A Deep Variational Approach to Clustering Survival Data

Laura Manduchi[†*], Ričards Marcinkevičs[†*] Michela C. Massi, Thomas Weikert, Alexander Sauter, Verena Gotta, Timothy Müller, Flavio Vasella, Marian Neidert, Marc Pfister, Bram Stieltjes, Julia E. Vogt*
*Department of Computer Science, ETH Zürich
✉: laura.manduchi@inf.ethz.ch, ricards.marcinkevics@inf.ethz.ch

ETH zürich · medical data science · ICLR

1 Survival Cluster Analysis

Given:

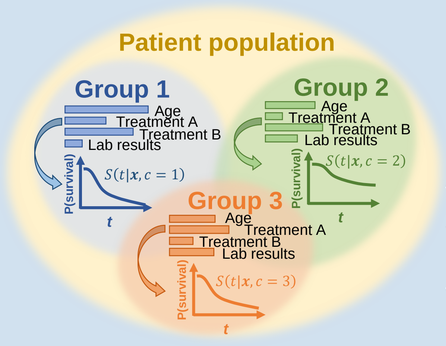
- data $\mathcal{D} = \{\boldsymbol{x}_i, t_i, \delta_i\}_{i=1}^N$
- $\boldsymbol{x}_i$ is a vector of *explanatory variables*
- t_i is the *time to a certain event*
- t_i is *censored* ($\delta_i = 0$) when its value is only partially known

Goal: discover **cluster-specific** relationships between *explanatory variables* $\boldsymbol{x}$ and *time-to-event* t, that is:

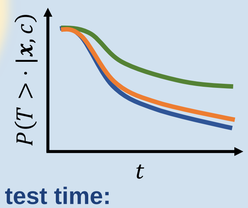
1. discover the unobserved *cluster assignment* c

2. model the *survival distribution*: $S(t|\boldsymbol{x}, c) = P(T > t|\boldsymbol{x}, c)$

Example: patient population consists of three clusters with different disease subtypes. Each group features different associations between the covariates $\boldsymbol{x}$ and survival time t



Groups ***might, but don't need*** *to* have *disparate* survival distributions:
cf. groups **1** & **3**



At training time:

- age
- treatment A
- treatment B
- lab results
- …
- **time-to-event,** t

$\}\ \boldsymbol{x}$

At test time:

- age
- treatment A
- treatment B
- lab results
- …
- ???

$\}\ \boldsymbol{x}$

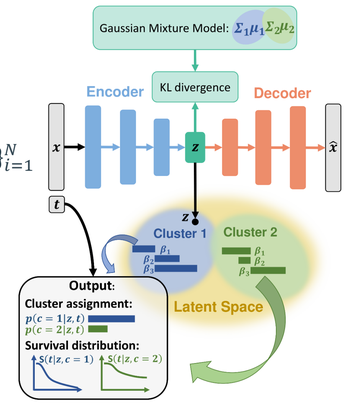
2 Motivation

- Exploratory analysis for survival data
- Discovery of heterogeneity in the relation between $\boldsymbol{x}$ and t
  - better understanding of the disease and its progression
  - *personalised* disease management
- Patient stratification by risk

3 Variational Deep Survival Clustering (VaDeSC)

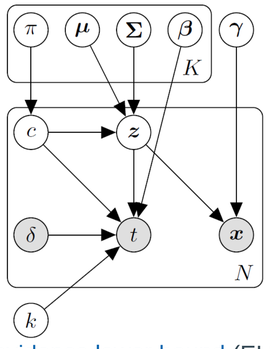
Model Overview:

- VAE with a Gaussian mixture prior
- *cluster-specific Weibull* survival distributions in the latent space
- maximise the joint likelihood of $\{\boldsymbol{x}_i, t_i\}_{i=1}^N$



Generative process:

1. *Cluster assignment* sampled from a categorical distribution**:**
$$c \sim p(c;\ \boldsymbol{\pi}) = \pi_c.$$
2. *Latent embedding* sampled from a cluster-specific Gaussian distribution:
$$\boldsymbol{z} \sim p(\boldsymbol{z}|c) = \mathcal{N}(\boldsymbol{z};\ \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c).$$
3. *Inputs* generated from a distribution parameterised by a neural net (the decoder
$$\boldsymbol{x} \sim p(\boldsymbol{x}|\boldsymbol{z};\ \boldsymbol{\gamma}).$$
4. *Time-to-event* sampled from a cluster-specific Weibull distribution:
$$t \sim p(t|\boldsymbol{z}, c;\ \{\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_K\}, k).$$



Evidence Lower Bound:

The likelihood is intractable! We maximise the evidence lower bound (ELBO):

$$\mathbb{E}_{q(\boldsymbol{z},c|\boldsymbol{x},t)}[\underbrace{\log p(\boldsymbol{x}|\boldsymbol{z};\ \boldsymbol{\gamma})}_{\text{reconstruction}} + \underbrace{\log p(t|\boldsymbol{z}, c;\ \boldsymbol{\beta}, k)}_{\text{survival}}] + \underbrace{D_{KL}\big(q(\boldsymbol{z}, c|\boldsymbol{x}, t)||p(\boldsymbol{z}, c;\ \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi})\big)}_{\text{clustering}}$$

The variational distribution factorises as follows

$$q(\boldsymbol{z}, c|\boldsymbol{x}, t) = \underbrace{q(\boldsymbol{z}|\boldsymbol{x})}_{\text{encoder}}\underbrace{q(c|\boldsymbol{z}, t)}_{\text{soft cluster assignments}}$$

parameterised by an NN

4 Experimental Setup

An aggregated **computed tomography** (CT) dataset from non-small cell lung cancer patients:

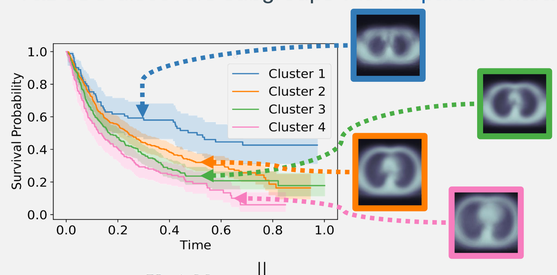
- **Basel**: 392 patients, University Hospital Basel, CH
- **Lung1**: 422 patients, Maastro clinic, NL
- **Lung3**: 89 patients, Maastro clinic, NL
- **NSCLC Radiogenomics**: 211 patients, Stanford University School of Medicine and Palo Alto Veterans Affairs Healthcare System, US

5 Results

VaDeSC is competitive at predicting survival times:

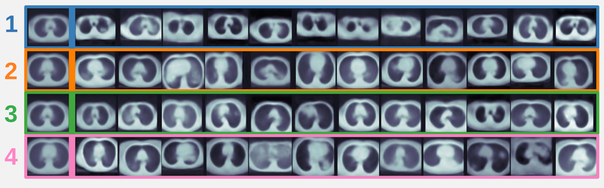
Method	CI	RAE_{nc}	RAE_{c}	CAL
Radiomics + Cox PH	**0.60±0.02**	—	—	—
Radiomics + Weibull AFT	**0.60±0.02**	**0.70±0.02**	0.45±0.03	1.26±0.04
DSM	*0.59±0.04*	0.72±0.03	**0.34±0.06**	*1.24±0.07*
VaDeSC (ours)	**0.60±0.02**	*0.71±0.03*	*0.35±0.05*	**1.21±0.05**

VaDeSC discovers subgroups with disparate characteristics:



Variable	1	2	3	4	p-val.
T. Vol., cm^3	43	36	40	63	$\leqslant$ **5e-2**
Age, yrs	62	69	67	70	**1e-3**
Female, %	36	19	38	23	**1e-3**
Smoker, %	67	94	87	100	0.12
M1, %	20	45	44	45	0.2
$\geq$ **T3**, %	10	29	35	31	0.7

Clusters are associated with tumour location:



6 Summary

A **deep probabilistic model** for clustering survival data:

- novel generative assumptions
- improved clustering and competitive time-to-event prediction
- a holistic representation learning view of survival data

Future work:

- interpretability/explainability
- multimodal survival analysis