

Published in final edited form as:

Med Image Anal. 2022 July 01; 79: 102465. doi:10.1016/j.media.2022.102465.

A multimodal deep learning model for cardiac resynchronisation therapy response prediction

Esther Puyol-Antón^{a,*}, Baldeep S. Sidhu^{a,b}, Justin Gould^{a,b}, Bradley Porter^{a,b}, Mark K. Elliott^{a,b}, Vishal Mehta^{a,b}, Christopher A. Rinaldi^{a,b}, Andrew P. King^a

^aSchool of Biomedical Engineering & Imaging Sciences, <https://ror.org/0220mzb33> King's College London, UK.

^bGuy's and St Thomas' Hospital, London, UK

Abstract

We present a novel multimodal deep learning framework for cardiac resynchronisation therapy (CRT) response prediction from 2D echocardiography and cardiac magnetic resonance (CMR) data. The proposed method first uses the 'nnU-Net' segmentation model to extract segmentations of the heart over the full cardiac cycle from the two modalities. Next, a multimodal deep learning classifier is used for CRT response prediction, which combines the latent spaces of the segmentation models of the two modalities. At test time, this framework can be used with 2D echocardiography data only, whilst taking advantage of the implicit relationship between CMR and echocardiography features learnt from the model. We evaluate our pipeline on a cohort of 50 CRT patients for whom paired echocardiography/CMR data were available, and results show that the proposed multimodal classifier results in a statistically significant improvement in accuracy compared to the baseline approach that uses only 2D echocardiography data. The combination of multimodal data enables CRT response to be predicted with 77.38% accuracy (83.33% sensitivity and 71.43% specificity), which is comparable with the current state-of-the-art in machine learning-based CRT response prediction. Our work represents the first multimodal deep learning approach for CRT response prediction.

Keywords

Multi-modality imaging; Cardiac resynchronisation therapy; Treatment response prediction; Multi-view deep learning

1 Introduction

Cardiac imaging techniques play a pivotal role in heart failure (HF) diagnosis, assessment of aetiology and treatment planning. Several modalities are available that are of relevance

This work is licensed under a [BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

*Corresponding author. esther.puyol_anton@kcl.ac.uk (E. Puyol-Antón).

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: EPA, BS, JG, BP, MKE, CAR, APK has nothing to declare. VM has received fellowship funding from Abbott, outside of the submitted work.

in patients with HF. Echocardiography is the first-choice imaging technique in the daily practice of cardiology as it is non-invasive, low-cost, easily available and provides most of the information required for the management and follow up of HF patients (Kirkpatrick et al., 2007). However, echocardiography has a number of limitations. First, image quality is heavily dependent on operator experience and expertise. Second, in normal clinical practice the images are two-dimensional (2D) meaning that geometrical assumptions are made in order to compute three-dimensional (3D) clinical metrics such as volumes and ejection fraction. Finally, echocardiography has limited spatial resolution with a relatively narrow field of view which can lead to poor endocardial definition.

Cardiac magnetic resonance (CMR) imaging is considered to be the gold standard in the evaluation of left ventricular (LV) function and is increasingly used in the assessment of HF due to its excellent temporal and spatial resolution and lack of ionizing radiation (Hundley et al., 2010). However, CMR is a time-consuming and expensive modality which requires significant technical expertise to operate, and hence has limited availability in some geographical areas.

Recently, machine learning (ML), and more specifically deep learning (DL) techniques have shown promising performance in a range of medical image analysis tasks (Litjens et al., 2017). In cardiac image analysis, most ML/DL techniques have considered the modalities of CMR and echocardiography in isolation. However, some works (Puyol-Antón et al., 2018; Puyol-Antón et al., 2017; Bruge et al., 2018) have shown that they can contain complementary information and so considering them in combination could have benefits in terms of performance. To date, these methods have been based upon traditional ML approaches such as multiview learning (Puyol-Antón et al., 2018).

However, recently, inspired by the success of DL methods in other applications, multimodal deep learning (MMDL) (Ramachandram and Taylor, 2017) has attracted significant research attention due to its ability to learn common features from multiple modalities, with the potential to exploit their natural strengths and reduce redundancies. In this paper we aim to use MMDL methods for prediction of response to Cardiac Resynchronisation Therapy (CRT), which is a common treatment for HF. Our aim is to use MMDL to produce an automated tool for CRT response prediction that utilises only echocardiography data as input at test time whilst also exploiting multimodal (CMR and echocardiography) data at training time.

1.1 Cardiac resynchronisation therapy

HF is a complex clinical syndrome associated with a significant morbidity and mortality burden. Cardiac remodeling is a pivotal process in the progression of HF, and it is defined as a change in the size, shape, or structure of one or more of the cardiac chambers. This remodelling can result in the development of dyssynchronous ventricular activation, which is often induced by electrical conduction delay in some regions of the LV and can lead to a decline in cardiac efficiency.

CRT is a common treatment for patients with heart failure with reduced ejection fraction (HFrEF) as it can restore LV electrical and mechanical synchrony. It has been shown

to increase quality of life, improve functional status, reduce hospitalisation, improve LV systolic function and reduce mortality in properly selected patients (Bristow et al., 2004; Cleland, 2005). While CRT is an effective therapy, approximately 30–40% of patients treated with CRT gain little or no symptomatic benefit from the treatment (Yancy et al., 2017; Ponikowski et al., 2016; McAlister et al., 2007; Parsai et al., 2009). The phenomenon of non-response to CRT is likely multi-factorial and related to patient selection criteria, CRT lead positioning and post-implant factors.

Current consensus guidelines (Authors/Task Force Members et al., 2013; Ponikowski et al., 2016) regarding selection for CRT focus on a limited set of patient characteristics including NYHA functional class, left ventricular ejection fraction (LVEF), QRS duration, type of bundle branch block, aetiology of cardiomyopathy and atrial rhythm (sinus, atrial fibrillation). The clinical research literature reveals a number of important insights into improving selection criteria, ranging from a lack of consensus regarding the definition of non-responders to technological limitations in the delivery of therapy.

Mullens et al. (2009) have previously described a post-implantation CRT optimisation clinic to investigate the causes of CRT non-response. They show that there were multiple common factors such as anemia, sub-optimal medical therapy, underlying narrow QRS duration and primary right ventricular dysfunction that could be identified pre-implantation and might help to improve outcomes and avoid implantation in unsuitable patients. Other factors that have been shown to be associated with increased response to CRT are strict left bundle branch block (LBBB) with type II contraction pattern (Jackson et al., 2014) and presence of septal flash (SF)¹ and apical rocking (Stankovic et al., 2016; Marechaux et al., 2016). Other predictors of non-response to CRT that have been identified include ischemic cardiomyopathy, extensive scar, presence of right bundle branch block, absence of mechanical dyssynchrony, and poor LV lead placement (i.e. in a sub-optimal location) (Linde et al., 2012). Despite more than 20 years of clinical development, a consensus definition of response and non-response to CRT has not been reached and it is necessary to better identify its causes for improving its results.

1.2 Related work

In this section, we provide an overview of the relevant literature on multimodal machine learning (Section 1.2.1) and the use of machine learning for CRT response prediction (Section 1.2.2).

1.2.1 Multimodal machine learning—Multimodal machine learning aims to build models that can process and relate information from multiple modalities. Compared to single modality machine learning techniques, learning from multimodal sources offers the possibility of capturing correspondences between modalities, reducing redundancies, and improving generalisation. Traditional multimodal machine learning approaches have included co-training algorithms (Brefeld and Scheffer, 2004; Muslea et al., 2000; Yang et al., 2012), co-regularisation algorithms (Kan et al., 2015; Sun, 2011), margin consistency

¹An early inward motion of the ventricular septum.

algorithms (Sun and Chao, 2013; Chao and Sun, 2016) and multiple kernel learning (MKL) (Gönen and Alpaydm, 2011).

Several different DL algorithms have been proposed for multi-modal learning. The most common are: (1) variants of the deep Boltzmann machine, which have been proposed to model the joint distribution from different modalities' data (Srivastava et al., 2012; Hu et al., 2013); (2) extensions of classical autoencoders to discover correlations between hidden representations of two modalities (Wang et al., 2015; Feng et al., 2014); and (3) non-linear extension of the Canonical Correlation Analysis (CCA) algorithm using deep neural networks (Andrew et al., 2013). However, in their originally proposed forms these models are not scalable to the number of features (pixels/voxels) typically present in medical images as they were based on fully-connected neural networks. Recently, some convolutional neural network (CNN)-based architectures that combine information from multiple sources for image and shape recognition have been proposed (Su et al., 2015; Yao et al., 2017; Wang et al., 2017). With the aim of directly utilising the CMR and echocardiography data, in this paper we employ a CNN-based architecture for deep multimodal classification.

1.2.2 CRT response prediction—A limited number of papers have investigated the use of ML to predict response to CRT. The literature can be mainly divided into approaches that use data from electronic health records (EHR) (Hu et al., 2019; Feeny et al., 2019; Kalscheur et al., 2018; Nejadeh et al., 2021; Ahmad et al., 2018; Bernard et al., 2015), approaches that use biomarkers derived from imaging data (Cikes et al., 2019; Bernard et al., 2015; Donal et al., 2019; Galli et al., 2021; Chao et al., 2012; Lei et al., 2019) and atlas-based approaches (Peressutti et al., 2017; Duchateau et al., 2010; Sinclair et al., 2018).

In the first category, the most common parameters used from the EHR are demographic information (e.g. sex, age, race), diagnosis codes (i.e. ICD9 and ICD10 codes), encounter information (i.e. visit type, length of stay), laboratory reports (e.g. lipids, glucose, creatinine), medication lists and cardiology reports (e.g. QRS duration, presence of LBBB, sinus rhythm). Hu et al. (2019) predicted CRT response in a cohort of 990 subjects using both structured and unstructured data from the EHR. The authors evaluated a variety of ML algorithms and showed that the gradient boosting classifier obtained the highest performance with a positive predictive value of 79%. Feeny et al. (2019) used a naive Bayes classifier with only 9 variables derived from the EHR and showed in a cohort of 455 subjects better CRT response prediction than current guidelines (area under the curve (AUC) = 0.7). Later, the same authors used Principal Components Analysis (PCA) followed by K-means clustering to predict CRT response using pre-and post-CRT 12-lead QRS waveforms (Feeny et al., 2020). Kalscheur et al. (2018) developed a random forest (RF) model for CRT patient survival prediction using 45 EHR features. Their model differentiated outcomes (AUC = 0.74) better than only current clinical discriminator features such as LBBB and QRS duration. Lei et al. (2019) showed that the three features of QRS duration, LBBB, and non-ischemic cardiomyopathy achieved the highest accuracy (84.81%) in identifying CRT responders when using a support vector machine (SVM) classifier. Finally, Ahmad et al. (2018) used a RF model to predict outcomes in 44,886 HF patients from the Swedish Heart Failure Registry. They also performed a cluster analysis to identify 4 distinct subgroups that differed significantly in outcomes and in response to therapeutics.

In the second category, the most common CMR and echocardiography-derived parameters used for CRT response prediction have been longitudinal, radial and circumferential strain (Cikes et al., 2019; Bernard et al., 2015; Donal et al., 2019; Chao et al., 2012) and right ventricular (RV) free wall strain and tricuspid annular plane systolic excursion (TAPSE) (Galli et al., 2021). For example, Cikes et al. (2019) used unsupervised MKL to combine EHR features with LV longitudinal strain derived from 2D echocardiography data to phenogroup patients with HF with respect to both outcomes and response to CRT. Chao et al. (2012) computed radial peak strain from 2D echocardiography data on a cohort of 26 CRT patients, and used it with a SVM classifier to identify CRT responders with 95.4% accuracy. Galli et al. (2021) used a k-medoid algorithm with Gower distance to identify 16 features with good prediction of CRT response (AUC = 0.81) and outcomes (AUC = 0.84) in a cohort of 193 CRT patients. Donal et al. (2019) combined EHR data with 2D echocardiography-derived parameters in a RF algorithm for predicting response to CRT in a cohort of 54 HF patients.

The works mentioned above focused on simple features derived from imaging data, and did not incorporate spatio-temporal information, which allows for a richer characterisation of cardiac function. Cardiac motion atlases have been previously used to exploit cardiac motion information from a cohort of subjects in a range of applications. Some works have used these atlases for CRT response prediction and to identify motion patterns that can be unique to CRT responders. Duchateau et al. (2010) built a cardiac motion atlas from 2D echocardiography data, and used it to detect septal flash and LV motion abnormalities in a cohort of CRT patients. Peressutti et al. (2017) built a cardiac motion atlas from CMR imaging data and used supervised MKL to combine motion and non-motion features to predict CRT response, achieving approximately 90% accuracy on a cohort of 34 patients. Sinclair et al. (2018) built a cardiac motion atlas based on a novel approach to compute strain at different spatial scales in the LV from CMR imaging. A combination of PCA and linear discriminant analysis was used for identifying the spatial scales at which myocardial strain was most strongly predictive of CRT response. An accuracy of 86.7% was achieved in identifying CRT responders in a cohort of 43 patients.

Until recently, DL had not been applied to predict CRT response. We have recently proposed such an approach in Puyol-Antón et al. (2020), which described a CMR-based pipeline based on a variational autoencoder (VAE) that allows CRT response prediction as well as the prediction of explanatory concepts to aid interpretability.

1.3 Contributions

In this paper we propose the first MMDL method for CRT response prediction. The method builds upon our recent work (Puyol-Antón et al., 2020) in which we proposed a DL framework for CRT response prediction based on CMR images, and extends it to exploit 2D CMR and 2D echocardiography data at training time. At test time the CRT response prediction is made using only the echocardiography data whilst taking advantage of the implicit relationship between CMR and echocardiography features learnt from the model. This is the first ML model to learn features from multimodal imaging data for CRT response prediction. In addition, with the exception of our preliminary work (Puyol-Antón et al.,

2020) it is the first DL model for CRT response prediction. The remainder of this paper is organised as follows. In Section 2, we describe details of the clinical data sets used for evaluation. In Section 3 we describe the novel MMDL framework developed for CRT response prediction. In Section 4 we present a thorough evaluation of the MMDL method, and Section 5 discusses the findings of this paper in the context of the literature and proposes potential improvements for future work.

2 Materials

Four data sets were used for the training and validation of the MMDL model, and these are described below:

1. **UK Biobank (UKBB):** This database contains only CMR data and is used for pre-training the CMR segmentation model (see Section 3.1). In this work, we use a cohort of 700 healthy subjects, where the LV endocardial and epicardial borders and the RV endocardial border were manually traced at end diastole (ED) and end systole (ES) frames using the cvi42 software (version 5.1.1, Circle Cardiovascular Imaging Inc., Calgary, Alberta, Canada). CMR imaging was performed using a 1.5 T Siemens MAGNETOM Aera (see Petersen et al. (2015) for further details of the image acquisition protocol).
2. **EchoNet-Dynamic** (Ouyang et al., 2020): This database contains only echocardiography data and is used for pre-training the echocardiography segmentation model (see Section 3.1). Apical-4-chamber echocardiography images were acquired by skilled sonographers using iE33, Sonos, Acuson SC2000, Epiq 5G, or Epiq 7C ultrasound machines in a cohort of 10,030 patients. For all subjects, the endocardial borders was manually traced at ED and ES frames. For further details of the image acquisition protocol see Ouyang et al. (2020).
3. **Guys and St Thomas NHS Foundation Trust (GSTFT):** This database contains paired CMR and echocardiography data for a cohort of 50 HF patients and a cohort of 50 CRT patients. The GSTFT HF database was used to train and validate the CMR and echocardiography segmentation models (see Section 3.1). The GSTFT CRT database was used to train and validate the MMDL algorithm. Both studies were approved by the London Research Ethics Committee (11/LO/1232), all patients provided written informed consent for participation in this study and the research was conducted to the Helsinki Declaration guidelines on human research. CMR imaging for the GSTFT HF and GSTFT CRT database was carried out on multiple scanners: Siemens Aera 1.5T, Siemens Biograph mMR 3T, Philips 1.5T Ingenia and Philips 1.5T and 3T Achieva. In this study, the cine CMR 4 chamber single-slice long-axis (1a4Ch) data were used, which had a slice thickness between 6 and 10 mm and an in-plane resolution between $0.92 \times 0.92\text{mm}^2$ and $2.4 \times 2.4\text{mm}^2$. For the GSTFT CRT cohort, 2D echocardiography imaging was acquired prior to CRT and at 6-months follow-up. For both cohorts (i.e. GSTFT HF and GSTFT CRT), the ultrasound machines used were Philips IE33 and EPIQ 7C (Phillips Medical Systems, Andover, MA,

USA) and the General Electric Vivid E9 (GE Health Medical, Horten, Norway), each equipped with a matrix array transducer. In this study, the apical 4Ch view was used for the development and evaluation of the MMDL model, and the apical 2Ch and 4Ch views were employed to estimate the left ventricular end-diastolic volume (EDV), end-systolic volume (ESV) and LVEF at 6-months follow-up for the GSTFT CRT cohort. The echocardiography images had an in-plane resolution between $0.26 \times 0.26 \text{mm}^2$ and $0.62 \times 0.62 \text{mm}^2$.

- 4. GSTFT CRT echocardiography database:** This database contains only echocardiography data for a cohort of 12 CRT patients who underwent upgrade of their device. These data are used to further test the proposed MMDL method in the intended clinical application of using only echocardiography data at test time. This cohort has similar 2D echocardiography image parameters to the GSTFT CRT cohort described above.

CRT volumetric response

For the GSTFT CRT database, all patients fulfilled the conventional criteria for CRT (see Section 1.1) and underwent CMR and 2D echocardiography imaging and clinical evaluation prior to CRT and at 6-months follow-up. For the GSTFT CRT echocardiography database, subjects only underwent 2D echocardiography imaging rather than both CMR and echocardiography. All patients were classified as responders or non-responders based on volumetric measures derived from 2D echocardiography acquired at the 6-months follow-up evaluation (Authors/Task Force Members et al., 2013). Patients were classified as responders if they had a reduction of 15% in LV ESV after CRT, and were classed as non-responders otherwise. From the GSTFT CRT cohort, there were 32/50 patients who were classified as responders to CRT and for the CRT echocardiography cohort, 7/12 patients who were classified as responders to CRT in this way. This information was used as the primary output label in training our proposed model.

3 Methods

An overview of the proposed MMDL framework is shown in Fig. 1. First, for each modality (i.e. CMR and echocardiography) a DL-based segmentation model is trained. Second, the latent spaces of these segmentation models are combined by the MMDL model and used for classification.

In the following, Section 3.1 briefly reviews the main steps involved in the automated segmentation of CMR and echocardiography images, Section 3.2 describes how the latent spaces of the segmentation models are extracted and combined and Section 3.3 introduces our MMDL classifier.

3.1 Automatic segmentation network

We used the ‘nnU-Net’ architecture (Isensee et al., 2021) for automatic segmentation of CMR and echocardiography images in all frames through the cardiac cycle. In comparison to the standard U-net, the ‘nnU-Net’ framework is a deep learning-based segmentation method that automatically configures itself based on the training database. This automatic

configuration can be grouped into three categories: 1) fixed, which corresponds to all design choices that do not require adaptation between datasets (i.e. architecture template) and are optimised for robust generalization on the datasets from the Medical Decathlon Segmentation Challenge; 2) rule-based, which contains a set of heuristic rules to infer data-dependent hyper-parameters of the pipeline; and (3) empirical parameters that help to select the best configuration based on the ensemble of different network configuration(s) and post-processing. Fig. 2 shows an overview of the ‘nnU-Net’ networks used for CMR and echocardiography data.

CMR cine la4Ch segmentation—The CMR segmentation model performed automated segmentation of the left ventricle blood pool (LVBP), left ventricular myocardium (LVMyo) and right ventricle blood pool (RVBP) from CMR la4Ch images. The network was trained with 1400 images from the UK Biobank database (ED and ES frames) and 200 images (ED and ES frames) from the GSTFT HF cohort.

Echocardiography apical 4Ch segmentation—The echocardiography segmentation model performed automated segmentation of the left ventricle blood pool (LVBP) from echocardiography apical 4Ch images. The model was pre-trained using the EchoNet-Dynamic database (Ouyang et al., 2020), which includes 20,060 echocardiography images with annotations (ED and ES frames). To take into account the inter-vendor differences in intensity distributions, the segmentation model was then fine-tuned using 300 images (multiple time points from 50 echocardiography scans) from the GSTFT HF cohort.

Implementation Details—Both segmentation models were trained and evaluated using a five-fold cross-validation on the training set. As in Isensee et al. (2021), the networks were trained for 1000 epochs, where one epoch is defined as an iteration over 250 mini-batches. The batch sizes were 33 and 10 respectively for the CMR cine la4Ch and echocardiography apical 4Ch segmentation models. Stochastic gradient descent with Nesterov momentum ($\mu=0.99$) and an initial learning rate of 0.01 was used for learning network weights. The loss function used to train the ‘nnU-Net’ model was the sum of cross-entropy and Dice loss. Data augmentation was performed on the fly and included techniques such as rotations, scaling, Gaussian noise, Gaussian blur, brightness, contrast, simulation of low resolution, gamma correction and mirroring. Please refer to Isensee et al. (2021) for more details of the network training.

3.2 Generation of the latent space

The two segmentation models were used to segment the CMR and echocardiography images from the GSTFT CRT cohort in all frames through the cardiac cycle. To correct for variation in acquisition protocols between vendors, all images were first temporally resampled to $T = 25$ frames per cardiac cycle using piecewise linear warping based on cardiac timings (Puyol-Antón et al., 2018). For each frame, the latent space of the segmentation network was stored and these were concatenated to generate a 2D matrix, in which rows correspond to the latent variables and columns to the different temporal frames - see Fig. 2 for an example case.

3.3 Multimodal deep learning (MMDL)

We used the 2D Deep Canonical Correlation Analysis (DCCA) algorithm (Andrew et al., 2013; Wang et al., 2017) followed by a SVM classifier to develop our MMDL model for CRT response prediction. DCCA extends the linear CCA model by projecting multiple views of the data to a common latent space using a deep model with multiple branches, each corresponding to one view (see Fig. 3). Here we consider CMR and echocardiography images as two views of the same object, i.e. the heart.

The 2D DCCA framework we employ is similar in concept to the configuration provided in Wang et al. (2017) and contains three parts:

1. **Nonlinear feature transformation:** The first three layers are convolutional layers (i.e. CNNs): two 8-dims, two 16-dims and two 32-dims followed by a ReLU activation function. The size of all the receptive fields is 3×3 and the stride in all the layers is 1. A max pooling layer is inserted between each pair of sequences. The last layer is a global pooling layer, which enforces correspondences between feature maps and categories. The outputs of the last layer are defined as L_1 and L_2 in Fig. 3, and are the inputs of the CCA layer.
2. **CCA layer:** This layer has the goal to jointly learn parameters for both views X_1 and X_2 to maximise $\text{corr}(f_1(X_1), f_2(X_2))$, where $f_1()$ and $f_2()$ are the nonlinear functions learnt by the networks from the previous step. We denote by θ_1 the vector of all parameters from the nonlinear feature transformation of the first view, and similarly for θ_2 . These are determined according to the following objective:

$$(\theta_1^*, \theta_2^*) = \underset{(\theta_1, \theta_2)}{\text{argmax}} \text{corr}(f_1(X_1; \theta_1), f_2(X_2; \theta_2)) \quad (1)$$

The parameters θ_1 and θ_2 of DCCA are trained to optimise this quantity using gradient-based optimisation. For the full derivation of the gradients we refer to the original DCCA paper by Andrew et al. (2013). The outputs of the CCA layer are V_1 and V_2 .

3. **Feature fusion** The fusion layer combines the outputs of the CCA layer (V_1 and V_2) as follows: $F_{\text{fusion}} = \alpha \cdot V_1 + \beta \cdot V_2$, where α and β are the fusion weights. In our experiments, in order to balance the composition of features, we set $\alpha = \beta = 0.5$.

The final part of the MMDL architecture is a binary SVM model with a radial basis function (RBF) kernel classifier, where the input is the output of the feature fusion layer and the output is a binary variable (i.e. CRT responders vs CRT non-responders). At test time, we can use either the output of the CMR or echocardiography layer (V_1 or V_2) to simulate the scenarios in which only one of the modalities is available.

Compared to the standard DCCA framework, where the nonlinear feature transformations are extracted using fully connected layers, we have replaced these layers by CNNs to take

into account that our input data are 2D and to exploit the power of feature learning using CNNs.

Implementation Details—The segmentation models and the MDDL model were trained independently, to maximise the anatomical accuracy (and hence interpretability) of the segmentation models. The different parameters of the MDDL model were optimised using a grid search strategy. The 2D CCA model was trained for 500 epochs, with a batch-size of 10 images (generated from the latent space of the segmentation networks, see Section 3.2). We used the Adam optimiser with a momentum set to 0.9. For the 2D CCA model the parameters that were optimised were the size of the output layer k_{DCCA} (range $\in [5, 10, 15, 20, 25, 30, 35]$) and the learning rate, lr (range $\in [0.01, 0.001, 0.0001]$). The SVM hyperparameters that we optimised were γ (range $\in [0.1, 0.01, 0.001, 0.0001]$) and the cost parameter C (range $\in [1, 10, 100, 1000]$). All hyperparameters were optimised using a 5-fold nested cross-validation using the GSTFT CRT database (see Section 4.1 for more details).

4 Experiments and results

Two sets of experiments were performed. The first set of experiments (see Section 4.2) aimed to validate the two segmentation models described in Section 3.1, while the second set of experiments (see Section 4.3) aimed to validate the proposed MDDL approach detailed in Section 3.3.

All experiments were carried out using the Python programming language with standard Python libraries Pytorch (Paszke et al., 2019) and scikit-learn. Before describing the experiments in detail, we first describe the evaluation measures and comparative approaches used.

4.1 Evaluation metrics and comparative approaches

For the segmentation networks, performance was evaluated using the Dice metric. For the MDDL model, a 5-fold nested cross-validation was used to validate its performance. In each nested fold, the predicted classes were stored and we then computed the overall classification balanced accuracy (i.e. the average of the accuracies obtained for each class individually), as well as the sensitivity (the proportion of CRT responders correctly classified) and the specificity (the proportion of CRT non-responders correctly classified). The balanced accuracy (BACC), sensitivity (SEN) and specificity (SPE) metrics are defined as:

- Balanced accuracy (BACC): $\frac{1}{2} \left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP} \right)$
- Sensitivity (SEN): $\frac{TP}{TP + FN}$
- Specificity (SPE): $\frac{TN}{TN + FP}$

where TP represents true positives, FP is false positives, FN is false negatives and TN is true negatives.

With the aim of evaluating the impact of using multimodal data for CRT response prediction, we compared the MMDL model with single modality classifiers. Due to the lack of prior work on image-based CRT response prediction, we chose to train and evaluate a number of different state-of-the-art CNN-based classifiers using the latent space matrices (see Fig. 2) of either the echocar-diography or CMR segmentation models. We evaluated six different classifiers (AlexNet, DenseNet, MobileNet, ShuffleNet, SqueezeNet and VGG) for CRT response prediction using the two different modalities' latent space matrices. Each network was trained for 200 epochs with binary cross entropy loss, to classify between CRT responders and non-responders. During training, data augmentation was performed on-the-fly using random translations (± 30 pixels), rotations ($\pm 90^\circ$), flips (50% probability) and scalings (up to 20%) to each mini-batch of images before feeding them to the network. The probability of augmentation for each of the parameters was 50%. A nested cross validation was used to select the optimal learning rate, lr (range $\in [0.01, 0.001, 0.0001]$) and the optimal CNN classification network. We consider as a baseline the best such CNN trained using echocardiography data only, as it represents the current state-of-the-art in the use of echocardiography data alone for CRT response prediction. Since CMR is considered to be the gold standard for analysis of cardiac function, we consider the best CNN technique trained using only CMR data as a reference technique. shows the BACC, SEN and SPE for the 6 classifiers for the CMR and echocardiography data.

In addition, we also compared the proposed approach with the DL-based baseline approach of our previous work (Puyol-Antón et al., 2020). Using the GSTFT CRT cohort, we first tested the baseline VAE model implemented in Puyol-Antón et al. (2020) using the short-axis (SAX) CMR data (as was the case in the original paper). Then, we trained the same model using the la4Ch CMR data. The first approach enabled a direct comparison with our previous framework, but using the GSTFT CRT cohort employed in this paper. Comparing the first and second approaches enabled us to assess the impact of using la4Ch data rather than SAX data for CRT response prediction. Finally, the second approach allowed a comparison of the technique described in Puyol-Antón et al. (2020) to the MMDL model proposed in this paper (using la4Ch CMR data).

4.2 Automated segmentation

For the CMR cine la4Ch segmentation model, the Dice metrics between automated and manual segmentations (only ED and ES frames, for which ground truths were available) were 0.97 for the LV blood pool, 0.92 for the LV myocardium and 0.96 for the RV blood pool, which is in line with previous published methods (Leng et al., 2018; Ruijsink et al., 2020). For the echocardiography apical 4Ch segmentation network, similar to Ouyang et al. (2020), the Dice metric between automated and manual segmentations was 0.96 for the LV blood pool (again, only ED and ES frames). The segmentations for other frames, where manual annotations were not available, were visually inspected by an expert cardiologist.

4.3 Evaluation of the multimodal deep learning framework

Table 1 shows the results of our experiments using the GSTFT CRT database and the GSTFT CRT echocardiography database.

As described in Section 4.1, for the MMDL model a nested cross validation was used to ensure unbiased validation; the optimal parameters were: size of latent space $k_{DCCA}=25$, learning rate $lr=0.01$, and the SVM parameters were $\gamma=0.01$ and $C=10$. The results of this method applied using CMR or echocardiography data are shown in the top row of Table 1 (i.e. *MMDL*).

As a baseline approach, we report results for the VGG network trained and applied using only echocardiography data. VGG obtained the best BACC and SEN values compared to the other CNN classification networks. The optimal learning rate for this approach was $lr=0.001$. The results of this approach are shown in the second row of Table 1 (i.e. *VGG_{Echo}*).

As a reference approach (i.e. CNN trained and applied using only CMR data), VGG was again the best performing model, with $lr=0.001$. These results are shown in the third row of Table 1 (i.e. *VGG_{CMR}*).

Finally, we compared to our previously published method (Puyol-Antón et al., 2020) which is trained and applied on CMR data only. In the original paper, the VAE model was trained using CMR SAX data. As discussed above, we evaluated this model using both SAX and la4Ch CMR data. These results are shown in the fourth and fifth rows of Table 1 (i.e. *VAE_{SAX}* and *VAE_{LAX}*).

Students t-tests (99% confidence) were used to compare the performance of the baseline approach with the other approaches.

The results show that the MMDL framework is capable of performing CRT response prediction using data from either view with similar accuracy, sensitivity and specificity. We assessed the statistical significance of the BACC using the proposed method compared to the performance of the baseline approach of *VGG_{Echo}* trained on echocardiography data and we show that our proposed method outperforms the baseline approach. Similar performance for the MMDL framework is achieved on the additional GSTFT CRT echocardiography validation set (see bottom rows of Table 1), offering additional evidence of the robustness of our proposed model. For the comparative approaches, we can see that the multimodal classifier algorithm has a higher accuracy compared to the single modality methods except for the VAE-based approach trained with SAX data (BACC 82.64% compared to 81.19%).

For the VAE-based approach, if we compare the models trained using SAX and la4Ch data, we can see that there is a slight decrease in accuracy when using la4Ch data. As mentioned in Section 4.1, the *VAE_{SAX}* results are only reported for comparison to our previous work (Puyol-Antón et al., 2020). However, this method cannot be directly compared to the proposed MMDL method as it is trained using different data, and paired echocardiography data would not be available for the SAX view. We hypothesize that the difference can be explained by the different regions covered by the two modalities: the SAX data focuses on the basal region of the heart where there is more motion, whilst the la4Ch images focus on a cross section of the heart which may not capture some motion that can be important for CRT response prediction.

From Table 1 we can also see that using only la4Ch CMR data we achieve the highest accuracy with the MMDL model. Based on these results, we conclude that the highest accuracy is achieved using our proposed multimodal deep learning framework, whether using only echocardiography data or CMR data as input. We also conclude that including the CMR data into the training (as well as the echocardiography data) improves the performance of the MMDL classifier.

5 Discussion and conclusion

We have proposed a novel MMDL method based on DCCA that has the ability to predict CRT response using only echocardiography data but at the same time taking advantage of the implicit relationship between CMR and echocardiography. To the best of our knowledge, this is the first time that multimodal imaging data has been used directly for CRT response prediction. Previous works have mainly focused on imaging-derived parameters or atlas-based approaches, where manual input is required. Compared to our previous work (Puyol-Antón et al., 2020), the proposed approach allows CRT response to be predicted using only the widely available and cheap modality of echocardiography. Some previous work has proposed a multimodal ML model that exploits echocardiography alone at test time (Puyol-Antón et al., 2018). Compared to this work, the proposed approach makes direct use of 2D CMR and 2D echocardiography imaging data, rather than the results of motion tracking algorithms, which reduces the complexity of building the spatio-temporal model and eliminates the effect of potential errors in the motion tracking. In addition, the proposed framework is fully automated while the framework proposed in Puyol-Antón et al. (2018) required manual delineations of the CMR and echocardiography data.

Our results showed that the use of the MMDL algorithm resulted in a statistically significant increase in classification accuracy compared to the use of only echocardiography or CMR data, and that our technique for prediction of CRT response (i.e. 83.33%/71.43% sensitivity/specificity) is comparable with the current state-of-the-art. For example, Peressutti et al. (2017) reported 100% sensitivity and 62.5% specificity for predicting CRT responders based on a combination of a cardiac motion atlas with non-motion data obtained from several sources. Sohal et al. (2014) reported 85% sensitivity and 82% specificity for predicting CRT responders based on volume-change systolic dyssynchrony index. Importantly, both of these approaches required significant manual intervention.

It is important to consider the clinical interpretation of sensitivity and specificity in the context of CRT response. Sensitivity refers to the proportion of CRT responders correctly identified and specificity refers to the proportion of CRT non-responders correctly identified. In our application, we are interested in achieving the highest possible sensitivity to ensure that there are no subjects that are denied a treatment that would have resulted in symptomatic benefit. Our technique achieved a sensitivity of 83.33%, meaning that only ~17% of such cases would have been missed, and specificity was 71.43%, meaning that almost three quarters of non-responders would be spared the unnecessary treatment, which is not without risk. To foster clinical translation and promote clinical trust, it is important that sensitivity is as high as possible, preferably 100%, and this will be the focus of future work.

Increasing the response rate to CRT has been the focus of many studies and the results to date have been conflicting. This is because CRT treatment planning is multi-factorial and related to patient selection criteria, CRT lead positioning and post-implant factors. Future work will aim to incorporate other clinical parameters such as pacing leads, presence of scar and septal flash to have a more complete and clinically useful pipeline.

Another area for future work is to validate the approach on a larger multi-centre database of CRT patients, to better evaluate the accuracy and robustness of the proposed framework. This would eventually involve establishing a clinical trial to validate the impact of the proposed model on clinical decision making in CRT patient selection.

Furthermore, we would like to test the same pipeline for predicting response to other types of treatment, such as pharmacology, and try to incorporate uncertainty techniques to be able to estimate confidences in CRT response prediction. This would move us closer to our vision of an interpretable tool that can be used to support the decision making of cardiologists when selecting HF patients for different types of treatment.

Data Statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: The UK Biobank data set is publicly available for approved research projects from <https://www.ukbiobank.ac.uk/>. The EchoNet-Dynamic dataset is a publicly available dataset of de-identified echocardiogram videos available at <https://echonet.github.io/echoNet/>. The GSTFT data set cannot be made publicly available due to restricted access under hospital ethics and because informed consent from participants did not cover public deposition of data.

Acknowledgments

This work was supported by the EPSRC (EP/R005516/1, EP/P001009/1), by core funding from the Wellcome/ EPSRC Centre for Medical Engineering (WT203148/Z/16/Z), and by the NIHR Cardiovascular MedTech Co-operative award to the Guys and St Thomas NHS Foundation Trust. This research was funded in whole, or in part, by the Wellcome Trust WT203148/Z/16/Z. For the purpose of open access, the author has applied a CC BY public copyright license to any author accepted manuscript version arising from this submission. EPA and APK acknowledge financial support from the Department of Health through the National Institute for Health Research (NIHR) comprehensive Biomedical Research Centre award to Guys & St Thomas NHS Foundation Trust in partnership with Kings College London. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, EPSRC, or the Department of Health. This research has been conducted using the UK Biobank Resource (application number 17806) on a GPU generously donated by NVIDIA Corporation.

Abbreviations

HF	Heart failure
CMR	Cardiac magnetic resonance
DL	Deep learning
CRT	Cardiac Resynchronisation Therapy
MMDL	Multimodal deep learning.

References

- Ahmad T, Lund LH, Rao P, Ghosh R, Warier P, Vaccaro B, Dahlström U, O'connor CM, Felker GM, Desai NR. Machine learning methods improve prognostication, identify clinically distinct phenotypes, and detect heterogeneity in response to therapy in a large cohort of heart failure patients. *J Am Heart Assoc.* 2018; 7 (8) e008081 [PubMed: 29650709]
- Andrew, G; Arora, R; Bilmes, J; Livescu, K. Deep canonical correlation analysis; International Conference on Machine Learning; 2013. 1247–1255.
- Authors/Task Force Members. Brignole M, Auricchio A, Baron-Esquivias G, Bordachar P, Boriani G, Breithardt O-A, Cleland J, Deharo J-C, Delgado V, et al. 2013 ESC guidelines on cardiac pacing and cardiac resynchronization therapy: the task force on cardiac pacing and resynchronization therapy of the european society of cardiology (ESC). developed in collaboration with the european heart rhythm association (EHRA. *Eur Heart J.* 2013; 34 (29) 2281–2329. [PubMed: 23801822]
- Bernard A, Donal E, Leclercq C, Schnell F, Fournet M, Reynaud A, Thebault C, Mabo P, Daubert J-C, Hernandez A. Impact of cardiac resynchronization therapy on left ventricular mechanics: understanding the response through a new quantitative approach based on longitudinal strain integrals. *J of the American Soc of Echocardio.* 2015; 28 (6) 700–708.
- Brefeld, U; Scheffer, T. Co-em support vector learning; Proceedings of the Twenty-first International Conference on Machine Learning; 2004. 16
- Bristow MR, Saxon LA, Boehmer J, Krueger S, Kass DA, De Marco T, Carson P, DiCarlo L, DeMets D, White BG, et al. Cardiac-resynchronization therapy with or without an implantable defibrillator in advanced chronic heart failure. *N top N Engl J Med.* 2004; 350 (21) 2140–2150.
- Bruge, S, Simon, A, Courtial, N, Betancur, J, Hernandez, A, Tavard, F, Donal, E, Lederlin, M, Leclercq, C, Garreau, M. Multi-Modality Imaging. Springer; 2018. 67–82.
- Chao G, Sun S. Consensus and complementarity based maximum entropy discrimination for multi-view classification. *Inf Sci (Ny).* 2016; 367: 296–310.
- Chao P-K, Wang CL, Chan HL. An intelligent classifier for prognosis of cardiac resynchronization therapy based on speckle-tracking echocardiograms. *Artif Intell Med.* 2012; 54 (3) 181–188. [PubMed: 21968204]
- Cikes M, Sanchez-Martinez S, Claggett B, Duchateau N, Piella G, Butakoff C, Pouleur AC, Knappe D, Biering-Sørensen T, Kutlyifa V, et al. Machine learning-based phenogrouping in heart failure to identify responders to cardiac resynchronization therapy. *Eur J Heart Fail.* 2019; 21 (1) 74–85. [PubMed: 30328654]
- Cleland J. Cardiac resynchronization-heart failure (CARE-HF) study investigators: the effect of cardiac resynchronization on morbidity and mortality in heart failure. *N Engl J Med.* 2005; 352: 1539–1549. [PubMed: 15753115]
- Donal E, Hubert A, Le Rolle V, Leclercq C, Martins R, Mabo P, Galli E, Hernandez A. New multiparametric analysis of cardiac dyssynchrony: machine learning and prediction of response to CRT. *JACC: Cardio Imaging.* 2019; 12 (9) 1887–1888.
- Duchateau, N; De Craene, M; Piella, G; Hoogendoorn, C; Silva, E; Doltra, A; Mont, L; Castel, MA; Brugada, J; Sitges, M; , et al. Atlas-based quantification of myocardial motion abnormalities: added-value for the understanding of CRT outcome?; International Workshop on Statistical Atlases and Computational Models of the Heart; 2010. 65–74.
- Feeny AK, Rickard J, Patel D, Toro S, Trulock KM, Park CJ, LaBarbera MA, Varma N, Niebauer MJ, Sinha S, et al. Machine learning prediction of response to cardiac resynchronization therapy: improvement versus current guidelines. *Circulation: Arrhythmia and Electrophysiology.* 2019; 12 (7) e007316 [PubMed: 31216884]
- Feeny AK, Rickard J, Trulock KM, Patel D, Toro S, Moennich LA, Varma N, Niebauer MJ, Gorodeski EZ, Grimm RA, et al. Machine learning of 12-lead qrs waveforms to identify cardiac resynchronization therapy patients with differential outcomes. *Circulation: Arrhythmia and Electrophysiology.* 2020; 13 (7) e008210 [PubMed: 32538136]
- Feng, F; Wang, X; Li, R. Cross-modal retrieval with correspondence autoencoder; Proceedings of the 22nd ACM International Conference on Multimedia; 2014. 7–16.

- Galli E, Le Rolle V, Smiseth O, Aalen J, Sade E, Hernandez A, Leclercq C, Duchenne J, Voigt J, Donal E. Importance of systematic right ventricular assessment in patients undergoing cardiac resynchronization therapy: a machine-learning approach. *Archives of Cardiovascular Diseases Supplements*. 2021; 13 (1) 59–60.
- Gönen M, Alpaydin E. Multiple kernel learning algorithms. *The Journal of Machine Learning Research*. 2011; 12: 2211–2268.
- Hu, H; Liu, B; Wang, B; Liu, M; Wang, X. Multimodal dbn for predicting high-quality answers in cqa portals; Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics; 2013. 843–847.
- Hu S-Y, Santus E, Forsyth AW, Malhotra D, Haimson J, Chatterjee NA, Kramer DB, Barzilay R, Tulskey JA, Lindvall C. Can machine learning improve patient selection for cardiac resynchronization therapy? *PLoS ONE*. 2019; 14 (10) e0222397 [PubMed: 31581234]
- Hundley WG, Bluemke DA, Finn JP, Flamm SD, Fogel MA, Friedrich MG, Ho VB, Jerosch-Herold M, Kramer CM, Manning WJ, et al. ACCF/ACR/AHA/NASCI/SCMR 2010 Expert consensus document on cardiovascular magnetic resonance: a report of the american college of cardiology foundation task force on expert consensus documents. *J Am Coll Cardiol*. 2010; 55 (23) 2614–2662. [PubMed: 20513610]
- Isensee F, Jaeger PF, Kohl SA, Petersen J, Maier-Hein KH. NnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods*. 2021; 18 (2) 203–211. [PubMed: 33288961]
- Jackson T, Sohal M, Chen Z, Child N, Sammut E, Behar J, Claridge S, Carr-White G, Razavi R, Rinaldi CA. A U-shaped type II contraction pattern in patients with strict left bundle branch block predicts super-response to cardiac resynchronization therapy. *Heart Rhythm*. 2014; 11 (10) 1790–1797. [PubMed: 24912138]
- Kalscheur MM, Kipp RT, Tattersall MC, Mei C, Buhr KA, DeMets DL, Field ME, Eckhardt LL, Page CD. Machine learning algorithm predicts cardiac resynchronization therapy outcomes: lessons from the COMPANION trial. *Circulation: Arrhythmia and Electrophysiology*. 2018; 11 (1) e005499 [PubMed: 29326129]
- Kan M, Shan S, Zhang H, Lao S, Chen X. Multi-view discriminant analysis. *IEEE Trans Pattern Anal Mach Intell*. 2015; 38 (1) 188–194.
- Kirkpatrick JN, Vannan MA, Narula J, Lang RM. Echocardiography in heart failure: applications, utility, and new horizons. *J Am Coll Cardiol*. 2007; 50 (5) 381–396. [PubMed: 17662389]
- Lei J, Wang YG, Bhatta L, Ahmed J, Fan D, Wang J, Liu K. Ventricular geometry-regularized qrsd predicts cardiac resynchronization therapy response: machine learning from crosstalk between electrocardiography and echocardiography. *Int J Cardiovasc Imaging*. 2019; 35 (7) 1221–1229. [PubMed: 31104177]
- Leng, S; Yang, X; Zhao, X; Zeng, Z; Su, Y; Koh, AS; Sim, D; Le Tan, J; San Tan, R; Zhong, L. Computational platform based on deep learning for segmenting ventricular endocardium in long-axis cardiac mr imaging; 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2018. 4500–4503.
- Linde C, Ellenbogen K, McAlister FA. Cardiac resynchronization therapy (CRT): clinical trials, guidelines, and target populations. *Heart Rhythm*. 2012; 9 (8) S3–S13. [PubMed: 22521934]
- Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, Van Der Laak JA, Van Ginneken B, Sánchez CI. A survey on deep learning in medical image analysis. *Med Image Anal*. 2017; 42: 60–88. [PubMed: 28778026]
- Marechaux S, Menet A, Guyomar Y, Ennezat P-V, Guerberai RA, Graux P, Tribouilloy C. Role of echocardiography before cardiac resynchronization therapy: new advances and current developments. *Echocardiography*. 2016; 33 (11) 1745–1752. [PubMed: 27562174]
- McAlister FA, Ezekowitz J, Hooton N, Vandermeer B, Spooner C, Dryden DM, Page RL, Hlatky MA, Rowe BH. Cardiac resynchronization therapy for patients with left ventricular systolic dysfunction: a systematic review. *JAMA*. 2007; 297 (22) 2502–2514. [PubMed: 17565085]
- Mullens W, Grimm RA, Verga T, Dresing T, Starling RC, Wilkoff BL, Tang WW. Insights from a cardiac resynchronization optimization clinic as part of a heart failure disease management program. *J Am Coll Cardiol*. 2009; 53 (9) 765–773. [PubMed: 19245967]

- Muslea, I; Minton, S; Knoblock, CA. Selective sampling with redundant views; AAAI/IAAI; 2000. 621–626.
- Nejadedh M, Bayat P, Kheirkhah J, Moladoust H. Predicting the response to cardiac resynchronization therapy (CRT) using the deep learning approach. *Biocybernetics and Biomedical Engineering*. 2021.
- Ouyang D, He B, Ghorbani A, Yuan N, Ebinger J, Langlotz CP, Heidenreich PA, Harrington RA, Liang DH, Ashley EA, et al. Video-based ai for beat-to-beat assessment of cardiac function. *Nature*. 2020; 580 (7802) 252–256. [PubMed: 32269341]
- Parsai C, Bijmens B, Sutherland GR, Baltabaeva A, Claus P, Marciniak M, Paul V, Scheffer M, Donal E, Derumeaux G, et al. Toward understanding response to cardiac resynchronization therapy: left ventricular dyssynchrony is only one of multiple mechanisms. *Eur Heart J*. 2009; 30 (8) 940–949. [PubMed: 19004844]
- Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L, et al. Pytorch: an imperative style, high-performance deep learning library. *arXiv preprint*. 2019. arxiv:1912.01703
- Peressutti D, Sinclair M, Bai W, Jackson T, Ruijsink J, Nordsletten D, Asner L, Hadjicharalambous M, Rinaldi CA, Rueckert D, et al. A framework for combining a motion atlas with non-motion information to learn clinically useful biomarkers: application to cardiac resynchronisation therapy response prediction. *Med Image Anal*. 2017; 35: 669–684. [PubMed: 27770718]
- Petersen SE, Matthews PM, Francis JM, Robson MD, Zemrak F, Boubertakh R, Young AA, Hudson S, Weale P, Garratt S, et al. UK Biobanks cardio-vascular magnetic resonance protocol. *J of cardio magnetic reso*. 2015; 18 (1) 1–7.
- Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JG, Coats AJ, Falk V, González-Juanatey JR, Harjola V-P, Jankowska EA, et al. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: the task force for the diagnosis and treatment of acute and chronic heart failure of the european society of cardiology (ESC) developed with the special contribution of the heart failure association (HFA) of the ESC. *Eur Heart J*. 2016; 37 (27) 2129–2200. [PubMed: 27206819]
- Puyol-Antón, E; Chen, C; Clough, JR; Ruijsink, B; Sidhu, BS; Gould, J; Porter, B; Elliott, M; Mehta, V; Rueckert, D; , et al. Interpretable deep models for cardiac resynchronisation therapy response prediction; International Conference on Medical Image Computing and Computer-Assisted Intervention; 2020. 284–293.
- Puyol-Antón E, Ruijsink B, Gerber B, Amzulescu MS, Langet H, De Craene M, Schnabel JA, Piro P, King AP. Regional multi-view learning for cardiac motion analysis: application to identification of dilated cardiomyopathy patients. *IEEE Trans Biomed Eng*. 2018; 66 (4) 956–966. [PubMed: 30113891]
- Puyol-Antón E, Sinclair M, Gerber B, Amzulescu MS, Langet H, De Craene M, Aljabar P, Piro P, King AP. A multimodal spatiotemporal cardiac motion atlas from MR and ultrasound data. *Med Image Anal*. 2017; 40: 96–110. [PubMed: 28646674]
- Ramachandram D, Taylor GW. Deep multimodal learning: a survey on recent advances and trends. *IEEE Signal Process Mag*. 2017; 34 (6) 96–108.
- Ruijsink B, Puyol-Antón E, Oksuz I, Sinclair M, Bai W, Schnabel JA, Razavi R, King AP. Fully automated, quality-controlled cardiac analysis from cmr: validation and large-scale application to characterize cardiac function. *Cardio Imaging*. 2020; 13 (3) 684–695.
- Sinclair M, Peressutti D, Puyol-Antón E, Bai W, Rivolo S, Webb J, Claridge S, Jackson T, Nordsletten D, Hadjicharalambous M, et al. Myocardial strain computed at multiple spatial scales from tagged magnetic resonance imaging: estimating cardiac biomarkers for CRT patients. *Med Image Anal*. 2018; 43: 169–185. [PubMed: 29112879]
- Sohal M, Duckett SG, Zhuang X, Shi W, Ginks M, Shetty A, Sammut E, Kozerke S, Niederer S, Smith N, et al. A prospective evaluation of cardiovascular magnetic resonance measures of dyssynchrony in the prediction of response to cardiac resynchronization therapy. *J of Cardio Magnetic Reso*. 2014; 16 (1) 1–12.
- Srivastava, N; Salakhutdinov, R; , et al. Multimodal learning with deep boltzmann machines; NIPS; 2012. 2

- Stankovic I, Prinz C, Ciarka A, Daraban AM, Kotrc M, Aarones M, Szulik M, Winter S, Belmans A, Neskovic AN, et al. Relationship of visually assessed apical rocking and septal flash to response and long-term survival following cardiac resynchronization therapy (PREDICT-CRT). *Eur Heart J -Cardio Imaging*. 2016; 17 (3) 262–269.
- Su, H; Maji, S; Kalogerakis, E; Learned-Miller, E. Multi-view convolutional neural networks for 3D shape recognition; *Proceedings of the IEEE International Conference on Computer Vision*; 2015. 945–953.
- Sun, S. Multi-view laplacian support vector machines; *International Conference on Advanced Data Mining and Applications*; 2011. 209–222.
- Sun, S; Chao, G. Multi-view maximum entropy discrimination; *Twenty-third International Joint Conference on Artificial Intelligence*; 2013.
- Wang, S; Huang, D; Wang, Y; Tang, Y. 2d-3d heterogeneous face recognition based on deep canonical correlation analysis; *Chinese Conference on Biometric Recognition*; 2017. 77–85.
- Wang, W; Arora, R; Livescu, K; Bilmes, J. On deep multi-view representation learning; *International Conference on Machine Learning*; 2015. 1083–1092.
- Yancy CW, Jessup M, Bozkurt B, Butler J, Casey DE, Colvin MM, Drazner MH, Filippatos GS, Fonarow GC, Givertz MM, et al. 2017 ACC/AHA/HFSA focused update of the 2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Failure Society of America. *J Am Coll Cardiol*. 2017; 70 (6) 776–803. [PubMed: 28461007]
- Yang, P, Gao, W, Tan, Q, Wong, K-F. Information-theoretic multi-view domain adaptation. *Association for Computational Linguistics*; 2012.
- Yao, S; Hu, S; Zhao, Y; Zhang, A; Abdelzaher, T. Deepsense: A unified deep learning framework for time-series mobile sensing data processing; *Proceedings of the 26th International Conference on World Wide Web*; 2017. 351–360.

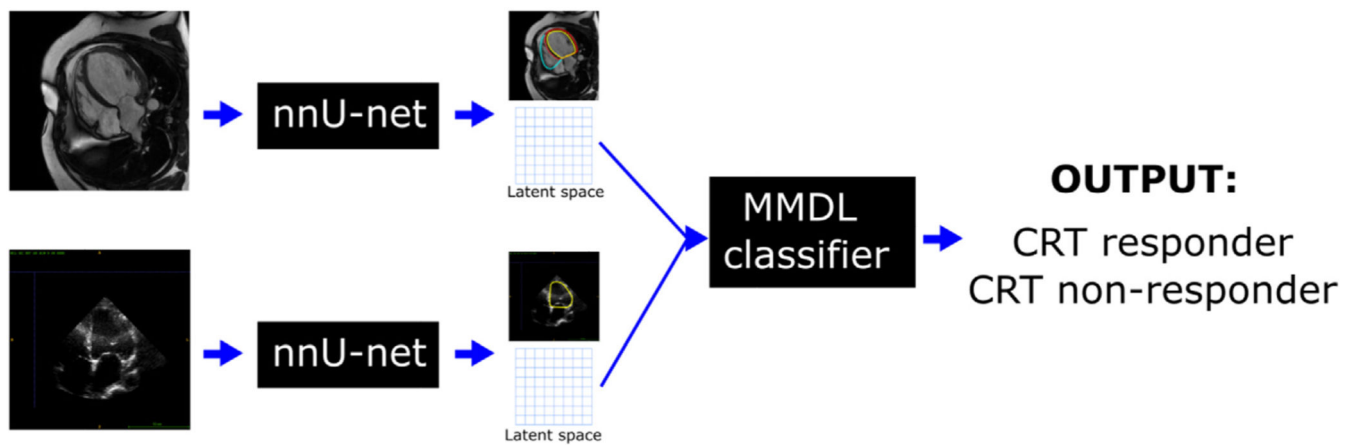


Fig. 1.

Overview of the proposed framework to combine 2D CMR and echocardiography data for CRT response prediction. First, the 'nnU-Net' architecture is used to extract segmentations of the heart over the full cardiac cycle from the two modalities. Next, a multimodal deep learning classifier is used for CRT response prediction, which combines the latent spaces of the 'nnU-Net' models from the two modalities. At test time, this framework can be used with 2D echocardiography data only, whilst taking advantage of the implicit relationship between CMR and echocardiography features learnt from the model.

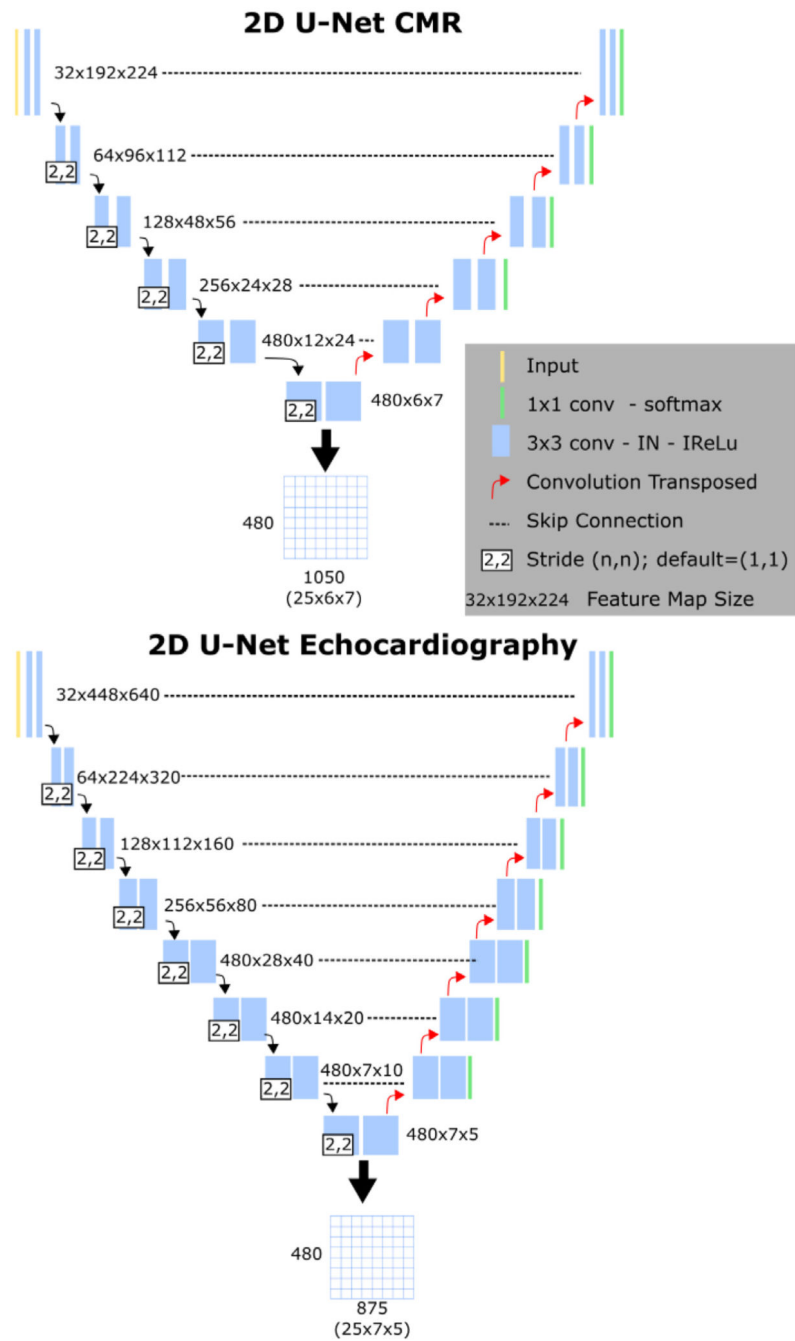


Fig. 2. The ‘nnU-Net’ architecture. Top: the segmentation network used for CMR cine la4Ch images. Bottom: the segmentation network used for echocardiography apical 4Ch images. Below each figure, the generation of the latent space matrices is shown.

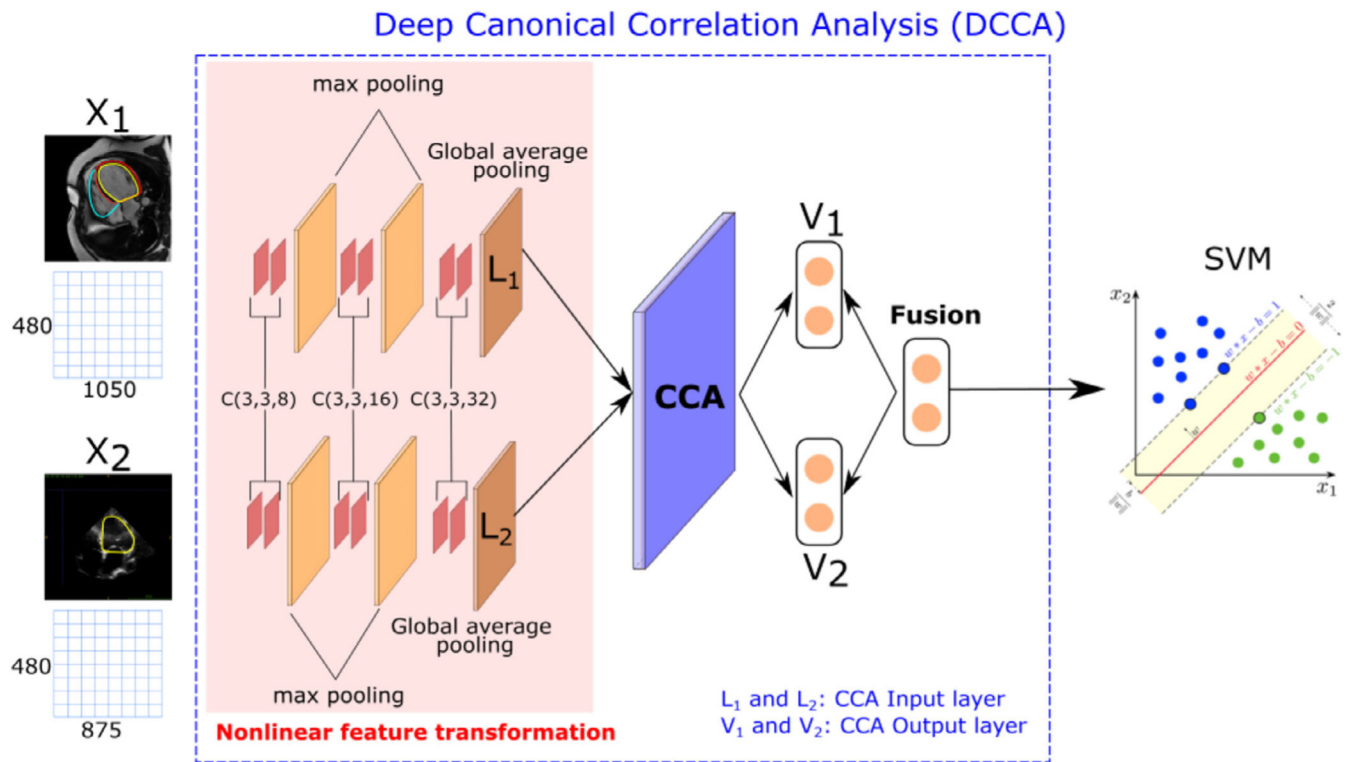


Fig. 3. Overview of the proposed MMDL method, which fuses the outputs of a DL model for each data view and applies a SVM classifier. $C(m, n, k)$ denotes a convolutional layer with an m -by- n receptive field and k channels.

Table 1

Balanced Accuracy (BACC), Sensitivity (SEN) and Specificity (SPE) of the proposed and comparative methods and Students *t*-test (99% confidence) results. The CMR and Echocardiography column headings indicate which data were used to apply the model. The first group corresponds to the results of the MMDL method for CRT response prediction, the second group corresponds to the results of the base-line method, the third group corresponds to the comparative methods and the last group corresponds to the results on the GSTFT CRT echocardiography database. An asterisk indicates a statistically significant improvement in accuracy over the base-line comparative approach. Bold-italic text indicates the method with the overall best classification accuracy, bold indicates the indicates the method with the highest classification accuracy for the la4Ch data, and dash indicates that this approach cannot be applied for this modality.

	CMR			Echocardiography		
	BACC(%)	SEN(%)	SPE(%)	BACC(%)	SEN(%)	SPE(%)
<i>1. Multimodal deep learning approach</i>						
MMDL	81.19*	86.21	76.19	77.38*	83.33	71.43
<i>2. Baseline approach (echocardiography-trained)</i>						
VGG _{Echo}	-	-	-	70.26	75.00	65.52
<i>3. Single modality (CMR-trained) approaches</i>						
VGG _{CMR}	71.98	68.97	75.00	-	-	-
VAE _{SAX}	82.64 *	87.50	77.78	-	-	-
VAE _{la4Ch}	78.30*	84.38	72.22	-	-	-
GSTFT CRT echocardiography						
	BACC(%)	SEN(%)	SPE(%)	BACC(%)	SEN(%)	SPE(%)
MMDL	-	-	-	72.86	85.72	60.00
VGG _{Echo}	-	-	-	55.71	71.43	40.00