ESC European Society of Cardiology

European Heart Journal - Imaging Methods and Practice (2024) **2**, qyae042
https://doi.org/10.1093/ehjimp/qyae042

ORIGINAL ARTICLE

Joint shape/texture representation learning for cardiovascular disease diagnosis from magnetic resonance imaging

Xiang Chen [1], Yan Xia[1], Erica Dall'Armellina [2], Nishant Ravikumar[1,†], and Alejandro F Frangi [3,4,5,6,7,8,9,10,*†]

[1]School of Computing, University of Leeds, Woodhouse, LS2 9JT Leeds, UK
[2]School of Medicine, University of Leeds, Woodhouse, LS2 9JT Leeds, UK
[3]Christabel Pankhurst Institute, The University of Manchester, Oxford Rd, M13 9PL Manchester, UK
[4]Department of Computer Science, School of Engineering, The University of Manchester, Oxford Rd, M13 9PL Manchester, UK
[5]Division of Informatics, Imaging, and Data Sciences, School of Health Sciences, The University of Manchester, Oxford Rd, M13 9PL Manchester, UK
[6]NIHR Manchester Biomedical Research Centre, Manchester Academic Health Science Centre, Oxford Rd, M13 9PL Manchester, UK
[7]Medical Imaging Research Center (MIRC), University Hospital Gasthuisberg, UZ Herestraat 49 - bus 7003, 3000 Leuven, Belgium
[8]Department of Cardiovascular Sciences, KU Leuven, UZ Herestraat 49 - box 911, 3000 Leuven, Belgium
[9]Department of Electrical Engineering, KU Leuven, Kasteelpark Arenberg 10 postbus 2440, 3001 Leuven, Belgium
[10]Alan Turing Institute, British Library, 96 Euston Rd., NW1 2DB London, UK

Received 23 October 2023; accepted after revision 9 April 2024; online publish-ahead-of-print 14 May 2024

Abstract

Aims

Cardiovascular diseases (CVDs) are the leading cause of mortality worldwide. Cardiac image and mesh are two primary modalities to present the shape and structure of the heart and have been demonstrated to be efficient in CVD prediction and diagnosis. However, previous research has been generally focussed on a single modality (image or mesh), and few of them have tried to jointly consider the image and mesh representations of heart. To obtain efficient and explainable biomarkers for CVD prediction and diagnosis, it is needed to jointly consider both representations.

Methods and results

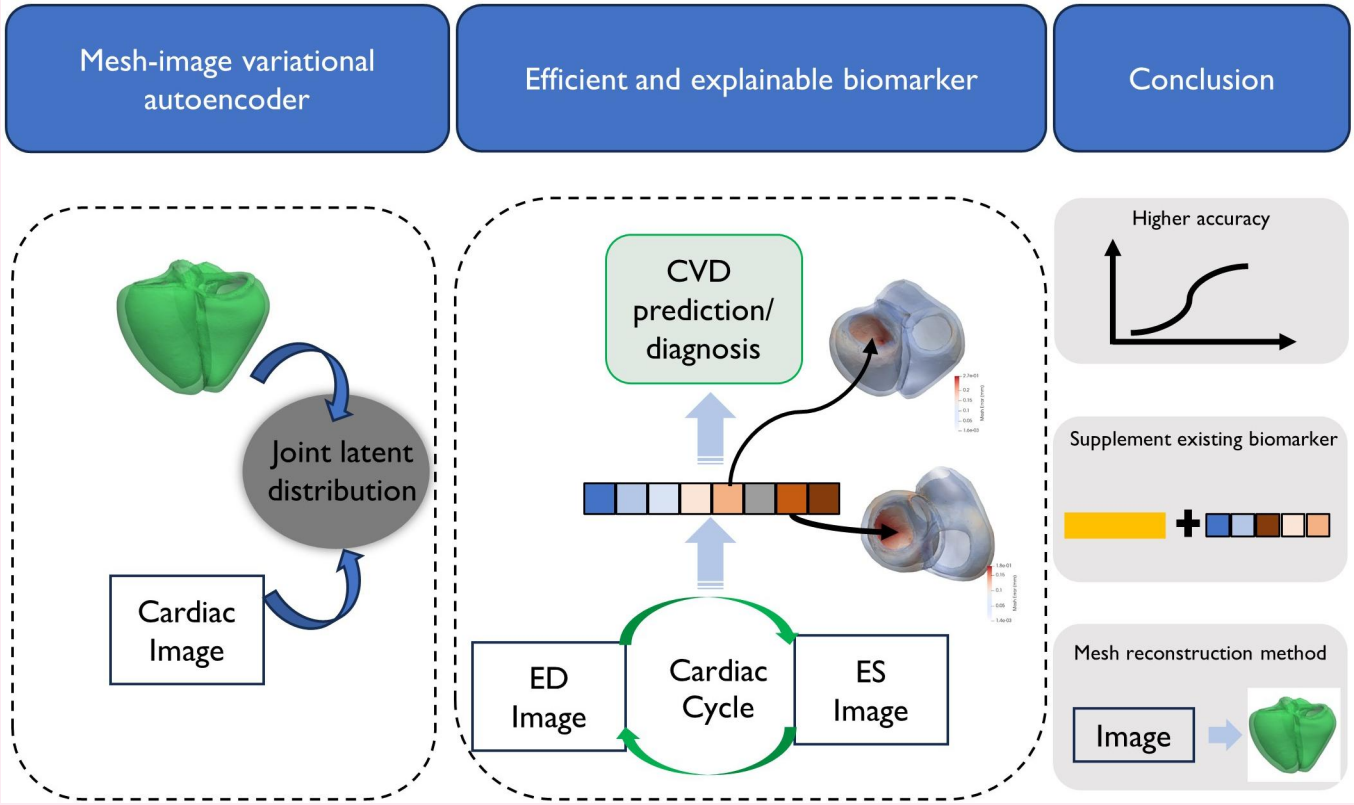
We design a novel multi-channel variational auto-encoder, mesh-image variational auto-encoder, to learn joint representation of paired mesh and image. After training, the shape-aware image representation (SAIR) can be learned directly from the raw images and applied for further CVD prediction and diagnosis. We demonstrate our method on data from UK Biobank study and two other datasets via extensive experiments. In acute myocardial infarction prediction, SAIR achieves 81.43% accuracy, significantly higher than traditional biomarkers like metadata and clinical indices (left ventricle and right ventricle clinical indices of cardiac function like chamber volume, mass, and ejection fraction).

Conclusion

Our mesh-image variational auto-encoder provides a novel approach for 3D cardiac mesh reconstruction from images. The extraction of SAIR is fast and without need of segmentation masks, and its focussing can be visualized in the corresponding cardiac meshes. SAIR archives better performance than traditional biomarkers and can be applied as an efficient supplement to them, which is of significant potential in CVD analysis.

* Corresponding author. E-mail: alejandro.frangi@manchester.ac.uk
† These authors are joint last authors.

Graphical Abstract



Keywords CVD diagnosis • CVD prediction • cardiac biomarker • multi-channel VAE • mesh reconstruction

Introduction

Cardiovascular disease (CVD) is the leading cause of global mortality. As non-invasive imaging is widespread in cardiology, deep learning (DL) combined with magnetic resonance (MR), computed tomography, and ultrasound are increasingly powerful in assessing and diagnosing heart-related diseases. MR is generally considered the gold standard among those image modalities due to its high contrast in anatomical structures and without ionising radiation.[1] Previous research has demonstrated the feasibility and efficiency of image-based diagnosis on various CVDs (e.g. heart failure and ischaemic heart disease).[2]

To analyse the CVDs from given images, numerous machine learning-based (ML) and DL-based approaches have been proposed, solving various tasks.[3] Automatic segmentation approaches[4] are widely studied to eliminate the time-consuming manual delineation work. Using predicted segmentation masks at end-diastole (ED) and end-systole (ES) of the cardiac cycle, clinical indices like left ventricle (LV) and right ventricle (RV) ejection fraction, ES and ED volume, and myocardial mass can be estimated. In addition, image registration can obtain the deformation fields between different time frames in the cardiac cycle for cardiac motion tracking and strain estimation.[5,6] Cardiac shape analysis based on the 3D cardiac mesh is also popular, which provides an intuitive way to observe and capture cardiac motion.[4]

Research on ML/DL-based CVD analysis can be roughly divided into three classes: direct disease diagnosis,[7] disease/survival prediction[2] (i.e. predict the probability of disease/death in a specific period from now), and association analysis between cardiac motion and diseases/genomes/other factors.[3] The direct disease diagnosis generally extracts feature descriptors from the original images/deformation fields/cardiac meshes and then uses classifiers [e.g. support vector machine (SVM)] for disease diagnosis.[7,8] For diagnosis, the extraction of biomarkers is essential, where some basic information (such as sex and age) and cardiac clinical indices derived from images/segmentation/deformation fields are generally used. Recently, DL-based approaches have outperformed traditional ML-based methods in various tasks.[5] However, few works have attempted to apply DL methods for direct cardiac disease diagnosis because of the lack of interpretability. Disease/survival prediction has similar feature extraction steps to disease diagnosis, aiming to predict the probability of getting CVDs in specific years or the survival time of CVD patients.[2] Instead of applying DL-based methods for classification/regression, previous studies[3] have explored using DL networks to predict cardiac motion phenotype from MR images and analysing its correlation to specific factors (e.g. genetic and environmental factors).

Current CVD prediction/diagnosis is generally based solely on the image-derived intensity or morphological properties. Cardiac MR images provide high-quality local details. However, limitations like large slice thickness (for cardiac cine MR image), slice misalignment, interference of background tissues, and inability to visualize 3D shape may weaken the interpretability and performance of CVD diagnosis. In contrast, a cardiac mesh can effectively represent the morphology of cardiac structures, provide more structure priors, and facilitate the assessment of cardiac motion,

while the reconstructed meshes (derived from cardiac images) may introduce inaccurate results on local details (due to the nature of mesh reconstruction). Therefore, a natural idea is to synergistically leverage the advantages of both representations to enhance subsequent prediction and diagnosis, attaining optimal performance.

To obtain explainable and efficient representations from cardiac MR images and improve CVD prediction/diagnosis performance, we propose a mesh-image variational auto-encoder (MIVAE) to learn the joint latent representations of cardiac meshes and cine MR images. After training, using images alone as input, the learned latent embedding of images, which we named shape-aware image representation (SAIR), is fed into ML classifiers for downstream tasks like CVD prediction and diagnosis. In addition, MIVAE can be applied as a mesh reconstruction approach by feeding images only as input. We demonstrate our method on UK Biobank (UKBB),[9] Automatic Cardiac Diagnosis Challenge (ACDC),[10] and Multi-Centre, Multi-Vendor & Multi-Disease Cardiac Image Segmentation Challenge (M&M)[11] datasets.

The contributions of this paper are summarized as follows,

- We propose a novel, MIVAE, for cross-modality data (mesh and image), comprising a convolutional and graph encoder–decoder. To the best of our knowledge, this is the first study to learn the joint latent variable from 3D cardiac images and surface meshes, and it is generic and can be applied to other organs.
- Our MIVAE learns a novel, efficient, robust cardiac feature representation, SAIR. We demonstrate that it can improve predictive performance than existing biomarkers and can supplement the latter in predictive diagnostic tasks.
- Using the learned MIVAE, we can automatically reconstruct cardiac biventricular meshes from MR images, providing a novel method for cardiac mesh reconstruction.

Related work

This work mainly relates to joint representation learning and CVD prediction/diagnosis.

Joint representation learning

Joint representation learning is a process of learning a parametric mapping from different data in multi-domains (e.g. images and text) to feature vectors/tensors in a shared latent space, aiming to apprehend the latent correlation between multi-modality data and extract more refined and valuable features. It has drawn attention from various cross-modality applications in different domains, such as population clustering[12] and disease diagnosis.[13,14] The architecture of joint representation learning may significantly differ due to input differences. Nevertheless, they generally encompass several distinct encoders, which encode data from different domains into the joint latent space. Corresponding to different inputs, the encoders are generally different [for example, convolutional layers for images, fully connected (FC) layers for vectors, and graph convolutional layers[15] for graph], aiming to convert the redundant input into a low-dimensional vector that can enhance downstream tasks. Multi-channel variational auto-encoder (MCVAE[13]) is a popular structure used in joint representation learning, including multiple encoder–decoder pairs, which encode data from different modalities into the same latent distribution. MCVAE can reconstruct missing channels when the input is incomplete, and the learned latent representations containing sufficient information from the input multi-modal data can be applied for subsequent analysis (e.g. disease diagnosis[13,14] and separating cell populations[12]).

Image-based CVDs diagnosis

To obtain accurate CVDs prediction/diagnosis, numerous ML/DL-based approaches[8] have been developed, feeding the features extracted from image and non-image data to a classifier/regressor. The non-image data (metadata) generally include demographic data (e.g. sex and age), conventional risk factors (e.g. smoking and hypertension), and other available data in the electronic health record. The classifiers can be ML classifiers like SVM and DL networks. CVD prediction and diagnosis models extract discriminative features/biomarkers representative of the patient data and their underlying target class of interest. The features can be divided into four categories: metadata, clinical indices, radiomic features,[8,16] and automatic features extracted by DL networks. The first is available in the electronic health record, while the rest are derived from the images. The details of metadata, clinical indices, and radiomic features can be found in *Figure 1*.

In most previous research,[17] metadata and clinical indices were widely used and achieved reasonable results. The clinical indices include LV ED volume, LV ES volume, LV ejection fraction, LV myocardium mass, RV ED volume, RV ES volume, and RV ejection fraction, generally computed based on the segmentation masks at ED and ES frames of cardiac cycle. Radiomic feature[8,18] is another popular feature derived from raw images and corresponding segmentation masks, referring to a collection of handcrafted features. It was originally used for cancer diagnosis[18] and recently applied in the diagnosis of CVDs.[8]

Recently, researchers have explored DL networks in CVD prediction/diagnosis, using automatic feature extraction instead of manual-designed feature extraction and achieving comparable results with that produced by cardiologists.[19] However, end-to-end classification/regression networks generally lack interpretability, as it is difficult to interpret learned features in a clinical sense and how they contributed to the diagnosis of CVDs. Besides, CVD refers to a group of complex diseases affecting cardiac structure and motion, generally requiring more than one frame for accurate diagnosis. It would bring a huge computation burden to incorporate all frames (each is a 3D image) of the cardiac cycle into a single network.

In this paper, we design a novel MCVAE to learn the joint latent representation of cardiac image- and shape-based features, where the impacts of each variable in the latent vector can be assessed by varying it and visualizing the resulting reconstructions. Using the learned image latent representation, SAIR, as a biomarker, we can achieve explainable CVD prediction/diagnosis. Different from radiomic features and clinical indices, the extraction of SAIR does not require detailed segmentation masks. Consequently, our proposed method does not propagate segmentation errors incurred to the learned features (which is the problem with clinical indices and radiomic features) and can fit more complex scenarios where segmentation masks may not be available.

Methods

In this section, we first introduce data preparation and the network architecture of MIVAE and then describe CVDs prediction and diagnosis with the SAIR learned from MIVAE.

Mesh-image variational auto-encoder

To learn the joint latent embedding of cardiac image and mesh, we design a MIVAE as shown in *Figure 2*. We follow two previous research[4,20] to prepare the input cardiac meshes and images of MIVAE (the corresponding details can be found in Supplementary data online, *Appendix A1*).

MIVAE is essentially an MCVAE,[13] consisting of two channels of encoder–decoder, the mesh channel and image channel, respectively. Given the input (i.e. cardiac mesh and image pairs), denoted as $\mathbf{x} = \{\mathbf{x}_{mesh}, \mathbf{x}_{img}\}$, the corresponding encoder (mesh and image encoder) would encode them into l-dimensional latent vectors $\mathbf{z}$. Subsequently, two corresponding decoders (mesh and image decoder) are applied to decode $\mathbf{z}$ to

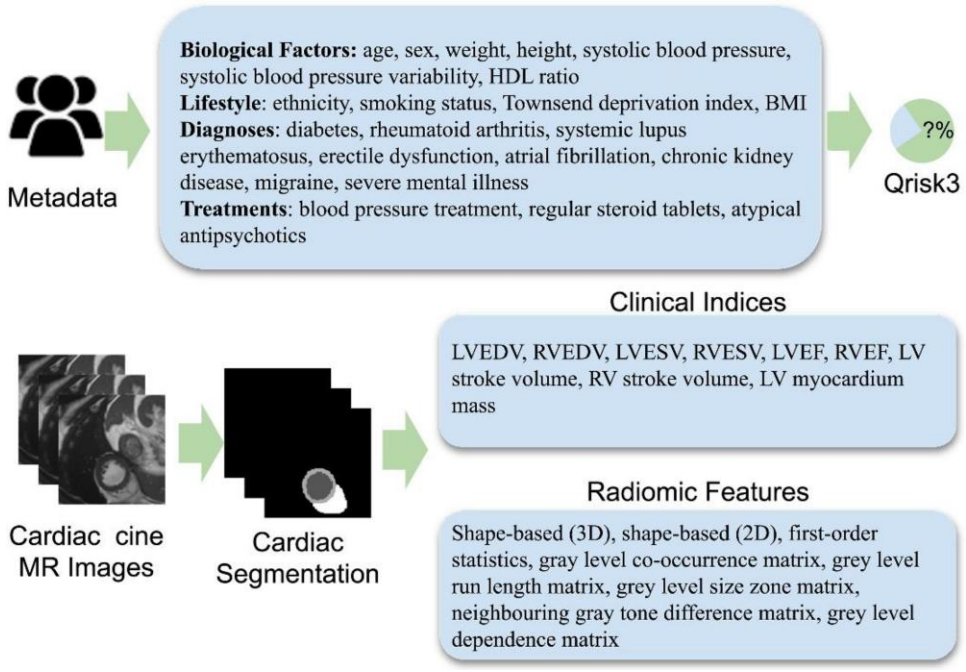


Figure 1 All traditional biomarkers used in this paper, including metadata, Qrisk3 (derived from metadata), clinical indices, and radiomic features. The cardiac magnetic resonance images were reproduced with the kind permission of the UK Biobank.

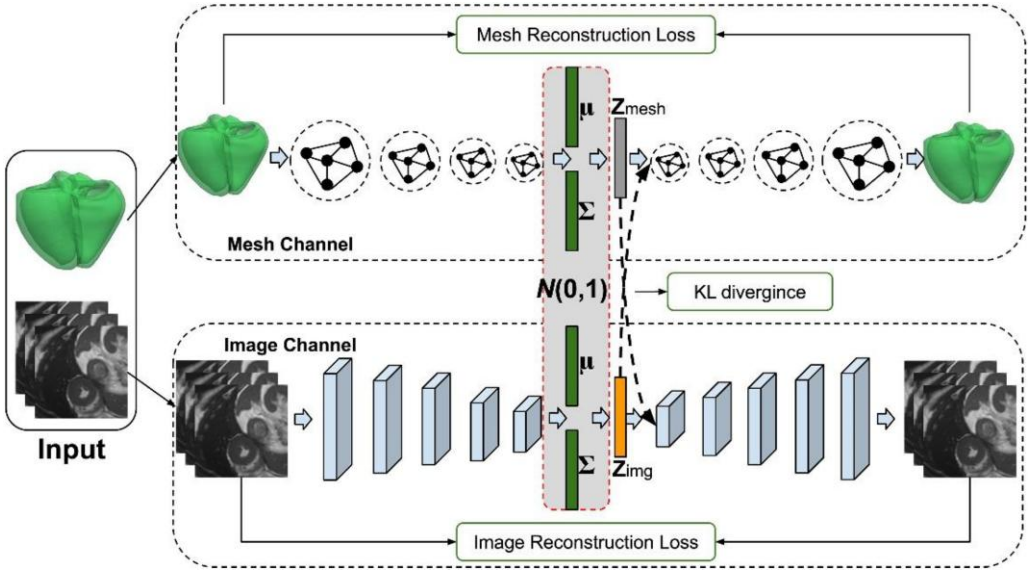


Figure 2 Schema of MIVAE. Our MIVAE includes two channels, the mesh encoder–decoder and image encoder–decoder, respectively. The latent variable $\mathbf{z}_{img}$ learned in the image channel (SAIR) is used for subsequent CVDs diagnosis. The cardiac MR images were reproduced by kind permission of UK Biobank©.

the reconstructed results, denoted as $\mathbf{x}' = \{\mathbf{x}'_{mesh}, \mathbf{x}'_{img}\}$. The generative process for the observation is formulated as

$$z \sim p(\mathbf{z}), \tag{1}$$

$$\mathbf{x}_c \sim p(\mathbf{x}_c|\mathbf{z}, \boldsymbol{\theta}_c), \quad \text{for } c \text{ in } \{\text{mesh}, \text{ img}\}, \tag{2}$$

where $p(\mathbf{z})$ is the prior distribution of latent vector $\mathbf{z}$ and $p(\mathbf{x}_c|\mathbf{z}, \boldsymbol{\theta}_c)$ is a likelihood distribution for the observations conditioned on the latent variable. The likelihood functions belong to a distribution family P parameterized by the set $\boldsymbol{\theta} = \{\boldsymbol{\theta}_{mesh}, \boldsymbol{\theta}_{img}\}$.

As deriving the posterior $p(\mathbf{z}|\mathbf{x}, \boldsymbol{\theta})$ is not always computable analytically, variational inference is used to compute an approximate posterior. We approximate the posterior distribution with $q(\mathbf{z}|\mathbf{x}_c, \boldsymbol{\phi}_c)$ (conditioned on single channel $\mathbf{x}_c$ and corresponding variational parameters $\boldsymbol{\phi}_c$), which belong to a distribution family Q parameterized by the set of parameters $\boldsymbol{\phi} = \{\boldsymbol{\phi}_{mesh}, \boldsymbol{\phi}_{img}\}$. Therefore, the

Table 1 Quantitative comparison of reconstruction performance

Methods	e_{mesh} (mm)	Average dice	LV dice	LVM dice	RV dice	HD (mm)	e_{image}
MIVAE (mesh as input)	0.67 ± 0.09	97.76 ± 0.37	98.01 ± 0.52	97.25 ± 0.50	98.01 ± 0.46	6.67 ± 3.28	0.172 ± 0.247
MIVAE (image as input)	3.56 ± 1.02	88.31 ± 3.07	**90.87 ± 2.79**	85.81 ± 3.70	88.24 ± 3.58	17.65 ± 9.65	0.155 ± 0.227
MCSI-Net[4] (image as input)	2.77 ± 1.23	90.28 ± 5.51	91.63 ± 5.75	88.76 ± 5.96	90.48 ± 5.19	14.35 ± 10.06	

The first and second rows are the results of MIVAE using only mesh or image as input, respectively. In the results of MIVAE using the image as the sole input, the bold highlights the results of MIVAE making no significant difference to MCSI-Net (*P* value larger than 0.05).

MIVAE is trained by maximizing the variational lower bound $\mathcal{L}(\boldsymbol{\theta}, \phi, \mathbf{x})$,

$$\mathcal{L}(\boldsymbol{\theta}, \phi, \mathbf{x}) = \mathbb{E}_c[L_c - \mathcal{D}_{\mathrm{KL}}(q(\mathbf{z}|\mathbf{x}_c, \phi_c)| \, |p(\mathbf{z}))], \tag{3}$$

where $\mathcal{D}_{\mathrm{KL}}$ is Kullback–Leibler divergence, used to impose a constraint enforcing each $q(\mathbf{z}|\mathbf{x}_c, \phi_c)$ to be as close as possible to the target posterior distribution. Here, L_c is the expected log-likelihood of decoding each channel from the latent representation of channel $\mathbf{x}_c$, generally formulated as

$$L_c = \mathbb{E}_{q(\mathbf{z}|\mathbf{x}_c,\phi_c)} \sum_{i=1}^{C} \ln \, p(\mathbf{x}_i|\mathbf{z}, \boldsymbol{\theta}_i). \tag{4}$$

As there is a mesh encoder–decoder in MIVAE, in addition to the log-likelihood, we further incorporate a mesh loss to ensure high-quality mesh reconstruction. The details about loss function can be found in Supplementary data online, *Appendix A2*.

Mesh channel encoder–decoder comprises mesh encoder and decoder that encodes the input mesh into a latent vector and then decodes it back to reconstruct the input 3D mesh. A mesh $\mathbf{M}(\mathbf{V}, \mathbf{F})$ is constructed by vertices $\mathbf{V}$ and faces $\mathbf{F}$. In this paper, all meshes are obtained by registering a template mesh, sharing the same faces. Therefore, we only need to predict the vertices of each mesh in the mesh encoder–decoder.

Both mesh encoder and mesh decoder are built using Chebyshev graph convolution[15] layers. The former is composed of four down-sampling blocks (sampling $\frac{1}{16}$, $\frac{1}{8}$, $\frac{1}{4}$ and $\frac{1}{4}$ points from the points in the previous layer, respectively), each comprising a graph convolution layer, a down-sampling layer, and an activation layer (exponential linear unit). Then a flatten operation followed by a FC layer is used to map the resulting features to an *l*-dimensional vector. The encoder predicts a distribution of an *l*-dimensional variable, parameterized by the mean $\boldsymbol{\mu}$ and standard variation $\boldsymbol{\sigma}$. Following the general variational auto-encoder (VAE),[13] the reparameterization trick is used, and consequently, an *l*-dimensional latent vector $\mathbf{z}_{\mathrm{mesh}}$ is sampled from the distribution. In the decoder, similarly, a FC layer followed by a reshape operation is utilized to recover the latent vector into a graph structure (i.e. shape like $N \times M$, where N is the number of points and M is the dimension of feature for each vertex). After that, corresponding to the encoder, four up-sampling blocks (comprising an up-sampling layer, a graph convolution layer, and an exponential linear unit activation layer) are used to up-sample the graph structure back to the original size of the input mesh.

Image channel encoder–decoder is a general convolution-based encoder–decoder. The input images are cardiac MR images in short-axis view, including a stack of slices. Due to the large slice thickness (the image spacing for cardiac MR images in UKBB is generally $1.8 \times 1.8 \times 10$ mm^3), we use 2D convolution instead of 3D convolution in the image encoder–decoder.

In image encoder, five down-sampling blocks are used, each comprising a convolution layer, batch-normalization layer, and activation layer (leaky rectified linear unit). Similarly, a flatten operation with an FC layer is used to turn the down-sampled features into an *l*-dimensional latent vector. Like mesh encoder–decoder, the reparameterization trick is also used, and an *l*-dimensional vector $\mathbf{z}_{\mathrm{img}}$ is sampled from the latent distribution. In the image decoder, the $\mathbf{z}_{\mathrm{img}}$ is fed into an FC layer followed by a reshape operation, turning into 4×4 feature maps. Corresponding to the image encoder, five up-sampling blocks, including the convolution layer, batch-normalization layer, and leaky rectified linear unit activation layer, are used to recover the feature back to the original images. The output layer is a convolution layer, followed by tanh() activation. Both $\mathbf{z}_{\mathrm{mesh}}$ and $\mathbf{z}_{\mathrm{img}}$ are fed into the mesh decoder and image decoder, predicting the reconstructed mesh and image of each channel.

CVD prediction/diagnosis

With MIVAE, we can extract the SAIR features for all the cardiac images from the cardiac cycle. Considering that most previous CVDs features/biomarkers (e.g. clinical indices) are extracted from cardiac images at ED and ES, we also compute SAIR at ED and ES frames of cardiac cycle for subsequent analysis. To demonstrate the efficiency of SAIR, we compare it with traditional CVDs biomarkers like clinical indices and radiomic features.

Using the learned SAIR as predictor, we can achieve CVD prediction/diagnosis with available classifiers. In this paper, we use SVM as the classifier, which outperforms random forest in our experiments. Limited by datasets, we evaluate the prediction performance of SAIR in UKBB and diagnosis performance in ACDC and M&M. Note that MIVAE are only trained on UKBB, without any re-training/fine-tuning on the external data. As the samples are limited in all three datasets, 10-fold cross-validation was used. In addition, the SAIR feature and radiomic features are both high-dimensional (over 500) feature vectors, which may lead to overfit on limited samples. Therefore, following Pujadas *et al.*,[8] we use sequential forward feature selection to identify the most important feature for CVD prediction/diagnosis in each feature vector.

Experiments

We demonstrate our method on three publicly available datasets, UKBB, ACDC, and M&M datasets (the implementation details and evaluation metrics can be found in Supplementary data online, *Appendices A3 and A4*).

Evaluation of image/mesh reconstruction

Before applying MIVAE to learn the joint latent representation of cardiac images and meshes, it is vital to ensure that the latent embedding can present sufficient information of input data, which can be reflected by the reconstruction quality. The quantitative results of the reconstruction are shown in *Table 1*. In MIVAE, each input channel has two outputs: the reconstructed image and reconstructed mesh, and after training, the trained MIVAE can only take one channel as input. Therefore, in the inference, we can use MR images/meshes alone as input and reconstruct the corresponding cardiac images and meshes. We find that image or mesh alone as input can reconstruct high-quality meshes. The reconstructed meshes using either meshes or images as input have low point-to-point errors to target meshes (same as input meshes), and even the meshes reconstructed from image input are with ~3.6 mm point-to-point error to target meshes. Applying the reconstructed meshes for cardiac MR image segmentation, the results of MIVAE using images as input are comparable with the MCSI-Net, with no significant difference in the LV dice score. Considering the nature of the variational auto-encoder and no ground-truth contour information is needed, SAIR captures sufficient information and is deemed suitable for subsequent CVD diagnosis.

In most realistic applications, there are only cardiac images available, without corresponding meshes. Our proposed MIVAE can reconstruct the corresponding accurate meshes from given images for multiple subsequent analysis tasks (e.g. segmentation, details can be found in Supplementary data online, *Appendix A5*). Meanwhile, the obtained latent embedding (SAIR) from MIVAE can be further applied for CVDs prediction and diagnosis.

Table 2 Quantitative comparison of AMI prediction results on UKBB between SAIR and the baseline features

Methods	Accuracy (%)	Precision (%)	F1(%)	Recall (%)	AUC
Metadata	65.79 ± 3.40	66.89 ± 3.66	65.17 ± 3.15	65.75 ± 3.17	66.88 ± 5.90
Qrisk3	65.41 ± 2.81	65.91 ± 2.91	65.26 ± 2.78	65.63 ± 2.80	69.91 ± 5.95
Qrisk3 + Metadata	67.67 ± 2.45	67.84 ± 2.29	67.62 ± 2.44	67.81 ± 2.29	71.69 ± 5.31
Clinical indices	57.52 ± 2.26	57.79 ± 1.98	57.35 ± 2.26	57.70 ± 2.01	61.14 ± 3.87
Radiomic features	79.47 ± 3.76	79.73 ± 3.63	79.38 ± 3.77	79.52 ± 3.74	82.50 ± 4.95
VAE	79.62 ± 2.41	79.63 ± 2.44	79.58 ± 2.41	79.66 ± 2.37	83.01 ± 3.88
SAIR	81.43 ± 2.93	81.47 ± 2.93	81.38 ± 2.91	81.46 ± 2.87	84.52 ± 3.68
All (without SAIR)	82.18 ± 2.73	82.54 ± 2.77	82.09 ± 2.72	82.23 ± 2.76	84.87 ± 4.73
All	83.38 ± 2.53	83.62 ± 2.64	83.32 ± 2.49	83.42 ± 2.50	85.10 ± 4.64

All (without SAIR) means combining all features (except SAIR) for classification, while all denotes using all features (including SAIR) in classification.

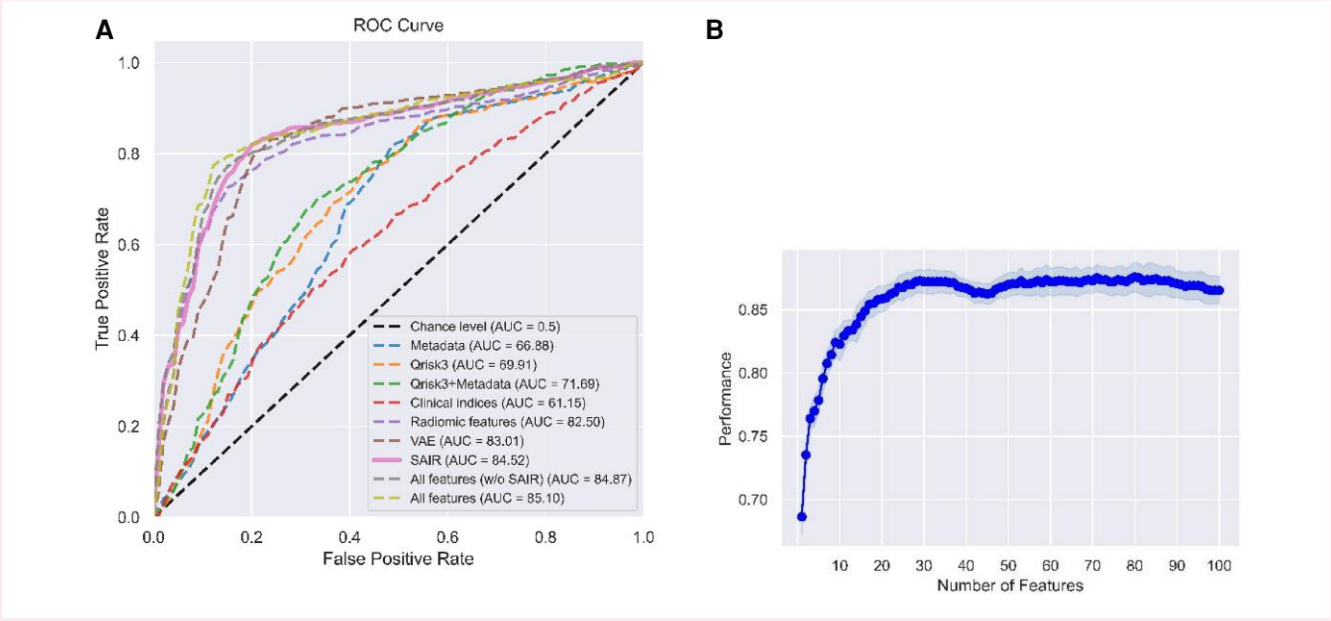


Figure 3 AMI prediction results on UKBB. (*A*) ROC curve of SAIR and baseline features. (*B*) Classification accuracy (mean with standard variation) with the increase of SAIR feature dimension. It achieves the best accuracy (87.59%) when selecting 80 features.

Acute myocardial infarction prediction on UKBB

With the learned SAIR from MIVAE, we implement CVD prediction/diagnosis by feeding it into SVM (using scikit-learn package, with radial basis function kernel and $C = 1.0$). We demonstrate its performance in the acute myocardial infarction (AMI) prediction, using cardiac MR images from UKBB. Here, we only use the SAIR at ED and ES frames of the cardiac cycle for each subject, as most traditional biomarkers (e.g. clinical indices) only use these two frames. We compare the performance of SAIR with traditional biomarkers used in CVD prediction, including metadata (following Carter *et al.*[21]), clinical indices, radiomic features, and Qrisk3[22] score (using Qrisk3 score as a feature), and explore the result of combining all features (with/without SAIR). To compare with DL-based approach, we also build a normal VAE to learn latent embeddings from images (denoted as VAE).

The rationale why we choose AMI is multifaceted. At first, AMI would lead to heart failure, which is the leading cause of death and disability globally, and thereby, a method to predict/prevent it in advance is of substantial clinical importance. Secondly, sufficient data regarding AMI in UKBB provide a solid basis for training and evaluating our method.

There are only 442 samples for the training and testing, which is a small number for SAIR (1024 dimensions) or radiomic features (2104 dimensions). Considering the curse of dimensionality, feature selection is needed. We apply a sequential forward feature selection method (following Pujadas *et al.*,[8] using Mlxtend package with 10-fold cross-validation) to select eight features from the given features, as the input to SVM. The classification results after feature selection are shown in *Table 2*, with the corresponding receiver operating characteristic (ROC) curve in *Figure 3A*. We observe that metadata performance surpasses clinical indices (65.79% vs. 57.52%). Qrisk3 score exhibits similar performance to metadata, as it is derived from the latter. The combination of metadata and Qrisk3 achieves a higher accuracy (67.67%) than the independent use of either one. Radiomic features outperform all traditional predictors, with 79.47% accuracy. The image feature learned by VAE leads to marginal higher accuracy than radiomics feature (79.62%). However, our learned SAIR performs better than all the single features (81.43%). Integrating all conventional features with SAIR (all features) yields improved AMI prediction accuracy than all features (without SAIR; 82.18% vs. 83.38%), which demonstrates that our learned SAIR can provide complementary information for traditional biomarkers. Statistical significance analysis is applied in all ROC curves, and we observe that SAIR significantly outperforms Metadata, Qrisk3, Qrisk3 + Metadata, and clinical indices, while it makes no significant difference to the performance using all features (0.05 as threshold).

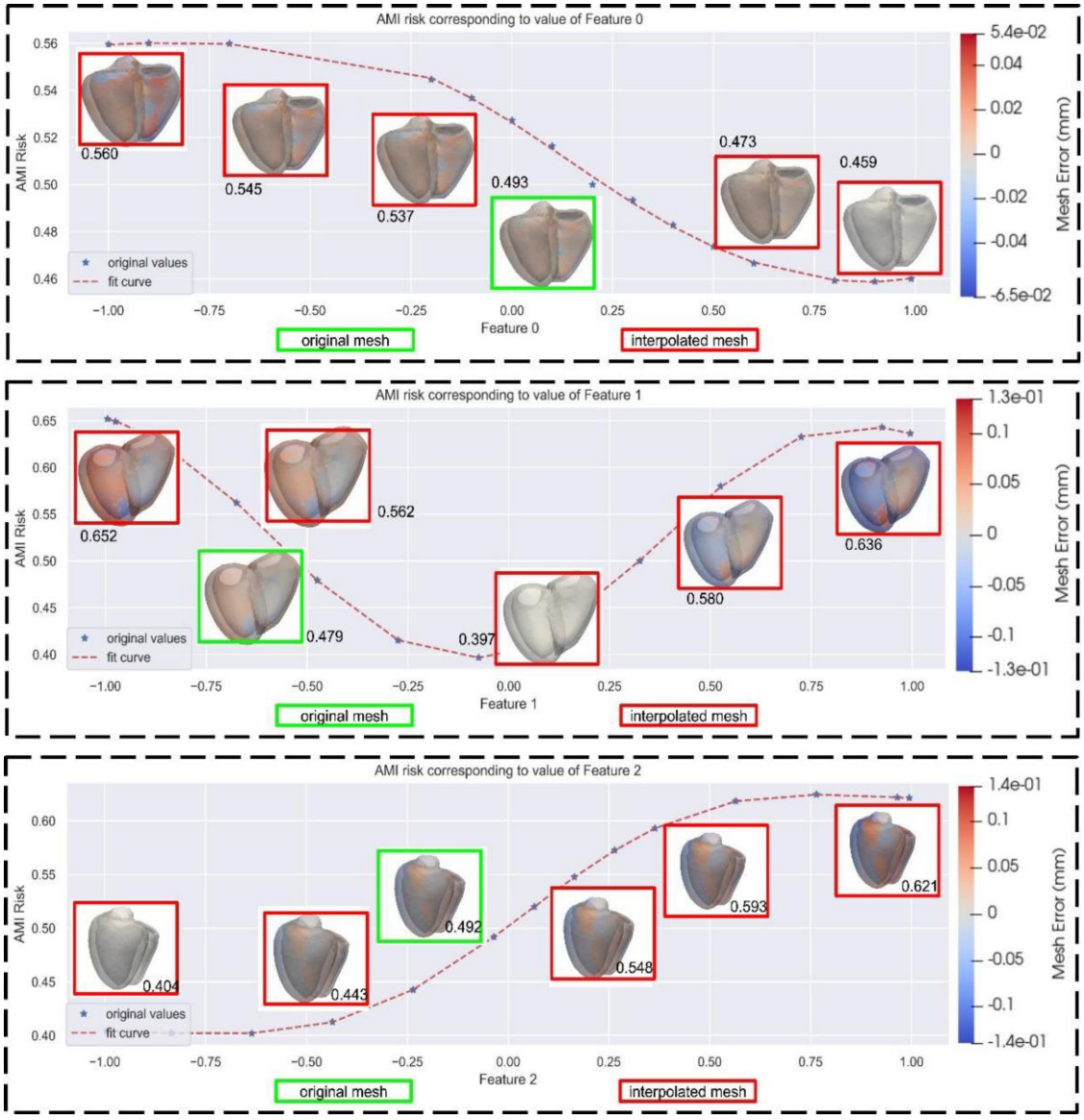


Figure 4 Visualization of the top three selected features. For each feature, we revised its value by interpolation (from −1.0 to 1.0, using standardization) and depicted the resultant reconstructed cardiac shapes alongside their corresponding AMI risks. Within each feature, the grey shape (the shape with the lowest AMI risk) is chosen as the reference shape. Variations in the other shapes are emphasized, with expansions marked in red and contractions in blue (as shown in the colour bar). The shapes in the red boxes are the interpolated shape, while the green box encloses the original shape of the patient's heart.

In addition, an accuracy curve along the SAIR feature dimension is also provided in *Figure 3B*. In the beginning, with more features, the classification accuracy increases. However, after 80 dimensions (where the accuracy is 87.59%), the accuracy tends to drop, which demonstrates the importance of feature selection. To understand how the SAIR features contribute to the CVDs diagnosis and the ventricular shape changes these features encode, we visualized the top three selected SAIR features (named as Features 0–2 for simplicity) by interpolation, with the corresponding shape changes and AMI risk in *Figure 4*. For each feature, we revised its value by interpolation (from −1.0 to 1.0, standardization is used) and depicted the resultant reconstructed cardiac shapes alongside their corresponding AMI risks. Within each feature, the grey shape (the shape with the lowest AMI risk) is chosen as the reference shape. Variations in the other shapes are emphasized, with expansions marked in red and contractions in blue (as shown in the colour bar). The shapes in the red boxes are the interpolated shape, while the green box encloses the original shape of the patient's heart. It can be found that Feature 0 predominantly affects the apex of the RV, Feature 1 alters the apex of the LV, and Feature 2 impacts the base of the LV.

To demonstrate the efficiency of our proposed method, we further demonstrate it on two other datasets, ACDC and M&M, without any fine-tune steps. The corresponding results can be found in Supplementary data online, *Appendix A6*, which demonstrates that our SAIR can further supplement the existing biomarkers. The images from ACDC and MM datasets are from multiple scanners and people with multiple diseases, closely resembling real-world clinical scenarios, which underscores the potential of our proposed method to contribute to current CVD diagnosis/prediction (with pre-processing and domain adaptation techniques to bridge the gap between real-world images and UKBB images).

Discussion

As a MCVAE, our proposed MIVAE can be applied to cross-domain reconstruction, reconstructing meshes from corresponding images or reconstructing images from corresponding meshes. In realistic scenarios, raw images are generally available, and therefore, our MIVAE can be used for image-to-mesh reconstruction. We have demonstrated that it can achieve comparable segmentation performance with the previous reconstruction approach. Still, there is a limitation for MIVAE, in that the predicted mesh of MIVAE is not aligned with the input images. Therefore, an additional approach like Xia *et al.*[4] to predict transformation parameters is required.

In CVD prediction, the learned SAIR shows significantly better results than traditional biomarkers in UKBB and better performance than specific traditional features/biomarkers in unseen images from other datasets. Although in non-UKBB data our SAIR performs worse than specific biomarkers, it does not require detailed ground-truth segmentation as radiomic features/clinical indices, resulting in more general and robust applications. In current experiments, we only plot the results of SAIR using ED and ES frames from cardiac cycle. Nevertheless, it is possible to incorporate more frames in the cardiac cycle to achieve better prediction/diagnosis performance. In our preliminary exploration, incorporating more frames of SAIR would not lead to higher prediction performance when selecting only eight features. However, more frames would lead to better performance when more features are selected. Besides CVDs prediction/diagnosis, our learned SAIR may also be applicable in finding potential sub-types of diseases using clustering.

Like other DL-based approaches, the main limitation of SAIR is the generalization of different datasets. While we can directly apply the trained MIVAE to extract SAIR of images from other sources, its performance would be weakened if the input images have significantly different appearances from images in UKBB. Pre-processing techniques like re-sampling and histogram matching can be applied to alleviate this decrease, and the exploration of domain adaptation or generalization techniques deserve to try.

Although several DL/ML approaches[8,17] have proposed to apply CMR imaging for CVD diagnosis/prediction, we proposed a novel biomarker considering the joint latent distribution of shape and image, SAIR, instead of transferring an existing biomarker from another domain to CVD prediction/diagnosis (e.g. radiomic features). While MIVAE offers a more accurate diagnostic approach for CVDs compared with traditional biomarkers, demonstrating the ability to illustrate correlations between regions of cardiac structures and CVDs, there are challenges to be addressed before its clinical application. Firstly, the proposed method remains trained on a relatively limited dataset. As other larger datasets become available, we would like to extend this work with replication studies that build more general evidence on its performance across real-world scenarios. Secondly, our study primarily focusses on AMI, representing only a subset of CVDs. Future work will expand our findings across a larger set of CVDs. Despite these limitations, this work establishes early evidence of the feasibility and a solid foundation for future research in this domain, serving as a valuable supplement to clinicians' decision-making processes.

CVDs constitute a complex group of disorders, making their prediction and diagnosis inherently challenging. DL approaches, when trained on large-scale data, are able to achieve reasonable and objective prediction/diagnosis for CVD analysis. Our proposed method can provide reliable prediction results while presenting the evidence by corresponding meshes, supporting the decision-making. There are three main potential applications at the moment: (i) With only cine MR images, our proposed method can predict the corresponding AMI risk, which is cheap and applicable in daily clinics to assess patients' risk. (ii) Our MIVAE can work as a novel mesh reconstruction approach, providing the cardiac structure to clinicians in a 3D view. (iii) Our method can be further applied to analyse the inherent correlation between CVDs and cardiac structures, as we have explored in *Figure 4*.

Conclusion

In this paper, to obtain efficient representations from cardiac MR images for subsequent CVD prediction and diagnosis, we designed a novel two-channel MCVAE, MIVAE, to learn joint latent representations of cardiac images and corresponding meshes, as a novel biomarker. After training, given MR images alone as input, our MIVAE can reconstruct high-quality biventricular meshes and learn SAIRs, useful for subsequent CVD prediction/diagnosis. Through experiments on UKBB, we demonstrated that the segmentation performance of our approach is comparable with previous approaches. The learned novel biomarker, SAIR, captures efficient representations from raw images for CVD prediction/diagnosis, leading to better performance than traditional biomarkers. Also, the learned SAIR feature captures information not contained within existing biomarkers (e.g. clinical indices), helping supplement existing biomarkers and improves predictive performance. We further demonstrated the robustness of SAIR on non-UKBB data (ACDC and M&M) and showed that it can enhance the performance of traditional predictors.

Supplementary material

Supplementary data are available at *European Heart Journal – Imaging Methods and Practice* online.

Consent

All data used in this study were obtained form the UKBB under the approved project protocols. UKBB obtained informed consent from all participants.

Funding

This research was conducted using the UK Biobank resource under Access Application no. 11350. A.F.F. was supported by the Royal Academy of Engineering under a RAEng Chair in Emerging Technologies (CiET1919/19) and by the Engineering and Physical Sciences Research Council under a Frontier Research Horizon Europe Guarantee Award (EP/Y030494/1). A.F.F. was also funded by the NIHR Manchester Biomedical Research Centre. The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the NIHR, or the Department of Health.

Conflict of interest: None declared.

Data availability

The data that support the findings of this study are available in the UK Biobank (UKBB).

Lead author biography

Xiang Chen received his B.S. degree in Electronics and Information Engineering in 2016 and the M.S. degree in Communication and Information System in 2019, both from Sichuan University, Chengdu, China. He has acted as the main participant in several projects in the areas of artificial intelligence (e.g. Identification of Unsound Kernels in Wheat). He was a member of the Image Information Institute of Sichuan University during his post-graduate studies. He joined the CISTIB research group

at the University of Leeds as a PhD student in September 2019 supervised by Prof. Frangi, Dr Ravikumar, and Dr Xia.

References

1. Khalil A, Ng S-C, Liew YM, Lai KW. An overview on image registration techniques for cardiac diagnosis and treatment. *Cardiol Res Pract* 2018;**2018**:1–15.
2. Bello GA, Dawes TJ, Duan J, Biffi C, De Marvao A, Howard LS *et al.* Deep-learning cardiac motion analysis for human survival prediction. *Nat Mach Intell* 2019;**1**:95–104.
3. Thanaj M, Mielke J, McGurk KA, Bai W, Savioli N, de Marvao A *et al.* Genetic and environmental determinants of diastolic heart function. *Nat Cardiovasc Res* 2022;**1**:361–71.
4. Xia Y, Chen X, Ravikumar N, Kelly C, Attar R, Aung N *et al.* Automatic 3D+t four-chamber CMR quantification of the UK biobank: integrating imaging and non-imaging data priors at scale. *Med Image Anal* 2022;**80**:102498.
5. Chen X, Diaz-Pinto A, Ravikumar N, Frangi AF. Deep learning in medical image registration. *Prog Biomed Eng* 2021;**3**:012003.
6. Chen X, Xia Y, Ravikumar N, Frangi AF. A deep discontinuity-preserving image registration network, International Conference on Medical Image Computing and Computer-Assisted Intervention. Strasbourg, France: Springer; 2021. p46–55
7. Sarmiento E, Pico J, Martinez F. Cardiac disease prediction from spatio-temporal motion patterns in cine-MRI, IEEE 15th International Symposium on Biomedical Imaging. Washington, DC, USA: IEEE; 2018. p1305–8.
8. Pujadas ER, Raisi-Estabragh Z, Szabo L, McCracken C, Morcillo CI, Campello VM *et al.* Prediction of incident cardiovascular events using machine learning and CMR radiomics. *Eur Radiol* 2022;**33**:3488–3500.
9. Petersen SE, Matthews PM, Francis JM, Robson MD, Zemrak F, Boubertakh R *et al.* UK biobanks cardiovascular magnetic resonance protocol. *J Cardiovasc Magn Reson* 2015; **18**:8.
10. Bernard O, Lalande A, Zotti C, Cervenansky F, Yang X, Heng P-A *et al.* Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: is the problem solved? *IEEE Trans Med Imaging* 2018;**37**:2514–25.
11. Campello VM, Gkontra P, Izquierdo C, Martín-Isla C, Sojoudi A, Full PM *et al.* Multi-centre, multi-vendor and multi-disease cardiac segmentation: the M&Ms challenge. *IEEE Trans Med Imaging* 2021;**40**:3543–54.
12. Ternes L, Dane M, Gross S, Labrie M, Mills G, Gray J *et al.* A multi-encoder variational autoencoder controls multiple transformational features in single-cell image analysis. *Commun Biol* 2022;**5**:1–10.
13. Antelmi L, Ayache N, Robert P, Lorenzi M. Sparse multi-channel variational autoencoder for the joint analysis of heterogeneous data, International Conference on Machine Learning. California: PMLR; 2019. p302–11.
14. Diaz-Pinto A, Ravikumar N, Attar R, Suinesiaputra A, Zhao Y, Levelt E *et al.* Predicting myocardial infarction through retinal scans and minimal personal information. *Nat Mach Intell* 2022;**4**:55–61.
15. Ranjan A, Bolkart T, Sanyal S, Black MJ. Generating 3D faces using convolutional mesh autoencoders, *Proceedings of the European Conference on Computer Vision (ECCV)*. Munich, Germany: Springer ; 2018. p704–720.
16. Van Griethuysen JJ, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V *et al.* Computational radiomics system to decode the radiographic phenotype. *Cancer Res* 2017;**77**:e104–7.
17. Juarez-Orozco LE, Knol RJ, Sanchez-Catasus CA, Martinez-Manzanera O, Van der Zant FM, Knuuti J. Machine learning in the integration of simple variables for identifying patients with myocardial ischemia. *J Nucl Cardiol* 2020;**27**:147–55.
18. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S *et al.* Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* 2014;**5**:1–9.
19. Kusunose K, Abe T, Haga A, Fukuda D, Yamada H, Harada M *et al.* A deep learning approach for assessment of regional wall motion abnormality from echocardiographic images. *Cardiovasc Imaging* 2020;**13**:374–81.
20. Chen X, Ravikumar N, Xia Y, Attar R, Diaz-Pinto A, Piechnik SK *et al.* Shape registration with learned deformations for 3D shape reconstruction from sparse and incomplete point clouds. *Med Image Anal* 2021;**74**:102228.
21. Carter AR, Gill D, Smith GD, Taylor AE, Davies NM, Howe LD. Cross-sectional analysis of educational inequalities in primary prevention statin use in UK biobank. *Heart* 2022; **108**:536–42.
22. Li Y, Sperrin M, van Staa T. R package 'QRISK3': an unofficial research purposed implementation of ClinRisk's QRISK3 algorithm into R. *F1000Res* 2020;**8**:2139.