Dynamic Outcomes-Based Clustering of Disease Trajectory in Mechanically Ventilated Patients

Emma Rocheteau*, University of Cambridge
Ioana Bica, DeepMind
Pietro Liò, University of Cambridge
Ari Ercole, University of Cambridge
*ecr38@cam.ac.uk

UNIVERSITY OF CAMBRIDGE

Introduction

1. Patients on mechanical ventilation are a highly heterogeneous group, with widely differing outcomes.
2. Temporal clustering based on *phenotype* and *outcomes*, would be greatly beneficial for the following reasons:

- The clusters could be used to create interpretable early warning systems to alert physicians of deteriorating patients.
- They could help to study and understand sub-types of disease trajectory.
- They could be used to categorise patients early on in intervention studies.

Methods

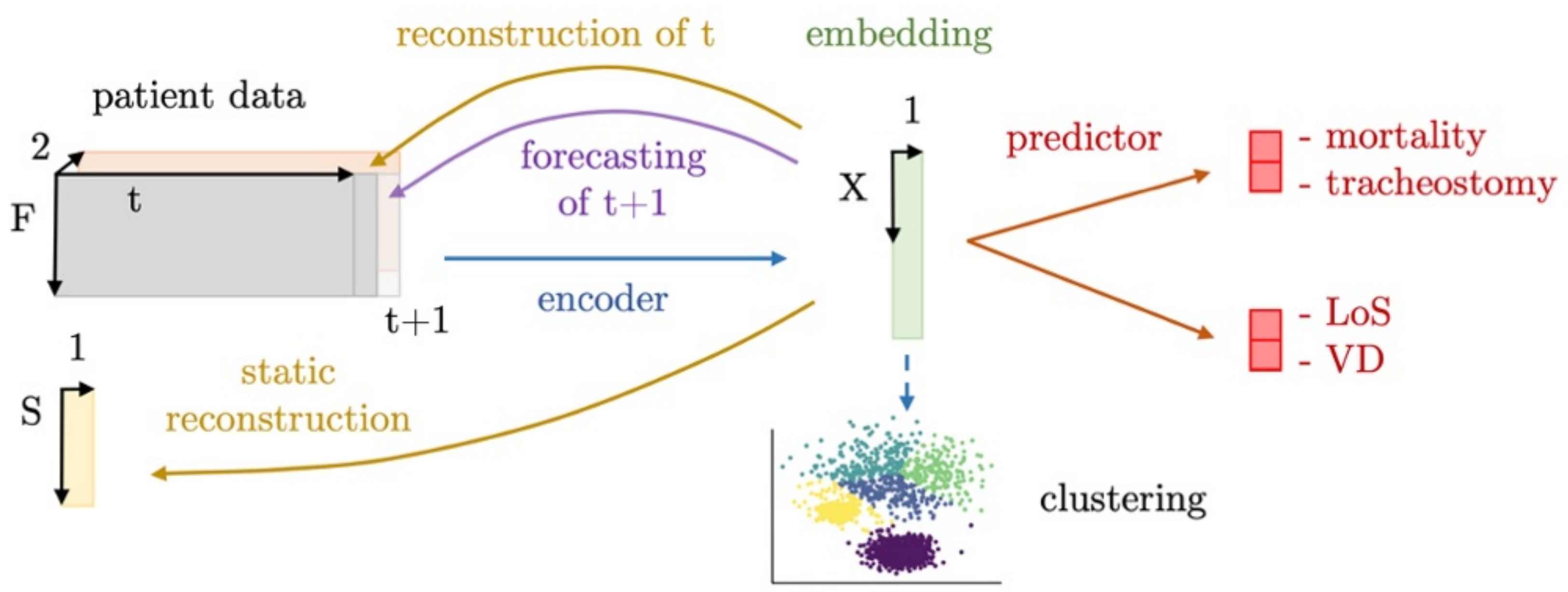

Figure 1: Overview of our model. The data (timeseries and static variables) are given to an encoder (LSTM, Transformer or TPC[1]) to produce an embedding (green). The embedding is trained using supervised tasks: mortality, tracheostomy risk, length of stay and ventilation duration (red); unsupervised tasks (yellow); and a forecasting task (purple). K-medoids clustering is used to produce the clusters.

Task Performance

The TPC model was the best performing model. An ablation study showed that the model did better when all of the tasks in Figure 1 were included.

Table 1: Encoder performance on the prediction tasks averaged over 5 independent training runs. The error margins are 95% confidence intervals. For mortality and tracheostomy, higher AUROC and AUPRC is better; for LoS and VD, lower MAD and MSLE is better. (a) shows the full multi-task setting as shown in Figure 1, (b) is a variational alternative to the full task setting. Statistically significant differences are indicated by daggers ($^{\dagger} = p < 0.05$, $^{\ddagger} = p < 0.001$). If the result is significantly better than the comparison models*, it is highlighted in **blue**, if it is significantly worse it is highlighted in **pink**. *In (a) the statistical testing compares the three model types, in (b) each model type is compared to its corresponding 'non-variational' model in table (a).

		In-Hospital Mortality		Tracheostomy		Length of Stay		Vent. Duration	
	Model	AUROC	AUPRC	AUROC	AUPRC	MAD	MSLE	MAD	MSLE
(a)	TPC	**$0.833 \pm 0.010^{\dagger}$**	**$0.644 \pm 0.013^{\ddagger}$**	**$0.804 \pm 0.007^{\ddagger}$**	**$0.507 \pm 0.020^{\dagger}$**	**$7.20 \pm 0.13^{\ddagger}$**	**$0.359 \pm 0.010^{\ddagger}$**	**$3.24 \pm 0.07^{\ddagger}$**	**$0.210 \pm 0.008^{\ddagger}$**
	Transformer	0.697 ± 0.012	0.434 ± 0.019	0.760 ± 0.012	0.419 ± 0.033	8.46 ± 0.07	0.495 ± 0.007	3.95 ± 0.20	0.256 ± 0.016
	LSTM	0.823 ± 0.002	0.608 ± 0.008	0.774 ± 0.002	0.473 ± 0.015	9.16 ± 0.06	0.663 ± 0.008	5.57 ± 0.04	0.681 ± 0.011
(b)	TPC	**$0.807 \pm 0.006^{\ddagger}$**	**$0.584 \pm 0.014^{\ddagger}$**	**$0.775 \pm 0.008^{\ddagger}$**	**$0.437 \pm 0.012^{\ddagger}$**	**$9.06 \pm 0.10^{\ddagger}$**	**$0.555 \pm 0.018^{\ddagger}$**	**$4.42 \pm 0.03^{\ddagger}$**	**$0.347 \pm 0.006^{\ddagger}$**
	Transformer	**$0.660 \pm 0.023^{\dagger}$**	**$0.373 \pm 0.039^{\dagger}$**	**$0.714 \pm 0.020^{\ddagger}$**	**$0.353 \pm 0.018^{\dagger}$**	**$9.42 \pm 0.27^{\ddagger}$**	**$0.623 \pm 0.020^{\ddagger}$**	**$4.63 \pm 0.27^{\ddagger}$**	**$0.359 \pm 0.030^{\ddagger}$**
	LSTM	**$0.803 \pm 0.004^{\ddagger}$**	**$0.555 \pm 0.006^{\ddagger}$**	**$0.748 \pm 0.005^{\ddagger}$**	**$0.411 \pm 0.010^{\ddagger}$**	**$10.2 \pm 0.1^{\ddagger}$**	**$0.813 \pm 0.016^{\ddagger}$**	**$5.95 \pm 0.04^{\ddagger}$**	**$0.775 \pm 0.007^{\ddagger}$**

Cluster analysis

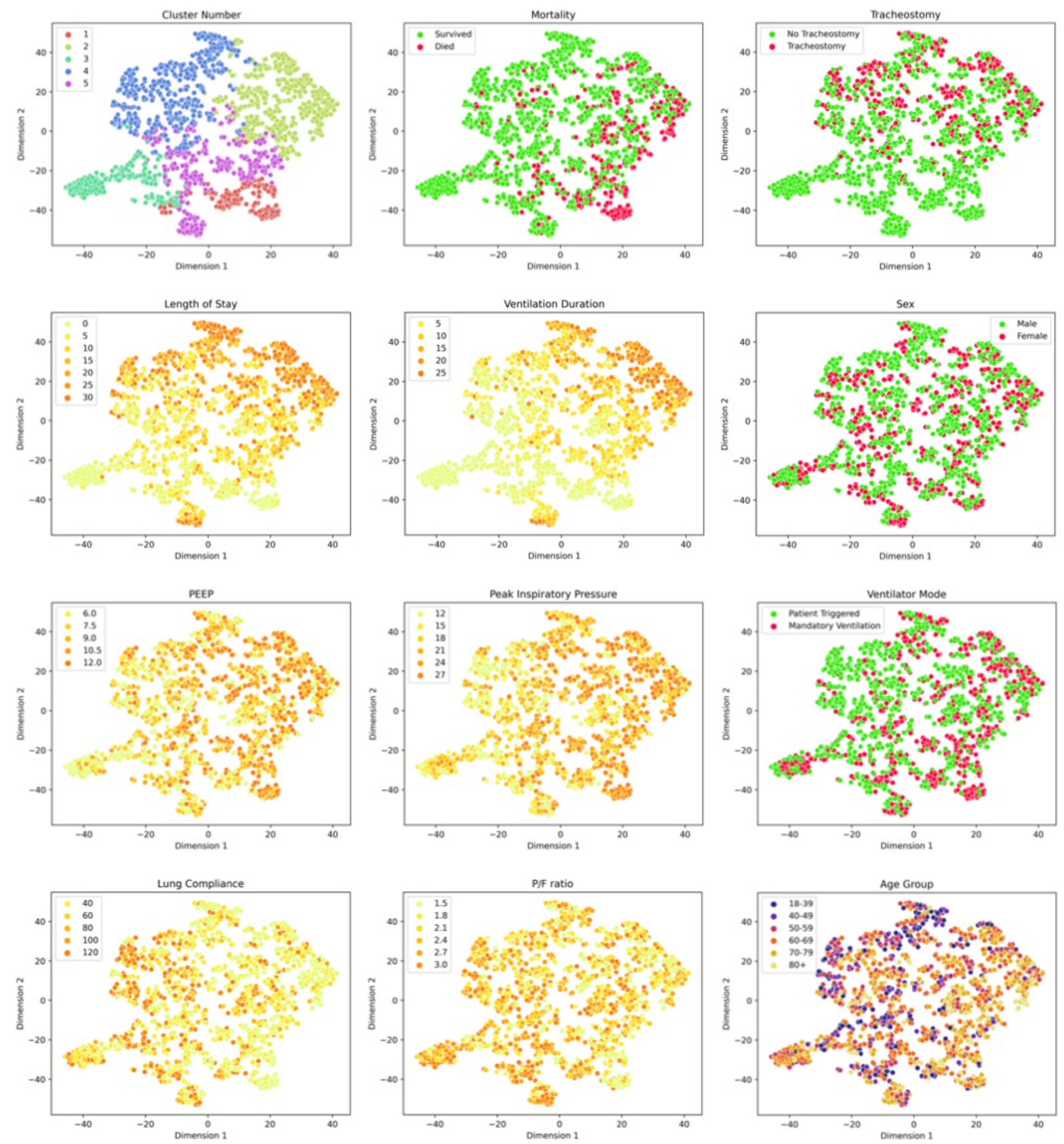

Figure 2: t-SNE plots of the learned embeddings of the TPC[1] model, plotted against different attributes. The top left plot shows the cluster assignments.

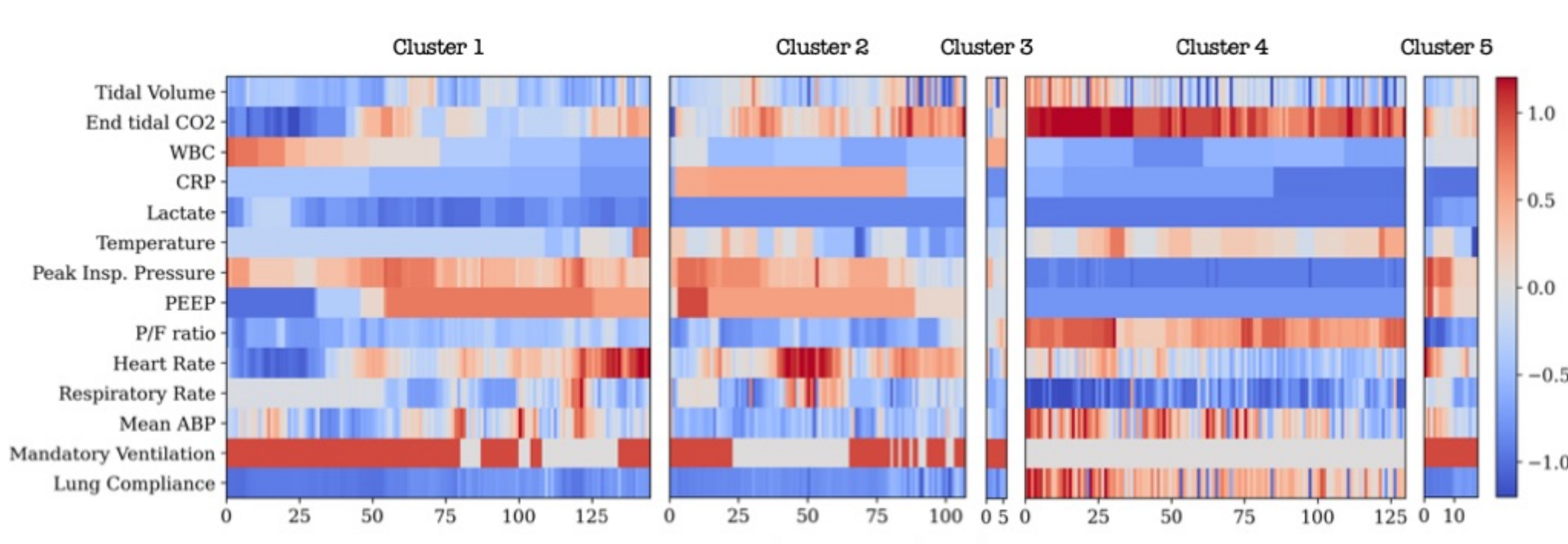

Figure 3: Raw timeseries from each of the 5 medoids resulting from the k-medoids algorithm. These can be considered the 'archetypal' patients for each cluster.

Table 2: Average outcomes by cluster $\pm$ 95% confidence intervals for the TPC model. Each patient has been classified into a primary cluster, which is the cluster that they spent the majority of their time in. LoS and VD are shown in days.

Cluster	Patients	Mortality (%)	Tracheostomy (%)	Length of Stay	Vent. Duration
1	232	72.0 ± 5.8	1.3 ± 1.5	3.8 ± 0.8	2.4 ± 0.3
2	133	34.6 ± 8.2	38.3 ± 8.4	30.0 ± 3.6	21.4 ± 2.2
3	1,292	1.9 ± 0.7	1.5 ± 0.7	2.8 ± 0.3	0.7 ± 0.0
4	347	4.0 ± 2.1	31.1 ± 4.9	22.0 ± 1.8	7.4 ± 0.9
5	227	26.0 ± 5.7	8.4 ± 3.6	13.0 ± 1.6	7.2 ± 0.9

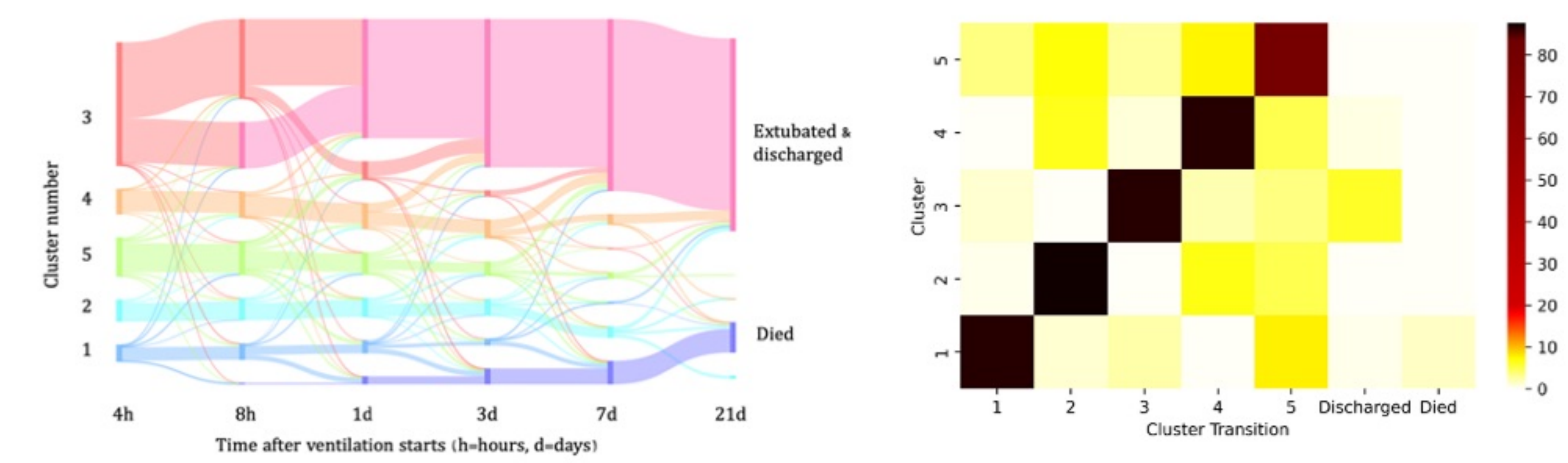

Figure 3: The temporal sankey plot and cluster transition matrix both show that the clusters are remarkably stable over time.

Cluster 1 *Acute life-threatening pulmonary injury*: Contains the sickest patients with a mortality of 72%. They have signs of severe respiratory distress.
Cluster 2 *Pulmonary critical illness*: Substantial mortality, long length of stay and ventilation duration. Very difficult to wean, hence the high tracheostomy rate.
Cluster 3 *Short stay*: Contains the healthiest patients. Most likely perioperative.
Cluster 4 *Critical illness (other)*: Long length of stay, but good lung parameters.
Cluster 5 *Acute critical illness (other)*: Poor outcomes, but lung injury not prominent.

Summary

1. The TPC[1] model significantly outperforms alternative temporal encoders on patient outcome prediction tasks.
2. It can be used to generate clinically meaningful and interpretable clusters with distinct phenotypes and outcomes.
3. Key aspects of the phenotypes are similar across choices of encoder.
4. The cluster assignment is remarkably stable over time, and membership is determined early on. This is particularly encouraging as a substrate for future intervention studies, because they rely on early phenotyping.
5. Stable transitions between clusters do occur but they are infrequent. Studying these transitions with a view towards understanding the cause of a change in prognosis is an important avenue for future work.

References

[1]Emma Rocheteau, Pietro Liò, Stephanie Hyland (2021). "Temporal Pointwise Convolutional Networks for Length of Stay Prediction in the Intensive Care Unit." In: Proceedings of the Conference on Health, Inference, and Learning, CHIL'21.