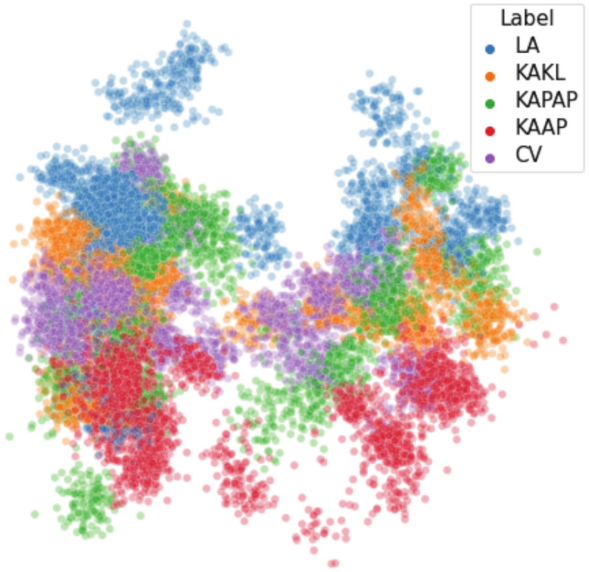
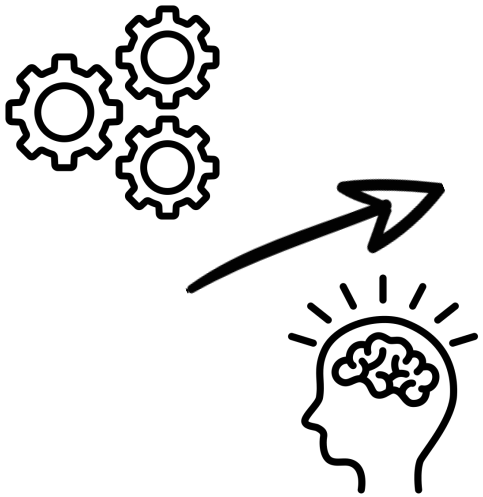
ETH zürich

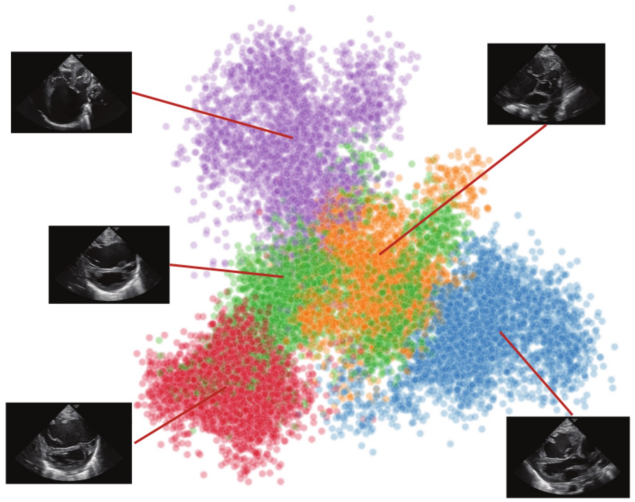
D INFK

medical data science

Informed Representation Learning with Deep Generative Models







ALPACA Workshop

22/11/2023

Laura Manduchi

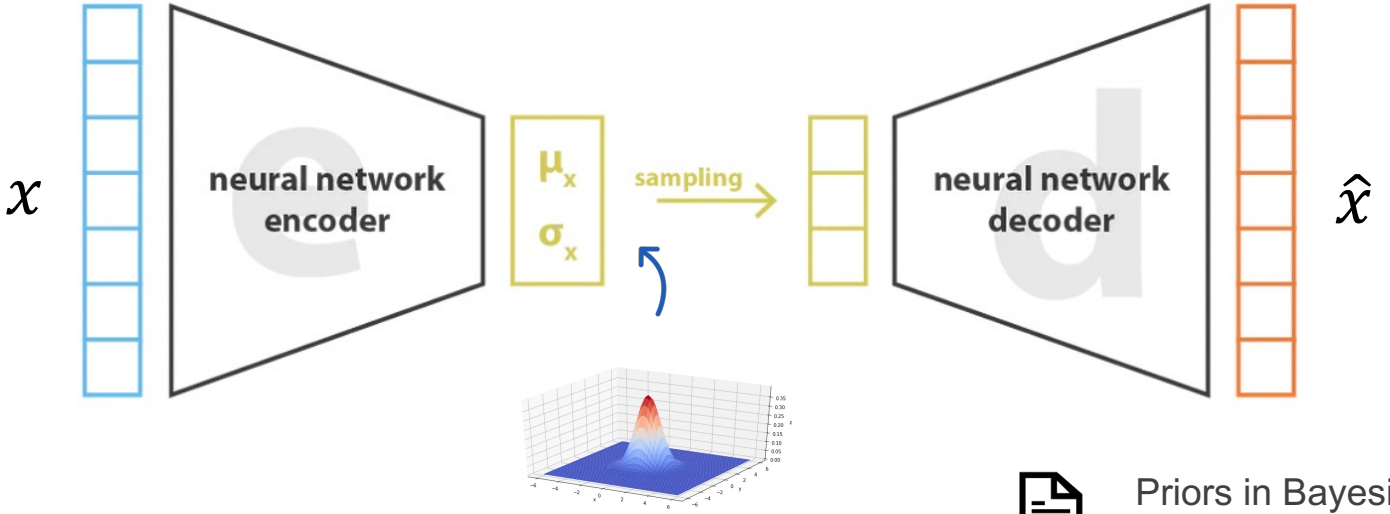
PhD fellow at ETH Zürich, Department of Computer Science

Representation Learning

- The process of **extracting meaningful patterns** from raw data to create representations that are;
  - Easier to visualize, understand, and process
  - Useful for subsequent learning task

Informed Representations of Data

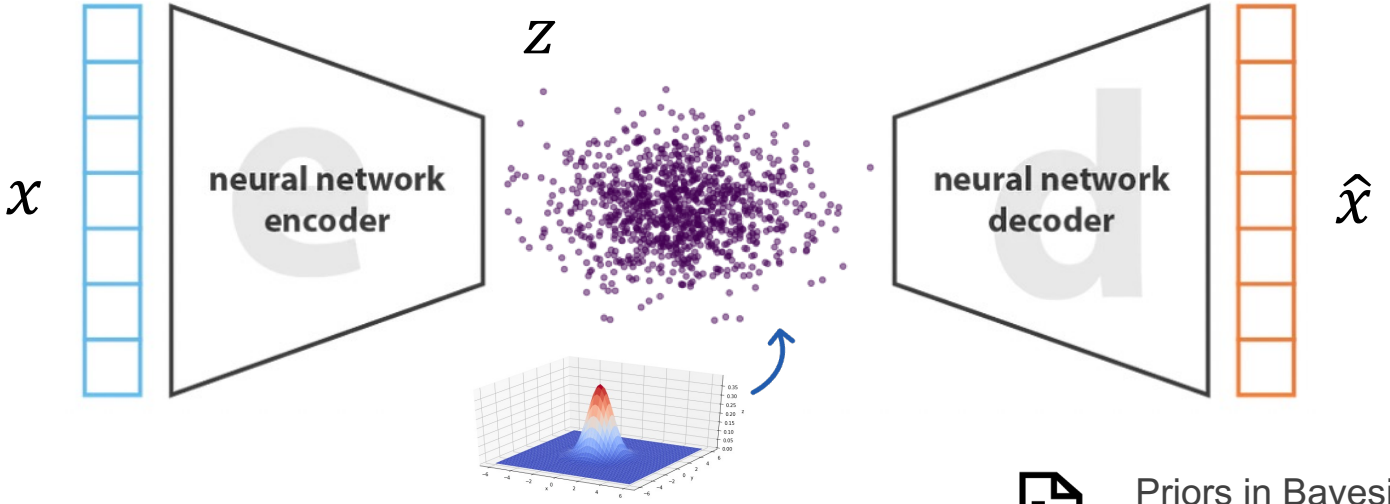
- Purely **data-driven approaches** requires a sufficiently diverse training data.
- **No-free-lunch theorem**: no universal representations for all tasks.
- How can we **integrate prior belief** ? → Bayesian latent variable models (VAEs)

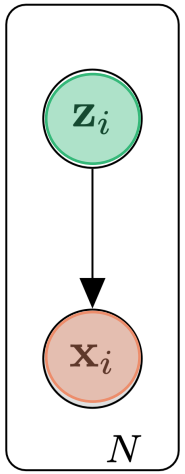


Priors in Bayesian Deep Learning: A Review. International Statistical Review 2021. F. Vincent

Informed Representations of Data

- Purely **data-driven approaches** requires a sufficiently diverse training data.
- **No-free-lunch theorem**: no universal representations for all tasks.
- How can we **integrate prior belief** ? → Bayesian latent variable models (VAEs)





$$\boldsymbol{z}_i \sim p(\boldsymbol{z}_i) = \mathcal{N}(\mathbf{0}, \mathbb{I})$$

$$\boldsymbol{x}_i \sim p_\theta(\boldsymbol{x}_i|\boldsymbol{z}_i) = \mathcal{N}(\boldsymbol{x}_i|\boldsymbol{\mu}_{\boldsymbol{x}_i}, \boldsymbol{\sigma}^2_{\boldsymbol{x}_i}\mathbb{I})$$

Priors in Bayesian Deep Learning: A Review. International Statistical Review 2021. F. Vincent

Informed Representations of Data

Prior & Probabilistic relations

Clustered representations:

- Categorical prior
- Hypergeometric prior
- Hierarchical prior

Physical constraints

- Periodicity in temporal data.
  - Variational Trajectory Models
  - Personalized Heart Meshes
- X-Ray Spectra Fitting

Weak supervision

- Similarity between samples
- Survival information of patients

Informed Representations of Data

Prior & Probabilistic relations

Clustered representations:

- Categorical prior
- Hypergeometric prior
- Hierarchical prior

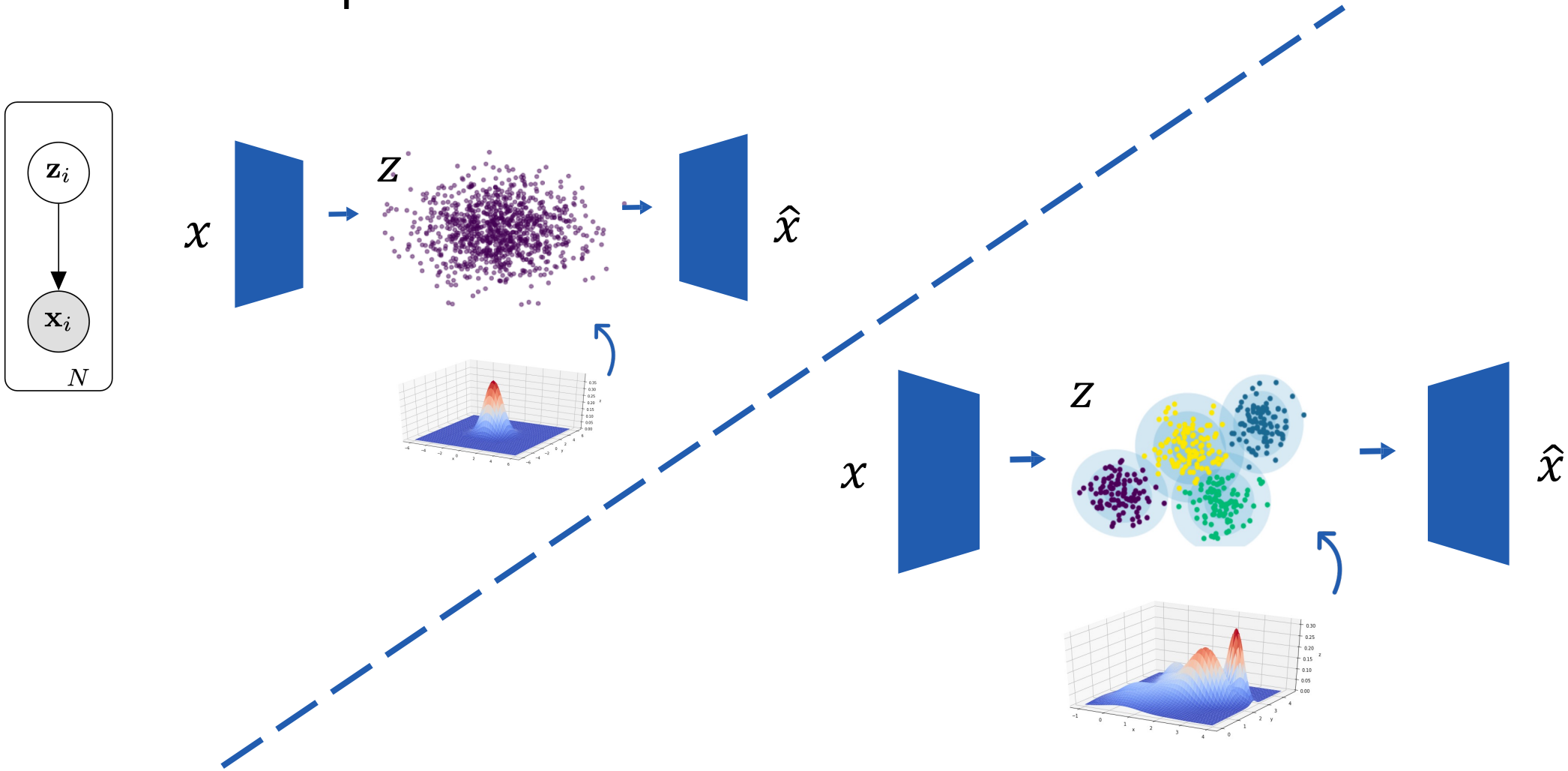
Physical constraints

- Periodicity in temporal data.
  - Variational Trajectory Models
  - Personalized Heart Meshes
- X-Ray Spectra Fitting

Weak supervision

- Similarity between samples
- Survival information of patients

Clustered Representations with VAEs

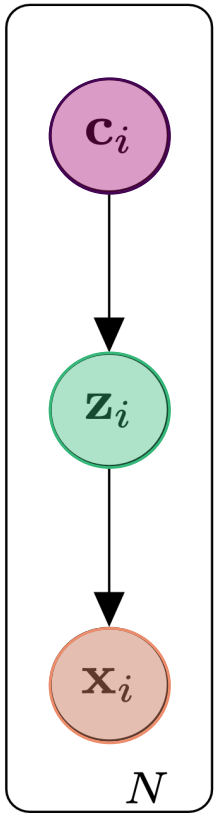


Gaussian mixture prior

Variational Deep Embedding: A Generative Approach to Clustering. IJCAI 2017. Jiang et al.

Clustered Representations with VAEs

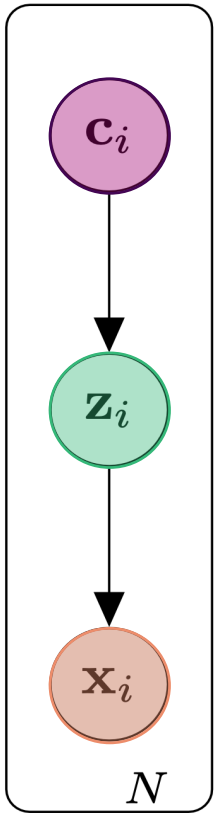
1. Cluster assignments: $\mathrm{c_i} \sim p(\mathrm{c_i})$
2. Latent embedding: $\boldsymbol{z}_i \sim p(\boldsymbol{z}_i|\mathrm{c_i}) = \mathcal{N}(\boldsymbol{z}_i|\boldsymbol{\mu}_{c_i}, \boldsymbol{\sigma}^2_{c_i}\mathbb{I})$
3. Explanatory raw variables : $\boldsymbol{x}_i \sim p_\theta(\boldsymbol{x}_i|\boldsymbol{z}_i) = \mathcal{N}(\boldsymbol{x}_i|\boldsymbol{\mu}_{x_i}, \boldsymbol{\sigma}^2_{x_i}\mathbb{I})$



Variational Deep Embedding: A Generative Approach to Clustering. IJCAI 2017. Jiang et al.

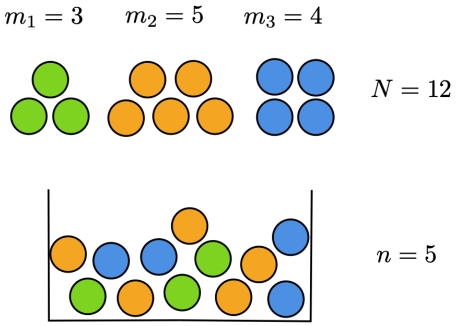
Clustered Representations with VAEs

1. Cluster assignments: $\mathrm{c_i} \sim p(\mathrm{c_i})$
2. Latent embedding: $\boldsymbol{z}_i \sim p(\boldsymbol{z}_i|\mathrm{c_i}) = \mathcal{N}(\boldsymbol{z}_i|\boldsymbol{\mu}_{c_i}, \boldsymbol{\sigma}^2_{c_i}\mathbb{I})$
3. Explanatory raw variables : $\boldsymbol{x}_i \sim p_\theta(\boldsymbol{x}_i|\boldsymbol{z}_i) = \mathcal{N}(\boldsymbol{x}_i|\boldsymbol{\mu}_{x_i}, \boldsymbol{\sigma}^2_{x_i}\mathbb{I})$



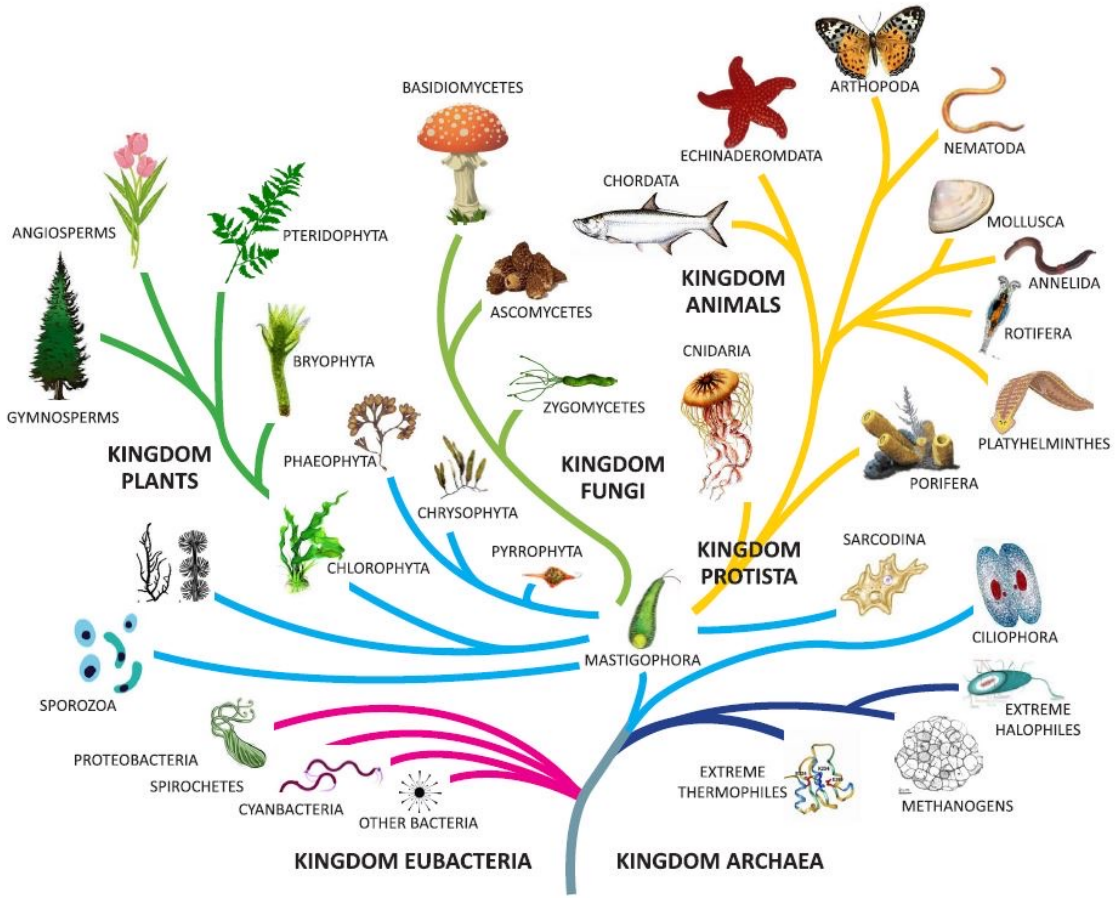
→ **Hypergeometric distribution** to model dependence between samples.

$$\boldsymbol{c} \sim p(\boldsymbol{c}) = \frac{1}{P_0}\prod_i^K \binom{m_i}{y_i} w_i^{x_i}, \ \mathbf{c} \in \mathbb{R}^N$$

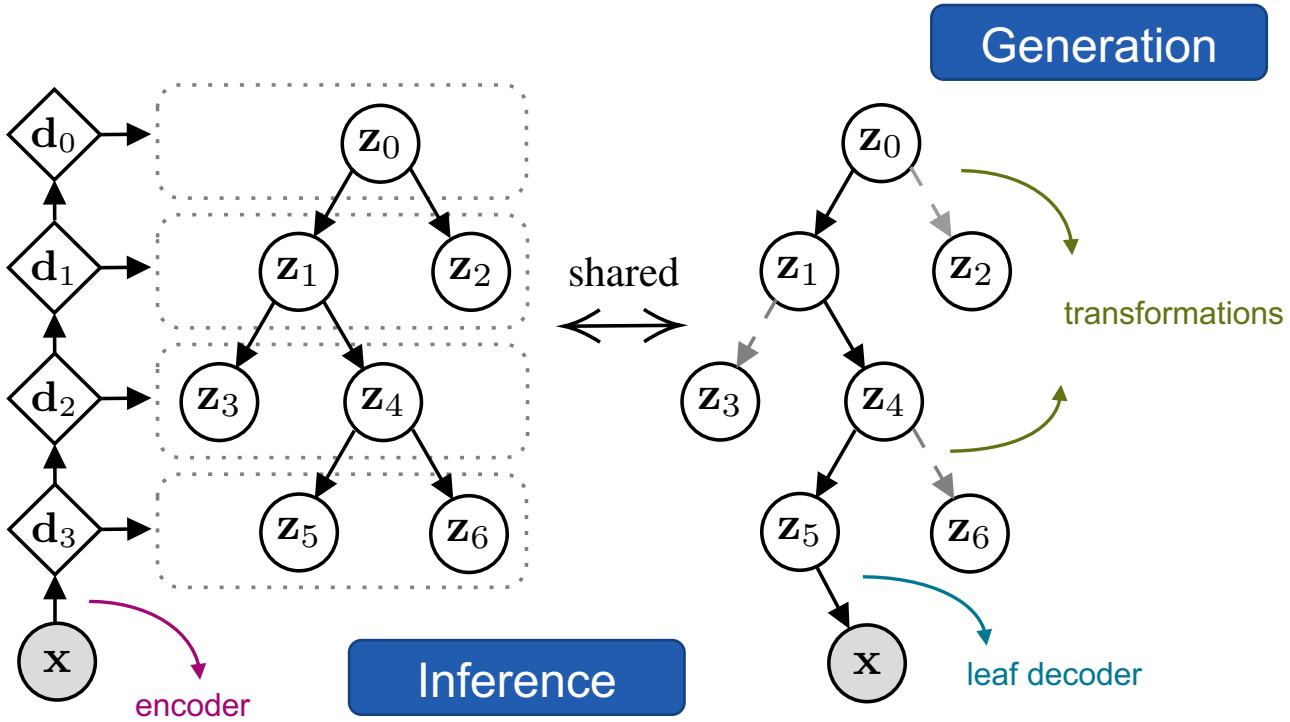


Beyond Flat Representations

- The human mind perceived the world in hierarchical structure.
- Detailed information on the level of similarity of data.
- Nested relationship between the data.



Tree Variational Autoencoders



- Variational autoencoder (VAE) with tree-based posterior.
- Learnt binary tree structure $\mathcal{T}$ with nodes $\mathbb{V}$ and edges $\mathcal{E}$.
- *Sample-specific* latent embeddings $\boldsymbol{z} = \{\boldsymbol{z}_0, \dots, \boldsymbol{z}_V\}$
  - gaussian distributions conditioned on the parent.
- *Sample-specific* decision paths $\mathcal{P}_l$
  - each decision is modeled by a Bernoulli distribution.

Optimization:

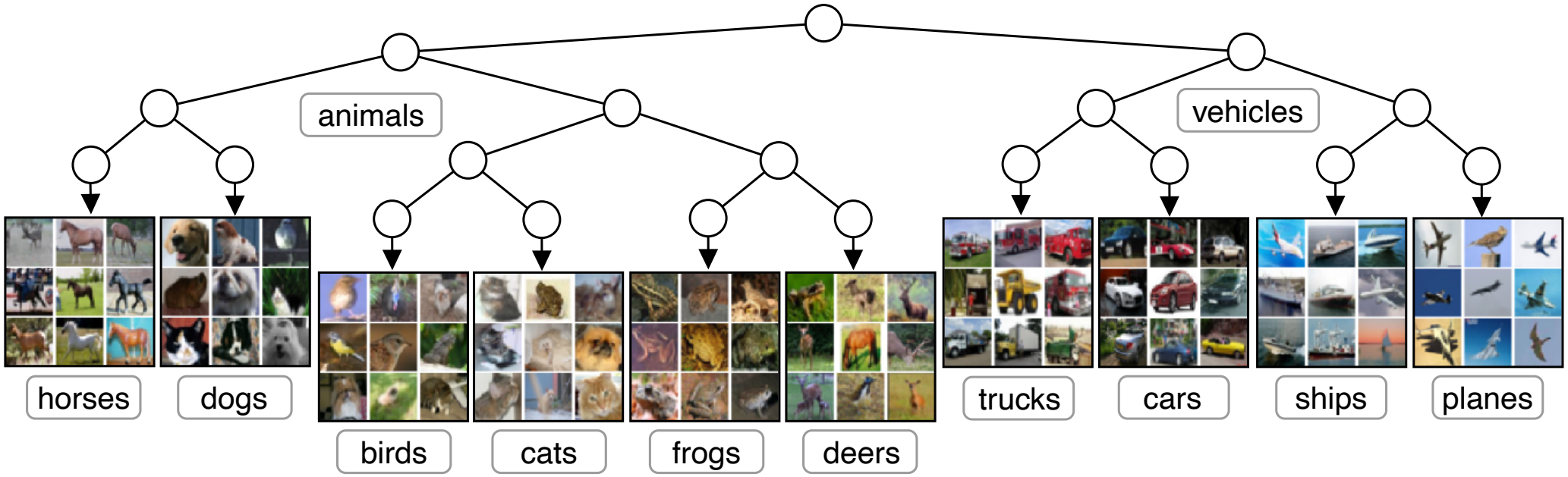
$$\mathcal{L}(\boldsymbol{x} \mid \mathcal{T}) := \mathbb{E}_{q(\boldsymbol{z},\mathcal{P}_l \mid \boldsymbol{x})}\left[\log p\left(\boldsymbol{x} \mid \boldsymbol{z}, \mathcal{P}_l\right)\right] - \mathrm{KL}\left(q\left(\boldsymbol{z}, \mathcal{P}_l \mid \boldsymbol{x}\right) \| p\left(\boldsymbol{z}, \mathcal{P}_l\right)\right)$$

Reconstruction loss

KL loss

Tree Variational Autoencoders. Spotlight NeurIPS 2023. **L. Manduchi**, M. Vandenhirtz , A. Ryser J. E. Vogt

Tree Variational Autoencoders



- Hierarchically divides samples according to their intrinsic characteristics.
- It adapts its architecture to discover the optimal tree to encode dependencies between latent variables.
- More flexible posterior & specialized leaf decoders.

Tree Variational Autoencoders. Spotlight NeurIPS 2023. **L. Manduchi**, M. Vandenhirtz , A. Ryser J. E. Vogt

Informed Representations of Data

Prior & Probabilistic relations

Clustered representations:

- Categorical prior
- Hypergeometric prior
- Hierarchical prior

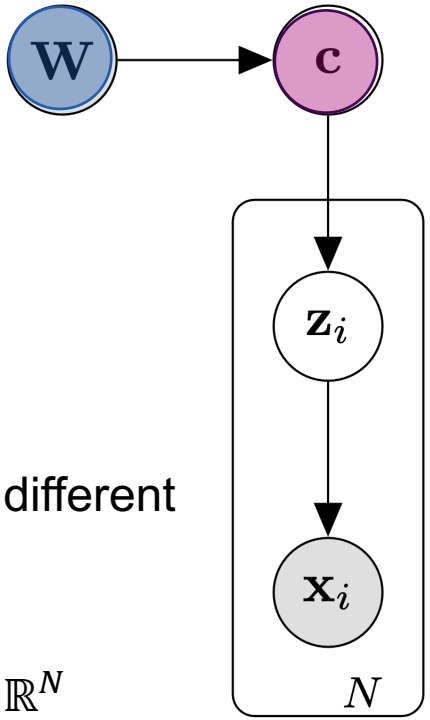
Physical constraints

- Periodicity in temporal data.
  - Variational Trajectory Models
  - Personalized Heart Meshes
- X-Ray Spectra Fitting

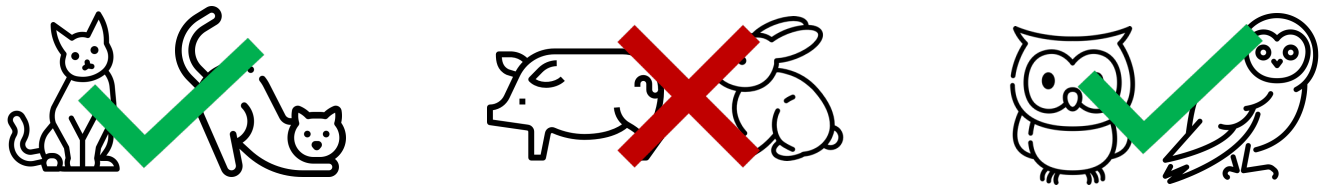
Weak supervision

- Similarity between samples
- Survival information of patients

Weak Supervision: Pairwise Information



- Access to similarity of a subset of datapoints **W**



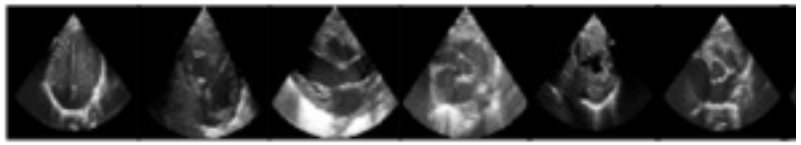
1. Prior information: $\boldsymbol{W} \in \mathbb{R}^{NxN}$ contains some instance-level annotiations with different noise levels.
2. Cluster assignments: $\mathbf{c} \sim p(\mathbf{c} \mid \boldsymbol{W}; \boldsymbol{\pi}) := \frac{1}{\Omega(\boldsymbol{\pi})} \prod_i \pi_{\mathbf{c}_i} h_i(\mathbf{c}, \boldsymbol{W}), \mathbf{c} \in \mathbb{R}^N$

$$\mathcal{L}_{\mathrm{C}}(\mathbf{X} \mid \boldsymbol{W}) := \mathbb{E}_{q_\phi(\mathbf{Z}|\mathbf{X})}\left[\log p_{\boldsymbol{\theta}}(\mathbf{X} \mid \mathbf{Z})\right] - D_{KL}\left(q_\phi(\mathbf{Z}, \mathbf{c} \mid \mathbf{X}) \| p(\mathbf{Z}, \mathbf{c} \mid \boldsymbol{W}; \boldsymbol{\nu}, \boldsymbol{\pi})\right)$$

Reconstruction term **Clustering term** **Conditional Gaussian Mixture**

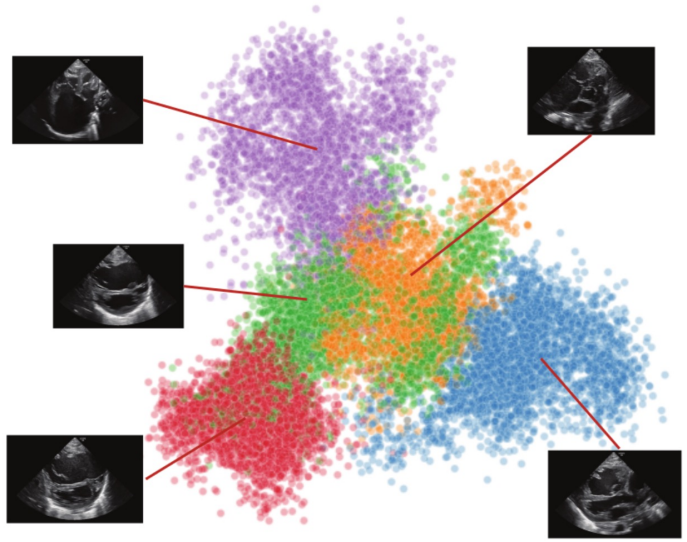
Deep Conditional Gaussian Mixture Model for Constrained Clustering. NeurIPS 2021.
L. Manduchi, K. Chin-Cheong, H. Michel, S. Wellmann, J. E. Vogt

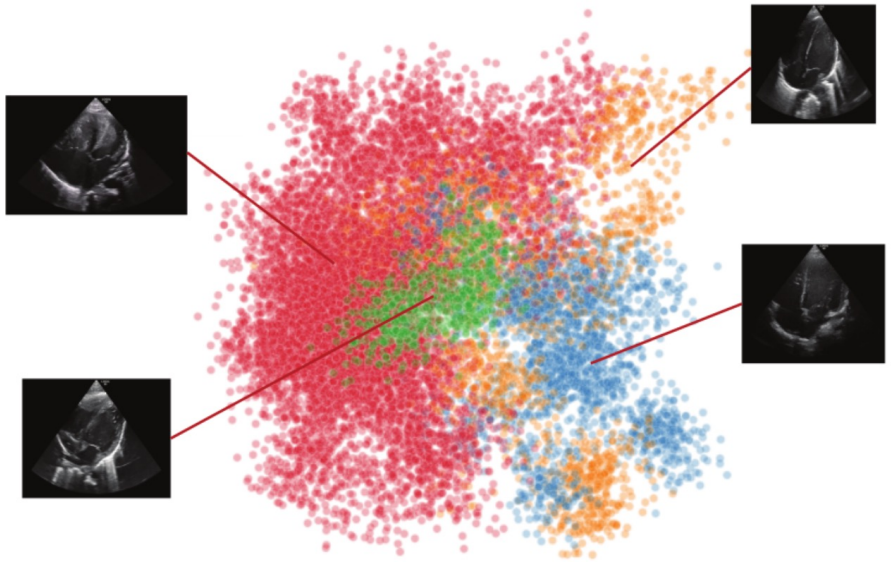
Clustering Heart Echocardiogram Videos



q: noise level

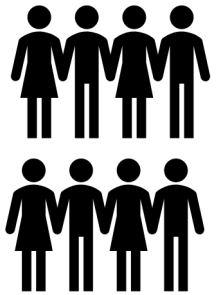
$q = 0.1$ $q = 0.2$ $q = 0.4$





Deep Conditional Gaussian Mixture Model for Constrained Clustering. NeurIPS 2021.
L. Manduchi, K. Chin-Cheong, H. Michel, S. Wellmann, J. E. Vogt

Weak Supervision: Survival Data

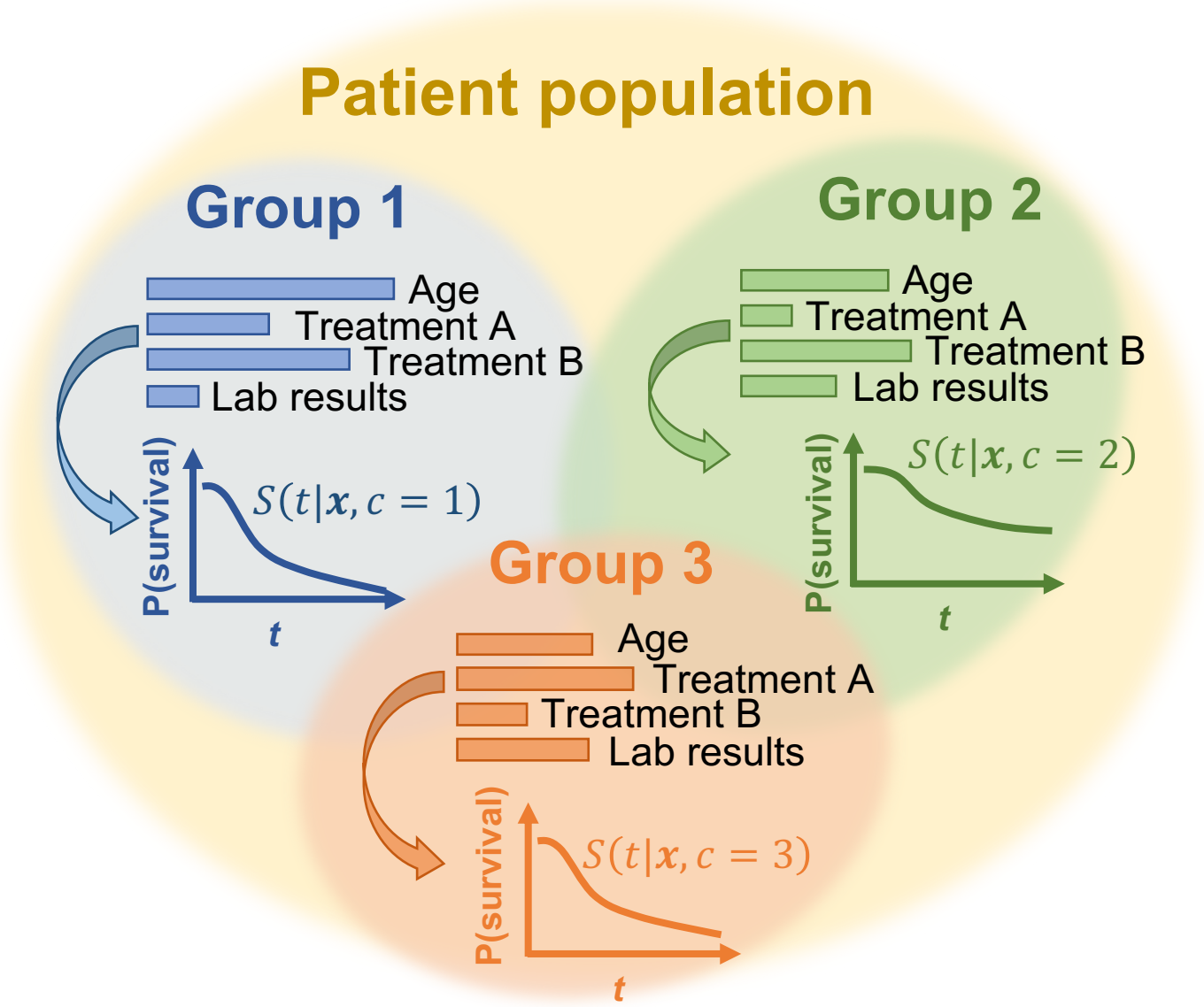


- Age
- Treatment A
- Treatment B
- Lab results

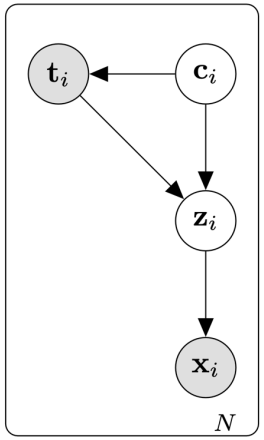
} **x**

- …
- **Time-to-event, t**

- Discover groups of patients characterised by different *associations* between the covariates $\boldsymbol{x}$ and survival time t
- Stratify patients by both risk and disease phenotype
- A more *personalised disease management*

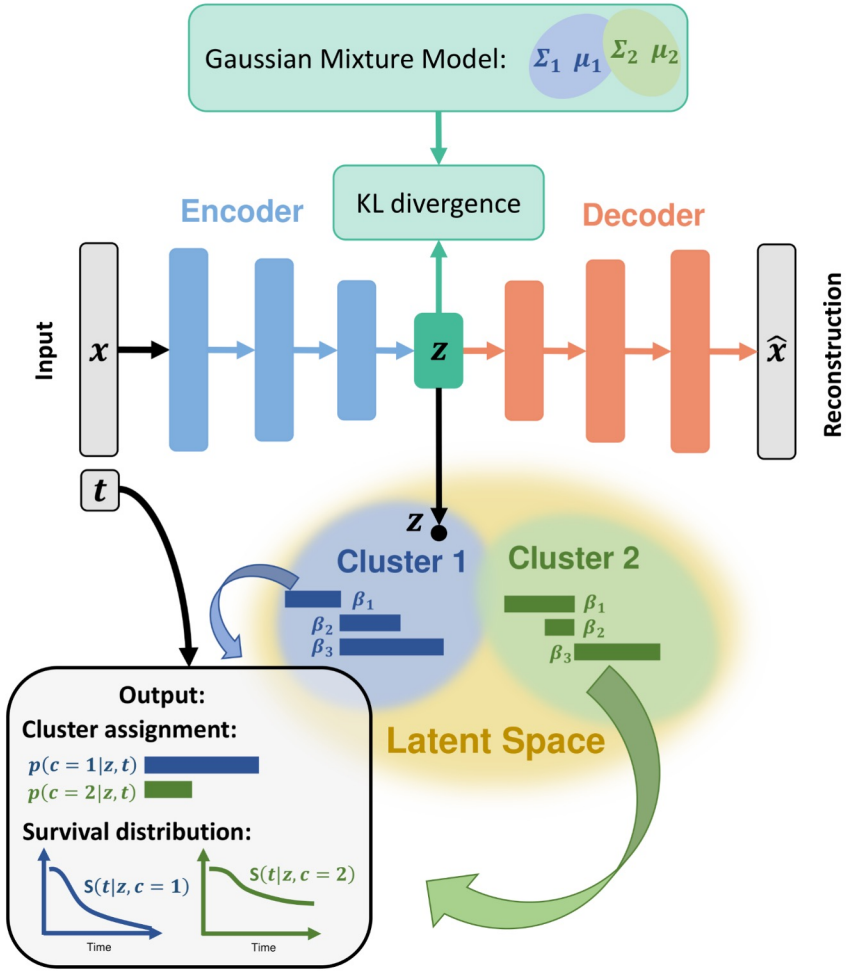


Variational Deep Survival Clustering (VaDeSC)



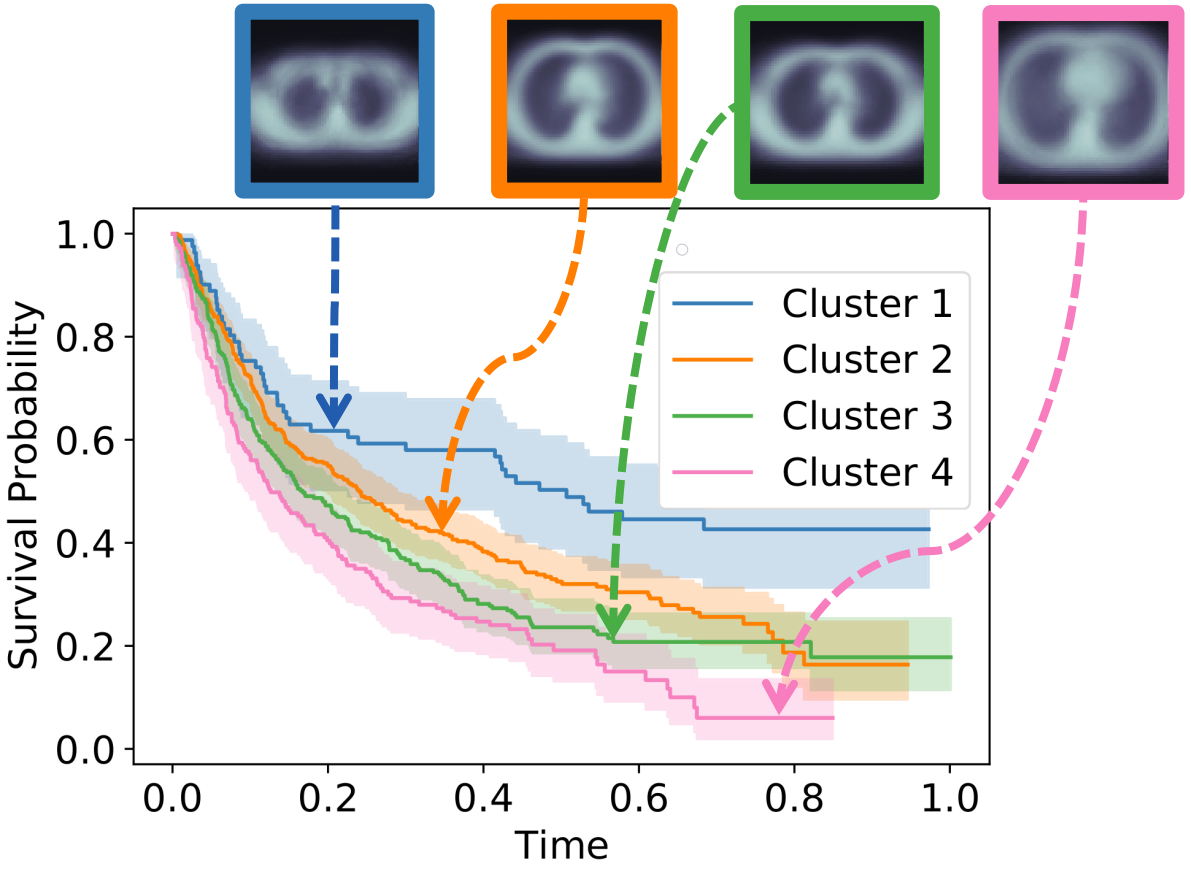
Overview of our model:

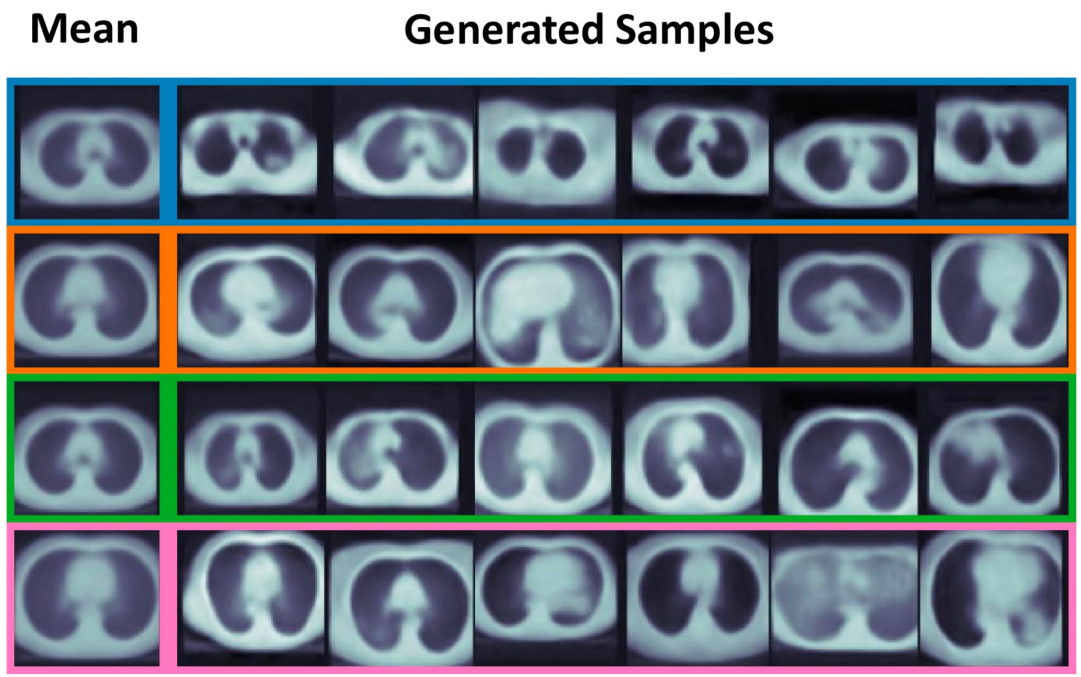
- based on a variational autoencoder
  - Gaussian mixture prior
  - clustering in the embedding space
- mixture of survival distributions
  - defined in the embedding space
  - Weibull with cluster-specific scales



A Deep Variational Approach to Clustering Survival Data. ICLR 2022
L. Manduchi, R. Marcinkevičs, M. C. Massi, T. Weikert, A. Sauter, V. Gotta, T. Müller, F. Vasella, M. C. N., M. P., B. S., J. E. Vogt

Clustering NSCLC Patients Using Computed Tomography Data





A Deep Variational Approach to Clustering Survival Data. ICLR 2022
L. Manduchi, R. Marcinkevičs, M. C. Massi, T. Weikert, A. Sauter, V. Gotta, T. Müller, F. Vasella, M. C. N., M. P., B. S., J. E. Vogt

ETH zürich D INFK medical data science

Informed Representations of Data

Prior & Probabilistic relations

Clustered representations:

- Categorical prior
- Hypergeometric prior
- Hierarchical prior

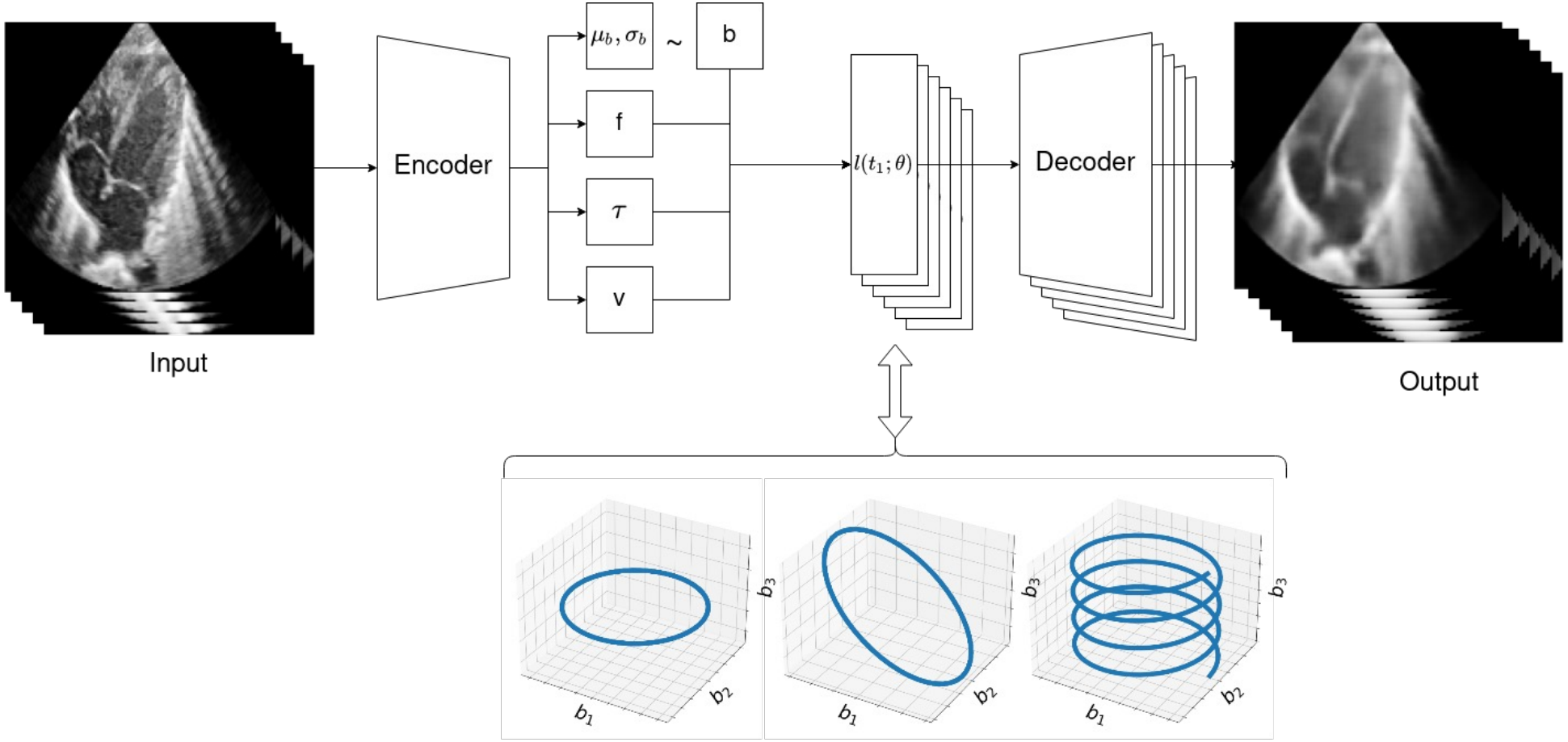
Physical constraints

- Periodicity in temporal data.
  - Variational Trajectory Models
  - Personalized Heart Meshes
- X-Ray Spectra Fitting

Weak supervision

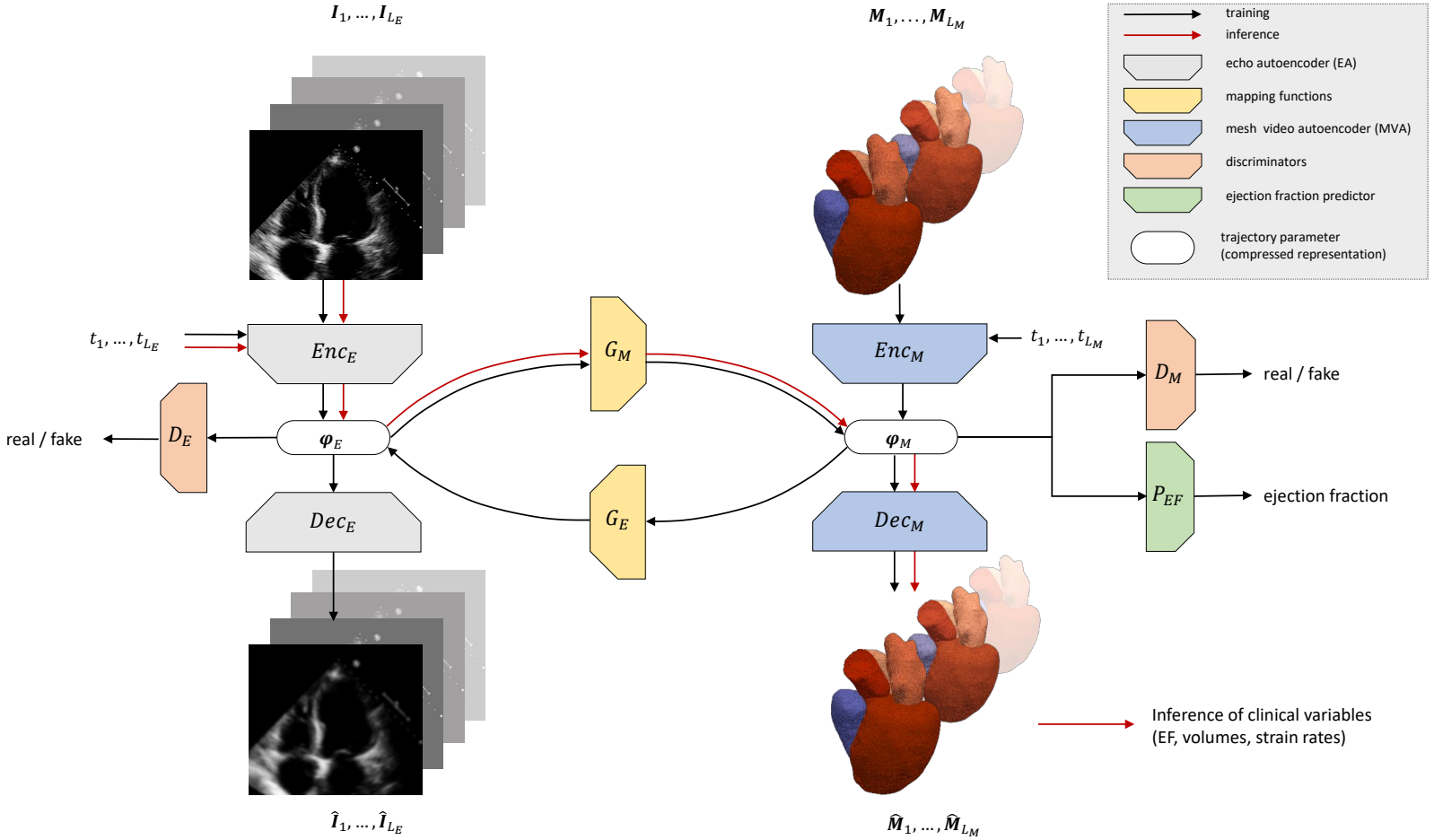
- Similarity between samples
- Survival information of patients

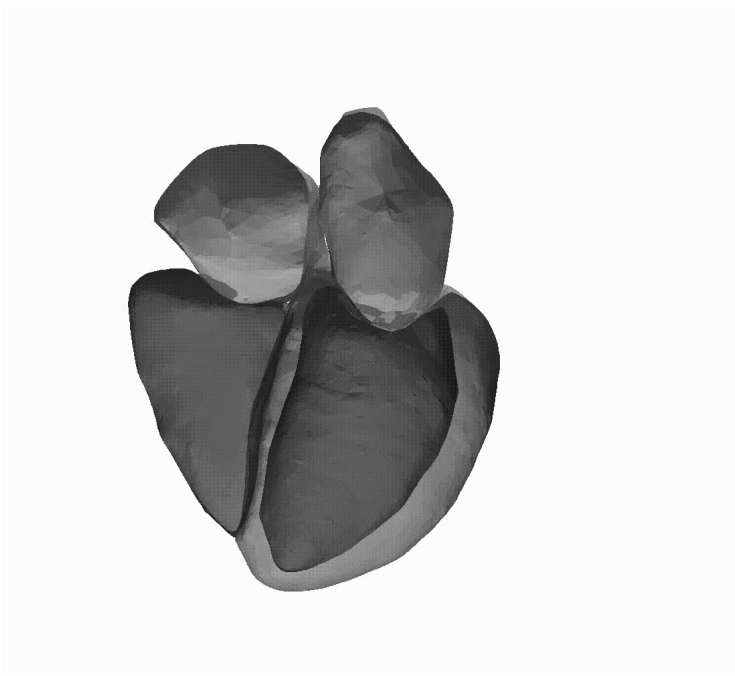
Variational Trajectories Models: model ciclicity



Anomaly Detection in Echocardiograms with Dynamic Variational Trajectory Models, MLHC, 2022.
A. Ryser, **L. Manduchi**, F. Laumer, H. Michel, S. Wellmann, J. E Vogt.

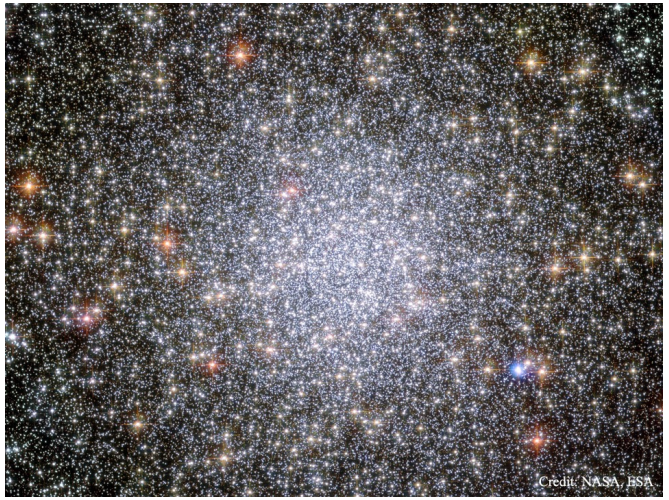
Beyond VAEs: Mesh Cycle Gans

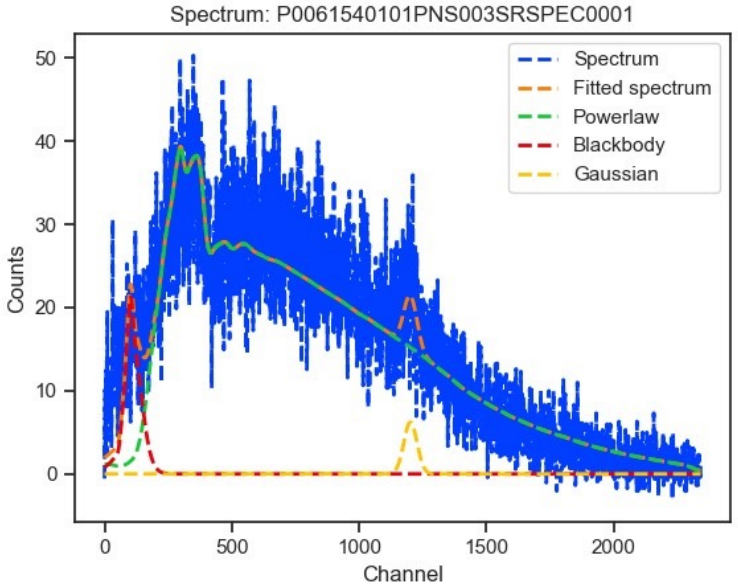


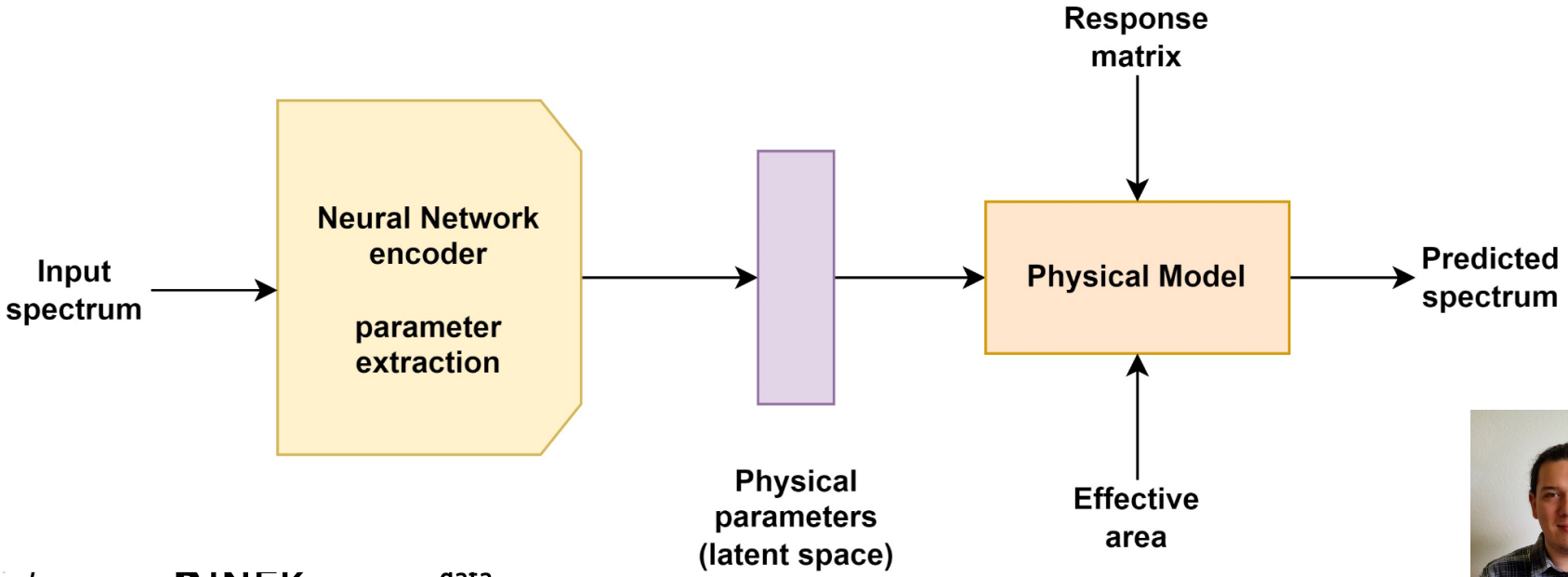


Weakly supervised inference of personalized heart meshes based on echocardiography videos, Medical Image Analysis, 2022.
F. Laumer, M. Amrani, **L. Manduchi**, A. Beuret, L. Rubi, A. Dubatovka, C. M. Matter, J. M. Buhmann,

X-ray Spectra Fitting of AGN







Cristi Blaga

Informed Representations of Data

Prior & Probabilistic relations

Clustered representations:

- Categorical prior
- Hypergeometric prior
- Hierarchical prior

Physical constraints

- Periodicity in temporal data.
  - Variational Trajectory Models
  - Personalized Heart Meshes
- X-Ray Spectra Fitting

Weak supervision

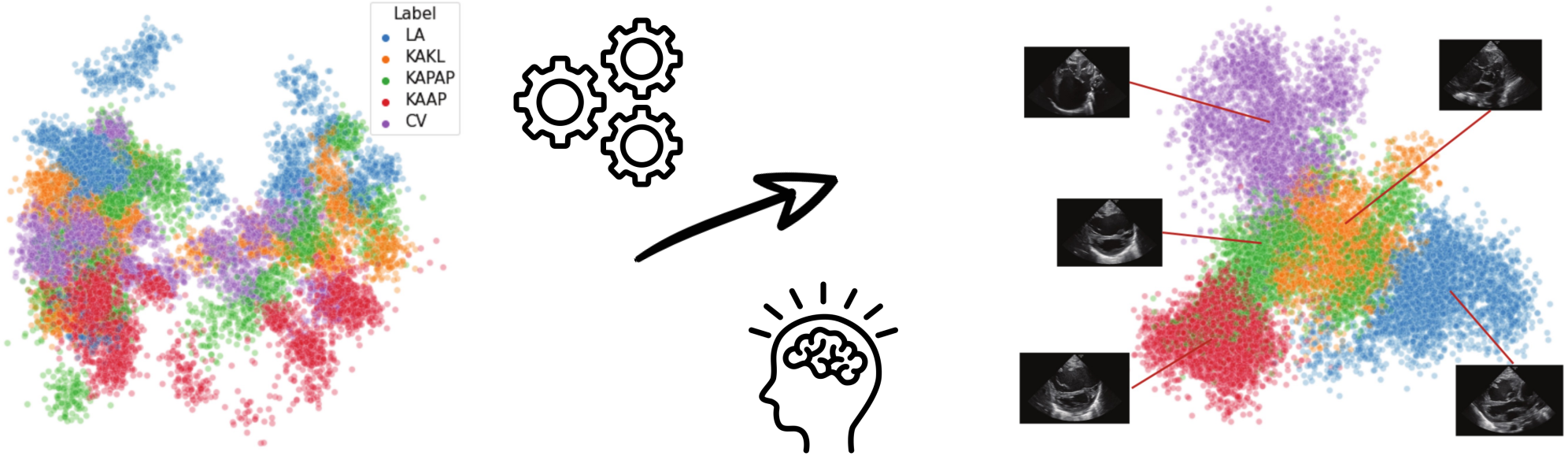
- Similarity between samples
- Survival information of patients

ETH zürich

D INFK

medical data science

Informed Representation Learning with Deep Generative Models



Laura Manduchi

ETH Zürich, Department of Computer Science