Hierarchical data integration with Gaussian processes: application to the characterization of cardiac ischemia-reperfusion patterns

Benoit Freiche, Gabriel Bernardino, Romain Deleat-Besson, Patrick Clarysse, Nicolas Duchateau

Hierarchical data integration with Gaussian processes: application to the characterization of cardiac ischemia-reperfusion patterns

Benoit Freiche, Gabriel Bernardino, Romain Deleat-Besson, Patrick Clarysse, Nicolas Duchateau

***Abstract*— Cardiac imaging protocols usually result in several types of acquisitions and descriptors extracted from the images. The statistical analysis of such data across a population may be challenging, and can be addressed by fusion techniques within a dimensionality reduction framework. However, directly combining different data types may lead to unfair comparisons (for heterogeneous descriptors) or over-exploitation of information (for strongly correlated modalities). In contrast, physicians progressively consider each type of data based on hierarchies derived from their experience or evidence-based recommendations, an inspiring approach for data fusion strategies.**

In this paper, we propose a novel methodology for hierarchical data fusion and unsupervised representation learning. It mimics the physicians' approach by progressively integrating different high-dimensional data descriptors according to a known hierarchy. We model this hierarchy with a Hierarchical Gaussian Process Latent Variable Model (GP-LVM), which links the estimated low-dimensional latent representation and high-dimensional observations at each level in the hierarchy, with additional links between consecutive levels of the hierarchy.

We demonstrate the relevance of this approach on a dataset of 1726 magnetic resonance image slices from 123 patients revascularized after acute myocardial infarction (MI) (first level in the hierarchy), some of them undergoing reperfusion injury (microvascular obstruction (MVO), second level in the hierarchy). Our experiments demonstrate that our hierarchical model provides consistent data organization across levels of the hierarchy and according to physiological characteristics of the lesions. This allows more relevant statistical analysis of myocardial lesion patterns, and in particular subtle lesions such as MVO.

***Index Terms*— Gaussian processes, Data hierarchy, Information fusion, Representation learning, Cardiac imaging, Myocardial infarction.**

The authors acknowledge the partial support from the French National Research Agency (ANR) through the LABEX PRIMES of Université de Lyon (ANR-11-LABX-0063) and the MIC-MAC project (ANR-19-CE45-0005), the Fédération Francaise de Cardiologie (MI-MIX project, Allocation René Foudon), the Université Claude Bernard Lyon 1 (AAP SENS 2022), and the Institut Universitaire de France.

B. Freiche, G. Bernardino, R. Deleat-Besson, P. Clarysse, N. Duchateau are with Univ Lyon, INSA-Lyon, Université Claude Bernard Lyon 1, UJM-Saint Etienne, CNRS, Inserm, CREATIS UMR 5220, U1294, F-69621, Lyon, France (e-mail: benoit.freiche@creatis.insa-lyon.fr).

G. Bernardino is also with the Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, Spain.

N. Duchateau is also with the Institut Universitaire de France (IUF), 75005 Paris, France.

I. Introduction

COMPLEX diseases can rarely be explained by a single information source (single type of measurement, imaging acquisition, or even modality), which only offers partial information about the patient's state. Physicians are actually able to merge several types of heterogeneous information in their mind to characterize the patients' condition, both for understanding purposes (unsupervised approach) and for diagnosis or prognosis (supervised approach) [1]. For better robustness and efficiency, they incorporate these pieces of information incrementally, with the order based on their experience (e.g. from simpler to more complex or more expensive measurements) or guidelines [2]. However, this merging of information remains very qualitative and subjective, hence the need for computational and reproducible methods.

Merging such information is still a major challenge for computational approaches [3], given the heterogeneous types and dimensionalities of data descriptors, in particular when processing high-dimensional imaging data. Making the data integration incremental (namely, following a hierarchy in the data descriptors) could be highly beneficial [4], in particular to limit the computational burden and potential unstabilities compared to strategies that merge all the data at once.

In this paper, we propose an original framework for the hierarchical integration of multiple types of high-dimensional data extracted from medical images, in the context of unsupervised representation learning. It can be seen as a substantial generalization of our previous work [5], which demonstrated that hierarchical data integration (with simple manifold learning) is relevant for such an application. Here, we rely on Gaussian processes (GP) to explicitly model dependencies between the data descriptors, while also considering the link between the latent representation and the observations, and potential priors on latent spaces or mappings. More specifically, we extend the framework of hierarchical Gaussian process latent variable models (GP-LVMs) [6] to construct such data hierarchies. Our scheme is different in the sense that a given latent representation (our lower level in the hierarchy) is used to guide another latent representation (our higher level) but not directly targetting a fusion of different information sources. Besides, in our approach each level in the hierarchy is connected to both observations and latent variables of data of potentially different complexity. Also, our approach notably includes an additional return GP to serve as regularization term for difficult

real data hierarchies.

We demonstrate its relevance on a medical problem of high prevalence in the general population, namely the characterization of cardiac ischemia-reperfusion patterns from delayed enhancement magnetic resonance imaging (MRI) images, using real data from a randomized clinical trial. This methodology allows more reliable and robust representations, compared to simpler variants of the hierarchy and single-descriptor representations, and more structured statistical analysis of myocardial lesion patterns.

Code and demo data are available at `https://github.com/creatis-myriad/HGP_LVM`.

II. Related work

A. Data fusion with representation learning

1) Unsupervised dimensionality reduction: Dimensionality reduction techniques allow representing useful information hidden in high-dimensional and complex imaging descriptors within a simpler data space, called latent space.

Among unsupervised approaches, manifold learning reduces dimensionality by exploiting the structure of the data space. It has been shown that most linear and non-linear manifold learning methods can be grouped into a common graph embedding framework [7]. The main idea is to build a similarity matrix whose elements encode a meaningful measure between the high-dimensional samples, and then perform the spectral decomposition of this matrix. The different distances used lead to distinct representations: for example, Isomap [8] operates globally with the geodesic distance, whereas diffusion maps [9] operate locally with the diffusion distance.

Auto-encoders (AEs) are deep neural networks that estimate a low-dimensional latent representation of a population while also optimally reconstructing each high-dimensional sample from this latent space. The encoding part is composed by layers of decreasing number of neurons, leading to a bottleneck, which is a low-dimensional representation of a sample in the latent space. From the latent space, the decoder reconstructs the input high-dimensional data. Variational auto-encoders (VAEs) [10] are a probabilistic extension of AEs. They associate to each sample a normal latent distribution instead of a vector of values, and both build a latent space and reconstruct samples by sampling from the learnt distribution. AEs have been very popular in medical image analysis, but they still suffer from difficulties in training and unstability, as well as limited interpretability and statistical properties of the estimated latent space, topics of active research in the recent years [11], [12].

GP-LVMs [13] are a complementary and promising framework to perform the low-dimensional embedding. They estimate a non-linear Gaussian mapping between observations and the latent space by explicitly modelling the relationship between observations and the latent representation. Their flexibility allows prior knowledge to be included in the model (e.g. dynamics [6]) and thus paves the ground for relevant extensions to multi-modal or multi-descriptor data, as discussed in the following subsections. GP-LVMs require less data than neural networks, but they struggle to scale to large datasets. A neural network version, the deep Gaussian Process [14], was developed for larger data sets.

2) Data fusion: Using multi-descriptor data provides richer information, but presents several difficulties, mainly linked to the high dimensionality and heterogeneity of the data descriptors, as direct concatenation leads to suboptimal results. To estimate a low-dimensional representation from multi-descriptor data, two types of methods have emerged: fusion and alignment [15].

Fusion methods merge the different data descriptors into a single latent space. Multiple kernel learning (MKL) [16] extends the general framework for manifold learning [7] by encoding pairwise affinities between samples into a different kernel matrix for each data descriptor, and simultaneously optimizing the low-dimensional embedding and the combination of such affinity matrices. In a comparable manner, the Similarity Network Fusion (SNF) algorithm [17] fuses such affinity matrices by an iterative diffusion process which amounts at finding a single graph that represents the whole dataset. Multi-channel variational auto-encoders (MCVAEs) [18]–[20] generalize the framework of VAEs to several channels, each associated to a different modality or data descriptor, controlling the encoding of each channel and its reconstruction. Similar to MCVAEs, Partial VAEs [21] achieve a relevant variable selection between several entries. This method aims to predict a quantity with the least amount of entries as possible, using a deep neural network able to handle missing variables and learn from partial observations.

Alignment methods, on the other hand, learn several latent spaces (one for each modality) while optimizing their matching. For example, multiple manifold learning (MML) [22], [23] encodes in a block matrix both modality-specific affinities between samples (diagonal blocks) and cross-modality affinities between samples (extra-diagonal blocks, which force the alignment of latent spaces).

In previous works, we demonstrated the relevance of fusion [24] and alignment [25] approaches to jointly analyze several types of high-dimensional descriptors extracted from cardiac images, and exploit the latent representation to characterize specific cardiac diseases. However, such approaches combine all the data descriptors simultaneously, which may be critical both in terms of efficiency and robustness, in particular for data descriptors of high and/or very different dimensionalities. In contrast, a hierarchical or incremental integration of such data has strong potential to overcome this.

3) Hierarchical data fusion: Within kernel-based approaches, iterative processes have been proposed to progressively evolve a consensus kernel matrix [26], [27], but these methods have only been demonstrated for the incremental incorporation of new samples (a concept close from curriculum learning [28]) and not for multi-modal data analysis. Within the framework of manifold learning, constraints close to those used to align several manifolds [22], [23] have been proposed to link consecutive levels in a hierarchy of data [29]. This method was demonstrated to consider multiple scales for the regional analysis of medical images.

As an alternative and more generic approach, GP-LVMs can be adapted to explicitly encode hierarchical relationships between different data levels [6]. The authors first illustrated this on time series of stickman movements with a two-step hierarchy where time is used as first level to guide the representation of the stickman's trajectory. They also elaborated on this idea to guide the representation of a moving articulated body by explicitly modeling the movement of separate organs as former levels in the hierarchy.

In previous work, we demonstrated the relevance of a two-step hierarchy based on manifold learning [29] to guide the analysis of image appearance in MRI data [5]. In this paper, we propose an original generalization based on Gaussian processes, which is both intuitive and provides flexibility for multiple data integration following a prescribed hierarchy.

B. Characterizing acute myocardial infarction patterns

Acute myocardial infarction occurs when the blood coming from one coronary artery is insufficient to meet the oxygen needs of the myocardium, usually because of obstructed coronary arteries. If ischemia is prolonged, it will lead to death of myocardium, which will trigger remodelling affecting cardiac shape and function. This may ultimately lead to severe complications and death. When detected early enough, the patient's heart is reperfused (revascularized e.g. by angioplasty). The restored supply in blood will ideally limit the expansion of the myocardial infarct, but the cells' ischemia defense mechanisms often lead to severe inflammation because of this sudden reflow of oxygen. Such reperfusion damages are called MVO or no-reflow phenomenon. The understanding of the ischemia-reperfusion mechanisms is a major clinical research topic [30], in particular to decide a suited revascularization therapy and monitor patient outcome. Notably, MVO presence increases with infarct transmurality and is an independent predictor of major adverse clinical outcome [31].

Medical imaging is of high value to study ischemic damages and the no-reflow phenomenon, usually using MRI [32], [33]. The MRI protocol starts with the acquisition of Native T1 and T2 maps. Then, a contrast agent (Gadolinium) is injected to the patient. Around 3-4 minutes after contrast agent injection, early Gadolinium enhancement (EGE) images are acquired to assess early MVO. Ten minutes after contrast agent injection, late Gadolinium enhancement (LGE) images are acquired to assess MI and late MVO (Fig. 1). Unfortunately, physicians lack tools to exploit the rich information available in the images. They simplify the assessment of myocardial lesions to simple scalar measurements extracted from the segmentations (infarct size and transmurality, mainly). This drastically limits the understanding of the complex mechanisms at stake here. Besides, specific care should be taken when analyzing medical imaging data that belong to a non-linear manifold, as for MI and MVO patterns. This is particularly critical for MVO compared to MI (Fig. 2), and even more considering that MVO patterns and evolution may be both spatially and temporally complex [34].

To substantially improve the understanding of ischemia-reperfusion mechanisms, it is necessary to integrate several types of data descriptors extracted from the images, of different levels of complexity, into the analysis of populations of patients with myocardial infarct. Our paper addresses this crucial aspect by means of an original representation learning method explicitly designed to consider hierarchies between multi-descriptor data.

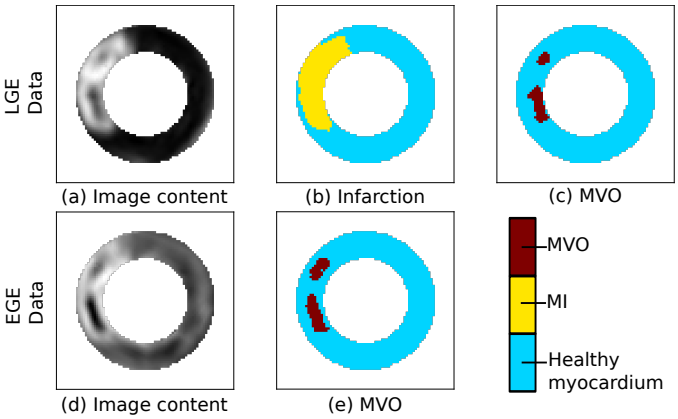

Fig. 1: The different types of images we may analyze with delayed enhancement cardiac MRI. Only myocardial contents are displayed, and represented on the same geometry, as for the analysis done in this paper. First row: LGE data: (a) LGE image, (b) MI segmentation in yellow, (c) MVO segmentation in dark red. Second row: EGE data: (d) EGE image, (e) MVO segmentation in dark red.

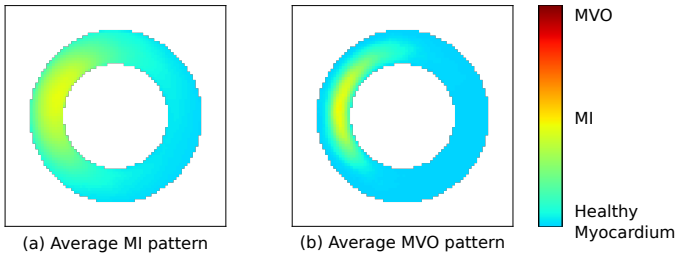

Fig. 2: Linear average of lesion patterns ((a) MI and (b) MVO) across all patients in the analyzed database (all lesions have been reoriented to the same coronary artery territory). The average pattern is blurry as both MI and MVO images originate from a non-linear manifold. This may not be too critical for MI, contrary to MVO which does not resemble at all a MVO lesion. This highlights the need for non-linear dimensionality reduction, as we perform in this paper.

III. METHODS

A. Data

1) Imaging data: We focus on myocardial lesions due to ischemia and reperfusion injuries in the context of acute myocardial infarction. Specifically, we focus on the data from a randomized clinical trial for which both early and late enhancement images were available: the Minimalist Immediate Mechanical Intervention (MIMI) study [35] (ClinicalTrials ID: NCT01360242). The study protocol was approved by the local ethics committee (IRB 2010-048), and complied with the Declaration of Helsinki and French laws. All subjects gave written informed consent. This study was originally designed to compare two reperfusion strategies: delayed and immediate stenting, the "delay" being of 24-48 hours. Image acquisition occurred 4 to 6 days after the medical intervention, so MVO is still visible both on EGE and LGE images, and the cardiac shape has not remodelled yet, which limits external factors that could affect the analysis.

Our dataset was comprised of the MRI images of 123 patients, with median resolution $1.5625 \times 1.5625 \times 5$ mm, on which the left ventricle (LV) ranged over 17 ± 2 slices. These images were cropped around the LV and resized to 80×80 pixels. Segmentations for the MI and MVO areas were obtained semi-automatically from commercial software (CVI42 v.5.1.0 Circle Cardiovascular Imaging, Calgary, Canada) by one experienced observer and controlled by two other experienced observers.

2) Spatial alignment: We identified the anterior LV-RV junction on each slice, and the apical and basal slices on each image stack, which allowed to parameterize the myocardium by defining radial, circumferential, and long-axis coordinates on each slice (all having values between 0 and 1), for each individual. Then, we used this parametrization to transport the image data and segmentations to a common geometry (linear interpolation for data on a scattered grid), to standardize the inputs before the machine learning [36].

Infarct patterns were assigned to one of the four following categories, corresponding to the coronary artery responsible for the myocardial lesions: left anterior descending (LAD)-proximal, LAD-mid, right coronary artery (RCA), left circumflex artery (LCX) (see Table I).

We did an additional spatial alignment step to benefit from a larger population size and focus on the infarct shape while limiting the influence of other less-relevant factors such as its location around the myocardium. To do so, all 3D infarct patterns were rotated along the circumference such that the average center of mass of all infarcts associated to a given coronary territory is aligned to the LAD center of mass. This still keeps a bit of angular variability (in particular, as we process 2D slices as we indicate in the following subsection), which serves to assess the relevance of our methods through the "Angle" parameter (Tab.II).

3) Analyzed slices: In the rest of this paper, the analysis is performed on 2D slices to speed up computations, in particular for experiments evaluating the hyperparameters role and choosing an optimal strategy. Besides, visualization is easier, a clear asset to compare the lesion patterns sometimes visible on few slices (as for MVO). We discarded the slices with no lesion, and focused the analysis on the remaining slices, considered as independent samples (to benefit from a larger population). As a result, our population consists of 1726 samples associated to 123 patients.

Culprit artery	Immediate stenting	Delayed stenting	Total
LAD mid	11	14	25
LAD proximal	12	8	20
LCX	8	9	17
RCA	34	27	61
TOTAL	65	58	**123**

TABLE I: Distribution of the studied population by type of treatment and culprit coronary artery.

MI lesion patterns are large, have somehow regular shapes, and partially overlap over different slices. In contrast, MVO patterns are small, irregular, and do not overlap. Their statistical analysis even with non-linear representation learning techniques is therefore much more challenging. However, MVO lesions are by definition contained within the MI infarct region, which means that the representation of MVO patterns could be guided by a previously obtained representation of MI. This is what we propose to investigate by formulating this relationship through a hierarchical process, within the framework of GP-LVMs.

B. GP-LVM

Let's denote $\mathbf{Y} = [\mathbf{y}_1, \cdots, \mathbf{y}_N]^T \in \mathbb{R}^{N \times D}$ the set of D-dimensional observations (the dimensionality D is usually large) for a population of N subjects. We look for a lower-dimensional representation $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_N]^T \in \mathbb{R}^{N \times d}$ with $d \ll D$, which we aim to model by a GP-LVM.

This model assumes that the observations $\mathbf{Y}$ can be computed from the latent variables, or drawn from a distribution depending on the latent variables $\mathbf{X}$ as:

$$P(\mathbf{Y} \mid \mathbf{X}) = \prod_{j=1}^{D} \mathcal{N}(\mathbf{y}_{:,j} \mid \mathbf{0}, \mathbf{K}_x), \tag{1}$$

where $\mathbf{y}_{:,j}$ is a column of the matrix $\mathbf{Y}$, and $\mathcal{N}(.|\mathbf{0}, \mathbf{K}_x)$ stands for a centered multivariate Gaussian distribution whose covariance is generated from the kernel matrix $\mathbf{K}_x = [k(\mathbf{x}_i, \mathbf{x}_j)] \in \mathbb{R}^{N \times N}$, which depends on the latent variables $\mathbf{X}$. The kernel we use here is a Gaussian radial basis function (RBF) kernel:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \omega^2 \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2l^2}\right) + \sigma^2 \delta_{ij}, \tag{2}$$

with ω and σ the scale parameters, l the lengthscale of the kernel, and δ_{ij} the Dirichlet delta, respectively. This assumes that all features come from the same distribution, which has the same dimension as the number of data-points N. The probability distributions mean that if we were to acquire a new pixel (i.e. a new column), then it can be sampled from $\mathbf{K}_x$. This approach has limits regarding the relations between neighbouring pixels. This can be improved using coregionalisation [37], [38], allowing to introduce prior knowledge

that nearby pixels should behave similarly, which we leave for future work.

We also assume that $\mathbf{X}$ has a Gaussian prior distribution:

$$P(\mathbf{X}) = \prod_{j=1}^{d} \mathcal{N}(x_{:,j} \mid \mathbf{0}, \mathbf{I}), \tag{3}$$

The latent space is estimated by looking for a maximum a posteriori solution, namely maximizing the log-likelihood with respect to $\mathbf{X}$:

$$\underset{\mathbf{X}}{\operatorname{argmax}} \log P(\mathbf{X} \mid \mathbf{Y}). \tag{4}$$

Using Bayes' rule, this can be rewritten as:

$$\underset{\mathbf{X}}{\operatorname{argmax}} \log \frac{P(\mathbf{Y} \mid \mathbf{X})P(\mathbf{X})}{P(\mathbf{Y})}, \tag{5}$$

Eq.(5) corresponds to the maximization of $\log P(\mathbf{Y} \mid \mathbf{X})P(\mathbf{X})$, as $P(\mathbf{Y})$ is constant with respect to $\mathbf{X}$. The minimization of $-\log(\cdot)$ is achieved in practice by a gradient descent algorithm (we used the L-BFGS algorithm).

C. Hierarchical GP-LVM

1) Simple hierarchical model: The previous scheme can be extended to several types of observations (in our case, several types of descriptors extracted from the images) in a hierarchical way, by setting a Gaussian process between the latent variables associated to each data descriptor (Fig. 3a). Let's consider two latent variables $\mathbf{X}_0$ and $\mathbf{X}_1$, with corresponding observations $\mathbf{Y}_0$ and $\mathbf{Y}_1$. $\mathbf{X}_0$ is here the first level of the hierarchy, $\mathbf{X}_1$ the second. This amounts at solving:

$$\underset{\mathbf{X}_0,\mathbf{X}_1}{\operatorname{argmax}} \log P(\mathbf{X}_0, \mathbf{X}_1 \mid \mathbf{Y}_0, \mathbf{Y}_1) \tag{6}$$

Again, using Bayes' rule, we obtain:

$$P(\mathbf{X}_0, \mathbf{X}_1 \mid \mathbf{Y}_0, \mathbf{Y}_1) = \frac{P(\mathbf{Y}_0, \mathbf{Y}_1 \mid \mathbf{X}_0, \mathbf{X}_1)P(\mathbf{X}_0, \mathbf{X}_1)}{P(\mathbf{Y}_0, \mathbf{Y}_1)}. \tag{7}$$

As $\mathbf{Y}_0$ and $\mathbf{Y}_1$ are independent, we also have:

$$P(\mathbf{Y}_0, \mathbf{Y}_1 \mid \mathbf{X}_0, \mathbf{X}_1) = P(\mathbf{Y}_0 \mid \mathbf{X}_0, \mathbf{X}_1)P(\mathbf{Y}_1 \mid \mathbf{X}_0, \mathbf{X}_1). \tag{8}$$

As $\mathbf{Y}_0$ is independent from $\mathbf{X}_1$, and $\mathbf{Y}_1$ is independent from $\mathbf{X}_0$ knowing $\mathbf{X}_1$:

$$P(\mathbf{Y}_0 \mid \mathbf{X}_0, \mathbf{X}_1) = P(\mathbf{Y}_0 \mid \mathbf{X}_0), \tag{9}$$

$$P(\mathbf{Y}_1 \mid \mathbf{X}_0, \mathbf{X}_1) = P(\mathbf{Y}_1 \mid \mathbf{X}_1). \tag{10}$$

Besides, we have:

$$P(\mathbf{X}_0, \mathbf{X}_1) = P(\mathbf{X}_1 \mid \mathbf{X}_0)P(\mathbf{X}_0). \tag{11}$$

Finally, combining equations (7), (8), (9), (10), (11) leads to:

$$P(\mathbf{X}_0, \mathbf{X}_1 \mid \mathbf{Y}_0, \mathbf{Y}_1) = \frac{P(\mathbf{Y}_0 \mid \mathbf{X}_0)P(\mathbf{Y}_1 \mid \mathbf{X}_1)P(\mathbf{X}_1 \mid \mathbf{X}_0)P(\mathbf{X}_0)}{P(\mathbf{Y}_0, \mathbf{Y}_1)}. \tag{12}$$

As for the classical GP-LVM, we define a Gaussian prior over $\mathbf{X}_0$. Also, $P(\mathbf{Y}_0, \mathbf{Y}_1)$ is constant with respect to $\mathbf{X}_0$ and $\mathbf{X}_1$, so solving the optimization problem (6) amounts at solving together the GPs $\mathbf{X}_0 \rightarrow \mathbf{Y}_0$, $\mathbf{X}_1 \rightarrow \mathbf{Y}_1$, and $\mathbf{X}_0 \rightarrow \mathbf{X}_1$.

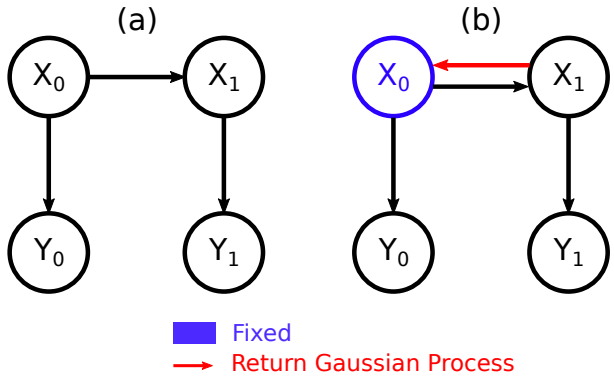

Fig. 3: (a) simple (Sec. III-C.1) and (b) improved (Sec. III-C.2) hierarchical GP-LVM models proposed in this paper. We train the improved model in two steps: first $\mathbf{X}_0$ is learnt from $\mathbf{Y}_0$, then it is fixed and we train $\mathbf{X}_1$.

2) Improved hierarchical model: In practice, the model presented in the previous section is difficult to optimize on real data. The optimization process leads to poorly organized latent spaces, which can even be diagonal in some attempts with redundant information on all dimensions. Thus, we made two substantial adaptations to this model, leading to the new model illustrated in Fig. 3b:

- We added a return Gaussian process between $\mathbf{X}_0$ and $\mathbf{X}_1$,
- We trained $\mathbf{X}_0$ first, and then $\mathbf{X}_1$ with $\mathbf{X}_0$ freezed, which amounts at forcing the hierarchical relationship between our data descriptors.

This amounts at solving:

$$P(\mathbf{X}_0, \mathbf{X}_1 \mid \mathbf{Y}_0, \mathbf{Y}_1) = \frac{P(\mathbf{Y}_0 \mid \mathbf{X}_0)P(\mathbf{Y}_1 \mid \mathbf{X}_1)P(\mathbf{X}_1 \mid \mathbf{X}_0)P(\mathbf{X}_0 \mid \mathbf{X}_1)P(\mathbf{X}_0)P(\mathbf{X}_1)}{P(\mathbf{Y}_0, \mathbf{Y}_1)}. \tag{13}$$

The broader alternate optimisation scheme would correspond to iteratively optimising $\mathbf{X}_0$ with fixed $\mathbf{X}_1$, and then optimising $\mathbf{X}_1$ with fixed $\mathbf{X}_0$. Our model is a step of this alternate optimisation scheme, which, if it converges, converges to a local minima of the joint function while corresponding to the concept of hierarchy we are looking for our application.

3) Hyperparameter tuning and training procedure: We trained our models with Python 3.7.6 using the gpflow library [39] (2.3.0 version).

Our proposed hierarchical model contains four GP-LVMs to optimize: two observation-based ($\mathbf{X}_0 \rightarrow \mathbf{Y}_0$ and $\mathbf{X}_1 \rightarrow \mathbf{Y}_1$), and two latent ones ($\mathbf{X}_0 \rightarrow \mathbf{X}_1$ and $\mathbf{X}_1 \rightarrow \mathbf{X}_0$). For each of them, there are three hyperparameters to optimize (ω, l, σ), totalling 12 hyperparameters. Each latent space has an additional hyperparameter, namely the standard deviation of the Gaussian prior, leading to 14 hyperparameters to optimize. Finally, on top of these GP-LVM-related hyperparameters, there are 4 additional hyperparameters controlling the relative weights of each GP-LVMs. In practice, we set the weight of $\mathbf{X}_0 \rightarrow \mathbf{Y}_0$ and $\mathbf{X}_1 \rightarrow \mathbf{Y}_1$ to 1, and use the same value β for

$\mathbf{X}_0 \rightarrow \mathbf{X}_1$ and $\mathbf{X}_1 \rightarrow \mathbf{X}_0$, which amounts at only determining one additional hyperparameter.

Regarding the training procedure, we initialize the latent spaces $\mathbf{X}_0$ and $\mathbf{X}_1$ with the first PCA vectors of their corresponding observations ($\mathbf{Y}_0$ and $\mathbf{Y}_1$, respectively). Then, we train the single-modality GP-LVMs $\mathbf{X}_0 \rightarrow \mathbf{Y}_0$ and $\mathbf{X}_1 \rightarrow \mathbf{Y}_1$, for which we alternatively optimize the hyperparameters and the latent space (for higher stability of the optimization). Finally, we train the two other GPs $\mathbf{X}_0 \rightarrow \mathbf{X}_1$ and $\mathbf{X}_1 \rightarrow \mathbf{X}_0$. For these, we re-use the hyperparameters values learnt for the single-modality GP-LVM and $\mathbf{X}_0$ is fixed, meaning that we only optimize their latent space $\mathbf{X}_1$.

We observed that results were unrobust when using several kernel lengthscales values for the hierarchy, so we only used for the hierarchy the mean of the lengthscales learnt with the single modality GP-LVM.

D. Validation

Our GP-LVM strategy performs unsupervised representation learning, which is challenging to validate. In this sense, we focused on the following measurements to quantify the quality of the learnt representation.

1) Distribution of samples in the latent space: We defined five physiological quantities from the MI and MVO segmentations, which correspond to global measurements usually performed by physicians to quantify myocardial lesions (Tab.II). These labels were used to quantify the quality of the estimated latent spaces, either by qualitatively examining the distribution of samples colored by a given label (Figs. 5, 7 and 9), or by quantifying the correlation of the latent dimensions with these labels (Fig. 9). These were estimated from the spatially aligned MI and MVO segmentations, taking advantage of the estimated local coordinates. Segmental transmurality was estimated as the average of MI segmentation values over sections of $15°$, and the total *transmurality* as the average of segmental transmuralities at sections where MI is present. *Angle* corresponds to the location of the MI center of mass across the circumference, with respect to the LAD center of mass. *ESL* corresponds to the proportion of pixels where MI is present near the endocardium (concretely, for radial coordinates lower than 0.25).

2) Reconstruction of high-dimensional images: We also examined the quality of samples reconstructed from the latent space, which reflects how much relevant information from the input data is actually encoded in the latent space. Contrary to VAEs, reconstruction is not part of the method. We therefore used a multi-scale kernel ridge regression [40], [41] from the latent space to the original data. This multi-scale strategy is more robust to local changes in the density of samples in the latent space, as observed on our data.

The experiments were performed using a leave-one-patient-out strategy for the reconstruction (namely, removing all the slices from a given patient out of the training), slices containing no MVO were not evaluated.

Label designation	Value
MI	Infarct size (% of the myocardium)
MVO	MVO size (% of the myocardium)
Transmurality (TM)	Average transmural extent of the infarct (%)
Angle	Location of the infarct center across the circumference (°)
Endocardial surface length (ESL)	Infarct length on the endocardial contour (% of the endocardium)

TABLE II: Physiological labels used to quantify the relevance of the latent space (details on their estimation are given in the main text).

For our application, we specifically examined:

- The reconstruction error as the L^2 distance between the reconstructed and the original samples (Fig. 4),
- The Dice coefficient between the reconstructed and the original samples (Fig. 8), after thresholding to 0.5 and 1.5 (the healthy myocardium, MI, and MVO regions being encoded in the original images as 0, 1, and 2, respectively),
- The high-dimensional lesion patterns encoded along the directions most correlated to the above-defined labels (Fig. 10), these directions being obtained by Orthogonal Partial Least Squares (O-PLS) [42].

IV. Experiments and Results

A. Hyperparameter choice

We trained all the models described in this section with the same hyperparameters. Once the hyperparameters were fixed after observation-based training described in subsection III-C.3, we trained a GP-LVM model either on MI or MVO images for a number of latent dimensions ranging from 1 to 30, aiming at finding a suitable number of dimensions for the hierarchical representation. Figure 4 displays the reconstruction error against the number of latent dimensions.

Looking at the median curves, we observed that respectively 60% (for MI) and 74% (for MVO) of the total decrease in reconstruction error has already occurred when reaching the sixth dimension. Being the beginning of a plateau for the MVO curve, we chose 6 dimensions for each of the latent spaces in the rest of this paper. We retained the same value for MI-only, MVO-only, and hierarchical latent spaces, as 12 dimensions may favor concatenation of information from each input source instead of the fusion we aim to achieve. The numerical hyperparameters retained for all experiments in the next sections are summarized in Tab. III.

B. Distribution of samples in the latent space

We applied the hierarchical GP-LVM depicted in Fig. 3b with $\mathbf{Y}_0$ = MI and $\mathbf{Y}_1$ = MVO, and set the weight of $\mathbf{X}_0 \rightarrow \mathbf{X}_1$ and $\mathbf{X}_1 \rightarrow \mathbf{X}_0$ to $\beta = 3$. The two first dimensions of the latent space $\mathbf{X}_1$ are displayed in Fig. 5 and colored by the quantity of MI or MVO in each slice. The space is rather well structured regarding both MI and MVO, reflecting the joint integration of these two types of information. The MI and MVO size are mostly reflected along the horizontal axis (corresponding to the first dimension of the latent space), the

Hyperparameter	**GP-LVM**			
	$\mathbf{X}_0 \rightarrow \mathbf{Y}_0$	$\mathbf{X}_1 \rightarrow \mathbf{Y}_1$	$\mathbf{X}_0 \rightarrow \mathbf{X}_1$	$\mathbf{X}_1 \rightarrow \mathbf{X}_0$
ω	0.10	0.23	0.10	0.23
l	6 values ranging from 0.55 to 1.24	6 values ranging from 0.11 to 1.73	0.99	0.89
σ	9×10^{-3}	1×10^{-6}	9×10^{-3}	1×10^{-6}
β	1	1	$\{1, 3, 5\}$	$\{1, 3, 5\}$

TABLE III: Hyperparameter values used in the different GP-LVMs. The standard deviations of the Gaussian prior for $\mathbf{X}_0$ and $\mathbf{X}_1$ were also learnt and amount at 2.9 and 3.7, respectively.

Model	GP-LVM	**Dimensionality involved**	**Description**
Standard slice	-	-	Prediction from a standard slice
MI-only	✓	6	GP-LVM on MI data
MVO-only	✓	6	GP-LVM on MVO data
Concatenation	✓	3+3 = 6	Concatenation of the 3-dimensional GP-LVMs on MI and MVO
MI+MVO	✓	6	GP-LVM on the sum of MI and MVO segmentation
Naive Fusion	✓	6	Two GP-LVMs sharing the same latent space, pointing to MI and MVO observations
Hierarchy (joint training)	✓	6	Hierarchical GP-LVM with $\mathbf{X}_0$ free
Hierarchy ($\mathbf{X}_0$ fixed)	✓	6	Hierarchical GP-LVM with $\mathbf{X}_0$ fixed (our method)
MCVAE	-	6	Sparse multi-channel VAE [18]
AR-VAE	-	6	VAE on MVO data regularized by MI attributes [43]

TABLE IV: The different methods tested to predict MI and MVO labels (Tab. II) from their latent space, evaluated in Fig. 8.

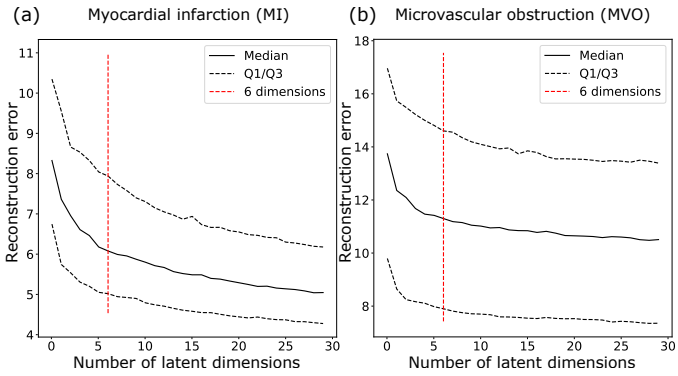

Fig. 4: Reconstruction error (median and first/third quartiles) against the number of latent dimensions for GP-LVMs on (a) MI-only and (b) MVO-only. The vertical red line indicates the number of dimensions retained.

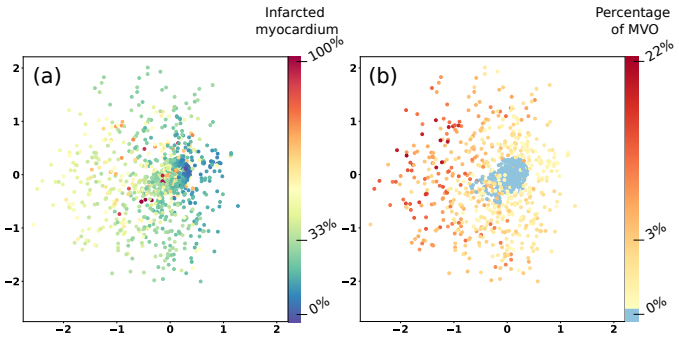

Fig. 5: First two latent dimensions for the hierarchical GP-LVM. The color code represents the amount of (a) MI and (b) MVO in each slice.

smallest infarcts being near the origin. This central part also corresponds to the slices without (or with extremely small) MVO. This figure is instructive but incomplete as we only display the first two dimensions of a 6-dimensional space. For instance, the almost entirely infarcted slices (in dark red) seem to be (mistakenly) spread around the central area of the space. However, these similar patterns are indeed grouped in the latent space, examining the sixth dimension confirms the relevant distribution of these samples against the others, as displayed in Figure 6. Note also that the color code only reflects the total amount of MI or MVO in each slice, while these patterns actually have more complex shapes (see Fig. 10).

We compared these results to those obtained from a single-modality GP-LVM trained either on MI or MVO segmentations independently, whose two first dimensions of each latent space are displayed in Fig. 7. At first sight, these two spaces represent well the type of data they respectively encode, with MI and MVO size smoothly arranged across these spaces. However, they have several defects:

- MVO samples may be side-by-side, while they shouldn't if they correspond to rather different MI patterns,
- Besides, as there is not an infinite amount of samples, several pairs of MVO patterns that do not overlap may be associated to the same distance, and (wrongly) mapped to close locations,
- Finally, slices with low or no MVO are all mapped to the same point at the center of the latent space (light blue point on the right part of Fig. 7b), without considering the associated MI pattern they may contain.

Focusing only on MVO without considering MI with a single-modality GP-LVM delivers an incomplete and therefore wrong representation of the lesions across the population, while building a hierarchy between MI (first level) and MVO (second level) allows overcoming the above-listed issues.

C. Accuracy of reconstructed samples

We complemented the qualitative observations of the latent spaces by quantifying the accuracy of reconstructed samples,

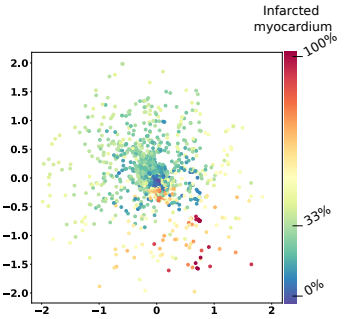

Fig. 6: The last two (5th and 6th) latent dimensions for the hierarchical GP-LVM. The color code represents the amount of MI. We observe that dark red points (i.e. biggest infarction patterns) are located on the same area of the latent space.

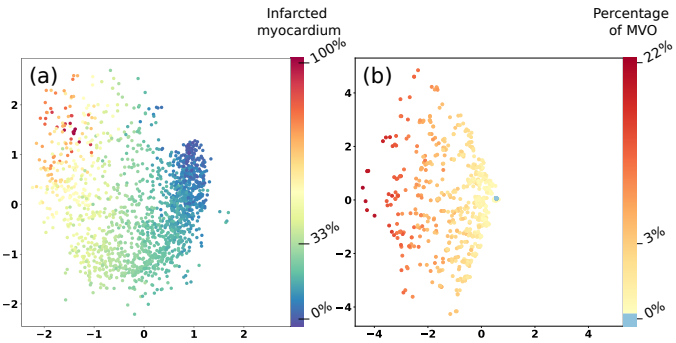

Fig. 7: First two latent dimensions for GP-LVMs on (a) MI-only and (b) MVO-only training. The color code is similar to Fig. 5) and corresponds to the amount of MI or MVO.

using the Dice coefficient, as explained in Sec. III-D.2. In particular, we compared our model with several variants listed in Tab. IV, the results for the Dice being displayed as boxplots in Fig. 8.

The figure also quantifies the accuracy of using a single representative slice for the MI and MVO segmentations of all subjects. For MI, this single slice corresponded to the pixel-wise median of all MI patterns. For MVO, this computation is not possible as these lesions are very small and do not often overlap. We therefore computed the pixel-wise incidence of MVO and thresholded this map at 20%, which resembled a representative MVO pattern.

Finally, we also evaluated two recent variants of VAEs for which code was publicly available:

- the sparse multi-channel VAEs (MCVAE) [18], which perform latent space alignment, with MI and MVO segmentations as input. They provide one latent code for each input, and reconstruct both MI and MVO for each of these latent codes.
- the attribute-based regularized VAEs (ARVAE) [43], which MVO segmentation as input and MI attributes (size, transmurality, and angle / see Tab. II) to regularize a given dimension of the latent space.

For both methods, the latent dimensionality was set to 6. Besides, given the polar geometry of the myocardium, we did not directly use the convolution-based encoders and decoders on the data in Cartesian coordinates. Instead, we unfolded the myocardium according to the radial and circumferential coordinates, leading to input images of 128×16 pixels. The VAEs consisted of 4 layers: 3 convolutional layers and 1 fully connected layer. The second convolutional layer had a *MaxPooling* of size $(2, 1)$ and the decoder architecture mirrored the encoder architecture.

Hierarchical approaches, with or without fixing $\mathbf{X}_0$, provided decent results for predicting MI and MVO segmentations for each slices. The different weight values for $\mathbf{X}_0 \rightarrow \mathbf{X}_1$ and $\mathbf{X}_1 \rightarrow \mathbf{X}_0$ offer different balances between MI and MVO. The latent space displayed in Fig. 5 corresponds to the second red boxplots in Fig. 8. Three main observations can be done from Fig. 8:

- The hierarchical spaces' results are close to the ones of the 6-dimensional spaces learnt on a single modality, while being able to predict both modalities,
- The other models tested here (including the VAEs extensions) failed to produce a joint representation of MI and MVO, and instead encoded either MI or MVO,,
- The strategy where $\mathbf{X}_0$ is not fixed slightly outperforms the one with $\mathbf{X}_0$ fixed, as it is more flexible.

The MCVAE and AR-VAE mapped all slices with no MVO to the same point, as for the single-modality GP-LVM on MVO data. By doing so, they are able to optimally reconstruct MVO data but they do not lead to a proper alignment of MVO and MI latent coordinates, and the MI reconstructed from the MVO latent coordinates correspond to an average MI pattern.

Of note, the encoder and decoder in the VAEs use convolutions, meaning that they better consider the image structure than our method, which still reaches better performance.

We did similar comparisons with the Hausdorff distance, which brings similar conclusions (not displayed here for the sake of brevity). Of note, the Hausdorff distance cannot be computed in case the reconstructed slices do not contain MVO, which happens to a much lower proportion for the hierarchical models (8% in arevage) compared to the fusion (31%) and single modality (39.5%) strategies.

D. Shuffling the first data level

We conducted an additional experiment to confirm that the better encoding achieved with the hierarchical framework was due to the existing link between the two modalities, and not to a too strong prior. To do so, $\mathbf{X}_0$ was shuffled, breaking any association between the two levels, thereby challenging the initial assumption. We estimated three latent spaces with different random rearrangements of the first modality, using exactly the same parameters as our reference hierarchical space ($\beta = 3$). Figures 9a and b display the 2-dimensional projection of one of the rearranged latent spaces, with respect to the quantity of MI and MVO. A similar arrangement is observed for the other two shuffled spaces (not displayed here). This means that the MVO quantity was learnt and that the MI pattern was respected only when there was no MVO in the slice, producing a messy space when looking away from the center. Figure 9c complements this by showing the correlations between each of the first 6 latent dimensions and the (shuffled) MI and MVO quantities.

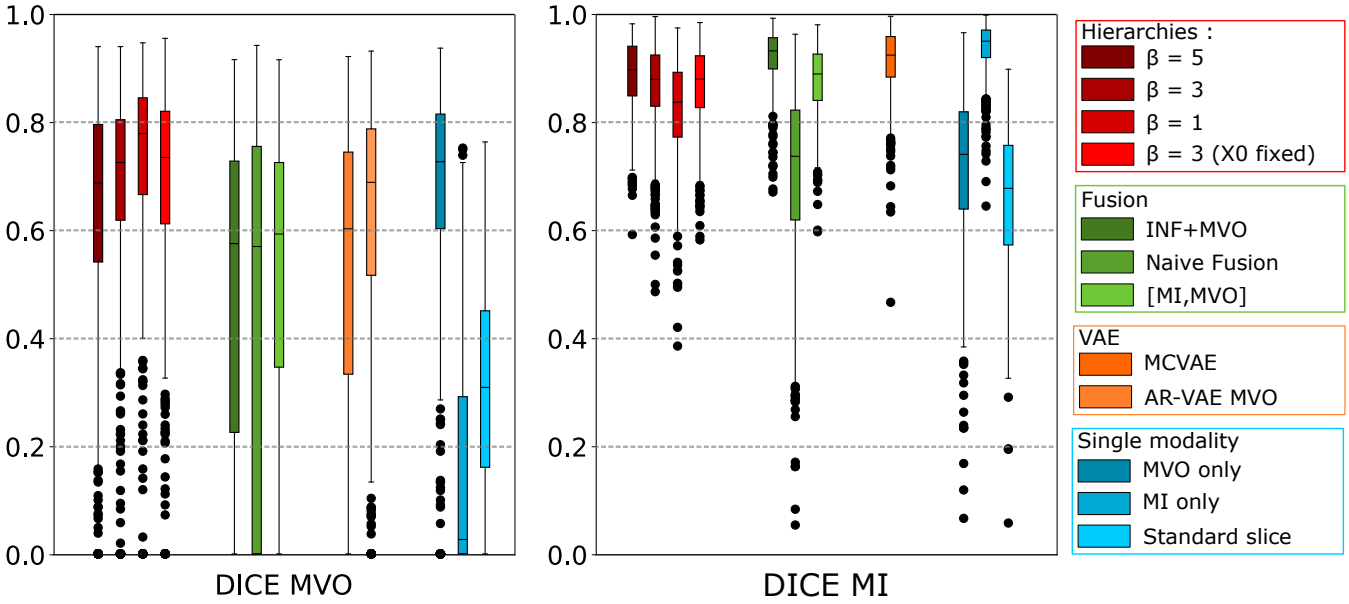

Fig. 8: Dice coefficient from the reconstruction of MVO (a) and MI (b) from a given latent space (see list of methods in Tab. IV).

Correlations are stronger globally for both modalities in the hierarchical space, and the impact of the fusion is significant. This is particularly true for MVO, showing that this information has been better learnt with the support of a structured hierarchical link. Even when the values are weak, the dimension correlates with both MI and MVO at the same time, indicating that the modalities are learnt jointly and that the hierarchical link has a real impact on the final representation for both quantities. The correlation map also supports the qualitative observation that the shuffled spaces exhibit a similar behavior overall. In all three shuffled spaces, the last dimensions are highly correlated with MI quantity, while MVO is correlated with the first dimension.

E. Physiological consistency of the latent space

Finally, we used the O-PLS algorithm to browse the latent space along the direction that best encodes a given physiological label (see list of labels in Tab. II). We reconstructed high-dimensional images corresponding to synthetic points sampled in a range of $+/-2\sigma$ along these directions, σ standing for the standard deviation of the projection of the latent space on such direction. For comparison purposes, we performed the same experiment for the concatenation of the 3-dimensional MVO and MI latent spaces.

The reconstruction from the hierarchical space (see Fig. 10) correctly reflects the label each direction is supposed to encode. This is particularly well rendered for ESL: the slice at $+2\sigma$ presents a large MI, very wide near the endocardium. In contrast, the reconstructions from the concatenation latent space for MI, transmurality and ESL all exhibit a similar behavior not specifically corresponding to the desired label.

We complemented these qualitative observations by quantifying the correlations between the projection of the latent coordinates along each attribute direction estimated with O-PLS, and the corresponding physiological labels (Fig. 10e and f). The figure confirms that the hierarchical space is better structured according to these physiological labels: projection

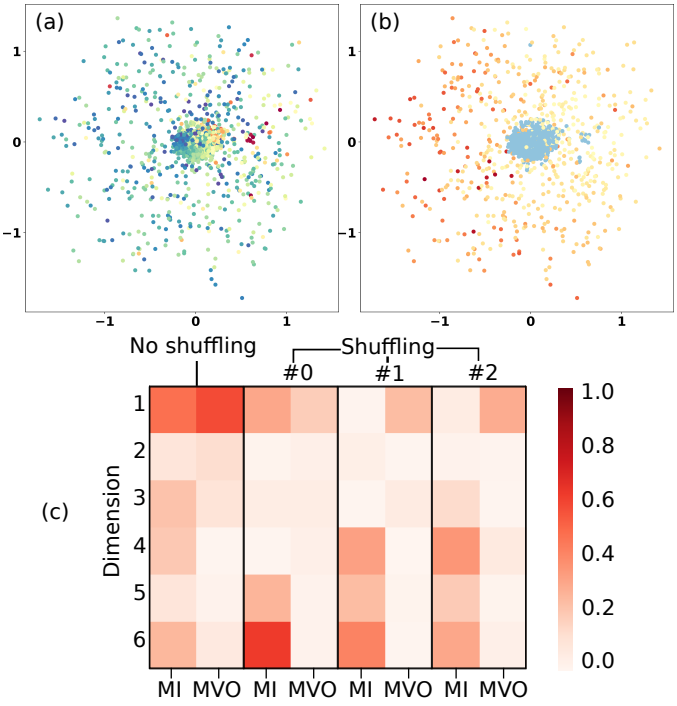

Fig. 9: Hierarchical latent space obtained when the first modality has been shuffled, breaking the hierarchical prior. (a) and (b) Estimated latent space colored either by MI or MVO. (c) Correlation between each dimension of the latent space and these two physiological labels, without (left) / with shuffling (three rightmost columns). Breaking the hierarchical prior leads to a decorrelation between the two labels and some redundancy across the dimensions, emphasizing the relevance of the hierarchical link in the data.

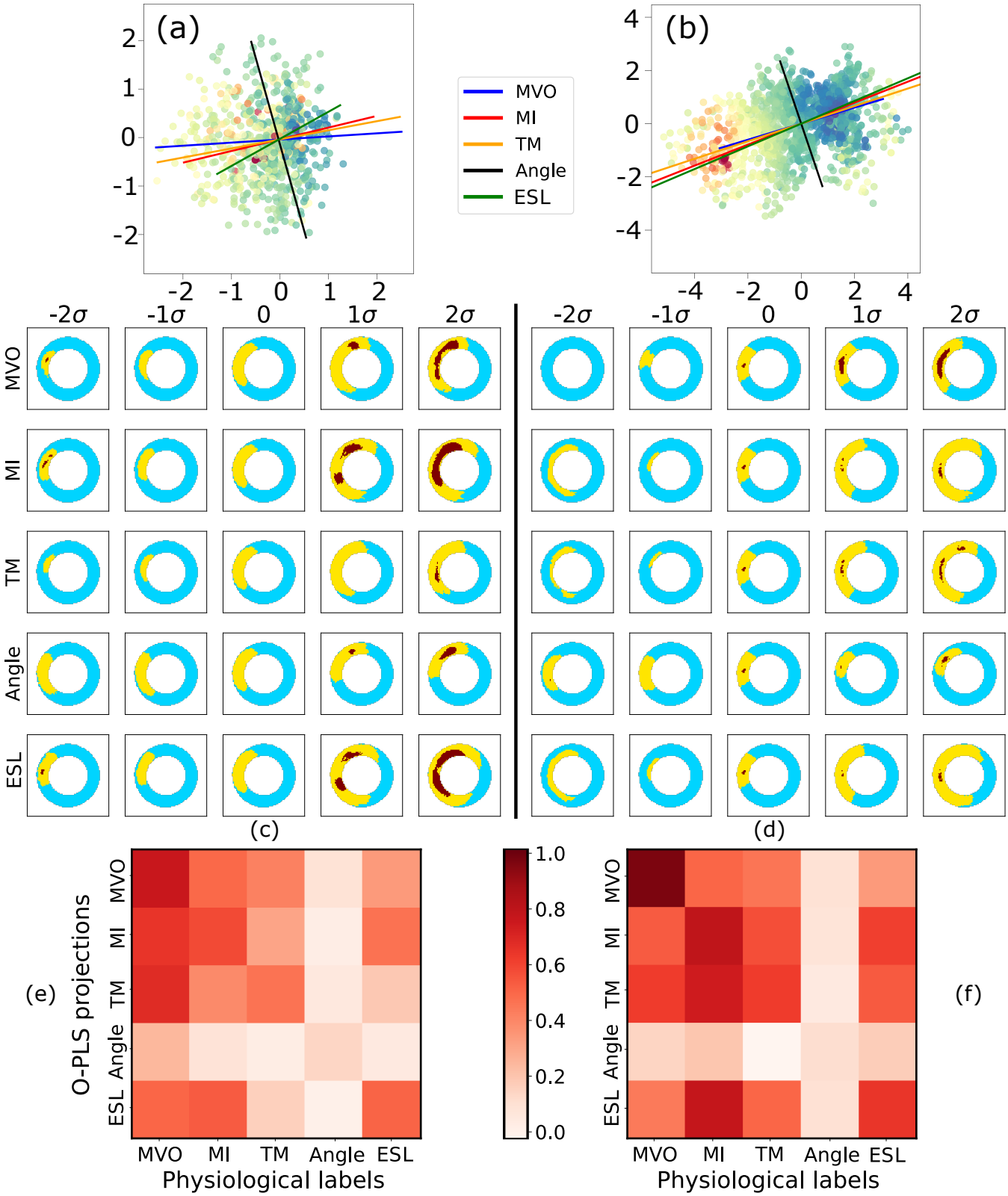

Fig. 10: Consistency of the latent space according to the physiological labels listed in Tab. II. (a) and (b) Two-dimensional projections of the hierarchical (left) and concatenation (right) latent spaces, colored by the MVO size. The colored straight lines stand for the main directions of the five physiological labels, estimated with O-PLS. (c) and (d) Reconstruction of high-dimensional images obtained by sampling along these axes, for the hierarchical and the concatenation spaces, respectively. The hierarchical space is able to reconstruct specific patterns from the labels, contrary to the concatenation space which does not distinguish between MI (2nd row), transmurality (3rd row) and ESL (5th row). (e) and (f) Correlations between the projection of the latent coordinates along each attribute direction estimated with O-PLS, and the corresponding physiological labels.

along the angle is mostly independent from the other labels, and each of the other projections is rather correlated to its corresponding label and related labels (given the existing correlations between these values, e.g. ESL is naturally related to MI); in contrast, the concatenation latent spaces shows limited representation of the angle, and higher correlations between each projection and all the other labels.

V. Discussion and conclusions

In this paper, we presented an original strategy for the hierarchical integration of high-dimensional descriptors extracted from medical images, potentially of heterogeneous types, in the context of unsupervised representation learning. The method is hierarchical, meaning that former levels in the (imposed) hierarchy of descriptors are used to guide the representation of next levels (potentially associated to more complex descriptors). More specifically, our method is based on GP-LVMs, and extends this framework to build a two-level hierarchy with specific considerations on the mutual link between the two levels. We demonstrated its relevance on real cardiac imaging data from delayed enhancement MRI, with the objective of better representing the distribution of lesion patterns related to acute infarction and reperfusion across a population. Our model was able to learn a relevant latent low-dimensional representation of these data, guided by a former level of hierarchy, allowing statistical analysis of the higher level data.

The two variants (with and without a fixed former level) of the proposed hierarchical model outperformed single-modality and fusion models on all evaluated aspects: the distribution of the latent space, its representation power, and the consistency with physiological labels used in clinical routine. The joint training model allows a better circulation of information between the two latent spaces $\mathbf{X}_0$ and $\mathbf{X}_1$, explaining slightly better representation power. For this model, the two latent spaces ($\mathbf{X}_0$ and $\mathbf{X}_1$) are different trade-offs between the representation of the observations $\mathbf{Y}_0$ and $\mathbf{Y}_1$. The hierarchical model with fixed $\mathbf{X}_0$ is interesting to mimic the way clinicians progressively integrate different levels of information: in our application, we would like to estimate a representation for MVO that benefits from former-level information (MI), but not to challenge the original MI representation. This is also a substantial asset when dealing with larger amounts of data and more hierarchical levels, which can be added to the hierarchy without having to retrain the whole model.

Our model is adaptable but this flexibility comes at a price: on real world data, it can be difficult to find the balance between each modality. In practice, if we let the model parameters free, the latent space is dominated either by the MI or MVO representations and fails to learn a decent trade-off between them. Besides, GP-LVMs provide an intuitive manner of representing the desired hierarchy, but they are challenging to optimize. To better control these aspects, we opted for fixing the hyperparameters before learning the embedding. In theory, these hyperparameters can be learnt by gradient descent, but this may require the design of an additional regularization term to constrain the trade-off between modalities. Nonetheless, the adaptability of the GP-LVM framework allows learning from much more complex data spaces and hierarchies. The major obstacle is then to draw the hierarchical relationships between numerous modalities. In this work, we chose a simple established hierarchy related to the data complexity to focus on its implementation with GP-LVMs. An alternative would be to learn the hierarchical order at the same time the embedding is learnt, which, although appealing, may be very challenging within this framework, but has been demonstrated with reinforcement learning for active modality selection on simpler data and for supervised problems [44].

Unsupervised learning is interesting to characterize the distribution of a population from complex descriptors, as performed in this study (characterizing specific lesion patterns from imaging data that clinicians over-simplify). Unsupervised methods are challenging to validate, given the lack of labels. In our context, we performed a thorough series of experiments to examine both the latent space and reconstructed data or physiological labels, which allowed understanding the behavior of the hierarchical model and demonstrated its better performance for our task. GP-LVMs can also be designed for supervised problems, meaning that our methodology could rather directly be applied to such problems.

Ischemia-reperfusion mechanisms are complex and evolve with follow-up. In this paper, we focused on a portion of the imaging data used by clinicians to characterize these patterns, namely LGE MRI segmentations, which reflect the shape of lesions from perfusion-related images. Although these segmentations were obtained by semi-automatic software and carefully controlled by experts, expanding this work to larger datasets would benefit from very recent developments in deep learning-based segmentation of MI and MVO from LGE images [45], [46]. A more complete characterization may also incorporate image appearance (and not just segmentations), and other imaging data related to myocardial tissue content (T1 and T2 images) and deformation (CINE or DENSE sequences), to be addressed in future work on larger cohorts with longitudinal data available.

Acknowledgments

The authors are grateful to the H2P platform[1] (CREATIS, Villeurbanne, France) for the data storage and management support, to Marco Lorenzi (INRIA Sophia-Antipolis, France) for preliminary discussion on Gaussian processes, and to Pierre Croisille, Magalie Viallon, and Lorena Petrusca (CREATIS, CHU Saint Etienne, France) for providing the imaging data and segmentations for the MIMI population.

References

[1] Sanchez-Martinez, S and Camara, O and Piella, G and Cikes, M and González-Ballester, MA and Miron, M and others, "Machine learning for clinical decision-making: Challenges and opportunities in cardiovascular imaging," *Front Cardiovasc Med*, vol. 8, p. 765693, 2022.

[2] Shekelle, PG, "Clinical practice guidelines: what's next?" *JAMA*, vol. 320, pp. 757–8, 2018.

[3] Weese, J and Lorenz, C, "Four challenges in medical image analysis from an industrial perspective," *Med Image Anal*, vol. 33, pp. 44–9, 2016.

[1] `http://humanheart-project.creatis.insa-lyon.fr/`

[4] Elman, JL, "Learning and development in neural networks: the importance of starting small," *Cognition*, vol. 48, pp. 71–99, 1993.

[5] B. Freiche, P. Clarysse, M. Viallon *et al.*, "Characterizing myocardial ischemia and reperfusion patterns with hierarchical manifold learning," *Proc. STACOM-MICCAI, LNCS*, pp. 66–74, 2021.

[6] N. Lawrence and A. Moore, "Hierarchical Gaussian process latent variable models," 2007.

[7] S. Yan, D. Xu, B. Zhang *et al.*, "Graph embedding and extensions: A general framework for dimensionality reduction," *IEEE Trans Pattern Anal Mach Intell*, vol. 29, pp. 40–51, 2007.

[8] J. Tenenbaum, V. Silva, and J. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, pp. 2319–23, 2000.

[9] R. Coifman and S. Lafon, "Diffusion maps," *Appl Comput Harmon Anal*, vol. 21, pp. 5–30, 2006.

[10] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *Proc. ICLR*, 2014.

[11] X. Liu, P. Sanchez, S. Thermos *et al.*, "Learning disentangled representations in the imaging domain," *Med Image Anal*, vol. 80, p. 102516, 2022.

[12] F. Leeb, S. Bauer, M. Besserve *et al.*, "Exploring the latent space of autoencoders with interventional assays," *Proc. NeurIPS*, vol. 35, pp. 21 562–74, 2022.

[13] N. Lawrence, "Gaussian process latent variable models for visualisation of high dimensional data," *Proc. NeurIPS*, 2003.

[14] A. Damianou and N. Lawrence, "Deep Gaussian processes," *Proc. AISTATS*, vol. 31, pp. 207–15, 2013.

[15] Y. Li, M. Yang, and Z. Zhang, "A survey on multi-view representation learning," *IEEE Trans Knowl Data Eng*, vol. 31, pp. 1863–83, 2019.

[16] Y. Lin, T. Liu, and C. Fuh, "Multiple kernel learning for dimensionality reduction," *IEEE Trans Pattern Anal Mach Intell*, vol. 33, pp. 1147–60, 2010.

[17] B. Wang, A. Mezlini, F. Demir *et al.*, "Similarity network fusion for aggregating data types on a genomic scale," *Nat Methods*, vol. 11, pp. 333–7, 2014.

[18] L. Antelmi, N. Ayache, P. Robert *et al.*, "Sparse multi-channel variational autoencoder for the joint analysis of heterogeneous data," *Proc. ICML*, pp. 302–11, 2019.

[19] Y. Shi, N. Siddharth, B. Paige *et al.*, "Variational mixture-of-experts autoencoders for multi-modal deep generative models," *Proc. NeurIPS*, pp. 15 718–29, 2019.

[20] A. Aguila, J. Chapman, and A. Altmann, "Multi-modal variational autoencoders for normative modelling across multiple imaging modalities," *Proc. MICCAI, LNCS*, vol. 14220, pp. 425–34, 2023.

[21] C. Ma, S. Tschiatschek, K. Palla *et al.*, "EDDI: efficient dynamic discovery of high-value information with partial VAE," *Proc. ICLR*, 2019.

[22] J. Valencia-Aguirre, A. Álvarez Meza, G. Daza-Santacoloma *et al.*, "Multiple manifold learning by nonlinear dimensionality reduction," *Proc. CIARP, LNCS*, vol. 7042, pp. 206–13, 2011.

[23] J. Clough, D. Balfour, G. Cruz *et al.*, "Weighted manifold alignment using wave kernel signatures for aligning medical image datasets," *IEEE Trans Pattern Anal Mach Intell*, vol. 42, pp. 988–97, 2020.

[24] S. Sanchez-Martinez, N. Duchateau, T. Erdei *et al.*, "Characterization of myocardial motion patterns by unsupervised multiple kernel learning," *Med Image Anal*, vol. 35, pp. 70–82, 2017.

[25] M. Di Folco, P. Moceri, P. Clarysse *et al.*, "Characterizing interactions between cardiac shape and deformation by non-linear manifold learning," *Med Image Anal*, vol. 75, p. 102278, 2022.

[26] P. Zhou, Y. Shen, L. Du *et al.*, "Incremental multi-view spectral clustering," *Knowl Based Syst*, vol. 174, pp. 73–86, 2019.

[27] H. Yin, W. Hu, Z. Zhang *et al.*, "Incremental multi-view spectral clustering with sparse and connected graph learning," *Neural Network*, vol. 144, pp. 260–70, 2021.

[28] Y. Bengio, J. Louradour, R. Collobert *et al.*, "Curriculum learning," *Proc. ICML*, pp. 41–8, 2009.

[29] K. Bhatia, A. Rao, A. Price *et al.*, "Hierarchical manifold learning for regional image analysis," *IEEE Trans Med Imaging*, vol. 33, pp. 444–61, 2014.

[30] G. Heusch, "Coronary microvascular obstruction: the new frontier in cardioprotection," *Basic Res Cardiol*, vol. 114, pp. 6–45, 2019.

[31] M. van Kranenburg, M. Magro, H. Thiele *et al.*, "Prognostic value of microvascular obstruction and infarct size, as measured by CMR in STEMI patients," *JACC Cardiovasc Imaging*, vol. 7, pp. 930–9, 2014.

[32] H. Bulluck, R. Dharmakumar, A. Arai *et al.*, "Cardiovascular magnetic resonance in acute st-segment–elevation myocardial infarction: recent advances, controversies, and future directions," *Circulation*, vol. 137, pp. 1949–64, 2018.

[33] C. Beijnink, N. van der Hoeven, L. Konijnenberg *et al.*, "Cardiac MRI to visualize myocardial damage after ST-segment elevation myocardial infarction: A review of its histologic validation," *Radiology*, vol. 301, pp. 4–18, 2021.

[34] S. Bekkers, S. Yazdani, R. Virmani *et al.*, "Microvascular obstruction: underlying pathophysiology and clinical diagnosis," *J Am Coll Cardiol*, vol. 20, pp. 1649–60, 2010.

[35] L. Belle, P. Motreff, L. Mangin *et al.*, "Comparison of immediate with delayed stenting using the Minimalist Immediate Mechanical Intervention approach in acute ST-segment-elevation myocardial infarction: the MIMI study," *Circ Cardiovasc Interv*, vol. 9, p. e003388, 2016.

[36] N. Duchateau, M. Viallon, L. Petrusca *et al.*, "Pixel-wise statistical analysis of myocardial injury in STEMI patients with delayed enhancement MRI," *Front Cardiovasc Med*, p. In press, 2023.

[37] E. Bonilla, K. Chai, and C. Williams, "Multi-task Gaussian process prediction," *Proc. NeurIPS*, vol. 20, 2007.

[38] M. Lorenzi, G. Ziegler, D. Alexander, and S. Ourselin, "Efficient Gaussian process-based modelling and prediction of image time series," *Proc. IPMI*, vol. 24, pp. 626–37, 2015.

[39] A. Matthews, M. van der Wilk, T. Nickson *et al.*, "GPflow: A Gaussian Process Library using TensorFlow ," *J Mach Learn Res*, vol. 18, pp. 1–6, 2017.

[40] A. Bermanis, A. Averbuch, and R. Coifman, "Multiscale data sampling and function extension," *Appl Comput Harmon Anal*, vol. 34, pp. 15–29, 2013.

[41] N. Duchateau, M. De Craene, M. Sitges *et al.*, "Adaptation of multiscale function extension to inexact matching. application to the mapping of individuals to a learnt manifold," *Proc. GSI, LNCS*, vol. 8085, pp. 578–86, 2013.

[42] J. Trygg and S. Wold, "Orthogonal projection to latent structures," *J Chemometr*, vol. 16, pp. 119–28, 2002.

[43] A. Pati and A. Lerch, "Attribute-based regularization of latent spaces for variational auto-encoders," *Neural Comput Appl*, pp. 4429–44, 2021.

[44] G. Bernardino, A. Jonsson, F. Loncaric *et al.*, "Reinforcement learning for active modality selection during diagnosis," *Proc. MICCAI, LNCS*, vol. 13431, pp. 592–601, 2022.

[45] Lalande, A and Chen, Z and Pommier, T and Decourselle, T and Qayyum, A and Salomon, M and others, "Deep learning methods for automatic evaluation of delayed enhancement-MRI. The results of the EMIDEC challenge," *Med Image Anal*, vol. 79, p. 102428, 2022.

[46] Viallon, M and Croisille, P and Bernard, O and Duchateau, N and Clarysse, P, "Myocardial segmentation with automated infarct quantification (MYOSAIQ) challenge," https://www.creatis.insa-lyon.fr/Challenge/myosaiq/, 2023.