Statistics and Computing (2026) 36:21
https://doi.org/10.1007/s11222-025-10772-x

ORIGINAL PAPER

Sparse Bayesian learning for label efficiency in cardiac real-time MRI

Anja Bach[4] · Achim Basermann[1] · Darius A. Gerlach[4] · Philipp Knechtges[1] · Jens Tank[4] · Raúl Tempone[2,3] · Felix Terhag[1,2]

Received: 15 April 2025 / Accepted: 4 November 2025
© The Author(s) 2025

Abstract

Cardiac real-time magnetic resonance imaging (MRI) is an emerging technology that images the heart at up to 50 frames per second, offering insight into the respiratory effects on the heartbeat. However, this method significantly increases the number of images that must be segmented to derive critical health indicators. Although neural networks perform well on inner slices, predictions on outer slices are often unreliable. This work proposes sparse Bayesian learning (SBL) to predict the ventricular volume on outer slices with minimal manual labeling to address this challenge. The ventricular volume over time is assumed to be dominated by sparse frequencies corresponding to the heart and respiratory rates. Moreover, SBL identifies these sparse frequencies on well-segmented inner slices by optimizing hyperparameters via type-II likelihood, automatically pruning irrelevant components. The identified sparse frequencies guide the selection of outer slice images for labeling, minimizing posterior variance. This work provides performance guarantees for the greedy algorithm. Testing on patient data demonstrates that only a few labeled images are necessary for accurate volume prediction. The labeling procedure effectively avoids selecting inefficient images. Furthermore, the Bayesian approach provides uncertainty estimates, highlighting unreliable predictions (e.g., when choosing suboptimal labels).

Keywords Sparse Bayesian learning · Label efficiency · Expectation maximization (EM) algorithm · Submodular set functions · Real-time magnetic resonance imaging (MRI)

1 Introduction

Cardiac magnetic resonance imaging (MRI) is the clinical reference standard for evaluating heart function, with ventricular volumes serving as key quantitative parameters. In clinical practice, these volumes are typically derived from segmentation of the ventricular cavity in each frame. Real-time MRI, an emerging technology, captures continuous data over multiple heartbeats at frame rates up to 50 frames per second (fps) Zhang et al. (2010); Salerno et al. (2017); Zhang et al. (2014). This enables observation of respiratory effects on cardiac function without requiring breath-holding, which is beneficial for vulnerable patient groups. However, the increased temporal resolution vastly expands the number of image frames per study, making manual annotation

✉ Felix Terhag
felix.terhag@dlr.de

Anja Bach
anja.bach@dlr.de

Achim Basermann
achim.basermann@dlr.de

Darius A. Gerlach
darius.gerlach@dlr.de

Philipp Knechtges
philipp.knechtges@dlr.de

Jens Tank
jens.tank@dlr.de

Raúl Tempone
tempone@uq.rwth-aachen.de

[1] Institute of Software Technology, High-Performance Computing, German Aerospace Center (DLR), Cologne, Germany

[2] Chair of Mathematics for Uncertainty Quantification, Department of Mathematics, RWTH Aachen University, Aachen, Germany

[3] Computer, Electrical and Mathematical Sciences and Engineering Division (CEMSE), King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia

[4] Institute of Aerospace Medicine, Cardiovascular Aerospace Medicine, German Aerospace Center (DLR), Cologne, Germany

Published online: 17 November 2025

especially laborious and time-consuming (Ramedani et al. 2025). A typical acquisition with 15 slices, 30 fps, and 10 seconds of recording produces 4,500 images to annotate. Manual segmentation of such volumes is infeasible in practice, highlighting the need for label-efficient strategies to derive reliable ventricular volume estimates.

Recent advances in automated segmentation, supported by publicly available cine MRI datasets (Bernard et al. 2018; Campello et al. 2021) and architectures such as U-Net (Isensee et al. 2021; Shelhamer et al. 2017), have improved analysis in conventional cine imaging. However, these models often fail to generalize to new acquisition protocols or rare pathologies (Bernard et al. 2018; Campello et al. 2021). For real-time MRI, the situation is even more challenging: to our knowledge, only a single public dataset exists, comprising data from one patient (Kunz et al. 2023). Segmentation performance is particularly unreliable in basal and apical slices (referred to as "outer slices" in the following), likely due to factors such as valve planes in the basal region and thin myocardium/small ventricular cavities at the apex. This is consistent with observations in cine MRI, where outer slices also exhibit reduced segmentation accuracy (Terhag et al. 2025). As a result, applying existing segmentation pipelines alone is insufficient to obtain reliable ventricular volume curves from real-time MRI.

In this study, we focus on real-time MRI of patients with univentricular hearts, a rare congenital condition in which only a single functional ventricle develops, using data collected in a clinical study (Müller et al. 2024). Multiple surgical interventions further alter cardiac geometry, resulting in highly variable and non-standard anatomy. These cases pose a particular challenge for automated segmentation, as existing datasets rarely include univentricular hearts. Neural networks, while powerful in standard cine MRI, rely on large and diverse training sets and thus struggle to generalize in this setting.

To address these challenges, we propose to model ventricular volume curves in real-time MRI using sparse Bayesian learning (SBL). The key observation is that ventricular volume over time is dominated by only a few frequencies, corresponding primarily to heart rate and respiration. SBL identifies these sparse components on well-segmented slices and transfers this information to slices where segmentation is unreliable. Crucially, the sparse prior allows us to determine which frames in a sequence are most informative for manual labeling: by selecting frames that minimize posterior uncertainty, we obtain reliable volume estimates with only a handful of expert annotations. This targeted labeling strategy contrasts with random or dense annotation and provides an explicit trade-off between annotation effort and predictive confidence. The resulting model reconstructs the full volume curve and quantifies its uncertainty, enabling clinicians to recognize when predictions are trustworthy.

SBL provides the methodological foundation of our approach (Tipping 2001). Originally developed in signal processing, it has been widely applied in domains such as remote sensing, where it identifies sparse frequency components from noisy, complex-valued measurements (Gemba et al. 2017; Gerstoft et al. 2016). In that literature, adaptive extensions often focus on optimizing sensor placement, for example in direction-of-arrival estimation where the question is which sensor positions provide the most informative measurements (Wang et al. 2018, 2023; Yang et al. 2013; Tang et al. 2024). In our setting, however, the slice positions are fixed and each slice represents a different part of the heart, while all time points of the sequence are available simultaneously. The central challenge is selecting which time points (frames) to label so as to best constrain the sparse frequency model.

A related line of work is active learning for medical image segmentation, where images are selected for annotation based on model uncertainty (Budd et al. 2021). Such methods, however, require large datasets and aim to improve segmentation networks with millions of parameters, whereas our problem involves estimating only a small number of amplitudes for sparse frequencies (roughly 10–15 parameters). Moreover, our goal is not to refine segmentation but to directly derive reliable ventricular volume curves, making conventional active learning strategies less applicable in this context.

To the best of our knowledge, this is the first work to apply SBL to cardiac real-time MRI and to exploit it for frame selection in a label-efficiency context. Specifically, we adapt SBL to real-valued MRI data and demonstrate its ability to recover sparse frequency components that characterize ventricular volume dynamics. Building on this foundation, we introduce a principled strategy for selecting frames to annotate by minimizing posterior uncertainty, with theoretical guarantees on the performance of the greedy selection algorithm. We further compare different selection criteria and show that our approach substantially outperforms random frame selection. Finally, we validate the method on real-time MRI data from patients with univentricular hearts, demonstrating that accurate ventricular volume curves can be reconstructed from only a few manual labels. The remainder of this paper is structured as follows: Section 2 details the adapted SBL algorithm and the optimization framework, Section 3 describes the clinical dataset and workflow, Section 4 presents the experimental evaluation, and Section 5 concludes with a discussion and outlook.

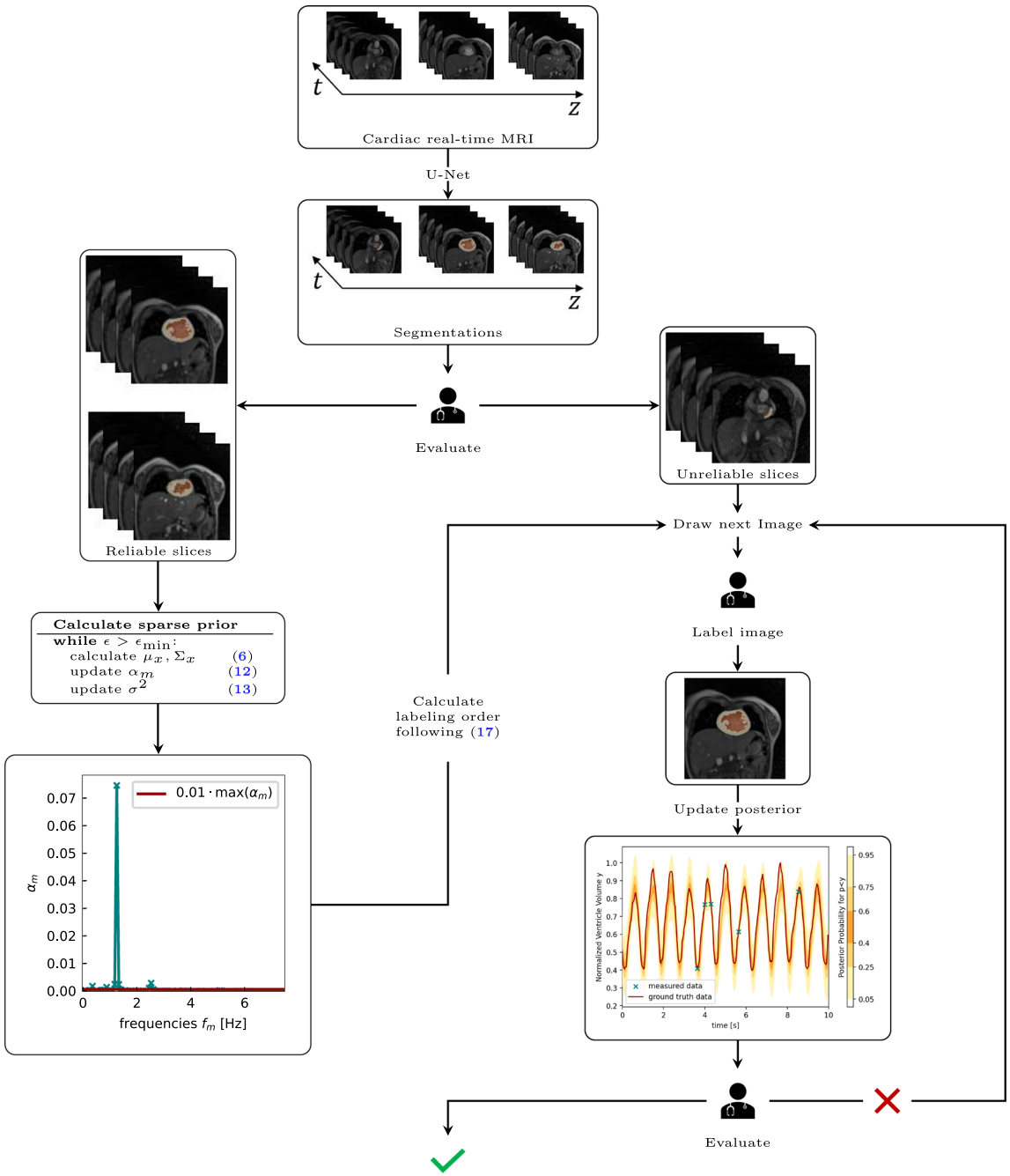


Fig. 1 Workflow of the proposed method. A neural network segments the MRI sequences, and a clinician performs a visual quality check to separate reliable from unreliable slices. The reliable slices provide sparse frequency priors, while only a few frames of the unreliable slices are manually labeled, selected via an SBL-based criterion. The posterior then reconstructs the full volume curve with uncertainty estimates, reported as expected deviation in ml. Example MRIs are short-axis views of univentricular hearts from Müller et al. (2024). For details, see Sec. 2.1.

2 Methods

2.1 Workflow overview

Figure 1 summarizes the proposed workflow. Starting from real-time MRI sequences, a neural network provides initial segmentations of all slices. A clinician then performs a visual quality check to distinguish slices with reliable segmentations from those with unreliable ones. Importantly, this decision is always made on the level of entire slices across all frames, not frame by frame. The reliable slices are used to identify the dominant sparse frequencies that characterize

ventricular volume dynamics. For slices with unreliable segmentations, the model leverages these frequencies to predict the volume curve. Only a small subset of frames is then manually annotated, with the selection guided by our SBL-based criterion for maximizing information gain. In practice, only the inner slices could be segmented reliably, with the exact number varying by patient; see Sec. 4 for details.

After each newly labeled frame, the posterior distribution is updated, which not only refines the predicted volume curve but also provides uncertainty estimates. While the optimization is based on the trace of the posterior covariance matrix, this quantity can be translated into clinically meaningful terms: for example, an expected mean or maximum deviation in ventricular volume in milliliters. This allows practitioners to monitor whether the current level of accuracy meets their requirements and to stop labeling once the uncertainty falls below a user-defined threshold. In this way, the workflow ensures both label efficiency and clinical interpretability.

2.2 Modeling

Cardiac real-time MRI retrieves video sequences of the beating heart due to a faster capture rate than cine MRI. Thus, for every slice $\ell \in [1, L]$, there are images for all distinct equidistant time steps $t_1, \ldots, t_N$. Only the ventricle volume and derived parameters, such as stroke volume, are of interest. The volume over time is assumed to be well described by a superposition of frequencies, comprising a subset of $\mathcal{F} = \{f_0 = 0, f_1, f_2, \ldots, f_M\}$, where the main frequencies are the unknown heart and respiratory rates. This work proposes the following linear model:

$$Y = AX + N, \tag{1}$$

with

- $Y \in \mathbb{R}^{N \times L}$, the real-valued measurement points $y_{k,\ell}$;
- $X \in \mathbb{R}^{2M+1 \times L}$, the real and imaginary parts of the complex amplitudes;
- $N \in \mathbb{R}^{N \times L}$, additive noise with $n_{k,\ell} \sim \mathcal{N}(0, \sigma^2)$;
- $A \in \mathbb{R}^{N \times 2M+1}$. the transfer matrix $A = [\mathbf{1}_N \; Re(B) \; Im(B)]$,
- with $\mathbf{1}_N$ an N-dimensional vector of 1s and $B \in \mathbb{C}^{N \times M}$ defined as follows,

$$B = \begin{bmatrix} \exp(-i2\pi f_1 t_1) & \ldots & \exp(-i2\pi f_M t_1) \\ \vdots & \ddots & \vdots \\ \exp(-i2\pi f_1 t_N) & \ldots & \exp(-i2\pi f_M t_N) \end{bmatrix}. \tag{2}$$

2.3 Finding sparse frequencies

The above model can describe measurements via superpositioned frequencies of $\mathcal{F}$. Only a few frequencies are assumed to influence the volume substantially. Furthermore, the influential frequencies (e.g., heart and respiratory rates) are assumed to be shared over all slices. Identifying these sparse frequencies is beneficial because the model predicts slices where only a few measurement points exist in time . Fewer amplitudes and phases must be found if fewer frequencies exist.

This work applies an approach similar to Gerstoft et al. (2016) to determine the sparse frequencies. The procedure is presented in Algorithm 1. In Gerstoft et al. (2016), the authors used SBL to localize the sparse sources from noisy signals in the signal processing application of estimating the directions of arrival of plane waves from sensor array data. They obtain observations from L different snapshots, whereas our method applies the volume over time for L different heart slices. One additional difference of the proposed approach is that the measurements in Gerstoft et al. (2016) are complex compared to real-valued measurements.

A Bayesian model is designed to use the SBL approach. With the assumed additive Gaussian noise of variance σ^2, the likelihood is

$$p(Y \mid X, \sigma^2) = \frac{\exp\left(-\frac{1}{2\sigma^2} \|AX - Y\|_F^2\right)}{(2\pi\sigma^2)^{NL/2}}. \tag{3}$$

Moreover, SBL achieves sparsity by employing a Gaussian zero-mean prior on the parameters X, where the variance of each parameter is considered a hyperparameter (Murphy 2023; Tipping 2001). This work employs the expectation maximization (EM) scheme to estimate the hyperparameters, where the expectation step calculates the expectation of the log-likelihood with respect to the current parameters (E-step). The following maximization step updates the parameters to maximize the expectation of the log-likelihood (M-step). As the EM iterations proceed, the algorithm identifies parameters relevant to explaining the observed data. The corresponding prior variances tend to approach zero for parameters not supported by data (i.e., irrelevant to the model). When a variance becomes small, the associated prior distribution effectively becomes a delta function centered at zero, "switching off" those parameters from the model (for a comprehensive explanation of this method, see Murphy 2023, Sec 15.2.8).

This work aims to determine the sparse frequencies that represent each slice simultaneously without limiting the phase for each frequency. The amplitude of frequency $f_{m,\ell}$, for $m \geq 1$ and slice ℓ is determined by $\sqrt{x_{m,\ell}^2 + x_{m+M,\ell}^2}$,

whereas the phase is determined by $\angle(x_{m,\ell}, x_{m+M,\ell})$.[1] The prior is shared over all slices because the frequencies are assumed to behave similarly over all slices. However, this method should not influence the phase because it is independent of the separate slices. The latter is a modeling choice because the time between capturing the slices is unknown, making it impossible to know the phase of one slice in relation to the next. Thus, the prior should be uniform with respect to the phases. This work introduces a slice-independent variance $\alpha_m > 0$ to achieve this uniformity . The prior $p_0(x_{0,\ell}, \alpha_0) = \mathcal{N}(0, \alpha_0)$ was selected for frequency $f_0 = 0$ and

$$p_m(x_{m,\ell} \mid \alpha_m) = p_{m+M}(x_{m+M,\ell} \mid \alpha_m) = \mathcal{N}(0, \alpha_m) \tag{4}$$

for $m > 0$. By setting $\gamma = [\alpha_0, \alpha_1, \ldots, \alpha_M, \alpha_1, \ldots, \alpha_M]^T$ and $\Gamma = \mathrm{diag}(\gamma)$, the prior of X simplifies to L independent multivariate normal distribution for each slice L

$$p(X \mid \alpha) = \prod_{\ell=1}^{L} \prod_{m=0}^{2M+1} p_m(x_{m,\ell} \mid \gamma_m) = \prod_{\ell=1}^{L} \mathcal{N}(0, \Gamma). \tag{5}$$

Combining the Gaussian prior and likelihood yields a Gaussian posterior $p(X|Y, \alpha, \sigma^2)$ with the mean and covariance given by

$$\begin{aligned} \mu_x &= \Gamma A^T \Sigma_y^{-1} Y \\ \Sigma_x &= \left(\frac{1}{\sigma^2} A^T A + \Gamma^{-1} \right)^{-1}, \end{aligned} \tag{6}$$

where $\Sigma_y = \sigma^2 I_N + A \Gamma A^T$ denotes the data covariance. This formulation reveals that the sparsity of μ_x controls the row sparsity. If entry $\alpha_m = 0$ for $m > 0$, it follows that $\gamma_m = \gamma_{m+M} = 0$, ensuring the posterior satisfying $p(x_m = x_{m+M} = 0|Y, \alpha_m = 0) = 1$. This approach makes both rows of X corresponding to frequency f_m zero.

2.3.1 Sparse Bayesian learning and hyperparameter estimation

The SBL approach relies on estimating the hyperparameters α_m and σ^2 with empirical Bayes, setting the hyperparameters that maximize the marginal likelihood. The marginal likelihood $p(Y|\alpha, \sigma^2)$ is calculated by treating the amplitudes X as a nuisance parameter and marginalizing over them Wipf and Rao (2007); Gerstoft et al. (2016), obtaining the following:

$$\begin{aligned} p(Y|\alpha, \sigma^2) &= \int p(y|X, \sigma^2) p(X|\alpha) dX \\ &= \frac{\exp\left(-\mathrm{tr}(Y^T \Sigma_y^{-1} Y)\right)}{\left(\pi^N \det \Sigma_y\right)^L}. \end{aligned} \tag{7}$$

The aim is to maximize the log marginal likelihood, obtaining the following cost function:

$$\begin{aligned} &\mathcal{L}(\alpha, \sigma^2) \\ &\quad = -\mathrm{tr}(Y^T \Sigma_y^{-1} Y) - L \log\det \Sigma_y \propto \log p(Y|\alpha, \sigma^2). \end{aligned} \tag{8}$$

Next, this cost function is derived with respect to α_m. Note that α_0 only occurs once in the diagonal matrix Γ, whereas α_m for $m > 0$ occurs at $\Gamma_{m,m}$ and $\Gamma_{m+M,m+M}$. Here, the focus is on $m > 0$, as the derivative with respect to α_0 can be calculated analogously. Hence, for $m > 0$,

$$\begin{aligned} \frac{\partial \Sigma_y^{-1}}{\partial \alpha_m} &= -\Sigma_y^{-1} \frac{\partial \Sigma_y}{\partial \alpha_m} \Sigma_y^{-1} \\ &= -\Sigma_y^{-1} a_m a_m^T \Sigma_y^{-1} - \Sigma_y^{-1} a_{m+M} a_{m+M}^T \Sigma_y^{-1} \end{aligned} \tag{9}$$

$$\begin{aligned} \frac{\partial \log\det\left(\Sigma_y\right)}{\partial \alpha_m} &= \mathrm{tr}\left(\Sigma_y^{-1} \frac{\partial \Sigma_y}{\partial \alpha_m} \right) \\ &= a_m^T \Sigma_y^{-1} a_m + a_{m+M}^T \Sigma_y^{-1} a_{m+M} \end{aligned} \tag{10}$$

yields the following derivative:

$$\begin{aligned} \frac{d\mathcal{L}(\alpha, \sigma^2)}{d\alpha_m} &= \mathrm{tr}\left(Y^T \Sigma_y^{-1} a_m a_m^T \Sigma_y^{-1} Y + Y^T \Sigma_y^{-1} a_{m+M} a_{m+M}^T \Sigma_y^{-1} Y \right) \\ &\quad - L \left(a_m^T \Sigma_y^{-1} a_m + a_{m+M}^T \Sigma_y^{-1} a_{m+M} \right) \\ &= ||Y^T \Sigma_y^{-1} (a_m + a_{m+M})||_2^2 \\ &\quad - L \left(a_m^T \Sigma_y^{-1} a_m + a_{m+M}^T \Sigma_y^{-1} a_{m+M} \right), \end{aligned} \tag{11}$$

where a_m denotes the mth column vector of A (cf. (Gerstoft et al. 2016)). The M-step uses the MacKay update rule introduced in MacKay (1992), obtained by setting the derivative (11) to zero and using a fixed-point equation. The update rule

$$\alpha_m^{\mathrm{new}} = \frac{||(\mu_x)_m + (\mu_x)_{m+M}||_2^2}{L \left(1 - \frac{1}{\alpha_m^{\mathrm{old}}} \left((\Sigma_x)_{m,m} + (\Sigma_x)_{m+M,m+M} \right) \right)} \tag{12}$$

is analogous to the results in Tipping (2001); Wipf and Rao (2007).

For the estimation of the hyperparameter σ^2, the update rule introduced in (Gerstoft et al. 2016, (27)) is employed, as the definition of the hyperparameter α_m does not influence this update. Thus, the update rule becomes

$$\left(\sigma^2\right)^{\mathrm{new}} = \frac{1}{N - K} \mathrm{tr}\left((I_N - A_{\mathcal{M}} A_{\mathcal{M}}^{+} S_y) \right), \tag{13}$$

[1] The notation $\angle(a, b)$ denotes the phase of the complex number $a + ib$.

where $S_y = YY^T/L$ represents the data sample covariance matrix. The matrix $A_{\mathcal{M}} = (a_0, a_{m_1}, \ldots, a_{m_K}, a_{m_1+M}, \ldots, a_{m_K+M})$ consists of column vectors of A with the indices $m_1, \ldots, m_K$ corresponding to the K largest values in α. The Moore–Penrose inverse of $A_{\mathcal{M}}$ is denoted by $A^+_{\mathcal{M}}$. The parameter $K << M$ can be selected with model-order selection criteria, as described in Gerstoft et al. (2016). The choice of K did not significantly influence the results.

Algorithm 1 Calculate the sparse prior

Initialize: here $\sigma_0^2 = 0.2, \alpha = \mathbf{1}_{M+1}, \epsilon_{\min} = 1e-4$
while $\epsilon > \epsilon_{\min}$ **do**
  $\gamma = [\alpha_0, \alpha_1, \ldots, \alpha_M, \alpha_1, \ldots, \alpha_M]^T, \quad \Gamma = \mathrm{diag}(\gamma)$
  Calculate μ_x, Σ_x with (6)
  Update α_m with (12)
  Update σ^2 with (13)
  $\epsilon = ||\alpha^{\text{new}} - \alpha^{\text{old}}||_1 / ||\alpha^{\text{old}}||_1$
end while

2.4 Minimizing labeling work

For each heart, there are slices with reliable volumes over time, whereas others have no information about the ventricle volumes. In these slices, practitioners rely on hand-labeled images. Labeling images by hand is tedious and expensive. The sparse priors obtained using the SBL approach introduced above minimize the labeling effort. Because the priors are sparse, only a few complex amplitudes $x_m + i x_{m+M}$ must be found.

Let J be the index set $J \subset \{1, \ldots, N\} = \Omega$, and $P_J \in \mathbb{R}^{N \times N}$ be the projection matrix, projecting all rows $i \in \Omega \setminus J$ to 0. When limited labeled data J exist, the model (1) reduces to

$$P_J Y = P_J A X + N,$$

The posterior of X is a Gaussian with mean μ_x and covariance Σ_x, similar to (6),

$$\begin{aligned} \mu_x^{(J)} &= \Gamma A^T P_J^T \left(\Sigma_y^{(J)}\right)^{-1} Y_J \\ \Sigma_x^{(J)} &= \left(\frac{1}{\sigma^2} A^T P_J A + \Gamma^{-1}\right)^{-1}, \end{aligned} \tag{14}$$

where $\Sigma_y^{(J)} = \sigma^2 I_{|J|} + P_J A \Gamma (P_J A)^T$ is the data covariance–this time with the projected matrices and $|J|$ labeled images. The prediction over the full time is obtained by multiplying the full matrix A with the posterior and adding Gaussian noise N. A posterior prediction is obtained, normally distributed with mean $\mu^{(J)}_{post_y}$ and covariance $\Sigma^{(J)}_{post_y}$, given by

$$\begin{aligned} \mu^{(J)}_{post_y} &= A\mu_x^{(J)} \\ \Sigma^{(J)}_{post_y} &= \sigma^2 I_N + A \Sigma_x^{(J)} A^T. \end{aligned} \tag{15}$$

The variance is a satisfactory measure of the distribution spread for a one-dimensional normal distribution . In this multivariate case, a few possible measures exist (Paindaveine 2008). The predominant considerations in the literature are the trace of the covariance matrix $\mathrm{tr}(\Sigma^{(J)}_{post_y})$ Dümbgen (1998); Visuri et al. (2003), the determinant $\det(\Sigma^{(J)}_{post_y})$ Dümbgen and Tyler (2005); Salibián-Barrera et al. (2006), or the largest eigenvalue $\lambda_1(\Sigma^{(J)}_{post_y})$ Randles (2000). The trace measures the total variance but does not account for the correlation. Nonetheless, the trace is computationally inexpensive. The determinant of the covariance matrix, as proposed in Paindaveine (2008), offers an alternative but may encounter numerical problems, particularly as $\det(\Sigma)$ approaches 0 in high-dimensional settings with small eigenvalues. A third option is to employ the largest eigenvalue of Σ, corresponding to the variance along the first principal component. This approach is also computationally more demanding than applying the trace of the covariance matrix.

This method results in the following optimization problem: Let $S : \mathcal{S}_N \to \mathbb{R}_0^+$ be a measure for the spread of a distribution, where $\mathcal{S}_N$ denotes the set of real-valued symmetric positive definite $N \times N$ matrices. For k labeled images, we want to determine

$$J^* := \arg\min_{|J|=k} S\left(\Sigma^{(J)}_{post_y}\right) \tag{16}$$

the indices J^* of the images, which need to be labeled to minimize the spread.

It might be desirable to let a practitioner iteratively label the next optimal image until practitioners are satisfied with the results. This approach leads to a slightly different optimization objective, lending itself to applying greedy algorithms. For a given index set $J \subset \{1, \ldots, N\}$, find

$$j^* := \arg\min_{j \in \{1,\ldots,N\}} S\left(\Sigma^{(J \cup j)}_{post_y}\right). \tag{17}$$

The selected images J^* and j^* do not depend on the measurements at those slices. Thus, the indices can be computed before the labeling process. However, they strongly depend on the computed sparse frequencies. The procedure is presented in Algorithm 2.

2.5 Optimality of the greedy algorithm

Finding the optimal solution of (16) is difficult due to the combinatorial explosion as k increases. There are already

more than 19 billion combinations to draw five images from $N = 300$ time steps. Thus, iteratively solving (17) using a greedy approach is a feasible alternative. The objective function is first reformulated to establish performance guarantees. Let $J \subseteq \Omega = \{1, \ldots, N\}$ and $\Sigma_y := \sigma^2 I_N + A\Gamma A^T$ the data covariance, we define the objective function as follows:

$$\begin{aligned} f(J) &= \operatorname{tr}\left(\Sigma_y\right) - \operatorname{tr}\left(\Sigma_{post_y}^{(J)}\right) \\ &= \operatorname{tr}(\Sigma_y) - \operatorname{tr}\left(\sigma^2 I_N + A\left(\frac{1}{\sigma^2} A^T P_J A + \Gamma^{-1}\right)^{-1} A^T\right). \end{aligned} \tag{18}$$

Maximizing this function is equivalent to minimizing $\operatorname{tr}\left(\Sigma_{post_y}^{(J)}\right)$. For brevity, this work is limited to investigating the trace because this measure performed best in the experiments. The following definitions are introduced:

Definition 1 Let Ω be a finite set and 2^Ω be the power set of Ω. A set function $f : 2^\Omega \to \mathbb{R}$ is monotone non decreasing if $f(\mathcal{X}) \leq f(\mathcal{Y})$ for all $\mathcal{X} \subseteq \mathcal{Y} \subseteq \Omega$.

Definition 2 The marginal gain of adding element j to set $\mathcal{X}$ is defined as follows:

$$f_j(\mathcal{X}) := f(\mathcal{X} \cup j) - f(\mathcal{X}). \tag{19}$$

Definition 3 A set function $f : 2^\Omega \to \mathbb{R}$ is submodular if

$$f_j(\mathcal{X}) \geq f_j(\mathcal{Y}) \tag{20}$$

for every subset $\mathcal{X} \subseteq \mathcal{Y} \subset \Omega$ and every $j \in \Omega \setminus \mathcal{Y}$.

The authors of Nemhauser et al. (1978) demonstrated that the greedy algorithm generates a solution at least $(1-1/e)$ of the optimal solution for submodular set functions. The objective function (18) is not submodular, as Figure 4 presents counterexamples. However, a weaker bound exists for (18). This work follows the definition from Hashemi et al. (2019).

Definition 4 The weak-submodularity constant c_f of a monotone non decreasing function f is defined as follows:

$$c_f := \max_{(\mathcal{X},\mathcal{Y},i)\in\tilde{\Omega}} f_i(\mathcal{Y})/f_i(\mathcal{X}), \tag{21}$$

where $\tilde{\Omega} = \{(\mathcal{X}, \mathcal{Y}, i) | \mathcal{X} \subseteq \mathcal{Y} \subset \Omega, i \in \Omega \setminus \mathcal{Y}\}$.

Note, that f is submodular if and only if $c_f \leq 1$ and is weakly submodular if and only if c_f is bounded.
For $J \subset \Omega$ and $i \in \Omega$ the marginal gain of the objective function (18) simplifies to

$$\begin{aligned} f_i(J) &= f(J \cup \{i\}) - f(J) \\ &= \operatorname{tr}\left(\Sigma_{post_y}^{(J)}\right) - \operatorname{tr}\left(\Sigma_{post_y}^{(J\cup\{i\})}\right). \end{aligned} \tag{22}$$

Moreover, we can prove the following lemma.

Lemma 5 *For f_i defined in (18), it holds $f_i(J) > 0$ for all $J \subset \Omega$ and $i \in \Omega \setminus J$.*

Proof If $J \subset \Omega$ and $i \in \Omega \setminus J$, it holds that

$$\begin{aligned} f_i(J) &= \operatorname{tr}\left(\Sigma_{post_y}^{(J)}\right) - \operatorname{tr}\left(\Sigma_{post_y}^{(J\cup\{i\})}\right) \\ &= \operatorname{tr}\left(A\left(\frac{1}{\sigma^2} A^T P_J A + \Gamma^{-1}\right)^{-1} A^T\right) \\ &\quad - \operatorname{tr}\left(A\left(\frac{1}{\sigma^2} A^T P_{J\cup\{i\}} A + \Gamma^{-1}\right)^{-1} A^T\right) \\ &= \operatorname{tr}\left(A\left(\left(\frac{1}{\sigma^2} A^T P_J A + \Gamma^{-1}\right)^{-1}\right.\right. \\ &\quad \left.\left. - \left(\frac{1}{\sigma^2} A^T P_{J\cup\{i\}} A + \Gamma^{-1}\right)^{-1}\right) A^T\right). \end{aligned} \tag{23}$$

This work defines $D := 1/\sigma^2 A^T P_J A + \Gamma^{-1}$, which is symmetric positive definite because Γ is a diagonal matrix with positive diagonal entries. For D and e_i, the standard basis vector in the ith direction, note that

$$\begin{aligned} \frac{1}{\sigma^2} A^T P_{J\cup\{i\}} A + \Gamma^{-1} &= D + \frac{1}{\sigma^2} A^T P_{\{i\}A} \\ &= D + A^T e_i e_i^T A \end{aligned} \tag{24}$$

With this formulation (24) and the Sherman–Morrison formula (Bartlett 1951), (23) can be written as follows:

$$\begin{aligned} f_i(J) &= \operatorname{tr}\left(A\left(D^{-1} - \left(D + A^T e_i e_i^T A\right)^{-1}\right) A^T\right) \\ &\underset{S.-M.}{=} \operatorname{tr}\left(A\left((1 + \frac{1}{\sigma^2} e_i^T A D^{-1} A^T e_i)^{-1} \frac{1}{\sigma^2} D^{-1} A^T e_i e_i^T A D^{-1}\right) A^T\right) \\ &= \left(1 + \frac{1}{\sigma^2} e_i^T A D^{-1} A^T e_i\right)^{-1} \\ &\quad \left(\frac{1}{\sigma^2} e_i^T A D^{-1} A^T A D^{-1} A^T e_i\right), \end{aligned} \tag{25}$$

where $\langle \cdot, \cdot \rangle_D$ can be written for the weighted inner product because D is a symmetric positive definite matrix. The marginal gain in (25) simplifies to

$$\begin{aligned} f_i(J) &= \left(1 + \frac{1}{\sigma^2} e_i^T A D^{-1} A^T e_i\right)^{-1} \left(\frac{1}{\sigma^2} e_i^T A D^{-1} A^T A D^{-1} A^T e_i\right) \\ &= \frac{1}{\sigma^2} \langle A D^{-1} A^T e_i, A D^{-1} A^T e_i \rangle / \left(1 + \frac{1}{\sigma^2} \langle A^T e_i, A^T e_i \rangle_{D^{-1}}\right) \\ &> 0. \end{aligned} \tag{26}$$

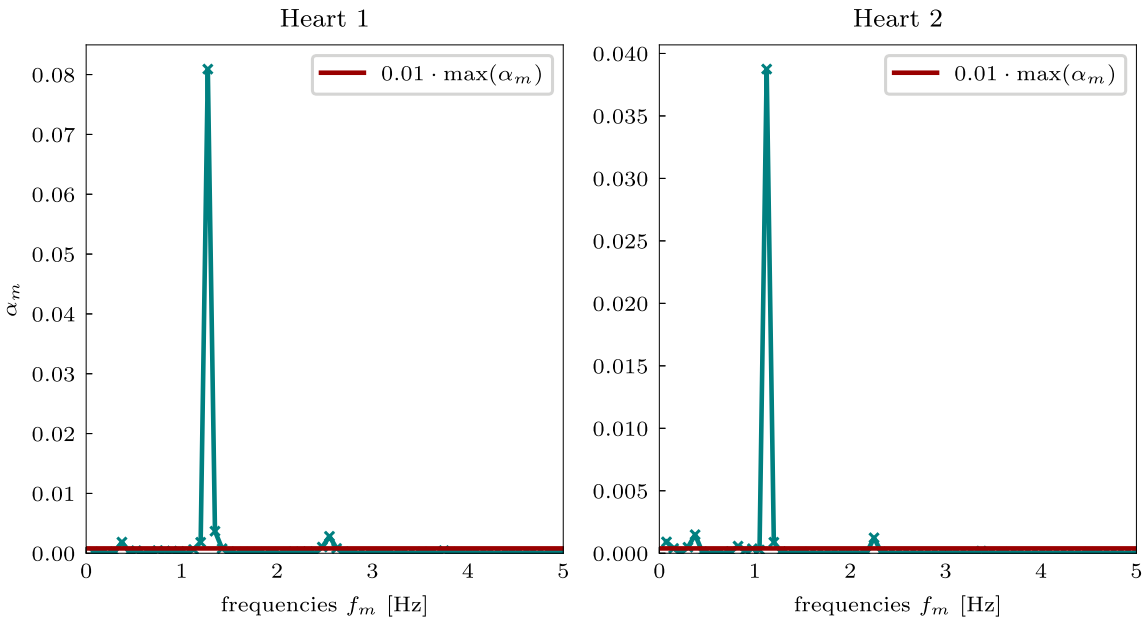


Fig. 2 Values for α_m, where $m > 0$, after the expectation maximization procedure of the SBL algorithm. The largest peak is at the heart rate, with smaller peaks at double the heart rate. The respiratory rate is about 0.25 Hz.

This result holds because the numerator is zero if and only if a zero row i exists in A, but by construction, the first entry of each row is 1. □

Theorem 6 *The objective function $f(J)$, as defined in (18), is a weakly submodular set function.*

Proof From Lemma 5, it follows directly that $f(J)$ is a monotone non-decreasing function. It is left to be shown, that the weak-submodularity constant c_f is bounded.

As Ω is a finite set, there exists

$$c_{max} := \max_{\mathcal{X} \subset \Omega, i \in \Omega \setminus \mathcal{X}} f_i(\mathcal{X}) \tag{27}$$

and

$$c_{min} := \min_{\mathcal{X} \subset \Omega, i \in \Omega \setminus \mathcal{X}} f_i(\mathcal{X}). \tag{28}$$

We know from Lemma 5, that $0 < c_{min} \leq c_{max}$; thus, $c_b = c_{max}/c_{min}$ exists, and the weak-submodularity constant c_f is bounded with

$$c_f = \max_{(\mathcal{X}, \mathcal{Y}, i) \in \tilde{\Omega}} f_i(\mathcal{Y}) / f_i(\mathcal{X}) \leq c_b, \tag{29}$$

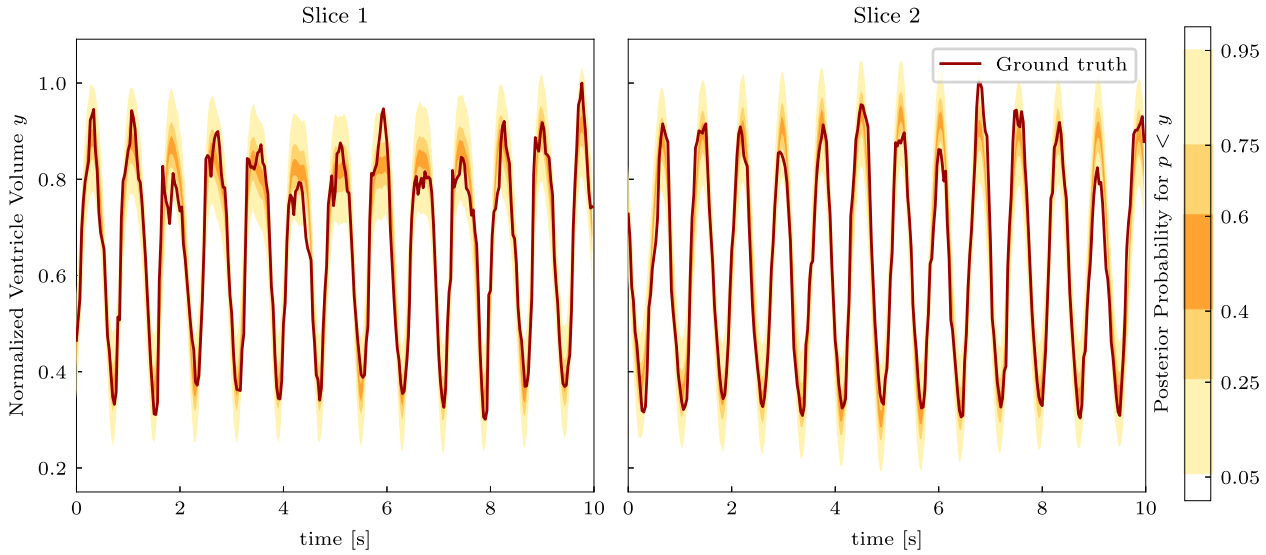
for $\tilde{\Omega} = \{(\mathcal{X}, \mathcal{Y}, i) | \mathcal{X} \subseteq \mathcal{Y} \subset \Omega, i \in \Omega \setminus \mathcal{Y}\}$. □

Based on Das and Kempe (2011); Hashemi et al. (2019), the solution obtained by the greedy algorithm is guaranteed to be within a factor of $(1 - e^{-1/c_f})$ of the optimal solution $f(J^*)$. Section 4 empirically examines the weak-submodularity constant.

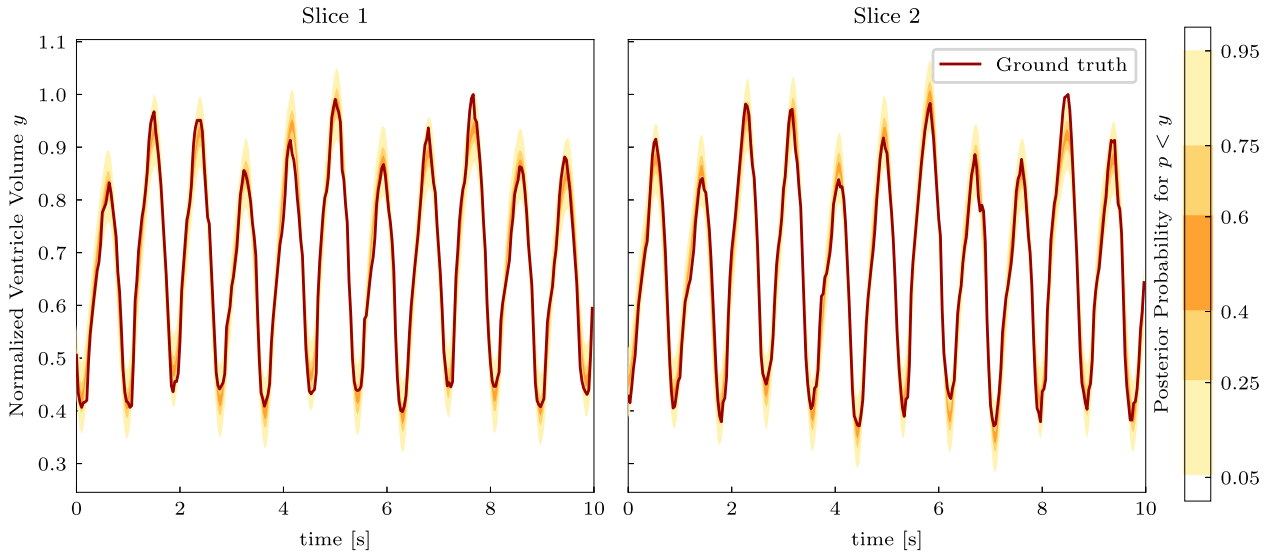
3 Application

This application involves real-time MRI scans of univentricular hearts, with videos of $N = 300$ images in each slice through the heart. Although the automatic segmentation in the intermediate slices is very reliable, it contains no information on the outer slices. This work is interested in the ventricle volume for each slice over time. Combining the results of Sections 2.3 and 2.4 can minimize the labeling effort in the outer slices. This work builds a linear model (1) to describe the volume over the slices and time. Section 2.3 employs the SBL algorithm to determine the prior empirically. The SBL algorithm ensures that the prior is sparse. Thus, the prior for many frequencies is close to a delta distribution around 0. The distributions are normal with variances of $\alpha_m \approx 0$. Moreover, α_m can be set to 0 for small α_m, and the corresponding columns can be removed from A and rows removed from X to circumvent numerical problems and reduce the computational expense of the problem. A practical method to set a threshold is $\epsilon_{thresh} = 0.01 \cdot \max_{m \in \{1,\ldots,M\}}(\alpha_m)$, as frequencies with amplitudes orders of magnitudes smaller than the most significant frequency have a negligible effect on the final volume.

A prior that assigns weight only to a few frequencies is beneficial when predicting slices without information about the volume because only the phase and amplitude of the most influential frequencies must be fit. Thus, only a few labeled images suffice to predict a whole slice. The sequence of images can be optimized for labeling by solving (16) for a fixed number of images. This problem is difficult due to



(a) Heart 1



(b) Heart 2

Fig. 3 Posterior distribution and ground-truth data for two slices of two hearts. (a) Heart 1 is much less regular. The posterior distribution reflects this, with a broader posterior.

the combinatorial explosion for larger numbers. Practically, (17) can be iteratively solved, and the best slice can be consecutively labeled until the practitioner is satisfied with the results. Section 2.5 proves that the solution of this approach is at most a factor of $(1 - e^{-1/c_f})$ away from the optimal solution to (16).

Algorithm 2 Efficient labeling and prediction

Require: $\Gamma = \text{diag}([\alpha_0, \alpha_1, \ldots, \alpha_M, \alpha_1, \ldots, \alpha_M])$ and σ^2 from Algorithm 1
Initialize: $J = \{\}$
1: **while** insufficient prediction **do**
2: $j^* := \arg\min_{j \in \{1,\ldots,N\}} S\left(\Sigma_{post_y}^{(J \cup j)}\right)$ with S from (30)
3: Manually label j^*
4: $J = J \cup j^*$
5: Calculate posterior $\mu_{post_y}^{(J)}, \Sigma_{post_y}^{(J)}$ with (15)
6: **end while**

4 Results

The MRI scans are from patients with univentricular hearts. This condition makes automatic segmentation incredibly challenging because their hearts usually differ significantly from the geometry of the hearts of other patients. Therefore, it is not possible to use publicly available datasets. The data contain MRI scans of two patients. The inner five slices for Heart 1 and the inner four slices for Heart 2 can be segmented reliably using a U-Net approach similar to Isensee et al. (2021). A medical expert confirmed the segmentation for those slices. The outer slices could not be segmented reliably, with mostly zero predictions.

4.1 Finding sparse frequencies

The frequencies $f_m = 3m/40$ Hz for $m = 0, \dots 100$ were employed to cover a broad range of possible frequencies. Although one of the ventricle volumes (Heart 1) is visibly more irregular than the other, the SBL algorithm reliably found sparse frequencies for both hearts. Figure 2 illustrates the values for α_m, $m > 0$, after the EM procedure of the SBL algorithm. For both hearts, the algorithm found seven frequencies higher than the threshold $0.01 \cdot \max_{m>0}(\alpha_m)$. By far, the most significant influence is the heart frequency. Minor influences include the respiratory rate at about 0.3 Hz and double the heart rate (i.e., , higher harmonics). Figure 3 depicts the posterior of the slices. The broader posterior for Heart 1 captures the more irregular heartbeat.

4.2 Minimizing labeling work

The labeling effort on the other slices can be minimized after empirically determining the prior on the well segmented intermediate slices. We used a greedy algorithm to search for the best next point to label. As objective functions, this work used the following:

$$\begin{aligned} S\left(\Sigma^{(J)}_{post_y}\right) &= \operatorname{tr}\left(\Sigma^{(J)}_{post_y}\right) \\ S\left(\Sigma^{(J)}_{post_y}\right) &= \sqrt[N]{\det\left(\Sigma^{(J)}_{post_y}\right)} \\ S\left(\Sigma^{(J)}_{post_y}\right) &= \lambda_1\left(\Sigma^{(J)}_{post_y}\right). \end{aligned} \tag{30}$$

The eigenvalues of the matrix became too small for a numerical calculation of $\det\left(\Sigma^{(J)}_{post_y}\right)$, which yields a value of 0. Therefore, the property that the determinant is the product of the eigenvalues was applied to compute the following:

$$S\left(\Sigma^{(J)}_{post_y}\right) = \sqrt[N]{\det\left(\Sigma^{(J)}_{post_y}\right)} = \prod_{i=1}^{N} \sqrt[N]{\lambda_i\left(\Sigma^{(J)}_{post_y}\right)}. \tag{31}$$

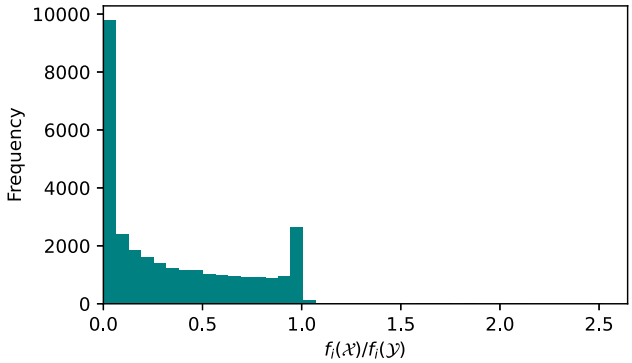


Fig. 4 Histogram of 30,000 random draws of $c_{(\mathcal{X},\mathcal{Y},i)}$, as defined in (33), to estimate the weak-submodularity constant. The highest obtained value is ≈ 2.52, and only 1.2% of the samples are greater than 1.

Section 2.5 demonstrates that, for the trace $\operatorname{tr}\left(\Sigma^{(J)}_{post_y}\right)$, the solution obtained by the greedy algorithm is guaranteed to be within a factor of $(1 - e^{-1/c_f})$ of the optimal solution, where c_f denotes the weak-submodularity constant

$$c_f := \max_{(\mathcal{X},\mathcal{Y},i)\in\tilde{\Omega}} f_i(\mathcal{Y})/f_i(\mathcal{X}), \tag{32}$$

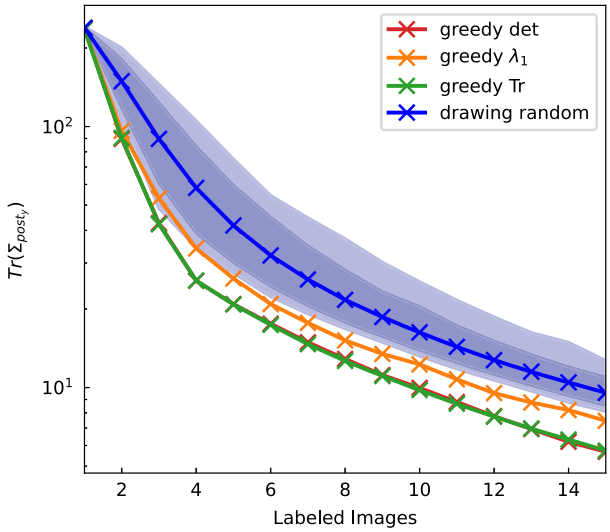
with $\tilde{\Omega} = \{(\mathcal{X}, \mathcal{Y}, i) | \mathcal{X} \subseteq \mathcal{Y} \subset \Omega, i \in \Omega \setminus \mathcal{Y}\}$ for $\Omega = \{1, \dots, 300\}$. Next, 30,000 samples were uniformly drawn for $i \in \Omega$, $\mathcal{Y} \in \Omega \setminus \{i\}$ and $\mathcal{X} \subset \mathcal{Y}$ to estimate c_f, and the following is calculated:

$$c_{(\mathcal{X},\mathcal{Y},i)} := f_i(\mathcal{Y})/f_i(\mathcal{X}) \tag{33}$$

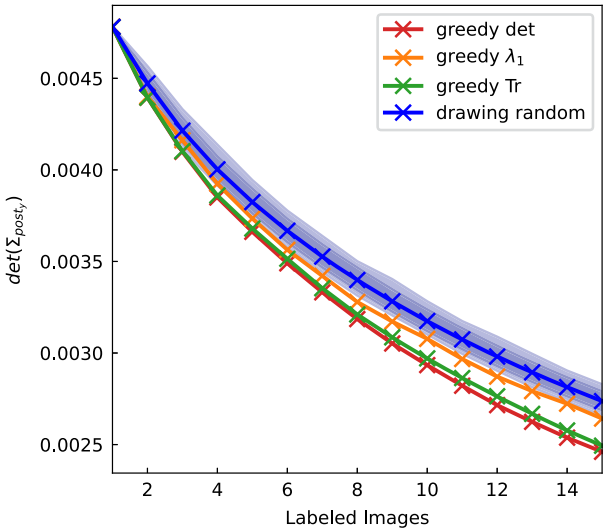
The highest value was ≈ 2.52, whereas only 1.2% of the samples were greater than 1 (see Figure 4). The values greater than 1 prove that the function is not a submodular set function, but the moderate value could suggest that c_f is within this order of magnitude.

We used jackknife resampling to evaluate the performance of the approach on the patient data because additional thoroughly segmented slices from the respective hearts were unavailable. Thus, slices were left out to determine the prior with sampling from the left-out slice. Despite the presence of visual obstructions, such as valves that may appear differently in certain heart diseases and pose challenges for neural network approaches to segment the images accurately, the most significant frequencies affecting the volume over time, such as respiratory and heart rates, are expected to remain consistent in those slices. Hence, the posterior predictions can be compared to the segmented images as ground truth.

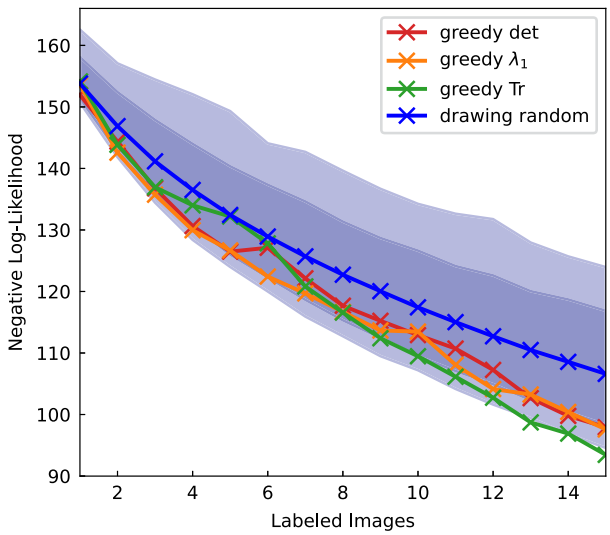
All three objective functions aim to reduce the posterior distribution spread. Figure 5a and 5b demonstrate that each objective function produces narrower posterior distributions than random sampling, suggesting that the greedy algorithm is a sensible approach. Appendix A compares all measures on



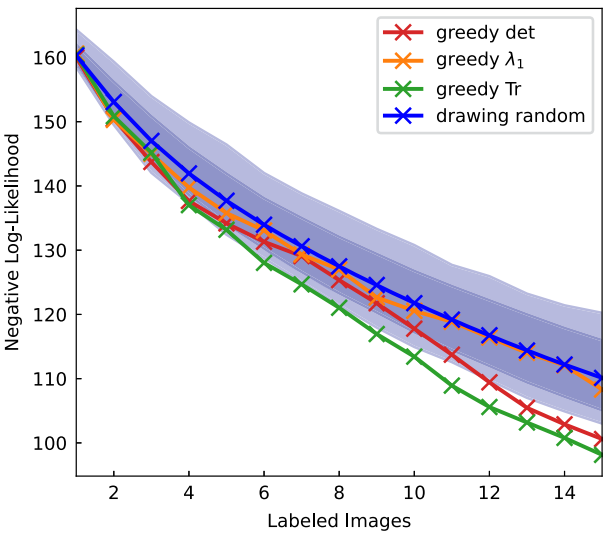
(a) Trace Heart 1



(b) Determinant Heart 1



(c) nll Heart 1

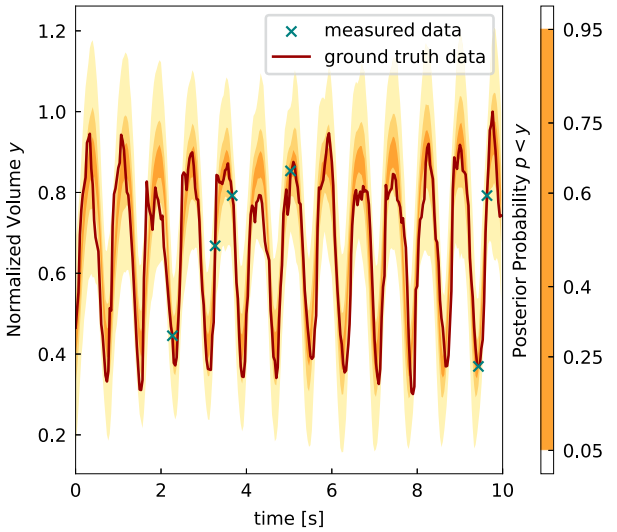


(d) nll Heart 2

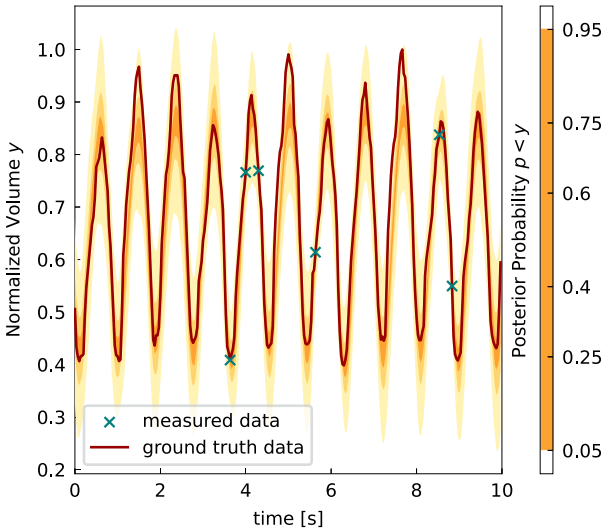
Fig. 5 Results of optimizing the posterior covariance. (a, b) Resulting covariance of the greedy approaches with 10,000 random draws. Shaded areas display the range of the random draws, where darker blue marks the inner 90%, and the blue curve represents the mean for (a) covariance trace, (b) determinant, and (c, d) resulting negative log-likelihood on Hearts 1 and 2. The wider band on Heart 1 reflects the higher irregularity in the heartbeat in this heart.

both hearts. The posterior distributions found by minimizing the trace and determinant of the posterior covariance matrix perform similarly, minimizing the largest eigenvalue yields a wider posterior. The first eigenvalue seems insufficient to describe the full spread of the posterior. More critical than the spread of the posterior distribution is for the posterior to represent the unknown data accurately. The jackknife sampling method enables a comparison of the posterior distributions with the ground-truth data, as illustrated in Figure 5c and 5d . The negative log-likelihood of the left-out slices using the greedy approach is compared with that of 10,000 random samples.

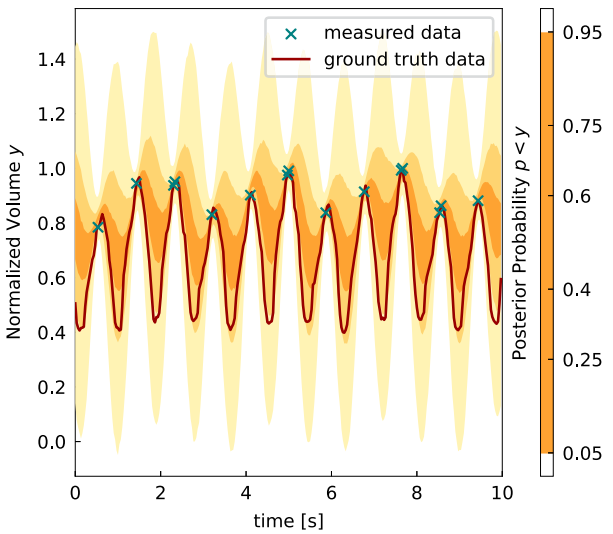
Among the three objective functions, $S\left(\Sigma^{(J)}_{post_y}\right) = \text{tr}\left(\Sigma^{(J)}_{post_y}\right)$ achieves the best results, as it is the only method to surpass random sampling across all samples with



(a) Five labeled images (Heart 1)



(b) Five labeled images (Heart 2)



(c) Greedy worst

Fig. 6 Predictive posterior for measured data depicted in the shaded yellow area for labeled images from unknown ground-truth data. The measured data is marked with a green x. (c) Results of 15 labeled images selected by a greedy worst approach on Heart 2. Compared to the result of six selected images by a greedy algorithm minimizing the trace on Heart 1 (a) or Heart 2 (b), the posterior variance is significantly smaller in (a) and (b) with fewer labeled images. The posterior also corresponds more accurately to the ground-truth data in those examples.

15 labeled images on both hearts. The determinant yields slightly worse results but still outperforms all random samples for more than 12 labeled images in Heart 2 and over 95% of the random samples in Heart 1 for 14 and 15 labeled images. Using the largest eigenvalue as an objective function yields the worst results, barely outperforming the average random sampling approach on Heart 2. This finding is also consistent with the observation that minimizing the largest eigenvalue fails to minimize the spread of the posterior to the same degree as the other two objectives.

Overall, the most important benefit of using these greedy algorithms is preventing the random selection of inefficient sequences, especially in the more irregular heart (Heart 1). The broader negative log-likelihood range of the random

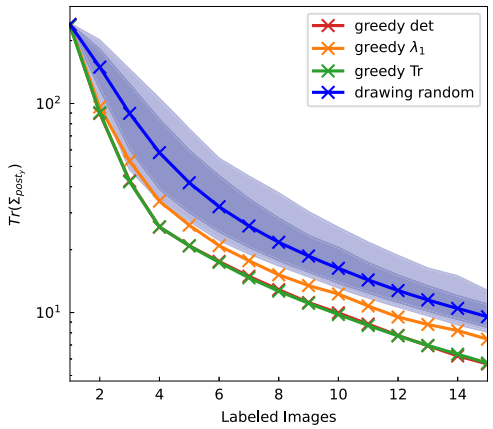
samples indicates the importance of an informed selection of labeled images. Drawing seven images according to the trace of the posterior distribution outperforms the worst random sample with 15 labeled images. For Heart 2, only nine labeled images are needed to outperform the worst random sample with 15 images. The importance of avoiding poor sequences can also be visually observed. The posterior distribution captures the ground truth for only five sample points selected with the greedy trace approach (see Figure 6a and 6b). In contrast , when selecting the sample images with a greedy worst approach, the posterior does not fit the data. However, the model also represents this uncertainty in a wide posterior distribution, providing a clear warning signal for the practitioner not to trust the predictions (Figure 6c).

5 Conclusion

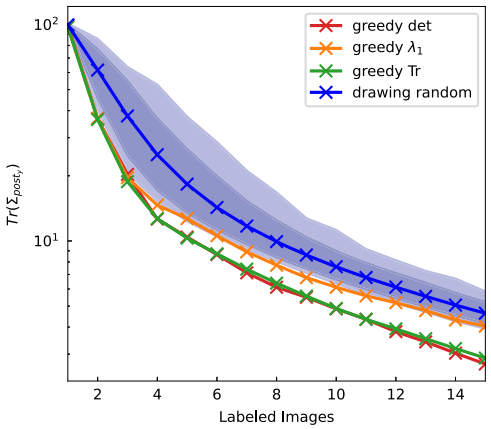
The SBL algorithm identifies the dominant frequencies in ventricle volumes over time from real-time MRI scans. Sparse frequency priors can efficiently label slices where automatic segmentation is inadequate. A greedy approach can be applied to select the optimal image for manual segmentation to minimize an objective function related to the posterior covariance spread. In this application, using the trace of the posterior covariance as an objective function is the least computationally expensive and yields the best results. Performance can be guaranteed for the greedy algorithm on the trace of the posterior covariance. The greedy approach stays within a factor of $(1 - e^{-1/c_f})$ of the optimal solution. Empirically, the highest value was $\approx$ 2.5, suggesting that c_f might also be in this order of magnitude. Testing this approach on real-life patient data demonstrated that the selection procedure minimized the posterior covariance spread. Moreover, the resulting models also accurately predicted the unknown data. Specifically, the model using 15 labeled images selected based on the trace of the posterior covariance consistently outperformed all models using randomly selected images, achieving the lowest negative log-likelihood for Hearts 1 and 2 across 10,000 random samples.

Appendix A Supplementary Figures

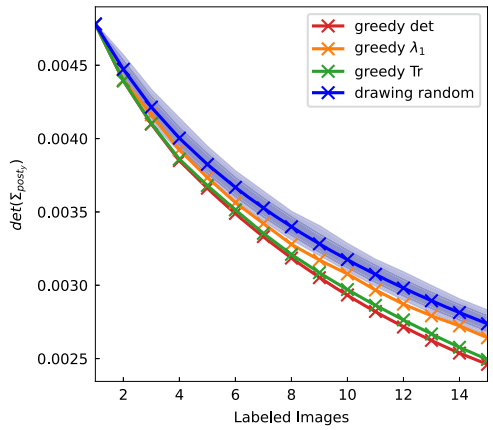
This work measures the posterior covariance spread with the trace, determinant, and largest eigenvalue of the posterior covariance matrix (see (30)). The greedy algorithm determined which images to label on the three mentioned measures. Figure 7 compares the performance of the selection schemes on both hearts for all three measures.



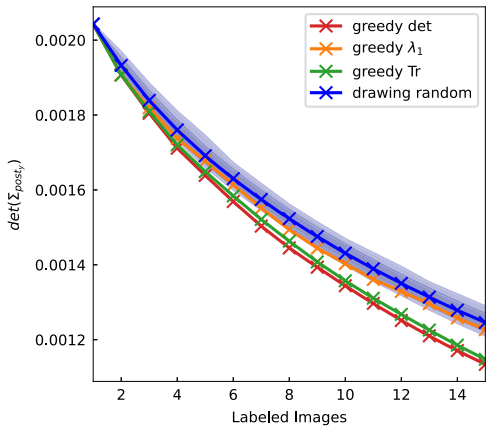
(a) Trace Heart 1



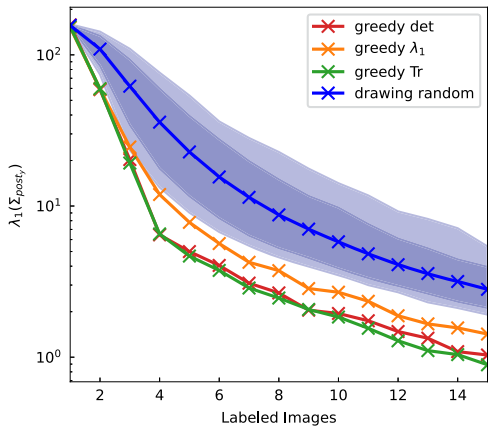
(b) Trace Heart 2



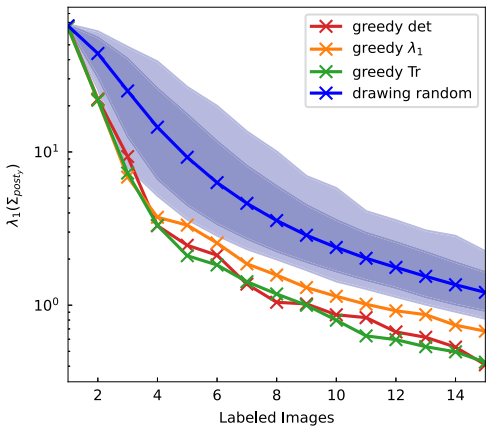
(c) Determinant Heart 1



(d) Determinant Heart 2



(e) Largest eigenvalue Heart 2



(f) Largest Eigenvalue Heart 2

Fig. 7 Results of optimizing the posterior covariance, comparing the covariance of the greedy approaches with 10,000 random draws. Shaded areas depict the range of random draws, where darker blue marks the inner 90%, and the blue curve is the mean. For comparison, the largest eigenvalue, determinant, and trace of the posterior covariance are plotted for both hearts.

Acknowledgements This work was performed as part of the Helmholtz School for Data Science in Life, Earth and Energy (HDS-LEE) and received funding from the Helmholtz Association of German Research Centres. This publication is partly supported by the Alexander von Humboldt Foundation and the King Abdullah University of Science and Technology (KAUST) Office of Sponsored Research (OSR) under Award No. OSR-2019-CRG8-4033. We would like to thank the whole Hypofon study team: Nicole Müller, Julian Alexander Härtel, Jan Schmitz, Ute Baur, Melanie von der Wiesche, Iris Rieger, Jon von Stritzky, Christopher Hart, Janina Bros, Benedikt Seeger, Emily Zollmann, Marijke Grau, Boris Dragutinovic, Laura-Maria de Boni, Jan-Niklas Hönemann, Wilhelm Bloch, Daniel Aeschbach, Eva-Maria Elmenhorst, Ulrike Herberg, Alena Hess, Moritz Schumann, Tobias Kratz, Jens Jordan and Johannes Breuer and the participants for the unique study data and Stiftung KinderHerz for making this study possible.

Author Contributions F.T. wrote the main manuscript text and prepared all figures. R.T., P.K., Ac.B. helped with mathematical modeling questions, design and evaluation. An.B., D.A.G., J.T. provided the data, helped interpreting the data, validated the neural network segmentations, and helped with medical intepretation. All authors reviewed the manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data Availability The data that support the findings of this study are not openly available due to reasons of sensitivity. The data was obtained by the HYPOFON study: https://doi.org/10.1161/CIRCULATIONAHA.123.067601

Declarations

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Bartlett, M.S.: An Inverse Matrix Adjustment Arising in Discriminant Analysis. Ann. Math. Stat. **22**(1), 107–111 (1951). https://doi.org/10.1214/aoms/1177729698

Bernard, O., Lalande, A., Zotti, C., Cervenansky, F., Yang, X., Heng, P.-A., Cetin, I., Lekadir, K., Camara, O., Gonzalez Ballester, M.A., Sanroma, G., Napel, S., Petersen, S., Tziritas, G., Grinias, E., Khened, M., Kollerathu, V.A., Krishnamurthi, G., Rohé, M.-M., Pennec, X., Sermesant, M., Isensee, F., Jäger, P., Maier-Hein, K.H., Full, P.M., Wolf, I., Engelhardt, S., Baumgartner, C.F., Koch, L.M., Wolterink, J.M., Išgum, I., Jang, Y., Hong, Y., Patravali, J., Jain, S., Humbert, O., Jodoin, P.-M.: Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: Is the problem solved? IEEE Trans. Med. Imaging **37**(11), 2514–2525 (2018). https://doi.org/10.1109/TMI.2018.2837502

Budd, S., Robinson, E.C., Kainz, B.: A survey on active learning and human-in-the-loop deep learning for medical image analysis. Med. Image Anal. **71**, 102062 (2021). https://doi.org/10.1016/j.media.2021.102062

Campello, V.M., Gkontra, P., Izquierdo, C., Martín-Isla, C., Sojoudi, A., Full, P.M., Maier-Hein, K., Zhang, Y., He, Z., Ma, J., Parreño, M., Albiol, A., Kong, F., Shadden, S.C., Acero, J.C., Sundaresan, V., Saber, M., Elattar, M., Li, H., Menze, B., Khader, F., Haarburger, C., Scannell, C.M., Veta, M., Carscadden, A., Punithakumar, K., Liu, X., Tsaftaris, S.A., Huang, X., Yang, X., Li, L., Zhuang, X., Viladés, D., Descalzo, M.L., Guala, A., Mura, L.L., Friedrich, M.G., Garg, R., Lebel, J., Henriques, F., Karakas, M., Çavuş, E., Petersen, S.E., Escalera, S., Seguí, S., Rodríguez-Palomares, J.F., Lekadir, K.: Multi-centre, multi-vendor and multi-disease cardiac segmentation: The M&Ms challenge. IEEE Trans. Med. Imaging **40**(12), 3543–3554 (2021). https://doi.org/10.1109/TMI.2021.3090082

Das, A., Kempe, D.: Submodular meets spectral: Greedy algorithms for subset selection, sparse approximation and dictionary selection, 1057–1064 (2011)

Dümbgen, L.: On Tyler's M-Functional of Scatter in High Dimension. Ann. Inst. Stat. Math. **50**(3), 471–491 (1998). https://doi.org/10.1023/A:1003573311481

Dümbgen, L., Tyler, D.E.: On the breakdown properties of some multivariate m-functionals. Scand. J. Stat. **32**(2), 247–264 (2005)

Gemba, K.L., Nannuru, S., Gerstoft, P., Hodgkiss, W.S.: Multi-frequency sparse Bayesian learning for robust matched field processing. The Journal of the Acoustical Society of America **141**(5), 3411–3420 (2017). https://doi.org/10.1121/1.4983467

Gerstoft, P., Mecklenbräuker, C.F., Xenaki, A., Nannuru, S.: Multisnapshot Sparse Bayesian Learning for DOA. IEEE Signal Process. Lett. **23**(10), 1469–1473 (2016). https://doi.org/10.1109/LSP.2016.2598550

Hashemi, A., Ghasemi, M., Vikalo, H., Topcu, U.: Submodular observation selection and information gathering for quadratic models. Proceedings of the 36th International Conference on Machine Learning **97**, 2653–2662 (2019)

Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J., Maier-Hein, K.H.: nnU-net: a self-configuring method for deep learning-based biomedical image segmentation. Nat. Methods **18**(2), 203–211 (2021). https://doi.org/10.1038/s41592-020-01008-z

Kunz, J., Ruschke, S., Heckel, R.: Free-breathing 2D Cartesian Cardiac MRI Datasets. IEEE Dataport. Published: May 23, 2023 (2023). https://doi.org/10.21227/f057-dw29

MacKay, D.J.C.: Bayesian interpolation. Neural Comput. **4**, 415–447 (1992)

Müller, N., Härtel, J.A., Schmitz, J., Baur, U., Wiesche, M., Rieger, I., Gerlach, D., Stritzky, J., Bach, A., Hart, C., Bros, J., Seeger, B., Zollmann, E., Grau, M., Dragutinovic, B., Boni, L.-M., Hónemann, J.-N., Bloch, W., Aeschbach, D., Elmenhorst, E.-M., Herberg, U., Hess, A., Schumann, M., Kratz, T., Jordan, J., Breuer, J., Tank, J.: Peripheral oxygenation and pulmonary hemodynamics in individuals with fontan circulation during 24-hour high-altitude exposure simulation. Circulation **149**(18), 1466–1468 (2024). https://doi.org/10.1161/CIRCULATIONAHA.123.067601

Murphy, K.P.: Probabilistic Machine Learning: Advanced Topics. MIT Press, Cambridge, MA (2023). http://probml.github.io/book2

Nemhauser, G.L., Wolsey, L.A., Fisher, M.L.: An analysis of approximations for maximizing submodular set functions-I. Math. Program. **14**(1), 265–294 (1978). https://doi.org/10.1007/BF01588971

Paindaveine, D.: A canonical definition of shape. Statistics & Probability Letters **78**(14), 2240–2247 (2008). https://doi.org/10.1016/j.spl.2008.01.094

Ramedani, M., Moussavi, A., Memhave, T.R., Boretius, S.: Deep learning-based automated segmentation of cardiac real-time mri in non-human primates. Comput. Biol. Med. **189**, 109894 (2025). https://doi.org/10.1016/j.compbiomed.2025.109894

Randles, R.H.: A Simpler, Affine-Invariant, Multivariate, Distribution-Free Sign Test. J. Am. Stat. Assoc. **95**(452), 1263–1268 (2000)

Salerno, M., Sharif, B., Arheden, H., Kumar, A., Axel, L., Li, D., Neubauer, S.: Recent advances in cardiovascular magnetic resonance. Circulation: Cardiovascular Imaging 10(6), 003951 (2017) https://doi.org/10.1161/CIRCIMAGING.116.003951

Salibián-Barrera, M., Aelst, S.V., Willems, G.: Principal components analysis based on multivariate mm estimators with fast and robust bootstrap. J. Am. Stat. Assoc. **101**(475), 1198–1211 (2006)

Shelhamer, E., Long, J., Darrell, T.: Fully convolutional networks for semantic segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **39**(4), 640–651 (2017). https://doi.org/10.1109/TPAMI.2016.2572683

Tang, T., Yang, C., Xie, T., Liu, Y., Xu, L., Chen, D.: FBC-SBL: Frequency band clustering sparse bayesian learning for off-grid wideband doa estimation with different frequency bands. IEEE Geoscience and Remote Sensing Letters 21, 1–5 (2024) https://doi.org/10.1109/LGRS.2024.3424540

Terhag, F., Knechtges, P., Basermann, A., Tempone, R.: Uncertainty quantification in machine learning based segmentation: A post-hoc approach for left ventricle volume estimation in mri. SIAM/ASA Journal on Uncertainty Quantification **13**(1), 90–113 (2025). https://doi.org/10.1137/23M161433X

Tipping, M.E.: Sparse bayesian learning and the relevance vector machine. J. Mach. Learn. Res. **1**, 211–244 (2001)

Visuri, S., Ollila, E., Koivunen, V., Möttönen, J., Oja, H.: Affine equivariant multivariate rank methods. Journal of Statistical Planning and Inference **114**(1), 161–185 (2003). https://doi.org/10.1016/S0378-3758(02)00469-X. (**C.R. Rao 80th Birthday Felicitation Volume, Part IV**)

Wang, Q., Zhao, Z., Chen, Z., Nie, Z.: Grid evolution method for doa estimation. IEEE Trans. Signal Process. **66**(9), 2374–2383 (2018). https://doi.org/10.1109/TSP.2018.2814998

Wang, Q., Yu, H., Li, J., Ji, F., Chen, F.: Adaptive grid refinement method for DOA estimation via sparse bayesian learning. IEEE J. Oceanic Eng. **48**(3), 806–819 (2023). https://doi.org/10.1109/JOE.2023.3235055

Wipf, D.P., Rao, B.D.: An empirical Bayesian strategy for solving the simultaneous sparse approximation problem. IEEE Trans. Signal Process. **55**(7), 3704–3716 (2007). https://doi.org/10.1109/TSP.2007.894265

Yang, Z., Xie, L., Zhang, C.: Off-grid direction of arrival estimation using sparse bayesian inference. IEEE Trans. Signal Process. **61**(1), 38–43 (2013). https://doi.org/10.1109/TSP.2012.2222378

Zhang, S., Uecker, M., Voit, D., Merboldt, K.-D., Frahm, J.: Real-time cardiovascular magnetic resonance at high temporal resolution: radial flash with nonlinear inverse reconstruction. J. Cardiovasc. Magn. Reson. **12**(1), 39 (2010). https://doi.org/10.1186/1532-429X-12-39

Zhang, S., Joseph, A.A., Voit, D., Schaetz, S., Merboldt, K.-D., Unterberg-Buchwald, C., Hennemuth, A., Lotz, J., Frahm, J.: Real-time magnetic resonance imaging of cardiac function and flow-recent progress. Quant. Imaging Med. Surg. **4**(5), 313 (2014). https://doi.org/10.3978/j.issn.2223-4292.2014.06.03

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.