Ultrasound in Medicine & Biology 49 (2023) 1996–2005

Contents lists available at ScienceDirect

Ultrasound in Medicine & Biology

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

Original Contribution

Automated Detection of Apical Foreshortening in Echocardiography Using Statistical Shape Modelling

Woo-Jin Cho Kim [a], Arian Beqiri [a,b], Adam J. Lewandowski [c], Angela Mumith [b], Rizwan Sarwar [b], Andrew King [a], Paul Leeson [b,c], Pablo Lamata [a,*]

[a] School of Biomedical Engineering and Imaging Sciences, King's College London, London, UK
[b] Ultromics Ltd., Oxford, UK
[c] Cardiovascular Clinical Research Facility, University of Oxford, Oxford, UK

ARTICLE INFO

Keywords:
2-D echocardiography
Apical foreshortening
Left ventricle

Objective: Automated detection of foreshortening, a common challenge in routine 2-D echocardiography, has the potential to improve quality of acquisitions and reduce the variability of left ventricular measurements. Acquiring and labelling the required training data is challenging due to the time-intensive and highly subjective nature of foreshortened apical views. We aimed to develop an automatic pipeline for the detection of foreshortening. To this end, we propose a method to generate synthetic apical-four-chamber (A4C) views with matching ground truth foreshortening labels.

Methods: A statistical shape model of the four chambers of the heart was used to synthesise idealised A4C views with varying degrees of foreshortening. Contours of the left ventricular endocardium were segmented in the images, and a partial least squares (PLS) model was trained to learn the morphological traits of foreshortening. The predictive capability of the learned synthetic features was evaluated on an independent set of manually labelled and automatically curated real echocardiographic A4C images.

Results: Acceptable classification accuracy for identification of foreshortened views in the testing set was achieved using logistic regression based on 11 PLS shape modes, with a sensitivity, specificity and area under the receiver operating characteristic curve of 0.84, 0.82 and 0.84, respectively. Both synthetic and real cohorts showed interpretable traits of foreshortening within the first two PLS shape modes, reflected as a shortening in the long-axis length and apical rounding.

Conclusion: A contour shape model trained only on synthesized A4C views allowed accurate prediction of foreshortening in real echocardiographic images.

Introduction

Cardiac imaging plays a prominent role in assessing ventricular morphology and function, with 2-D echocardiography (2DE) being the first choice for non-invasive imaging due to its availability, portability and absence of ionising radiation. Despite the existing guidelines for standard image acquisitions in 2DE, most often established and maintained by the American Society of Echocardiography (ASE) [1], echocardiographic examinations of the heart are challenging, as the process is highly operator dependent and image quality varies according to the level of operator expertise.

The apical-four-chamber (A4C) view is crucial for many echocardiographic screenings, but without adequate training and expertise, operators will struggle to orient the probe to ensure all key cardiac measurements are accurate and precise. Even experienced operators may face difficulties acquiring an optimal A4C view due to factors such as obstruction by the patient's ribs, low image contrast and other non-ideal image conditions [2–5]. This commonly leads to an acquisition in which the ultrasound (US) probe is not positioned directly on the apex of the heart, and thus the view is *foreshortened*. Most often the plane is oriented at an oblique angle, from the anterior apex to the posterior wall of the atria, leading to a geometric distortion of the left ventricle (LV). The misrepresented LV appears with a deceptively shorter long-axis, combined with a falsely thickened apex, leading to potentially biased estimates of systolic function (*e.g.,* volumes calculated using Simpson's bi-plane method [6]).

Despite recognition of the impact of foreshortening on cardiac disease diagnosis [7,8], there is limited quantitative research on the extent of misdiagnosis and the differentiation between foreshortened-caused rounded LV geometries and those caused by pathological influences on the apex, such as apical aneurysms. Previous studies have reported that foreshortened views result in a significant reduction in the long-axis dimension, with a decrease of nearly 10% in comparison to optimized views [9]. This decrease was accompanied by significantly lower end-systolic (ES) and end-diastolic (ED) LV volumes, resulting in a median relative difference in ejection fraction (EF) of up to 6.9%. Additionally,

* Corresponding author. 5th Floor Becket House, 1 Lambeth Palace Road, London SE1 7EU, UK.
E-mail address: pablo.lamata@kcl.ac.uk (P. Lamata).

https://doi.org/10.1016/j.ultrasmedbio.2023.05.003
Received 18 November 2022; Revised 16 April 2023; Accepted 4 May 2023

global longitudinal strain (GLS) measurements from the emerging technique of speckle tracking 2DE were also shown to be heavily impacted, with endocardial GLS measuring 23.2% higher in foreshortened views.

Automated detection of foreshortening can overcome these limitations, increase confidence of acquisition and reduce inter- and intra-observer variability of LV volume, EF and GLS measurements. The potential application of this technology would enable echocardiographers to be made aware that their corresponding quantitative metrics may be subject to additional error and then attempt a reacquisition.

However, a major practical roadblock in the study of foreshortening is the lack of labelled training data, as reliable manual annotations require extensive domain expertise and are very time-consuming to produce. The absence of an established and clearly defined tolerable range of foreshortening further complicates the curation of such data. This has led to limited research efforts aimed at developing methods for foreshortening detection, primarily through the use of 3-D echocardiography (3DE) images that allow controllable slicing to produce views with arbitrary foreshortening [10]. However, limitations in spatial resolution and poor contrast-to-noise ratio of 3DE raise concerns about the accuracy of the myocardial delineations used as reference volumes, especially in the context of foreshortening detection where an accurate representation of the LV geometry is crucial. Additionally, it is worth noting that the proposed measurement of foreshortening relied solely on tracking the apical position and long-axis dimensions, without considering the LV geometry as a whole.

Recent advances in 3-D imaging techniques and digital twin technologies have propelled the development of highly sophisticated and realistic biophysical models of the heart [11]. Although initially developed to simulate electromechanical events of cardiac function, these models may also provide a valuable source of high-quality structural information when built using the excellent contrast and spatial resolutions of computer tomography (CT) images.

In this work, we solve the foreshortening labelling problem by leveraging the rich structural information contained in 3-D anatomical models. More specifically, we propose a novel pipeline for generating 2-D foreshortened apical views from CT-based 3-D cardiac meshes, obviating the need to acquire and label large numbers of echocardiographic images. The *in silico* data set is then used to build a regression model between the anatomical variation present in the data and the presence of foreshortening. We demonstrate the utility of this pipeline in predicting foreshortening in an independent data set of real 2DE images.

Methods

A schematic diagram of the proposed approach is presented in Figure 1. First, idealised 2-D LV A4C echocardiographic views were generated from 3-D anatomical four-chamber models with a wide range of plausible foreshortening and viewing angles. Second, a partial least squares (PLS) regression model was built to capture the relationship between the anatomical variation of generated synthetic LV endocardial contours and the corresponding amount of foreshortening. Third, a logistic regression model was built, entirely synthetically, for the purpose of foreshortening prediction. Finally, without any further training, we tested the utility of the foreshortening detection model by evaluating it in real echocardiographic images that were fully manually segmented and automatically curated (*i.e.,* outlier removal).

Synthesis of foreshortened views from 3-D anatomical models

One thousand four-chamber cardiac models, reconstructed from a statistical shape model (SSM) built from CT images and available from the literature [12], were used to study the anatomical variation associated with foreshortening (additional details of the reconstruction process are given in Appendix S1 [online only]).

Idealised echocardiographic views were generated by intersecting the LV wall from the cardiac meshes with the plane corresponding to the standard A4C view. According to the ASE guidelines, the A4C view is defined as the plane intersecting the LV apex, the mitral and the tricuspid valve centres [13], and these were anatomical landmarks that were identified in the 1000 meshes using an automatic segmentation algorithm [14].

Foreshortening was then simulated by varying the optimal A4C orientation during the generation of contours by two factors: an apical angular offset F_θ with respect to a fixed mitral valve (MV) point and a rotational offset V_θ around its long axis (Fig. 2a). These choices were motivated by the fact that valves are a major structural landmark defining a standard A4C view and thus reasonably robustly captured by an operator, and that therefore foreshortening is less likely to be caused by

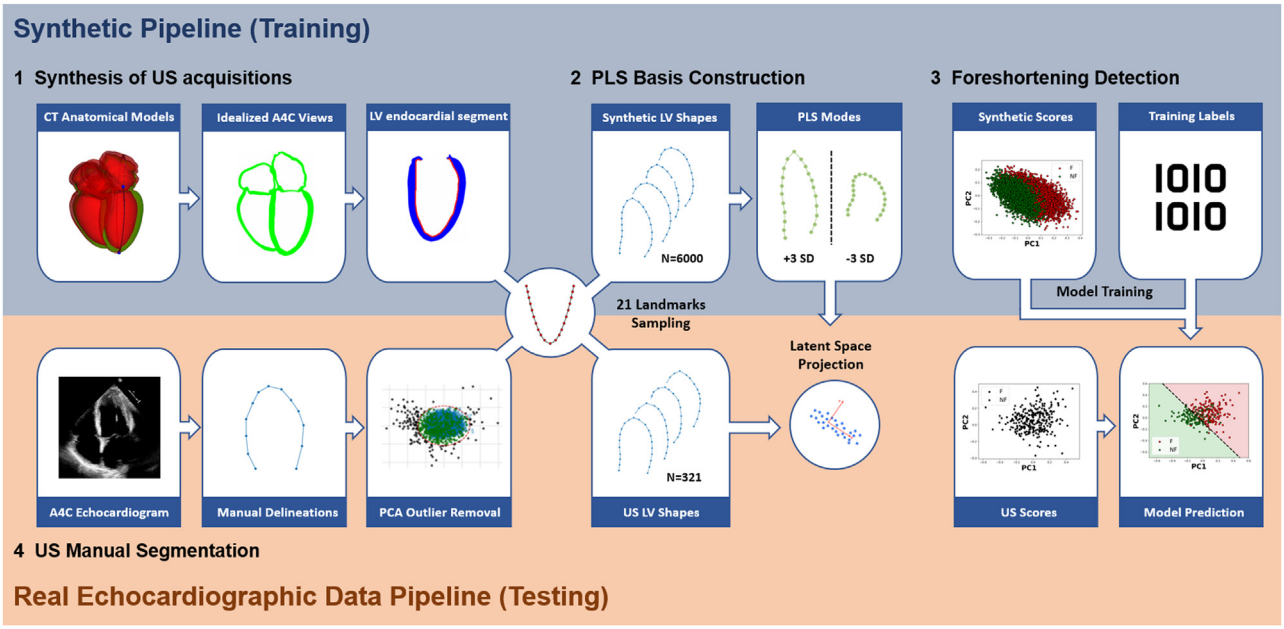

Figure 1. Overview of the proposed pipeline for extraction of the anatomical variation related to foreshortening and to use the learned features for foreshortening detection in real ultrasound images. A4C, apical four-chamber; CT, computed tomography; F, foreshortened; LV, left ventricle; N, number of LV contours; NF, non-foreshortened; PCA, principal components analysis; PLS, partial least squares; SD, standard deviation; US, ultrasound.

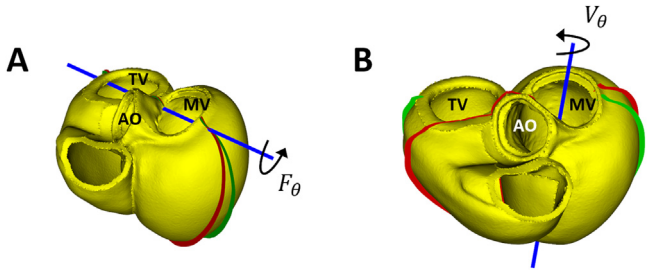

Figure 2. Simulation of foreshortened apical view acquisitions in 3-D anatomical cardiac models, in which the *green plane* represents the ideal apical four-chamber (A4C) plane and the *blue lines* represent the axes of rotation. (a) Foreshortened plane (*red*) that is tilted away by an apical offset F_θ around the axis of rotation (defined by the mitral and tricuspid valve centres). The ideal A4C plane has been rotated from the true apex (*black dot*) towards the aortic outflow tract (AOT) while keeping the mitral valve centre. (b) Offset plane (*red*) that is a result of a rotational offset V_θ along the vertical axis of the A4C view that contains the mitral centre. The ideal A4C plane has been rotated towards the apical three/five chamber view (A3C/A5C).

missing the valve centres—the MV centre containing the A4C planes was therefore fixed during the experiments. Finally, the apical offsets F_θ were generated with a distribution biased towards the A4C planes shifting to the aortic outflow tract, since this is the most common cause of foreshortening [15], and the rotational offset V_θ was only generated anteriorly towards the aortic valve, as the opposite direction is less plausible (see Appendix S1 for a more detailed explanation).

Generated ideal and offset view planes were used to cut through the 3-D models and generate the A4C views and their corresponding 2-D LV endocardial contours. The morphology of each endocardial contour was then encoded with a set of 21 landmarks that were sampled equidistantly from the most apical node; we will refer to this set of points as the A4C endocardial landmarks, $A4C_{EL}$. The Basal Oriented Disks interpretation of Simpson's bi-plane rule [6] was used to compute the percentage error in LV volume ($ErrVol_{\%}$) for each $A4C_{EL}$, which was the variable that served to define foreshortening with a threshold of 7% error in the volume computation (Appendix S1). Following the labelling process, the generated data set consisted of a balanced split of 3251 foreshortened (F) and 2749 not-foreshortened (NF) $A4C_{EL}$ from the 1000 anatomies.

PLS regression between $A4C_{EL}$ and foreshortening

We aimed to find the linear modes of anatomical variation in the $A4C_{EL}$ (*i.e.,* PLS modes) that predict the presence of foreshortening. The quest is for the latent variable—which will have an interpretation through the visualization of the PLS modes—that discriminates between the foreshortened (F) and non-foreshortened (NF) groups of contours [16].

$A4C_{EL}$ were first aligned by a rigid registration (*i.e.,* translation and rotation) of the contours using the Generalized Procrustes algorithm [17]. PLS was subsequently performed on the matrix consisting of the vectorized aligned shapes and the vector of foreshortening labels (encoded as the binary classification of F/NF).

Let us denote by $X = (x_1, \ldots, x_N)$ the matrix comprising the N sets of $A4C_{EL}$ after shape pre-processing and mean centering, and $Y = (y_1, \ldots, y_N)$ the vector corresponding to the binary foreshortening variable. Each 2-D shape is a vector of size $2L$ where L is the total number of landmarks (2 dimensions, x and y, per landmark). To ensure that the obtained modes correlate with the prediction task, the PLS model extracts a set of K latent vectors from the input data X that are most correlated with the labels Y. This is achieved by the following simultaneous decomposition of X and Y:

$$X \approx TP^T \tag{1}$$

$$Y \approx UQ^T \tag{2}$$

where the columns of P are the K PLS shape components after decomposition (latent vectors), Q is the coefficient loading matrix, and the rows of T and U are the score vectors in the column space of X and Y respectively (Fig. 3). The solution to the decomposition above is obtained

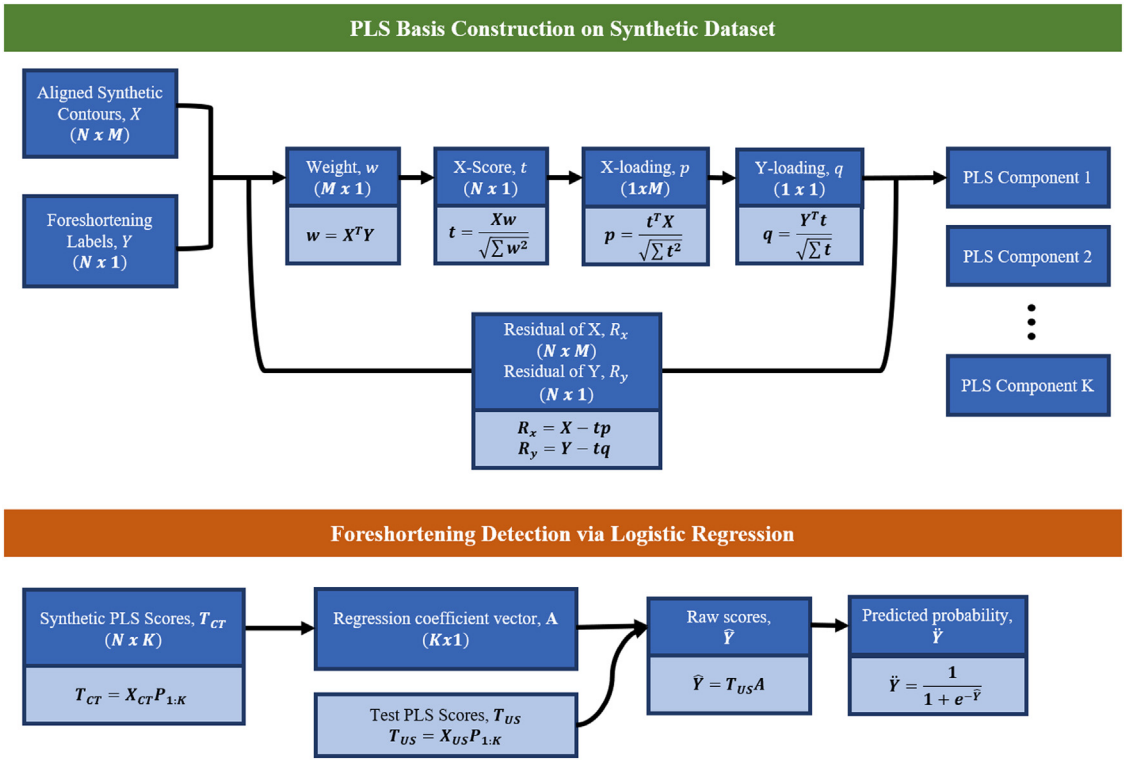

Figure 3. Flowchart outlining the procedural steps of the partial least squares (PLS) regression model. The construction of the PLS basis is performed iteratively where the residual matrices represent the variability in the data that has not been accounted by the preceding PLS decomposition.

iteratively through the non-linear iterative PLS algorithm. At every iteration, a new shape component p^i is extracted, and the decomposition is finalized so as to maximize the covariance between T and U [18].

Foreshortening detection combining PLS scores through logistic regression

The PLS components extracted after the decomposition in eqn (1) and eqn (2) are used to predict foreshortening. Using the notion of optimal linear reconstruction, each shape vector x^i is approximated as a linear combination of K shape components:

$$\sim X = TP^T_{1:K} \tag{3}$$

$$T = XP \tag{4}$$

where $\sim X$ is the approximated data and the columns of $P_{1:K}$ are the first K shape components of the PLS loading matrix P. For the reconstruction to be non-trivial, K is set smaller than the initial number of shape features $2L$. Each row vector t^i in the score matrix equates to the weights of the above linear combination and can be interpreted as the amount of each component present in a particular shape x^i. Using the PLS scores, we obtain a predicted score of $\hat{y}$ through the following logistic regression model:

$$\hat{y} = \sum_{i=0}^{N} t_i^T A \tag{5}$$

$$\ddot{y} = p(\hat{y}) = \frac{1}{1 + e^{-\hat{y}}} \tag{6}$$

where A is a vector of regression coefficients, and p is a sigmoidal function that maps score $\hat{y}$ into real-valued probabilities $\ddot{y}$ between 0 and 1 (between NF and F). A detailed flowchart outlining the various steps of our PLS regression method can be found in Figure 3.

Echocardiographic real views for external evaluation

Our US cohort was derived from the WASE-COVID study: a clinical investigation of 870 patients with acute COVID-19 infection who had undergone transthoracic echocardiograms (TTE) [19]. Automatic view classification and cardiac phase detection was performed using EchoGO v2 [20]—a deep learning–based cardiac image interpretation tool—and LV endocardial contours were then manually delineated from A4C views at end-diastole (ED). The LV endocardial landmarks ($A4C_{EL}$) from each manual contour were automatically sampled using the same 21-landmark extraction process as in the CT cohort to generate the data for testing the automated identification of foreshortening.

As this data set included aberrant/implausible echo views related to movement or other artefacts during acquisition, we additionally developed an automated process to identify and remove "anomalous cases" (Fig. 4). The principal components were derived from the synthetic cohort, and an anomalous shape was identified when it was found to lie outside the range of variation observed in the CT LV shapes (*i.e.,* outside ± 3 standard deviations (SD) away from the mean of the first 10 PCA modes).

Identification of foreshortening in the external validation data set

To define presence or absence of foreshortening in the images, seven observers were randomly assigned to manually label the corresponding A4C sequences as F or NF. Additionally, in an effort to study the degree of subjectivity in the labelling process of real images, inter-observer agreement of foreshortening labels was assessed on a subset of the uncurated US cohort, consisting of 29 A4C sequences annotated by a total of 22 observers. Eight observers with more than 50% of missing labels were excluded (drop of 31 labels), leading to the final 14 observers (44 F vs. 331 NF labels in the 29 A4C sequences) used in the analysis.

Benchmark of geometric descriptors linked to foreshortening

A collection of geometric descriptors known to be linked to foreshortening was generated from each endocardial contour of the real A4C views to provide a benchmark against the new PLS-based prediction model and help interpret what the model was capturing. The metrics were length, mid-cavity width, LV sphericity ratio (LV_{SR}), apical sphericity index (A_{SI}) and basal slanting angle (B_θ). The definition of each metric is illustrated in Figure S1 (online only).

Benchmark and ablation experiments

To establish a baseline in model performance, we conducted an ablation study in which we systematically replaced various components of the regression pipeline, including the type of model, data set and label, to determine their contributions to the overall performance of the classifier. This approach was necessary because there was no existing reference work against which to compare our results.

Four experiments were designed and implemented for this purpose. First, we tested the ability of the logistic regression of PLS features when trained to predict length, as a surrogate of the actual foreshortening. Second, we established a baseline of the prediction of foreshortening with conventional geometric descriptors (random forest classifier using length, width, LV_{SR}, A_{SI} and B_θ). Subsequently, we tested the ability to train the logistic regression, not with the synthetic but with the real data, by using a 75–25 stratified training and testing split of the original and curated US cohorts.

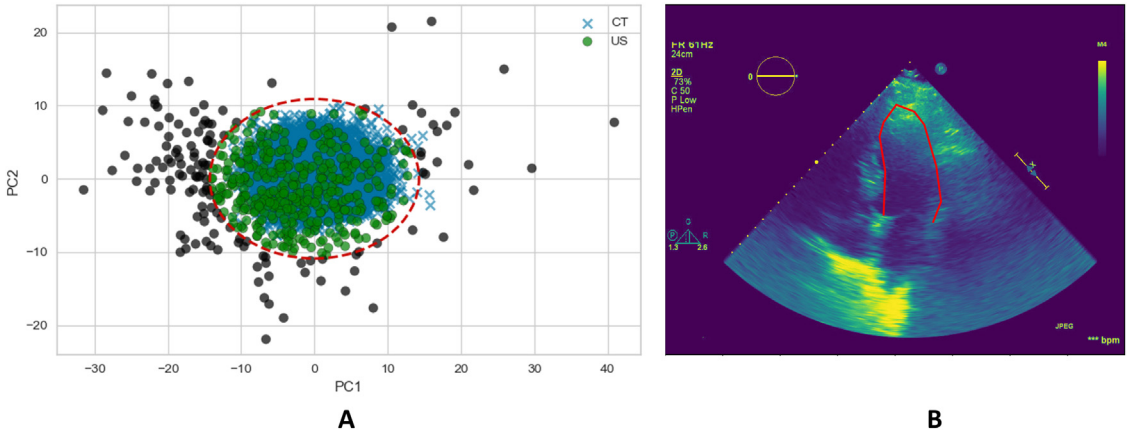

Figure 4. Data curation results. (a) Illustration of outlier removal in the two first principal components (PC). Note that removal acts in the 10 first PC dimensions, not only these two. (b) Exemplary data set of failed manual segmentation due to poor image quality.

Results

Learning anatomical variation that predicts foreshortening

Foreshortening was predicted in the training cohort with high efficacy (sensitivity of 85% and specificity of 87%, with an area under the receiver operating characteristic curve [AUC] of 0.95). The first 11 PLS shape components accounted for a cumulative sum of 94.1% and 91.2% of the total variance in X (anatomy) and Y (foreshortening), respectively (Fig. 5c). PLS1 (29.9% and 25.7% of the total variance in X and Y, respectively) primarily accounted for changes in length, apical pointiness and septal wall curvature (Fig. 5a). PLS2 and PLS3 displayed more subtle changes in

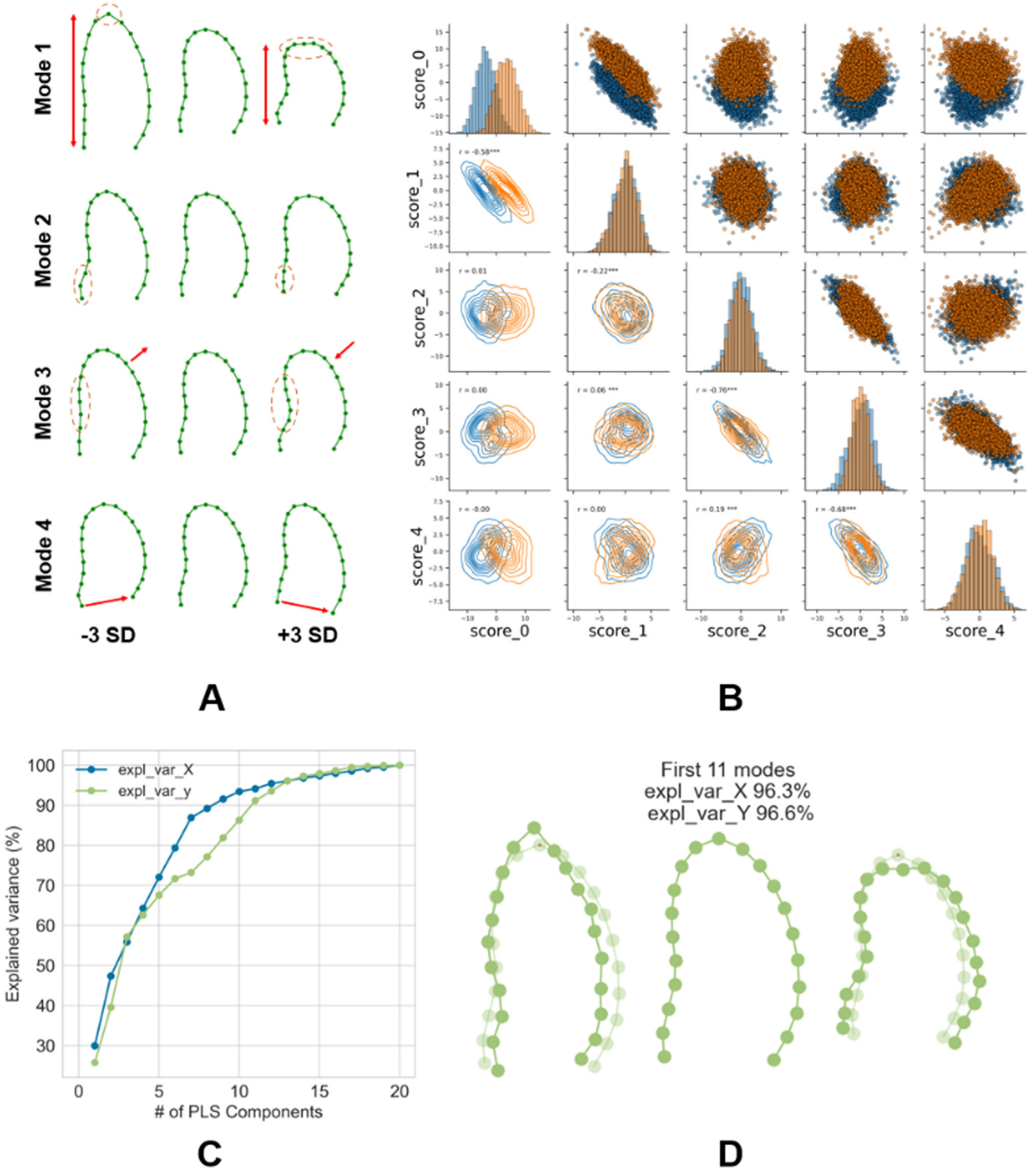

Figure 5. (a) Left ventricle (LV) shape modes of variations 1−3 showing significant anatomical differences observed at ±3 standard deviations (SDs) from the mean shape (centre column). *Red arrows* focus on the location and direction of change for key anatomical features related to foreshortening (*arrow* pointing towards the shape indicates decrease in feature size, whereas an *arrow* pointing outwards from the shape indicates an increase in feature size). (b) Pair plot of the partial least squares (PLS) shape scores: kernel density estimation (KDE) plots in the *lower triangle*, histogram distributions in the *diagonal* and scatter points in the *upper triangle*. (c) The cumulative explained variance ratio of each PLS mode in the computed tomography cohort. (d) The final shape mode constructed by applying the first 11 PLS shape modes to the mean LV shape (centre); left and right are ±3 SDs of the modes respectively overlayed on top of the mean LV shape (opaque).

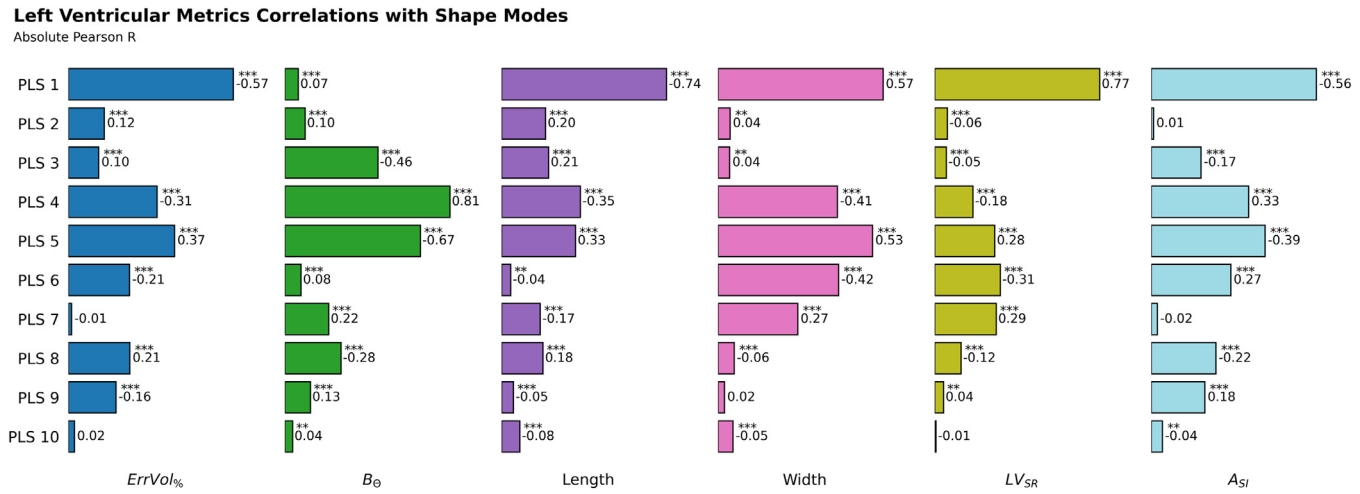

Figure 6. Bar graph of the Pearson correlation coefficients. Every grouped bar represents the estimated correlations between a shape-characterizing metric and the partial least squares (PLS) shape score values in the training set.

morphology, most notably variations in septal wall curvature. PLS4 captured notable differences in basal slanting angles. There were no easily interpretable features in subsequent modes 5 to 11. The pair-plot of correlations between PLS scores showed that the first two modes almost entirely separated F and NF clusters in the virtual cohort, with only a slight overlap between the two histogram distributions (Fig. 5b).

Interpretation of PLS modes

To better interpret the PLS modes, their correlation with geometrical descriptors in the synthetic cohort is provided in Fig. 6. PLS1 is the mode that best captured the changes in length, width, LV_{SR} and A_{SI}, and PLS4 and PLS5 were the ones that best captured B_θ. The continuous outcome metric that defines foreshortening classification, $ErrVol_{\%}$, was best correlated with PLS1, then PLS5 and PLS4. All geometrical descriptors showed a reasonable level of correlation ($r \sim 0.7$) with the main cause of foreshortening, the apical offset F_θ, but not with width ($r = -0.57$) and B_θ ($r = 0.16$) (Fig. 7). PLS1 scores of the testing set showed similar levels of correlation with the length, width and LV_{SR} as in the training set, with absolute r scores of 0.80, 0.72 and 0.79, respectively (Fig. 8a–e).

External cohort: curation and inter-observer disagreement

From the initial 870 patients of our study cohort, the A4C view was usable in 722 exams (83% of enrolled patients), while 523 had a complete process of automatic phase detection, LV manual contouring and annotations (drop of 199 exams). Finally, the data curation process discarded 202 contour shapes that were not consistent with the range of variability observed within the CT cohort, leading to the final set of 321 LV contour shapes (144F, 177NF).

There was evidence of significant subjectiveness in the labelling process of real images. Gwet's inter-observer agreement coefficient (AC1) produced an unstable and over-estimated agreement of 76.1% (95% CI: 66.2%–86.0%) due to the purportedly low prevalence of foreshortening occurrences in this subset, reflected in the severely imbalanced marginal totals of the F class (Table 1). This is also known as the *prevalence paradox* [21]. Accordingly, category-specific agreements (SA) were estimated as a more stable and representative measure of accordance between observers [22], for which the F-class SA score was found to be significantly lower than that of the NF class: 16.4% (95% CI: 15.7% –17.1%) and 82% (95% CI: 81.3%–83.8%) for the NF and F classes, respectively.

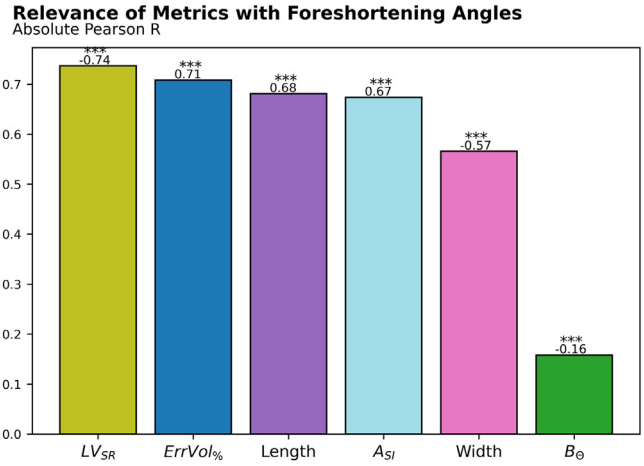

Figure 7. Pearson correlations between the geometrical descriptors and the ground truth foreshortening angles in the training set.

External validation: PLS model to predict foreshortening

The synthetically trained foreshortening detection model achieved an AUC of 0.84, and 0.83, 0.84 and 0.82 in accuracy, sensitivity and specificity, respectively, using a decision threshold of 0.61 in the probability of eqn (6). See the confusion matrix in Figure 9 and Figure 8f.

Benchmark and ablation results

Four experiments were designed, and results are reported in Table 2. The availability of the ground truth volumes (*i.e.*, $ErrVol_{\%}$), or the labels of foreshortening, is one of the practical bottlenecks with real images. First, the performance of the logistic regression of PLS features dropped to an AUC of 0.75 from 0.84 on the curated US cohort when trained to predict length. Second, the prediction of foreshortening with conventional geometric descriptors using a random forest classifier led to an AUC of 0.73 on the curated US cohort. Third, model evaluation of the logistic regressor trained with the real data, not the synthetic data, showed a decrease in performance (AUC of 0.68) in the curated US cohort and an almost random classifier performance (AUC of 0.54) when using the original US cohort. All experiments reported a severe drop of performance when evaluated in the original and not the curated US cohort.

Discussion

The morphology of the LV endocardium learned from synthetic anatomies can be used to detect foreshortening in real clinical data.

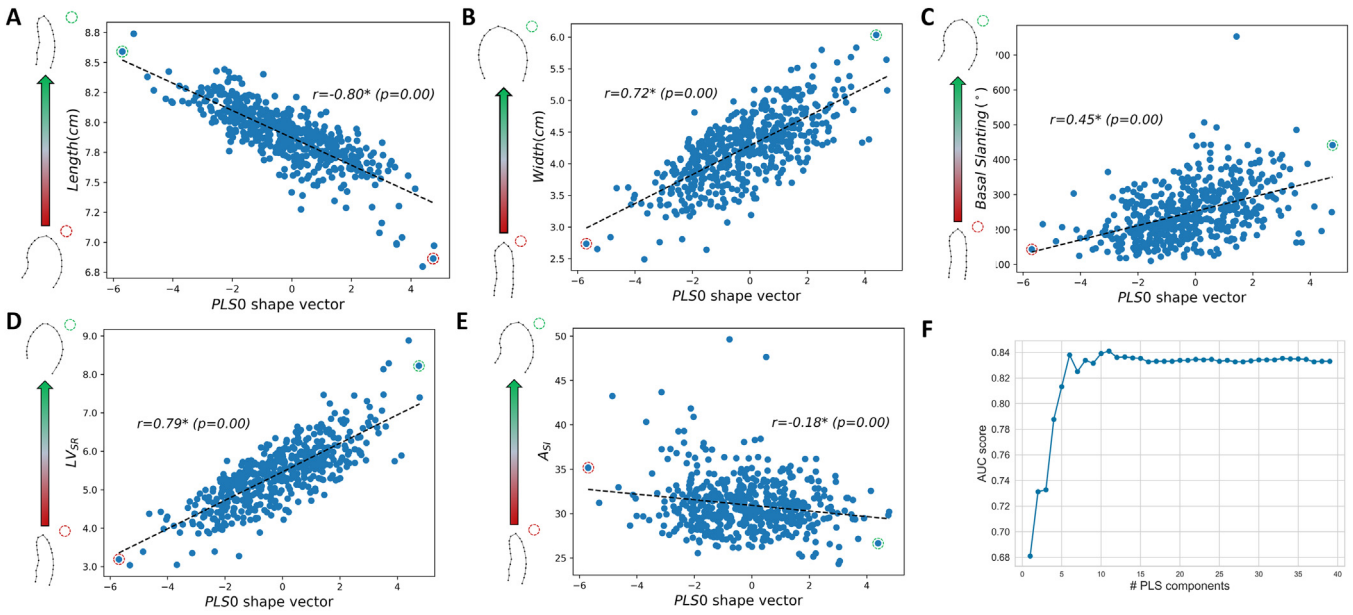

Figure 8. (a–e) Correlation of geometric descriptors with the first partial least squares (PLS) shape scores on the real echocardiographic left-ventricular (LV) contours. (f) Plot of number of PLS components vs. area under the receiver operating characteristic curve (AUC) score.

Foreshortening labelling is a challenging task for the human observer, and synthetic generation of high-quality ground truth anatomical and foreshortening data is a valid strategy to tackle it. Results reveal a potential for automated methodologies to improve quality control and thus reduction of error margins in echocardiography.

Strengths: maximising the overlap between synthetic variability and curated data

Both acquisition issues in real US data and failures from automatic or manual contouring will lead to aberrant cases of the LV endocardial morphology. This was a reasonably prevalent phenomenon in our real-world data set from the WASE-COVID study due to the front-line nature of the acquisitions [23]. Aberrant cases will introduce spurious sources of variability not reflective of the underlying true morphology and will thus hinder the extraction of the desired traits of foreshortening. On the other hand, the annotation of F/NF by visual inspection of a recording is a challenging and arguably ill-posed task for a human observer, as experimentally verified in this work. These are the fundamental reasons why there are not known well-curated and annotated data sets to address the problem of foreshortening, which motivated the quest to synthesise such

Table 1
Performance of the various cohorts, methods and metrics used for labelling of foreshortening

Training	Model form	Features	Variable predicted on	Tested on	AUC
CT cohort	PLS_{11}–Log	PLS scores	$ErrVol_{\%}$	CT cohort	0.95
				Original US cohort	0.67
				Curated US cohort	**0.84**
CT cohort	PLS_{11}–Log	PLS scores	Length	CT cohort	0.92
				Original US cohort	0.60
				Curated US cohort	**0.75**
CT cohort	Random forest	Geometrics	$ErrVol_{\%}$	CT cohort	0.91
				Original US cohort	0.55
				Curated US cohort	**0.73**
Curated US cohort	PLS_{11}–Log	PLS scores	Expert labels	CT cohort	0.75
				Original US cohort	0.60
				Curated US cohort	**0.68**
Original US cohort	PLS_{11}–Log	PLS scores	Expert labels	CT cohort	0.53
				Original US cohort	0.57
				Curated US cohort	0.54

AUC, area under the receiver operating characteristic curve; CT, computed tomography; PLS, partial least squares; US, ultrasound.
The random forest classifier was used to predict foreshortening using solely the geometric descriptor features as input to the model (length, width, LV_{SR} and A_{SI}), using the number of estimators as 100. PLS_{11}–Log refers to a PLS model with the first 11 shape modes for feature extraction, followed by a logistic regressor for predicting foreshortening. For the classification using left-ventricular long axis (Length), a threshold equivalent to the average length measurements across all cases was used to dichotomize the labels into foreshortened and non-foreshortened. For additional findings on the performance of the random forest model, refer to Table S1 (online only). Values in boldface are significant.

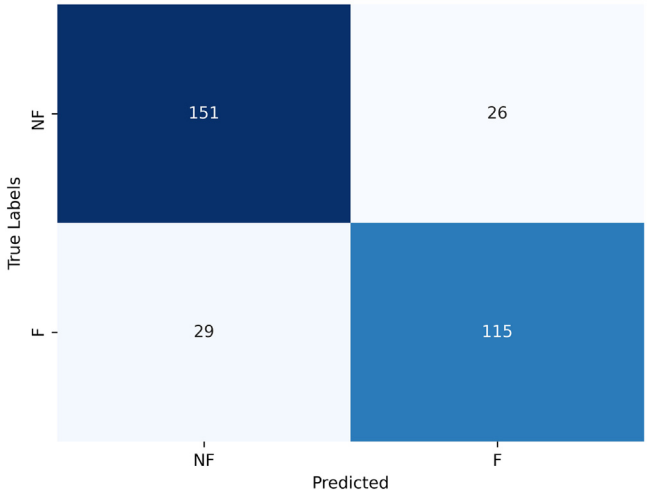

Figure 9. Confusion matrix of the foreshortening classification task, with an area under the receiver operating characteristic curve of 0.84. F, foreshortened; NF, non-foreshortened.

data. The key question is the extent to which the synthetic data generated were representative of the real problem.

Our approach was to systematically extract a clean data set of 2-D LV contours with controlled foreshortening conditions from ground truth CT anatomical models, free from any acquisition, segmentation and annotation errors. However, it was equally important to curate our data set of echocardiographic LV shapes through an outlier detection method that discarded anomalous cases using PCA. This data curation step proved to be necessary to improve the classification process: anomalous shape instances, mostly from noisy apical views, introduced a significant bias towards severely foreshortened shapes, leading to incorrect inferences and mis-classifications.

An alternative approach to maximise the required overlap between synthetic and real data would be the use of segmented contours from synthetic US images generated from the CT anatomies. The rationale would be to include the sources of segmentation uncertainty in the real data into the learning of the PLS features that predict foreshortening. Further research is needed to find the optimal learning strategy of synthetic foreshortening data.

The risks of synthetic data

It is important to stress, however, that the CT anatomies used were sampled from an SSM built from a cohort of asymptomatic subjects without any known cardiac condition; as such, it is not representative of the larger human anatomical variability. Diseased hearts exhibit high morphological variability and can manifest structural features that overlap with the geometrical changes of foreshortening (*e.g.,* apical hypertrophy), making the sources of anatomical variation more difficult to identify [24].

The overlap between the synthetic and real features may also be affected by the fact that there are specific anatomical details in the CT cardiac meshes that are not consistent with the shapes observed in the US cohort. Most notably is the basal region of the LV: 2DE images frequently render a blurry detail because of the depth, whereas CT-based reconstructed meshes show a high level of detail near the mitral annulus. This discrepancy in basal structure may have affected the mapping of anatomical variation from the virtual to the US domain.

Table 2
Evaluation of inter-rater agreement among 14 observers on a subset of the un-curated ultrasound data set

	Class label	Reported findings, n (%)	SA (95% CI)	Gwet AC1 (95% CI)
29 clips	F	44 (10.8%)	**16.4 (15.7, 17.1)**	76.1 (66.2, 86.0)
	NF	331 (81.5%)	82.0 (81.3, 83.8)	
21 clips[a]	F	32 (10.9%)	**14.4 (13.6, 15.3)**	77.0 (52.0, 98.7)
	NF	243 (83.7%)	82.8 (81.8, 83.8)	
8 clips[b]	F	12 (10.7%)	**21.8 (18.7, 24.9)**	75.7 (64.3, 87.1)
	NF	88 (78.6%)	79.9 (77.0, 82.8)	

CI, confidence interval; F, foreshortened; Gwet AC1, Gwet's coefficient of agreement for categorical data; NF, non-foreshortened; PCA, principal components analysis; SA, specific percentage of agreement.

[a] Agreement on 21 out of the 29 clips not discarded by PCA outlier removal.

[b] Agreement on 8 out of the 29 clips discarded by PCA outlier removal.

Beyond the A4C endocardial model

This work reports good performance in the detection of foreshortening only using the endocardial contour at end-diastole, without any further information of the epicardium or major cardiac structures, and without temporal information. Other researchers have shown the importance of the left-atrium size relative to the LV in quantifying foreshortening severity [25,26]. The epicardial contour and temporal domain have also been shown to be useful in detecting the extra thickening of the apical region [10].

More importantly, the model considered only the A4C view, which is the first view to be acquired from the apical window. The probe placement at the A4C defines the pivot for all other apical views, and it is therefore crucial to detect foreshortening at this acquisitional stage; consequently, if the interpreting clinician fails to pre-emptively identify foreshortening, its adverse effects can propagate to other views further beyond the A4C.

Construct validity: interpretation of the PLS features

The model was tested on an independent data set of real echocardiographic images and achieved a high classification accuracy while still being able to provide interpretable and clinically relevant features associated with foreshortening. In the first PLS shape mode, we recognized aspects of long-axis shortening and a rounding of the apex, which was corroborated by the strong correlation of PLS1 scores with LV_{SR} and length measurements. This particular mode, which explained the largest portion of shape variance, was in line with the morphological traits commonly described in foreshortened views [9]. A_{SI} and $ErrVol_{\%}$ were also moderately effective in explaining PLS1 shape variations, although with lower correlation than LV_{SR} and length. This suggests that area and volume measurements captured superfluous shape information that de-emphasized the changes in long-axis lengths. Basal slanting was also a prevalent feature in PLS4, which indicated a non-perpendicular alignment between the mitral valve planes and the long-axis direction in the cardiac meshes. This has been previously reported in studies assessing the challenges of mitral valve interventional techniques [27].

Although length and LV_{SR} could be used as *ad hoc* labels of foreshortening due to their high correlation with the ground truth F_θ angles, $ErrVol_{\%}$ values provided a direct correspondence with foreshortened views that were detrimental to LV volume and EF measurements, which were also derived from the Simpson's bi-plane formula. Additionally, length-derived labels, despite needing only a single A4C view, and thus a much simpler metric to measure than $ErrVol_{\%}$, cannot be used to distinguish between naturally rounded and foreshortened LVs, as both structures share overlapping traits. This is demonstrated by the drop in classification performance when using length measurements to predict foreshortening.

Are humans really able to detect foreshortening?

Our inter-reliability study revealed an exceptionally poor inter-observer agreement of the F-class annotations (Table 2). The agreements

were based on manual annotations by 14 observers across 29 A4C clips and, despite the limited range of appearances and characteristics of US images captured within this small subset, the analysis showcases the high degree of subjectivity in determining foreshortening in real echocardiographic images.

Only 8 out of these 29 studies were automatically discarded by our outlier removal step, suggesting that the poor reproducibility between observers is not so much caused by insufficient image quality, but rather an inherent difficulty in inferring probe placement solely from a 2-D perspective. The manual annotations were performed in conjunction with the A2C and A3C views; however, there was no access to quantitative systolic measurements such as LVEF to confirm the severity of foreshortening. Additionally, although foreshortening was recorded in accordance with clinical practice [15], there is currently no widely accepted threshold for determining the extent of discrepancy in long-axis length, relative to the other views, that warrants labelling a view as foreshortened. Notably, this discrepancy may not be readily apparent if all three views are concurrently foreshortened, a common scenario considering the A4C constitutes the pivot for the A2C and A3C views.

On the contrary, in the 3-D cardiac meshes of this study, foreshortening had a clean mathematical definition based on the threshold level of 7% of the $ErrVol_{\%}$. The lack of a defined clinical threshold for foreshortening, along with huge human variability and our associated small sample size to cope with it, were the main reasons the regression model suffered from a lack of generality, most notably when the anatomical variation associated with foreshortening was learned from the US data set (Table 1).

Study limitations and future work

This study does not solve the problem of foreshortening detection caused by the reasons discussed previously, which are mainly the risks of the synthetic data and the challenge of annotation and curation of real data. Therefore, it remains to be determined whether the foreshortening features and detection rate reported here will remain across larger patient populations and cardiovascular conditions.

In the context of PLS analysis, the interpretation of PLS scores by visual inspection can be challenging, particularly when the variables exhibit high collinearity [28,29], as is the case with the geometric descriptors, most of which are area- and length-derived measurements. Univariate analysis was deemed appropriate for interpreting the relationship between each descriptor and PLS score separately, enabling a more comprehensive understanding of each descriptor's contribution to the model. Nonetheless, future work should explore multivariate analysis, which considers the complex interdependencies among descriptors, in contrast to its univariate counterpart.

Additionally, in the interest of providing better interpretability of our model, we opted to employ a simple linear PLS regressor that is less complex than current state of the art classifiers. This decision was motivated by the need to mitigate the risk of overfitting, which was of particular concern due to the limited sample size of our testing set. However, considering the availability of a larger data set, forthcoming research should employ more intricate models, including ensemble of classifiers, to optimize performance.

The inter-observer variability analysis was conducted on a data set that exhibited significant imbalance, primarily due to a limited number of clips that were annotated by multiple observers. As a result, the variability outcomes may not generalize to the complete balanced US data set. Although augmenting the number of multi-annotated clips could potentially alleviate the class imbalance, the annotation of real data sets represents a considerable undertaking, and we are unaware of any openly accessible data sets of this nature.

Lastly, future work should also explore the use of more anatomical and temporal traits (*e.g.,* epicardial or atrial contours, additional views) and challenge the ability to predict foreshortening as a continuous variable beyond the binary, and currently somewhat arbitrary, classification task between F/NF.

Conclusion

We have presented a method to train a classifier to predict the presence of foreshortening, with zero manual labelling or acquisitions required and using only end-diastolic A4C LV contours. The classifier achieved an AUC of 0.84 in a curated external cohort of real 2DE images and provided clinically interpretable features of foreshortening. This novel method could simplify and improve future detection of foreshortening.

Acknowledgments

The authors acknowledge funding from Ultromics Limited, the EPSRC Centre for Doctoral Training in Medical Imaging (EP/L015226/1) and the Wellcome/EPSRC Centre for Medical Engineering (WT203148/Z/16/Z). P.La. holds a Wellcome Trust Senior Research Fellowship (209450/Z/17/Z).

Conflict of interest

W.-J.C.K. was a PhD student funded at a 50% level by Ultromics Ltd., and is currently an employee as A.B. A.M. and R.S. are. P. Leeson is the Academic Founder and Non-Executive Director of Ultromics.

Data availability statement

The source code and anatomical models are available to other researchers. Source code: https://github.com/woojinchokimm/foreshortening_detection.git. Anatomical models (virtual LV endocardial contours with matching ground truth foreshortening data): https://sandbox.zenodo.org/record/1109115#.YzRvCnbMI_M.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.ultrasmedbio.2023.05.003.

References

[1] Lang RM, Badano LP, Mor-Avi V, Afilalo J, Armstrong A, Ernande L, et al. Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. J Am Soc Echocardiogr 2015;28:1–39.e14.
[2] Flachskampf FA, Myrianthefs P, Beyer R, Myrianthefs PM. Echocardiography and thoracic ultrasound, 1. Oxford, UK: Oxford University Press; 2017.
[3] Kurt M, Shaikh KA, Peterson L, Kurrelmeyer KM, Shah G, Nagueh SF, et al. Impact of contrast echocardiography on evaluation of ventricular function and clinical management in a large prospective cohort. J Am Coll Cardiol 2009;53:802–10.
[4] Nagata Y, Yuichiro K, Takeshi O, Kyoko O, Akemi N, Yutaka O, et al. Impact of image quality on reliability of the measurements of left ventricular systolic function and global longitudinal strain in 2D echocardiography. Echo Res Pract 2018;5:28–39.
[5] Tsang W, Lang RM, Kronzon I. Transthoracic echocardiography tomographic views. In: American Society of Echocardiography, editor. ASE's comprehensive echocardiography, 3rd ed. Philadelphia, PA: Elsevier Publishing; 2021. p. 56–9.
[6] Kim WJC, Beqiri A, Lewandowski AJ, Puyol-Antón E, Markham DC, King AP, et al. Beyond Simpson's rule: accounting for orientation and ellipticity assumptions. Ultrasound Med Biol 2022;48:2476–85.
[7] Amzulescu MS, Slavich M, Florian A, Goetschalckx K, Voigt JU. Does two-dimensional image reconstruction from three-dimensional full volume echocardiography improve the assessment of left ventricular morphology and function? Echocardiography 2013;30:55–63.
[8] Tarr A, Stöbe S, Trache T, Kluge JG, Varga A, Pfeiffer D, et al. The impact of foreshortening on regional strain—a comparison of regional strain evaluation between speckle tracking and tissue velocity imaging. Ultraschall Med 2013;34:446–53.
[9] Ünlü S, Duchenne J, Mirea O, Pagourelias ED, Bézy S, Cvijic M, et al. Impact of apical foreshortening on deformation measurements: a report from the EACVI-ASE Strain Standardization Task Force. Eur Heart J Cardiovasc Imaging 2020;21:337–43.
[10] Smistad E, Ostvik A, Salte IM, Melichova D, Nguyen TM, Haugaa K, et al. Real-time automatic ejection fraction and foreshortening detection using deep learning. IEEE Trans Ultrason Ferroelectr Freq Control 2020;67:2595–604.
[11] Corral-Acero J, Margara F, Marciniak M, Rodero C, Loncaric F, Feng Y, et al. The 'Digital Twin' to enable the vision of precision cardiology. Eur Heart J 2020;41:4556–64.

[12] Rodero C, Strocchi M, Marciniak M, Longobardi S, Whitaker J, O'Neill MD, et al. Virtual cohort of 1000 synthetic heart meshes from adult human healthy population, <https://zenodo.org/record/4506930>; 2021 [accessed 08.09.22].
[13] Schiller NB, Shah PM, Crawford M, DeMaria A, Devereux R, Feigenbaum H, et al. Recommendations for quantitation of the left ventricle by two-dimensional echocardiography. American Society of Echocardiography Committee on Standards, Subcommittee on Quantitation of Two-Dimensional Echocardiograms. J Am Soc Echocardiogr 1989;2:358–67.
[14] Zheng Y, Barbu A, Georgescu B, Scheuering M, Comaniciu D. Four-chamber heart modeling and automatic segmentation for 3-D cardiac CT volumes using marginal space learning and steerable features. IEEE Trans Med Imaging 2008;27:1668–81.
[15] Leeson P, Augustine D, Mitchell ARJ, Becher H. Transthoracic anatomy and pathology: chambers and vessels. OSH echocardiography. Oxford, UK: Oxford University Press; 2012. p. 191–341.
[16] Lee LC, Liong CY, Jemain AA. Partial least squares-discriminant analysis (PLS-DA) for classification of high-dimensional (HD) data: a review of contemporary practice strategies and knowledge gaps. Analyst 2018;143:3526–39.
[17] Gower JC. Generalized Procrustes analysis. Psychometrika 1975;40:33–51.
[18] Rosipal R, Krämer N. Overview and recent advances in partial least squares. In: Saunders C, Grobelnik M, Gunn S, Shawe-Taylor J, editors. Subspace, latent structure and feature selection. SLSFS 2005. Lecture Notes in Computer Science, 3940. Berlin/Heidelberg, Germany: Springer; 2006. p. 34–51.
[19] Karagodin I, Carvalho Singulane C, Woodward GM, Xie M, Tucay ES, Tude Rodrigues AC, et al. Echocardiographic correlates of in-hospital death in patients with acute COVID-19 infection: the World Alliance Societies of Echocardiography (WASE-COVID) Study. J Am Soc Echocardiogr 2021;34:819–30.
[20] Upton R, Mumith A, Beqiri A, Parker A, Hawkes W, Gao S, et al. Automated echocardiographic detection of severe coronary artery disease using artificial intelligence. JACC Cardiovasc Imaging 2022;15:715–27.
[21] Feng GC. Factors affecting intercoder reliability: a Monte Carlo experiment. Qual Quant 2013;47:2959–82.
[22] Cicchetti DV, Feinstein AR. High agreement but low kappa: II. resolving the paradoxes. J Clin Epidemiol 1990;43:551–8.
[23] Asch FM, Descamps T, Sarwar R, Karagodin I, Singulane CC, Xie M, et al. Human versus artificial intelligence–based echocardiographic analysis as a predictor of outcomes: an analysis from the World Alliance Societies of Echocardiography COVID Study. J Am Soc Echocardiogr 2022;35 1226–37.e7.
[24] Turvey L, Augustine DX, Robinson S, Oxborough D, Stout M, Smith N, et al. Transthoracic echocardiography of hypertrophic cardiomyopathy in adults: a practical guideline from the British Society of Echocardiography. Echo Res Pract 2021;8:G61–86.
[25] Koka AR, Yau J, van Why C, Cohen IS, Halpern EJ. Underestimation of left atrial size measured with transthoracic echocardiography compared with 3D MDCT. AJR Am J Roentgenol 2010;194:W375–81.
[26] Kebed KY, Addetia K, Lang RM. Importance of the left atrium. Heart Fail Clin 2019;15:191–204.
[27] Loghin C, Loghin A. Role of imaging in novel mitral technologies—echocardiography and computed tomography. Ann Cardiothorac Surg 2018;7:799–811.
[28] Kim JH. Multicollinearity and misleading statistical results. Korean J Anesthesiol 2019;72:558–69.
[29] Yoo W, Mayberry R, Bae S, Singh K, Peter He Q, Lillard JW. A study of effects of multicollinearity in the multivariable analysis. Int J Appl Sci Technol 2014;4:9–19.