play a smaller role and it is the functional information that is driving the differentiation of health from disease, likely lower RV function, abnormal RV and LV flow patterns, and abnormal septal motion. Studying larger cohorts may allow higher accuracy with input of high-quality data.

**Limitations**

Landmarks were manually drawn to register the patient sequences to one another. Fully automated registration of sequences, e.g., via automated landmark detection using deep learning, is an area for future work that can further reduce manual user input to just a visual check to ensure accurate landmark detection and make corrections if necessary for quality assurance.

Whilst typical interobserver variability is removed using our automatic machine learning method, quantification of the impact of scan-to-scan variability is an area for future research. Only four-chamber images and mid-chamber short-axis cine images were used for this study. Further work includes taking as input multislice cardiac volumetric data and 4D flow data, in addition to flow imaging, which may improve diagnostic accuracy. Further works also include testing this approach in unselected patients with suspected PH and validating the approach in a second cohort, in addition to validation in a multi-vendor or multicentre setting. In addition, trialling such an approach in other challenging diagnostic and prognostic cardiac scenarios would be of value. Factoring image quality score into the approach is also an important area for further research to explore the potential of machine learning in low-quality cases that are particularly challenging for humans.

**Conclusion**

A tensor-based machine learning approach has been developed to analyse CMR images without manual image segmentation. It has holistically identified predictive and interpretable temporal and image features and allows rapid and accurate diagnosis of PAH.

**Funding**

This work was supported by Wellcome (215799/Z/19/Z and 205188/Z/16/Z), EPSRC (EP/R014507/1), NIHR (NIHR-RP-R3-12-027), and MRC (MR/M008894/1). The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the National Institute for Health Research and the Department of Health. D. Capener was partly funded by an unrestricted research grant from Bayer.

**Conflict of interest:** none declared.

**References**

Machine learning CMR approach


