Cardiac Shape Priors for Equivariant Neural Field Representations of Cine MRI

Sacha E. Buijs[1]✉, Fleur V. Y. Tjong[2,3], Erik J. Bekkers[4], Soufiane Ben Haddou[5,6*], and Thijs P. Kuipers[7,8*]

[1] Informatics Institute, University of Amsterdam, Amsterdam, The Netherlands
[2] Amsterdam UMC Heart Center, Department of Clinical and Experimental Cardiology, Amsterdam University Medical Centers, Amsterdam Cardiovascular Sciences, University of Amsterdam, Netherlands
[3] Department of Medicine, Cardiovascular Institute, Stanford University School of Medicine, Stanford, CA, USA.
[4] Amsterdam Machine Learning Lab, University of Amsterdam, The Netherlands
[5] QurAI group, Informatics Institute, University of Amsterdam, The Netherlands
[6] Amsterdam Cardiovascular Sciences, Amsterdam UMC, The Netherlands
[7] Department of Biomedical Engineering and Physics, Amsterdam UMC, The Netherlands
[8] Department of Radiology and Nuclear Medicine, Amsterdam UMC, The Netherlands
*These authors contributed equally to this work
✉ Corresponding author: sacha.buijs@student.uva.nl

Abstract. Cardiac shape has been shown to contain prognostic information beyond standard scalar clinical measures. Yet, cine MRI representation learning often relies on image intensities alone. Therefore, we incorporate cardiac shape as a prior into Equivariant Neural Fields (ENFs), which represent cine MRI images as continuous, localized, and geometrically grounded latent pointclouds. To obtain cardiac shape representations, we introduce HeartSDF, a latent set signed distance field model that encodes segmentation-derived left ventricular and myocardial anatomy as continuous 3D geometry. We study two ways of using cardiac shape in ENFs: an explicit prior, where HeartSDF latents are added to the ENF representation, and an implicit prior, where reconstruction is focused on the cardiac region by utilizing segmentations during training and inference. Experiments on UK Biobank show that HeartSDF captures global cardiac geometry and that cardiac shape provides a useful prior for clinical endpoint classification. Both explicit and implicit priors improve ENF-based classification, demonstrating the value of incorporating anatomical structure into neural field based cine MRI representations.

Keywords: Equivariant Neural Fields · Shape Representation · Cardiac Risk Stratification.

1 Introduction

Cardiovascular disease is one of the leading causes of death worldwide [1]. A major challenge for the treatment of cardiac disease is that pathologies can remain asymptomatic until a severe event occurs. For some patients, the first manifestation of the disease is cardiac arrest, which may result in sudden cardiac death (SCD) [1]. Patients considered at high risk for cardiac arrest can receive an implantable cardioverter defibrillator (ICD), which monitors the heart rhythm and terminates life-threatening ventricular arrhythmias [10]. However, identifying which patients benefit from ICD therapy remains difficult [15]. Current selection criteria rely heavily on left ventricular ejection fraction (LVEF), despite its limited prognostic value [15, 16]. Since ICD implantation is costly and associated with complications, improved risk stratification is needed [2].

Cardiac shape provides a clinically meaningful source of information for cardiac risk stratification; abnormalities in ventricular size, wall thickness, and chamber geometry are associated with ventricular arrhythmia and SCD risk [2, 19]. LVEF provides a partial measure of cardiac shape and is therefore widely used in risk stratification. However, LVEF and other standard scalar measures that are used in clinical practice, such as ventricular volumes and chamber diameters, do not capture full 3D cardiac geometry. This is an important limitation, as 3D left ventricular (LV) shape has been shown to contain prognostic information beyond these standard clinical measures [2, 19]. This motivates the need for richer cardiac shape representations in risk stratification.

Cine magnetic resonance imaging (MRI) is widely used for both image-based risk assessment and segmentation-derived cardiac shape analysis, as it is non-invasive, avoids ionizing radiation, and provides high soft-tissue contrast throughout the cardiac cycle [20]. Deep learning methods have increasingly been used to learn prognostic representations from cine MRI [16, 17, 34]. Most approaches rely on grid-based image encoders and learn directly from image intensities. Neural fields offer an alternative by modeling images as continuous functions over spatial or spatio-temporal coordinates [8, 31]. Equivariant Neural Fields (ENFs) extend this idea by representing an image with a localized latent pointcloud, where each latent has a spatial position and context vector [30]. This gives ENFs a geometrically grounded representation space that is interpretable, localized, and equivariant. Recent work has shown that ENFs can reconstruct cine MRI and learn representations that are competitive with ResNet for downstream cardiac endpoint classification [12, 31]. However, ENF representations are inferred from image intensities alone and do not explicitly encode cardiac anatomy. In this work, we build on Wiers et al. [31] by incorporating cardiac shape as a prior in ENF-based cine MRI representation learning, guiding the representations towards anatomy rather than image intensities alone.

Shape priors are well established in medical image analysis, where they are used to encourage anatomically plausible segmentations, guide reconstruction, and learn compact anatomical representations for prediction or generation [3–5, 11, 25, 29, 35]. Chen et al. [6] demonstrate that cardiac shape can improve cine MRI representation learning, but their approach relies on discrete meshes

and grid-based image representations. In contrast, we incorporate a continuous cardiac shape prior into the localized, geometrically grounded representation space of ENFs.

Neural fields provide such a continuous way to model shape [21, 22, 33]. In this work, we represent cardiac shape as a signed distance field (SDF), which assigns each point in space its signed distance to the cardiac surface [22]. Building on VesselSDF [18], a 3DShape2VecSet-style latent set model [33], we introduce HeartSDF to encode segmentation-derived cardiac shapes into compact latent sets and decode them as continuous SDFs. This representation is naturally compatible with ENFs, since both the image and shape models represent a sample using latent sets.

In this work, we investigate whether cardiac shape priors improve ENF-based cine MRI representations. We incorporate shape in two complementary ways: first, by using HeartSDF latents derived from segmentation-based LV and myocardial (MYO) shapes as explicit priors for ENF reconstruction, and second as an implicit prior, by sampling reconstruction coordinates directly from the cardiac region to allocate ENF capacity to the heart rather than the full image grid. Our contributions are threefold: (i) We extend ENF-based cine MRI representation learning with explicit and implicit cardiac shape priors; (ii) We represent cardiac anatomy using a continuous SDF, adapting a latent set shape model to LV and MYO shapes; (iii) We evaluate the learned representations on clinical endpoint classification and show that incorporating cardiac shape, either implicitly or explicitly, consistently improves downstream performance over the unconstrained full-image ENF baseline. The implicit prior often performs competitively with the explicit prior, while the added value of explicit shape information on top of the implicit prior is endpoint-dependent.

2 Method

Our method consists of three components: HeartSDF, a 3D cardiac shape model that represents cardiac shape as an SDF; a 2D ENF that learns cine MRI image representations; and a clinical endpoint classifier operating on the resulting latent pointclouds (see Figure 1).

2.1 Data and Preprocessing

In order to train and evaluate these components, we use short-axis cine MRI scans and pathology labels from UK Biobank [27], following the cohort construction of [31]. The final cohort contains 1,739 subjects and includes healthy subjects and three downstream endpoints: LVEF $< 40\%$ (284 subjects), cardiomyopathy (CMP) (313 subjects), and SCD (304 subjects). For the LVEF endpoint, manually derived LVEF measurements from UK Biobank were used rather than the automated UK Biobank inline measurements. Each cine MRI contains 50 temporal frames with a typical in-plane resolution of 1.8×1.8 mm and an 8 mm slice thickness. Cardiac segmentations are obtained with CINeMA [9].

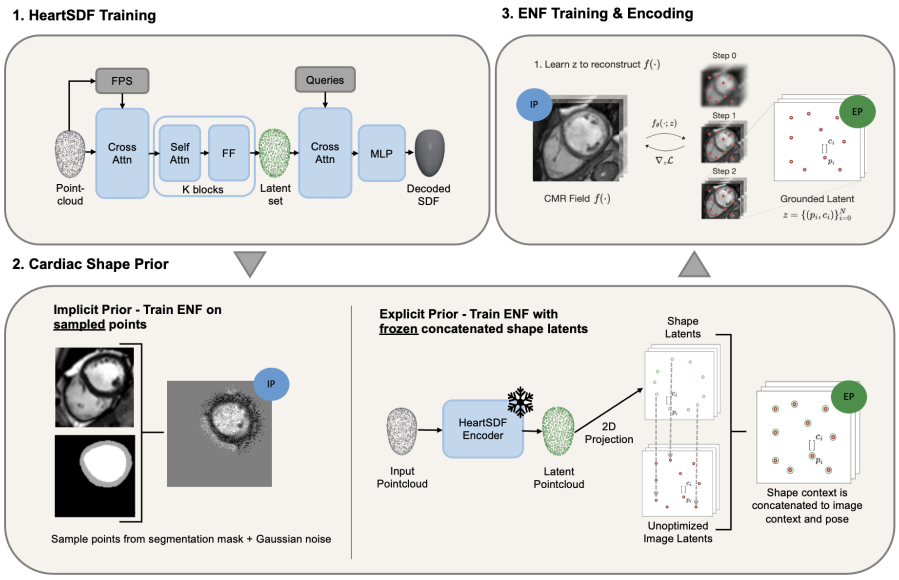


Fig. 1: Overview of the proposed cardiac shape prior framework. (1) HeartSDF learns an SDF representation of segmentation-derived cardiac shape. (2) The cardiac shape prior is incorporated into ENFs either implicitly (IP, blue) through restricting reconstruction to the cardiac region, or explicitly (EP, green) through HeartSDF shape latents. (3) The ENF represents each cine MRI image as a grounded latent pointcloud by optimizing image contexts to reconstruct the image field.

We use the end-diastolic (ED) and end-systolic (ES) frames, where ED is the first timeframe and ES is the frame with the smallest LV volume. Although the source data consist of cine MRI sequences, the experiments in this work use only the ED and ES frames and therefore do not model temporal dynamics across the full cardiac cycle. Images are cropped around the heart, min-max normalized per 2D slice, and coordinates are normalized between $[-1, 1]^2$, following Wiers et al. [31]. Cardiac shapes are constructed from segmentation slices, which are first center-of-mass aligned, inspired by Sander et al. [25]. The aligned segmentations are then converted to meshes and transformed to physical space. To preserve relative differences in heart size across participants, all shapes are normalized to $[-1, 1]^3$ using a single global scale factor determined by the largest heart in the dataset, and subsequently processed into watertight surfaces. For HeartSDF training, we sample surface points from each mesh and generate off-surface points by perturbing these samples with Gaussian noise.

2.2 Equivariant Neural Field Cine MRI Representation

ENFs are used as the image representation model in this work. They build on neural fields and conditional neural fields (CNFs). Neural fields represent data as continuous functions that map coordinates to values, allowing images to be queried at arbitrary locations rather than stored only on a fixed grid [32]. CNFs extend this idea by representing multiple fields with one function through a conditioning variable $\mathbf{z} \in \mathbb{R}^K$. An image is then represented as a function

$$f_\theta(\mathbf{x}, \mathbf{z}) : \mathbb{R}^D \times \mathbb{R}^K \to \mathbb{R}, \tag{1}$$

where $\mathbf{x} \in \mathbb{R}^D$ denotes an image coordinate and $f_\theta(\mathbf{x}, \mathbf{z})$ the corresponding intensity value given the conditioning variable. This conditioning variable can subsequently be used for downstream tasks [8]. A limitation of CNFs is that this latent is global, meaning that locality and spatial relations are not explicitly preserved in the representation [30]. ENFs overcome this limitation by replacing the global conditioning variable with a geometrically grounded, localized latent pointcloud [30].

Geometric Grounding The ENF representation for a 2D cine MRI slice is given by

$$\mathbf{z}_{\text{img}} = \{(\mathbf{p}_{\text{img},i}, \mathbf{c}_{\text{img},i})\}_{i=1}^N, \tag{2}$$

where $\mathbf{p}_{\text{img},i} \in \mathbb{R}^2$ denotes the latent position and $\mathbf{c}_{\text{img},i}$ is the corresponding image context vector, which stores local appearance information [30].

In this work, translation-equivariant ENFs are utilized. ENFs are constructed to satisfy the *steerability condition* which ensures a translation of the input image corresponds to the same translation of the latent poses. For a translation $\mathbf{t} \in \mathbb{R}^2$ and translation operator T_t the steerability condition is given by

$$f_\theta(T_\mathbf{t}^{-1}\mathbf{x}; \mathbf{z}_{\text{img}}) = f_\theta(\mathbf{x}; T_\mathbf{t}\mathbf{z}_{\text{img}}). \tag{3}$$

Specifically, the latent poses are transformed with the image, while context vectors are unchanged:

$$T_\mathbf{t}\mathbf{z}_{\text{img}} = \{(\mathbf{p}_{\text{img},i} + \mathbf{t}, \mathbf{c}_{\text{img},i})\}_{i=1}^N. \tag{4}$$

To enforce translation equivariance, the ENF does not depend on absolute image coordinates and latent poses, but on their relative relation $\mathbf{x} - \mathbf{p}_i$. For translation equivariance, this relative relation is

$$\mathbf{a}(\mathbf{x}, \mathbf{p}_{\text{img},i}) = \phi(\mathbf{x} - \mathbf{p}_{\text{img},i}), \tag{5}$$

where ϕ is a Gaussian random Fourier feature (RFF) embedding, which reduces the low-frequency bias of neural fields [23,28]. As this relation is preserved under translations, this ensures translation equivariance.

Locality The ENF conditions the underlying function on the latent pointcloud through cross-attention. For a query coordinate $\mathbf{x}_m$, the output is computed as

$$f_\theta(\mathbf{x}_m; \mathbf{z}) = \mathbf{W}_o \sum_{i=1}^{N} \mathrm{att}(\mathbf{x}_m, \mathbf{z}_i)\mathbf{v}\left(\mathbf{a}(\mathbf{x}_m, \mathbf{p}_i), \mathbf{c}_i\right), \tag{6}$$

where

$$\mathrm{att}(\mathbf{x}_m, \mathbf{z}_i) = \underset{i}{\mathrm{softmax}}\left(\frac{\mathbf{q}(\mathbf{a}(\mathbf{x}_m, \mathbf{p}_i))^\top \mathbf{k}(\mathbf{c}_i)}{\sqrt{d_k}} + \mu_\sigma(\mathbf{x}_m, \mathbf{p}_i)\right). \tag{7}$$

Here, $\mu_\sigma(\mathbf{x}_m, \mathbf{p}_i)$ is the log-domain form of a Gaussian spatial window that enforces locality,

$$\mu_\sigma(\mathbf{x}_m, \mathbf{p}_i) = -\sigma_{\mathrm{att}} \left\|\mathbf{x}_m - \mathbf{p}_i\right\|^2. \tag{8}$$

The Gaussian window, parametrized by σ_{att}, enforces interactions between queries and latent points to remain local; queries far from latents receive lower attention weights. For computational efficiency, each query coordinate only attends to its k nearest latent poses [30].

Training and Inference The ENF is trained to reconstruct image intensities on a set of sampled coordinates $\mathcal{X}$:

$$\mathcal{L}_{\mathrm{recon}} = \mathrm{MSE}\left(f_\theta(\mathcal{X}; \mathbf{z}_{\mathrm{img}}), y(\mathcal{X})\right), \tag{9}$$

where $f_\theta(\mathcal{X}; \mathbf{z}_{\mathrm{img}})$ and $y(\mathcal{X})$ denote the predicted and ground-truth intensities evaluated at all coordinates in $\mathcal{X}$ and MSE denotes the mean squared error.

The ENF has a decoder-only architecture, and is therefore trained using Model Agnostic Meta Learning [7], as in Wiers et al. [31]. During the inner loop, the decoder parameters are fixed, and only the image context vectors $\mathbf{c}_{\mathrm{img},i}$ are optimized. During the outer loop, the decoder parameters are updated based on the reconstruction loss from the inner loop optimization. At inference time, the decoder is kept fixed, and ENF latents are obtained by optimizing the context vectors for a small number of inner-loop steps.

2.3 HeartSDF Cardiac Shape Representation

We represent cardiac shape using HeartSDF, a supervised cardiac adaptation of VesselSDF [18]. Separate models are trained for the LV and MYO shapes to account for their different geometric complexity. The input is a surface pointcloud sampled from a segmentation-derived cardiac mesh. HeartSDF maps each input pointcloud to a latent set with a variational autoencoder [14]. Latent points are selected via farthest point sampling (FPS) and embedded into the full-resolution pointcloud via cross-attention. An SDF decoder then predicts signed distances at arbitrary 3D query coordinates by attending to this latent set. Input coordinates are encoded using RFF embeddings [24, 28]. HeartSDF is trained on

signed distances computed from segmentation-derived meshes, rather than using the self-supervised surface, normal, and Eikonal losses of VesselSDF [18]. For surface and off-surface query coordinates $\mathcal{Q}$ the training objective is

$$\mathcal{L}_{\text{HeartSDF}} = \text{MSE}(\tilde{d}(\mathcal{Q}, \mathbf{z}), d(\mathcal{Q})) + \lambda_{\text{KL}} \mathcal{L}_{\text{KL}}, \tag{10}$$

where $d(\mathcal{Q})$ denotes the ground truth distances for all query coordinates and $\tilde{d}(\mathcal{Q}, \mathbf{z})$ the prediction. Here, $\mathcal{L}_{\text{KL}}$ regularizes the variational latent distribution. After training, the encoder provides a patient-specific 3D cardiac shape representation which is used as an explicit prior in ENF training.

2.4 Cardiac Shape Priors

We compare two cardiac shape priors for ENF representation learning: an explicit prior, where HeartSDF shape representations are incorporated during ENF training; and an implicit prior, where reconstruction is restricted to the cardiac region.

Explicit Prior HeartSDF provides 3D shape representations while the ENF operates on 2D slices. Therefore, we first select the shape latents closest to each slice and project them onto the image plane. The projected HeartSDF latents are then concatenated with the ENF latents by matching their spatial positions using linear-sum assignment with a Euclidean distance cost. For each pair, we concatenate the image and shape context vectors while keeping the image pose:

$$\mathbf{z}_{\text{concat}} = \{(\mathbf{p}_{\text{img},i}, [\mathbf{c}_{\text{img},i} \mid \mathbf{c}_{\text{shape},j}])\}_{i=1}^{N}. \tag{11}$$

During ENF training and inference, the HeartSDF latents are fixed, and only the image context vectors are optimized.

Implicit Prior The implicit prior restricts reconstruction queries to the cardiac region without the use of HeartSDF latents. We sample query coordinates from the union of the LV and MYO masks and add Gaussian noise to include nearby background intensities, and obtain non-integer intensities by bilinear interpolation. Since neural fields can be queried at arbitrary coordinates, this focuses ENF capacity on the cardiac region. For the remainder of the research we will refer to this implicit prior as the Cardiac Sampling setting (CS). In this setting, ENF latent poses are initialized with FPS over the sampled cardiac coordinates. We compare CS to the full image reconstruction objective (FI). In this case the latents are placed in a regular grid over the image. CS uses the union of the LV and MYO masks to define the cardiac sampling region, whereas the explicit prior experiments consider LV and MYO shape representations separately; a combined LV+MYO explicit prior is not evaluated in this study.

2.5 Clinical Endpoint Classification

For clinical endpoint classification, we train a Platonic Transformer [13] classifier with Rotary Position Embeddings [26], which enables translation equivariance. The classifier takes latent pointcloud representations as input, combining the ED and ES representations from either the ENF or the shape model.

Image-based Classification To combine the timeframes, we match ED and ES latent representations in the same manner as the image and shape representations (section 2.4). The resulting representation contains one pose and the concatenated ED-ES context. We use the middle three short-axis slices, following Wiers et al. [31]. After ED-ES concatenation is performed per slice, the three slices are combined into a single pointcloud by adding a relative slice coordinate:

$$\mathbf{p}_i^{2D} = (x, y) \quad \rightarrow \quad \mathbf{p}_i^{3D} = (x, y, z), \qquad z \in \{-1, 0, 1\}. \tag{12}$$

These representations are then used as input to the classifier.

Shape-only Classification For shape-only classification, we use the 3D HeartSDF latent representations directly, after matching and concatenating the ED and ES timepoints. The resulting 3D latent pointcloud is passed to the same Transformer classifier.

Training and Evaluation For each endpoint, we split the dataset at the patient level into balanced training, validation, and test sets using a 60/20/20 ratio. For LVEF, CMP, and SCD, the training/validation/test splits contain 170/57/57, 187/63/63, and 182/61/61 patients, respectively. The ENF and HeartSDF representation models and downstream classifiers use different data splits, such that classifier test subjects may appear during representation model training. However, ENF and HeartSDF are trained without clinical endpoint labels, and therefore no label leakage occurs between representation learning and downstream classification. For each endpoint and representation type, we perform a Bayesian hyperparameter search over the learning rate and weight decay, using 10 optimization trials. Both parameters are sampled log-uniformly, with the learning rate ranging from 10^{-4} to 10^{-3} and weight decay from 10^{-3} to 10^{-1}. Model selection is based on validation Area Under the Receiver Operating Characteristic Curve (AUROC) using 3-fold cross-validation, after which the selected model is evaluated on the held-out test set.

3 Results and Evaluation

This section describes the experimental setup of the ENF and HeartSDF, after which the results on reconstruction performance and clinical endpoint classification are discussed.

3.1 Experimental Setup

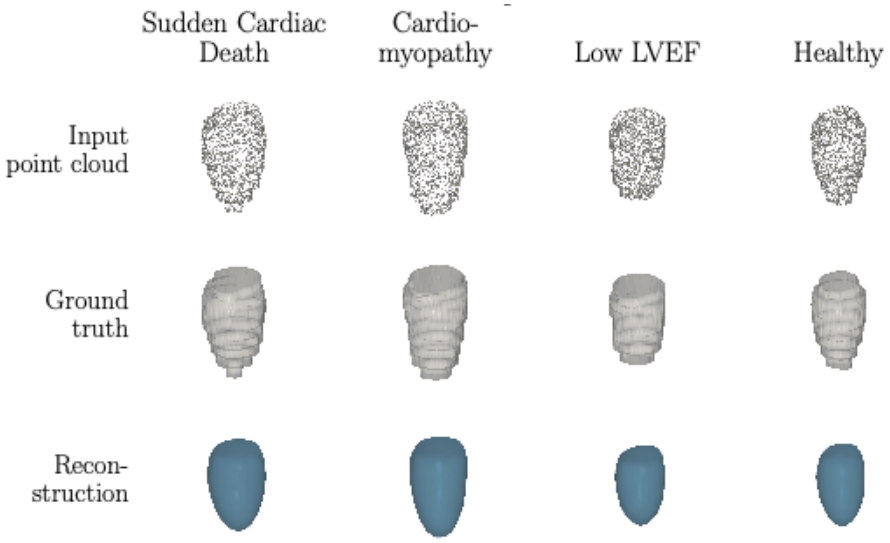
The ENF and HeartSDF models are evaluated separately. Image reconstruction is assessed using the peak signal-to-noise ratio (PSNR), while shape reconstruction is evaluated qualitatively and quantitatively through the Chamfer Distance (CD) and the 95th percentile Hausdorff Distance (HD-95). The quality of all learned representations is evaluated on clinical endpoint classification. This includes HeartSDF shape representations and ENF representations from the FI and CS settings, with and without explicit shape priors. For the three clinical endpoints, we report the AUROC. This is compared against a ResNet baseline [12].

ENF The ENF representations consist of 64 latent points with a dimensionality of 16. We use an 80/10/10 patient-level split, resulting in 1,391 training patients, 174 validation and 174 test patients. The ENF is trained on the middle 6 slices for both ED and ES timepoints, yielding 16,692 training samples. The decoder has a hidden and attention dimensionality of 128 and 3 attention heads. We use $k = 4$ nearest-neighbor attention and a Gaussian window with scale 2. The ENF is trained for 2 epochs with a batch size of 16. Each inner loop the image context latents are initialized uniformly with 1/16 and updated for 3 inner steps with a learning rate of 60.0. The decoder parameters are updated with full second-order gradients using Adam with learning rate 5×10^{-4}. For both the outer and inner-loop learning rates we use 50 warm-up steps. In the FI setting, all 8,100 image coordinates are used as queries. In the CS setting, we sample 6,480 coordinates from the shared MYO and LV mask, corresponding to 80% of the FI query coordinates, adding Gaussian noise with a standard deviation of 5 pixels.

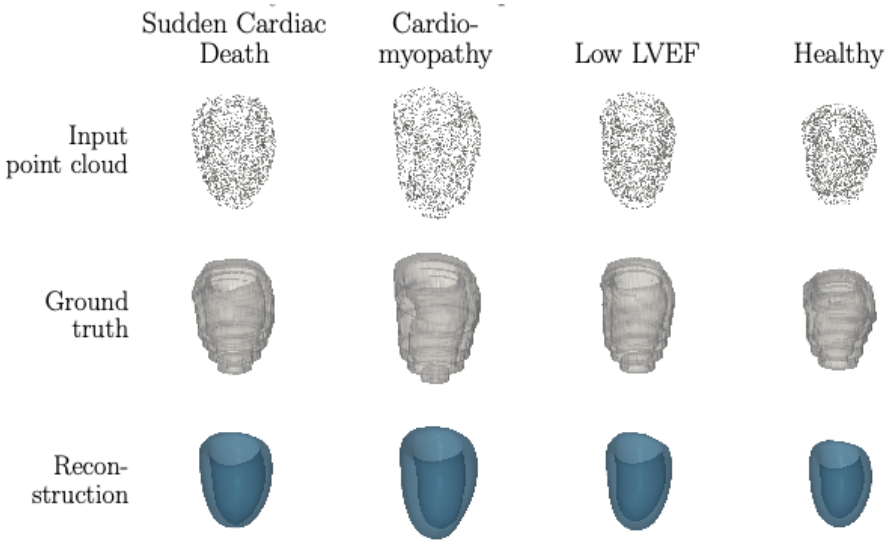
HeartSDF We sample 4096 coordinates from the shape surface, of which 2048 serve as input to the encoder and 2048 serve as surface query coordinates. Additionally we sample 1024 off-surface query coordinates. For MYO, half of the off-surface coordinates are sampled from the interior whenever available to better represent negative SDF values. The encoder downsamples the input to 512 latent points, and the decoder predicts SDF values for all surface and off-surface queries. LV and MYO models use the same architecture, consisting of six self-attention blocks, and differ only in the RFF sigma, which is set to 0.3 for LV and 0.5 for MYO. The autoencoder is trained for 10 epochs with a batch size of 16, using 14,416 training, 1,902 validation and 1,902 test samples with a patient-level split including both ED and ES. We use a linear learning-rate schedule from 1×10^{-6} to 1×10^{-4} over 5 epochs and back to 1×10^{-6}, a regularization weight of 1×10^{-3}, and data augmentation with rotations up to ± 0.2 radians and translations up to ± 0.1.

3.2 Reconstruction of Cardiac Shapes

HeartSDF produces smooth, anatomically plausible shapes that follow the global geometry of the ground truth (Figure 2). The model reconstructs continuous



(a) LV reconstructions



(b) MYO reconstructions

Fig. 2: Shape reconstructions for the LV and MYO.

shapes from sparse slice-based pointclouds, smoothing slice-induced artifacts and handling missing regions such as the apex in an anatomically plausible manner. However, fine anatomical detail is lost. The reconstructions preserve pathology-

Table 1: Shape reconstruction performance on the test set. Values are reported as mean ± standard deviation across test samples. CD is reported in normalized coordinate units ($\times 10^3$), while HD95 is reported in mm.

Structure	**CD** ($\times 10^3$)	**HD95 (mm)**
LV	2.55 ± 0.67	6.60 ± 0.88
MYO	2.48 ± 0.69	6.69 ± 1.30

related morphology, with SCD and CMP cases appearing larger and more dilated than healthy controls, whereas low-LVEF shapes are less visually distinct.

The qualitative results are supported quantitatively with comparable LV and MYO reconstruction performance in CD and HD95 (Table 1). Overall, HeartSDF captures global cardiac anatomy while smoothing fine details and completing missing regions.

Table 2: Reconstruction performance for full image reconstruction and cardiac sampling. The performance is denoted with PSNR in dB. The best performing models are highlighted. PSNR is computed over the respective reconstruction coordinate sets: the full image grid for FI and the sampled cardiac region for CS.

Model	**Full image PSNR**	**Cardiac Sampling PSNR**
No explicit prior	**27.174**	**27.538**
LV prior	24.727	24.353
MYO prior	24.856	25.072

3.3 Reconstruction of Cine MRI Images

The ENF without an explicit prior achieves the highest PSNR in both the FI and CS settings (Table 2). Because FI and CS PSNRs are evaluated over different coordinate supports, their absolute values are not directly comparable. Nevertheless, CS yields numerically higher PSNR values than FI for the no-explicit-prior and MYO-prior models. While this difference cannot be interpreted as a direct improvement in reconstruction quality, it is consistent with the hypothesis that restricting reconstruction to the cardiac region reduces the complexity of the reconstruction task. Within both reconstruction settings, adding an explicit shape prior reduces PSNR relative to the corresponding no-explicit-prior model under our two-epoch training budget. Qualitatively, CS concentrates reconstruction capacity on the heart while ignoring surrounding anatomy (Figure 3).

3.4 Clinical Endpoint Classification

We evaluate all learned representations on downstream clinical endpoint classification. This includes HeartSDF shape latents, ENF latents without an explicit

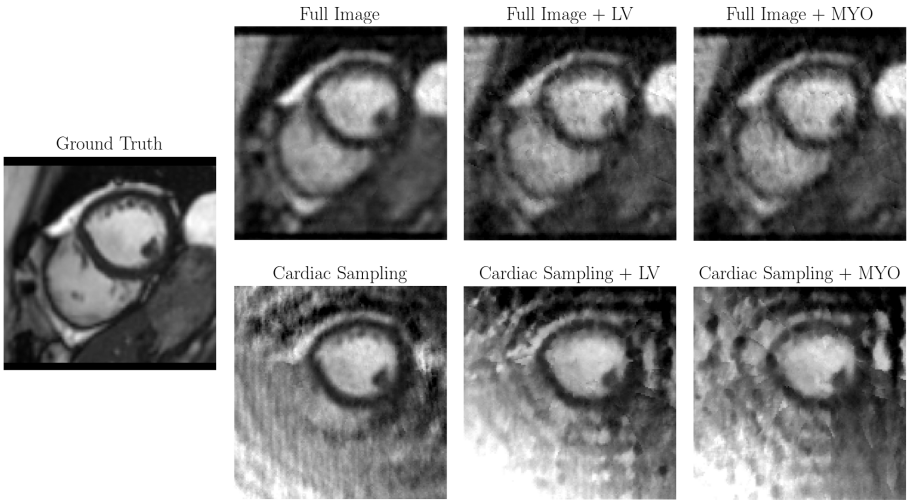


Fig. 3: ENF image reconstructions.

prior from the FI and CS settings, and ENF latents combined with explicit LV or MYO shape priors. The performance is compared against a ResNet baseline [12].

The shape-only baselines show that cardiac geometry contains prognostic information (Table 3). LV shape performs close to ResNet for LVEF prediction, consistent with the dependence of LVEF on LV cavity volume. MYO shape performs strongly for CMP, suggesting that MYO geometry and wall structure are informative for this endpoint. For SCD, all methods remain close to chance, although LV shape improves over the FI-only representation.

CS substantially improves the ENF without an explicit prior over FI reconstruction across all endpoints, with the largest gain for CMP. This indicates that restricting reconstruction to the cardiac region provides a strong implicit shape prior and yields more informative image representations. CS without an explicit prior achieves the best overall CMP performance and approaches the ResNet baseline for LVEF prediction.

Explicit shape priors are most beneficial in the FI setting. Adding LV or MYO shape improves FI performance across all endpoints. For SCD, FI + LV shape has the highest mean AUROC, although all methods remain relatively close to chance and the differences should be interpreted cautiously at this sample size. In the CS setting, the added value of the explicit prior is smaller and more endpoint-dependent: MYO shape slightly improves LVEF, LV shape improves SCD while remaining near chance, and neither improves over CS-only for CMP.

Overall, these results suggest that a cardiac shape prior can improve ENF representations, although the benefit depends on the prior formulation and clinical endpoint.

Table 3: Clinical endpoint classification AUROC results. Values are reported as mean ± standard deviation across the three cross-validation folds. The representations with the highest mean AUROC are highlighted. The CS representations require segmentations at inference.

Representation	**LVEF** $< 40\%$	**CMP**	**SCD**
ResNet	$\mathbf{0.9745} \pm 0.0023$	0.6478 ± 0.0101	0.5500 ± 0.0079
FI - no explicit prior	0.8908 ± 0.0255	0.5732 ± 0.0177	0.5305 ± 0.0116
CS - no explicit prior	0.9673 ± 0.0137	$\mathbf{0.7643} \pm 0.0107$	0.5889 ± 0.0510
LV shape only	0.9710 ± 0.0093	0.6850 ± 0.0052	0.5791 ± 0.0356
FI + LV shape	0.9686 ± 0.0099	0.7173 ± 0.0201	$\mathbf{0.6033} \pm 0.0192$
CS + LV shape	0.9707 ± 0.0086	0.7540 ± 0.0141	0.6011 ± 0.0334
MYO shape only	0.9453 ± 0.0075	0.7398 ± 0.0569	0.5576 ± 0.0422
FI + MYO shape	0.9491 ± 0.0177	0.6464 ± 0.0520	0.5445 ± 0.0293
CS + MYO shape	0.9738 ± 0.0117	0.7495 ± 0.0140	0.5597 ± 0.0365

4 Conclusion and Discussion

This work investigated whether cardiac shape priors improve ENF-based cine MRI representations, either through an explicit prior from HeartSDF shape representations or an implicit prior introduced by sampling coordinates from the cardiac region. Overall, the results show that cardiac shape can improve downstream clinical endpoint classification and that HeartSDF captures global cardiac geometry. However, under the training budget used in this study, the current strategy for incorporating the explicit shape prior does not improve image reconstruction.

The HeartSDF reconstructions show that the model can represent cardiac shapes derived from segmentation in a plausible way. The reconstructed LV and MYO surfaces preserve global anatomy and handle segmentation-induced artifacts and missing information. However, they smooth over fine anatomical detail, which may be explained by the low-frequency bias of neural fields.

Under our two-epoch ENF training budget, explicit shape priors reduce image reconstruction PSNR in both the FI and CS settings. This suggests that the ENF decoder does not effectively use the additional shape information for the reconstruction task. This is potentially caused by a harder initialization strategy due to the concatenated shape representations or by the dimensionality mismatch between 2D images and 3D shapes. As a result, the decoder may rely primarily on the image latents and treat the fixed shape latents as noise.

Despite this, the downstream results show that cardiac geometry contains clinically relevant information, suggesting that reconstruction quality should not be interpreted as a direct proxy for the clinical relevance of the learned representations. In the FI setting, explicit LV and MYO shape priors improve over the ENF without an explicit prior for all endpoints. This supports the motivation behind this work that cardiac shape provides information that is not fully

captured by unconstrained image reconstruction. For SCD, these improvements remain modest, with AUROCs close to chance, and do not provide strong evidence of improved SCD prediction. These results should therefore be interpreted as a proof of concept for anatomically informed cine CMR representations. The benefit of the explicit shape prior depends strongly on the baseline representation. When the implicit CS prior is used, the ENF without an explicit prior yields better results, indicating that restricting reconstruction to the cardiac region indeed acts as a strong implicit anatomical prior. In this setting, adding HeartSDF representations gives only small and task-dependent improvements.

Several limitations remain. First, slice alignment is based on the center of mass, which is a simple approximation and may remove relevant anatomical variation. As such, utilizing more sophisticated shape representation approaches that account for the difficulties in cine MRI reconstruction may improve the results further. Second, the current shape prior implementation strategy may not be best suited for the ENF. As the strategy does not guarantee that the ENF decoder uses the shape latents, the model may learn to ignore them. As such, other prior strategies should be explored. A natural direction is to train the ENF and HeartSDF models jointly, so that image and shape latents are learned in a shared representation space. Last, the ENF models were trained under limited compute constraints and on a relatively small set, which may limit representation quality and the robustness of the downstream results. Extending this to more data and more training iterations could help capture the fine details in the image and improve the representations overall.

Therefore, future work should focus on tighter coupling between image and shape representations, better shape representations, and more compute. Additional extensions could include inferring shape directly from image latents or incorporating other structures and modalities.

In conclusion, cardiac shape is a useful prior for ENF-based cine MRI representation learning. The explicit shape priors improve full image ENF representations, and cardiac sampling provides a strong implicit anatomical prior. Future neural field representations of cardiac MRI should therefore incorporate anatomical structure more directly.

Acknowledgments. Dr. Tjong has grants from the Dutch Heart Foundation (Dekker 03-005-2025-0129), Dutch Research Council (NWO Rubicon 452019308, NWO AINed XS NGF.1609.243.045), Amsterdam Cardiovascular Sciences and Health Holland. This research has been conducted using the UK Biobank Resource under application number 24711.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Adabag, A.S., Luepker, R.V., Roger, V.L., Gersh, B.J.: Sudden cardiac death: epidemiology and risk factors. Nature Reviews. Cardiology **7**(4), 216–225 (Apr 2010). https://doi.org/10.1038/nrcardio.2010.3

2. Balaban, G., Halliday, B.P., Hammersley, D., Rinaldi, C.A., Prasad, S.K., Bishop, M.J., Lamata, P.: Left ventricular shape predicts arrhythmic risk in fibrotic dilated cardiomyopathy. Europace **24**(7), 1137–1147 (Dec 2021). https://doi.org/10.1093/europace/euab306, https://pmc.ncbi.nlm.nih.gov/articles/PMC9301973/
3. Beetz, M., Banerjee, A., Grau, V.: Reconstructing 3d cardiac anatomies from misaligned multi-view magnetic resonance images with mesh deformation u-nets. In: Geometric Deep Learning in Medical Image Analysis. pp. 3–14. PMLR (2022)
4. Beetz, M., Banerjee, A., Li, L., Camps, J., Rodriguez, B., Grau, V.: 3D cardiac shape analysis with variational point cloud autoencoders for myocardial infarction prediction and virtual heart synthesis. Computerized Medical Imaging and Graphics **124**, 102587 (Sep 2025). https://doi.org/10.1016/j.compmedimag.2025.102587, https://www.sciencedirect.com/science/article/pii/S0895611125000965
5. Beetz, M., Corral Acero, J., Banerjee, A., Eitel, I., Zacur, E., Lange, T., Stiermaier, T., Evertz, R., Backhaus, S.J., Thiele, H., et al.: Interpretable cardiac anatomy modeling using variational mesh autoencoders. Frontiers in cardiovascular medicine **9**, 983868 (2022)
6. Chen, X., Xia, Y., Dall'Armellina, E., Ravikumar, N., Frangi, A.F.: Joint shape/texture representation learning for cardiovascular disease diagnosis from magnetic resonance imaging. European Heart Journal - Imaging Methods and Practice **2**(1), qyae042 (Jan 2024). https://doi.org/10.1093/ehjimp/qyae042, https://academic.oup.com/ehjimp/article/doi/10.1093/ehjimp/qyae042/7672911
7. Finn, C., Abbeel, P., Levine, S.: Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks. In: Proceedings of the 34th International Conference on Machine Learning. pp. 1126–1135. PMLR (Jul 2017), https://proceedings.mlr.press/v70/finn17a.html
8. Friedrich, P., Bieder, F., Cattin, P.C.: Medfuncta: Modality-agnostic representations based on efficient neural fields. arXiv e-prints pp. arXiv–2502 (2025)
9. Fu, Y., Bai, W., Yi, W., Manisty, C., Bhuva, A.N., Treibel, T.A., Moon, J.C., Clarkson, M.J., Davies, R.H., Hu, Y.: A versatile foundation model for cine cardiac magnetic resonance image analysis tasks. arXiv preprint arXiv:2506.00679 (2025)
10. Glikson, M., Friedman, P.A.: The implantable cardioverter defibrillator. The Lancet **357**(9262), 1107–1117 (2001)
11. Haddou, S.B., Alvarez-Florez, L., Bekkers, E.J., Tjong, F.V., Amin, A.S., Bezzina, C.R., Išgum, I.: Synthesis of late gadolinium enhancement images via implicit neural representations for cardiac scar segmentation. arXiv preprint arXiv:2602.11942 (2026)
12. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
13. Islam, M.M., Anand, R., Wessels, D.R., Kruiff, F.d., Kuipers, T.P., Ying, R., Sánchez, C.I., Vadgama, S., Bökman, G., Bekkers, E.J.: Platonic Transformers: A Solid Choice For Equivariance (Oct 2025). https://doi.org/10.48550/arXiv.2510.03511, http://arxiv.org/abs/2510.03511, arXiv:2510.03511 [cs]
14. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114 (2013)
15. Kolk, M.Z., Ruiperez-Campillo, S., Deb, B., Bekkers, E.J., Allaart, C.P., Rogers, A.J., Van Der Lingen, A.L.C., Alvarez Florez, L., Isgum, I., De Vos, B.D., et al.:

Optimizing patient selection for primary prevention implantable cardioverter-defibrillator implantation: utilizing multimodal machine learning to assess risk of implantable cardioverter-defibrillator non-benefit. Europace **25**(9), euad271 (2023)
16. Kolk, M.Z., Ruipérez-Campillo, S., Wilde, A.A., Knops, R.E., Narayan, S.M., Tjong, F.V.: Prediction of sudden cardiac death using artificial intelligence: Current status and future directions. Heart Rhythm **22**(3), 756–766 (2025)
17. Krebs, J., Mansi, T., Delingette, H., Lou, B., Lima, J.A., Tao, S., Ciuffo, L.A., Norgard, S., Butcher, B., Lee, W.H., et al.: Cine cardiac magnetic resonance to predict ventricular arrhythmia (certainty). Scientific reports **11**(1), 22683 (2021)
18. Kuipers, T.P., Konduri, P.R., Bekkers, E.J., Marquering, H.: Self-supervised synthetic cerebral vessel tree generation using semantic signed distance fields. In: Medical Imaging with Deep Learning (2025)
19. Mauger, C.A., Gilbert, K., Suinesiaputra, A., Bluemke, D.A., Wu, C.O., Lima, J.A., Young, A.A., Ambale-Venkatesh, B.: Multi-ethnic study of atherosclerosis: relationship between left ventricular shape at cardiac mri and 10-year outcomes. Radiology **306**(2), e220122 (2022)
20. Menchón-Lara, R.M., Simmross-Wattenberg, F., Casaseca-de-la Higuera, P., Martín-Fernández, M., Alberola-López, C.: Reconstruction techniques for cardiac cine mri. Insights into imaging **10**(1), 100 (2019)
21. Mescheder, L., Oechsle, M., Niemeyer, M., Nowozin, S., Geiger, A.: Occupancy Networks: Learning 3D Reconstruction in Function Space (Apr 2019). https://doi.org/10.48550/arXiv.1812.03828, http://arxiv.org/abs/1812.03828, arXiv:1812.03828 [cs]
22. Park, J.J., Florence, P., Straub, J., Newcombe, R., Lovegrove, S.: DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation (Jan 2019). https://doi.org/10.48550/arXiv.1901.05103, http://arxiv.org/abs/1901.05103, arXiv:1901.05103 [cs]
23. Rahaman, N., Baratin, A., Arpit, D., Draxler, F., Lin, M., Hamprecht, F., Bengio, Y., Courville, A.: On the spectral bias of neural networks. In: International conference on machine learning. pp. 5301–5310. PMLR (2019)
24. Rahimi, A., Recht, B.: Random features for large-scale kernel machines. Advances in neural information processing systems **20** (2007)
25. Sander, J., De Vos, B.D., Bruns, S., Planken, N., Viergever, M.A., Leiner, T., Išgum, I.: Reconstruction and completion of high-resolution 3D cardiac shapes using anisotropic CMRI segmentations and continuous implicit neural representations. Computers in Biology and Medicine **164**, 107266 (Sep 2023). https://doi.org/10.1016/j.compbiomed.2023.107266, https://linkinghub.elsevier.com/retrieve/pii/S001048252300731X
26. Su, J., Ahmed, M., Lu, Y., Pan, S., Bo, W., Liu, Y.: Roformer: Enhanced transformer with rotary position embedding. Neurocomputing **568**, 127063 (2024)
27. Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., et al.: Uk biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS medicine **12**(3), e1001779 (2015)
28. Tancik, M., Srinivasan, P., Mildenhall, B., Fridovich-Keil, S., Raghavan, N., Singhal, U., Ramamoorthi, R., Barron, J., Ng, R.: Fourier features let networks learn high frequency functions in low dimensional domains. Advances in neural information processing systems **33**, 7537–7547 (2020)
29. Tilborghs, S., Bogaert, J., Maes, F.: Shape constrained cnn for segmentation guided prediction of myocardial shape and pose parameters in cardiac mri. Medical Image Analysis **81**, 102533 (2022)

30. Wessels, D.R., Knigge, D.M., Papa, S., Valperga, R., Vadgama, S., Gavves, E., Bekkers, E.J.: Grounding Continuous Representations in Geometry: Equivariant Neural Fields (Feb 2025). https://doi.org/10.48550/arXiv.2406.05753, http://arxiv.org/abs/2406.05753, arXiv:2406.05753 [cs]
31. Wiers, J., Wessels, D.R., Arts, L., Ruiperez-Campillo, S., Kolk, M., Tjong, F., Bekkers, E.J.: Geometry-aware cardiac mri representation learning with equivariant neural fields. In: Medical Imaging with Deep Learning (2026)
32. Xie, Y., Takikawa, T., Saito, S., Litany, O., Yan, S., Khan, N., Tombari, F., Tompkin, J., Sitzmann, V., Sridhar, S.: Neural fields in visual computing and beyond. In: Computer graphics forum. vol. 41, pp. 641–676. Wiley Online Library (2022)
33. Zhang, B., Tang, J., Niessner, M., Wonka, P.: 3DShape2VecSet: A 3D Shape Representation for Neural Fields and Generative Diffusion Models (May 2023). https://doi.org/10.48550/arXiv.2301.11445, http://arxiv.org/abs/2301.11445, arXiv:2301.11445 [cs]
34. Zhang, Y., Hager, P., Liu, C., Shit, S., Chen, C., Rueckert, D., Pan, J.: Towards cardiac mri foundation models: Comprehensive visual-tabular representations for whole-heart assessment and beyond. Medical Image Analysis p. 103756 (2025)
35. Zotti, C., Luo, Z., Lalande, A., Jodoin, P.M.: Convolutional neural network with shape prior applied to cardiac mri segmentation. IEEE journal of biomedical and health informatics **23**(3), 1119–1128 (2018)