T-DPSOM - An Interpretable Clustering Method for Unsupervised Learning of Patient Health States

Laura Manduchi
ETH Zurich
Zürich, Switzerland

Matthias Hüser
ETH Zürich
Zürich, Switzerland

Martin Faltys
ETH Zürich
Zürich, Switzerland

Julia Vogt
ETH Zürich
Zürich, Switzerland

Gunnar Rätsch
ETH Zürich
Zürich, Switzerland

Vincent Fortuin
ETH Zürich
Zürich, Switzerland

ABSTRACT

Generating interpretable visualizations of multivariate time series in the intensive care unit is of great practical importance. Clinicians seek to condense complex clinical observations into intuitively understandable critical illness patterns, like failures of different organ systems. They would greatly benefit from a low-dimensional representation in which the trajectories of the patients' pathology become apparent and relevant health features are highlighted. To this end, we propose to use the latent topological structure of Self-Organizing Maps (SOMs) to achieve an interpretable latent representation of ICU time series and combine it with recent advances in deep clustering. Specifically, we (a) present a novel way to fit SOMs with probabilistic cluster assignments (PSOM), (b) propose a new deep architecture for probabilistic clustering (DPSOM) using a VAE, and (c) extend our architecture to cluster and forecast clinical states in time series (T-DPSOM). We show that our model achieves superior clustering performance compared to state-of-the-art SOM-based clustering methods while maintaining the favorable visualization properties of SOMs. On the eICU data-set, we demonstrate that T-DPSOM provides interpretable visualizations of patient state trajectories and uncertainty estimation. We show that our method rediscovers well-known clinical patient characteristics, such as a dynamic variant of the Acute Physiology And Chronic Health Evaluation (APACHE) score. Moreover, we illustrate how it can disentangle individual organ dysfunctions on disjoint regions of the two-dimensional SOM map.

ACM Reference Format:
Laura Manduchi, Matthias Hüser, Martin Faltys, Julia Vogt, Gunnar Rätsch, and Vincent Fortuin. 2021. T-DPSOM - An Interpretable Clustering Method for Unsupervised Learning of Patient Health States. In *ACM Conference on Health, Inference, and Learning (ACM CHIL '21), April 8–10, 2021, Virtual Event, USA.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3450439.3451872

ACM CHIL '21, April 8–10, 2021, Virtual Event, USA

ACM ISBN 978-1-4503-8359-2/21/04…$15.00
https://doi.org/10.1145/3450439.3451872

1 INTRODUCTION

Intensive care units (ICUs) treat patients with severe or life-threatening illnesses and injuries, which require constant care and supervision. A critically ill patient is closely monitored by multiple devices and regular blood samples. At the same time, the patient is also subject to multiple treatments. The physicians' task is to integrate these high-dimensional monitoring and treatment data into a comprehensive picture of the patient's current and likely future health state and to then decide on further treatment.

Traditional clustering methods, such as hierarchical clustering and latent class analysis, have been applied to physiological measurements to identify clinically relevant patient states and to discover new physiological relationships unknown to the physicians. Detecting and defining comprehensive patient sub-phenotypes is an important step in personalized medicine. It could also facilitate future treatment studies by allowing for more homogeneous patient cohorts [37]. More granular sub-phenotypes and intuitive visualizations of such patient health clusters could ultimately aid the physicians' central task of understanding the patient's health state and future trajectories [5, 15, 24].

Traditional clustering methods, however, have often poor performance on high-dimensional, complex, real-world data sets such as clinical time series, due to the inefficiency of similarity measures used. To overcome this issue, dimensionality reduction and feature transformation have been successfully applied to obtain a low-dimensional representation of the raw data that is easier to cluster [23]. Recently, traditional feature transformation methods, such as PCA [50] or Kernel methods [21], have been replaced by Deep Neural Networks (DNNs), due to their inherent property of highly non-linear transformation. The combination of DNNs and clustering methods have been proved to ease the clustering process and to substantially increase their performance [1, 52]. In such methods, however, the clustered feature points lie in a latent space that cannot generally be easily visualized or interpreted by humans, and not much effort is directed towards investigating the relationship between clusters.

In contrast, the Self-Organizing Map (SOM) [25] is a clustering method that provides such an interpretable representation. It produces a low-dimensional (typically 2-dimensional), discretized representation of the input space by inducing a flexible neighborhood structure over the clusters. As a result, it spatially groups together clusters that are similar and it permits pattern recognition in datasets that would otherwise be too complex to visualize, obtaining a clear overview of the state of the patients, and detecting

Table 1: Overview of related approaches and our proposed methods.

Model	SOM structure	Probabilistic	Clustering performance	Temporal model
DEC / IDEC [18, 52]	✗	✓	✓	✗
SOM-VAE [11]	✓	✗	✗	✓
DESOM [10]	✓	✗	✗	✗
DPSOM (ours)	✓	✓	✓	✗
T-DPSOM (ours)	✓	✓	✓	✓

subsets that might be missed otherwise. The SOM has been successfully applied to various medical settings and has been proven to be an excellent tool in the visualization of high-dimensional medical data [16, 19, 39] and in medical imaging [33, 40]. While the SOM is particularly effective for data visualization [30], only a few methods have attempted to combine it with DNNs. Moreover, as we will show in the experiments, their performances are subpar compared to modern deep clustering methods.

To address the above issues, we present a novel way of fitting SOMs with probabilistic cluster assignments, which we call Probabilistic SOM (PSOM). We then propose a deep architecture, the Deep Probabilistic SOM (DPSOM), which jointly trains a VAE and a PSOM to achieve an interpretable discrete representation while exhibiting state-of-the-art clustering performance. Instead of hard assignments of data points to clusters, our model uses centroid-based probability distributions. It minimizes their Kullback-Leibler divergence against auxiliary target distributions, while enforcing a SOM-friendly space. We further extend this model to support time series, yielding the temporal DPSOM (T-DPSOM). We demonstrate that the T-DPSOM serves as a useful tool to understand and track patient health states in the ICU.

Our main contributions are:

- DPSOM, a deep clustering architecture that combines a VAE with a novel SOM-based clustering objective (PSOM).
- T-DPSOM, an extension of this architecture to time series, improving clustering performance and enabling temporal forecasting.
- A thorough empirical assessment of our proposed models, showing superior performance on static image data and medical time series from the ICU.
- Illustrations of how T-DPSOM can be used to cluster patients into different sub-phenotypes and potentially gain better understanding of disease patterns and individual patient health states.

2 RELATED WORK

2.1 Applied clustering and visualization for the critically ill patient

Personalized medicine encourages a more singular approach to patients, treating each individual according to their specific history. This is becoming increasingly widespread, not only in the ICU [47, 49]. To provide personalized medicine, the different phenotypes of diseases have to be understood, and the patient's phenotype has to be correctly identified. Historically, clinical phenotypes were the most important. In the last years, data-driven approaches became more popular. Simple versions of such approaches (Latent Class Analysis, Hierarchical Clustering) now exist for almost all major diseases in the ICU [37] such as sepsis [15, 38], ARDS [4, 6, 24], and AKI [2]. Already in 2010, Buchman et al. [5] described the potential of clustering to understand the patient in his "physiological state space" and discussed the potential for using this method to describe current, past, and future patient states. Knox et al. [24] used SOM to visualize subtypes of patients with sepsis intuitively. Tscholl et al. [43] show that better visualizations of complex medical data can improve the understanding of the patient state, increase situational awareness, and reduce the workload. Even though this was only shown for directly visualizing vital parameters.

2.2 Self-organizing maps

Self-organizing maps have been widely used as a means to visualize information from large amounts of data [41]. They can be seen as a form of clustering in which the centroids are connected by a topological neighborhood structure [9]. Since their inception [25], several variants have been proposed to enhance their performance and scope. The adaptive subspace SOM, ASSOM [26], for example, combines PCA and SOMs to map data into a reduced feature space. Tokunaga and Furukawa [42] combine SOMs with multi-layer perceptrons to obtain a modular network. Liu et al. [30] proposed the Deep SOM (DSOM), an architecture composed of multiple layers similar to deep neural networks. Although there exist several methods tailored to representation learning on time series [12–14], only few models present extensions of the SOM optimized for temporal data. Examples are the Temporal Kohonen map [7], its improved version Recurrent SOM [32], as well as Recursive SOM [48]. Probabilistic versions of SOM include [8, 31] as well as the generative topographic map [3].

2.3 Deep SOM-based models

While there exist previous efforts to endow VAEs with a hierarchical latent space [17, 27, 46], to the best of our knowledge, only two approaches used deep generative models in combination with a SOM structure in the latent space. The SOM-VAE model [11], inspired by the VQ-VAE architecture [44] (which itself was later extended by Razavi et al. [36]), uses an AE to embed the input data points into a latent space and then applies a SOM-based clustering loss on top of this latent representation. Even though it prominently features a VAE in its name as well as model description, in practice it uses a Dirac δ-distribution and therefore hard assignments of data points to cluster centroids [11]. It also uses a uniform prior over cluster assignments, such that the KL-term

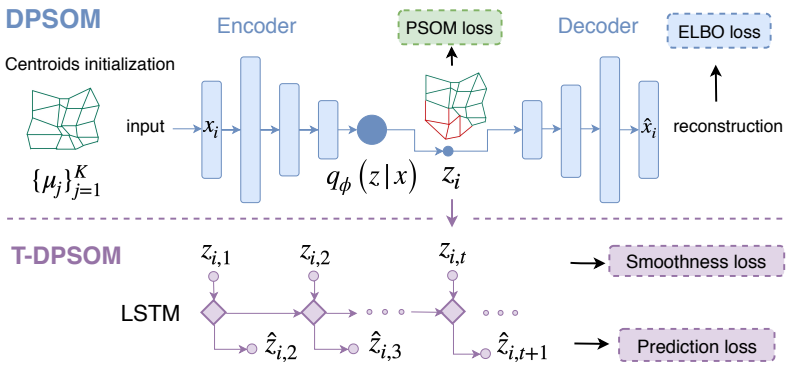


Figure 1: Model architectures of DPSOM and its temporal extension T-DPSOM. A data point x_i is mapped to a continuous embedding z_i using a VAE. In T-DPSOM, the embeddings $z_{i,t}$ for $t = 1, \ldots, T$ are connected by an LSTM, which predicts the embedding z_{t+1} of the next time step.

is dropped from the loss, thus effectively turning the used model into a standard autoencoder. Moreover, it employs a Markov model for the temporal dynamics. Both of these design choices yield inferior expressivity compared to our proposed method. The Deep Embedded SOM, DESOM [10], improved the previous model by using a Gaussian neighborhood window with exponential radius decay and by learning the SOM structure in a continuous setting. Both methods extract a topologically interpretable neighborhood structure and yield promising results in visualizing state spaces. However, those works did not include empirical comparisons to state-of-the-art deep clustering techniques. Moreover, they do not allow for a probabilistic interpretation of the cluster assignments. A concise overview of the differences between our proposed models and the related approaches is shown in Table 1.

3 PROBABILISTIC CLUSTERING WITH THE DPSOM

Given a set of data samples $\{x_i\}_{i=1,\ldots,N}$, where $x_i \in \mathbb{R}^d$, the goal is to partition the data into a set of clusters $\{S_j\}_{j=1,\ldots,K}$, while retaining a topological structure over the cluster centroids.

The proposed architecture is presented in Figure 1. The input vector x_i is embedded into a latent representation z_i using a VAE. This latent vector is then clustered using PSOM, our novel SOM clustering strategy for probabilistic cluster assignments. The VAE and PSOM are trained jointly to learn a latent representation, with the aim of improving the clustering performance. To model time series, we propose an architecture extension called T-DPSOM.

3.1 Background

A self-organizing map is comprised of K nodes $M = \{m_j\}_{j=1}^K$ tied by a neighborhood relation, where the node m_j corresponds to a centroid μ_j in the input space. Given a random initialization of the centroids, the SOM algorithm randomly selects an input x_i and updates both its closest centroid μ_j and its neighbors to move them closer to x_i. For a complete description of the SOM algorithm, we refer to the Appendix.

The Cluster Assignment Hardening (CAH) method has been recently introduced by the DEC model [52] and was shown to perform well in the latent space of AEs [1]. Given an embedding function $z_i = f(x_i)$, it uses a Student's t-distribution (S) as a kernel to measure the similarity between an embedded data point z_i, and a centroid μ_j. It improves the cluster purity by forcing the distribution S to approach a target distribution T:

$$s_{ij} = \frac{\left(1 + \left\|z_i - \mu_j\right\|^2 / \alpha\right)^{-\frac{\alpha+1}{2}}}{\sum_{j'} \left(1 + \left\|z_i - \mu_{j'}\right\|^2 / \alpha\right)^{-\frac{\alpha+1}{2}}}, \tag{1}$$

$$t_{ij} = \frac{s_{ij}^\kappa / \sum_{i'} s_{i'j}}{\sum_{j'} s_{ij'}^\kappa / \sum_{i'} s_{i'j'}}, \tag{2}$$

where α is the degrees of freedom of the Student's distribution. As it has been shown that learning α might be superfluous [45], we set it to 10 for all experiments. By taking the original distribution S to the power of κ and normalizing it, the target distribution puts more emphasis on data points that are assigned a high confidence and thus reduces the entropy of the distribution. Over the course of training, this lets the distribution approach a discrete cluster assignment (hence "hardening"). We follow Xie et al. [52] in choosing $\kappa = 2$, which leads to larger gradient contributions of points close to cluster centers, as they show empirically. The resulting clustering loss is defined as the Kullback-Leibler divergence between the target distribution, T, and the original distribution S:

$$\mathcal{L}_{\text{CAH}} = D_{KL}(T \| S) = \sum_{i=1}^{N} \sum_{j=1}^{K} t_{ij} \log \frac{t_{ij}}{s_{ij}}. \tag{3}$$

3.2 PSOM: Probabilistic SOM clustering

We propose a novel clustering method called Probabilistic SOM (PSOM), which extends the CAH method to include a SOM neighborhood structure over the centroids. We achieve this by combining (3) with a new objective function $\mathcal{L}_{\text{S-SOM}}$ (Soft SOM loss) to get an interpretable representation. This function maximizes the similarity between each data point and the neighbors of the closest centroids.

Therefore, it acts on soft cluster assignments, but still yields the qualitative behaviour of the SOM algorithm. The objective is presented in the following.

Given the set of K nodes, $M = \{m_j\}_{j=1}^{K}$, we define the neighborhood function as $N(j) = \{n_z(j)\}_{z=1}^{Z}$ for $j \in \{1, \dots, K\}$, where $n_z(j)$ returns the z-th neighbor's index of m_j. We require that $\cup_j m_{n_z(j)} = M$, which can for instance be achieved by using a toroid geometry for the map with one neighbor in each direction (the setting used in our experiments). Each node corresponds to a centroid μ_j in the latent space. We use s_{ij}, defined in (1), as the probability that data point z_i belongs to cluster centroid μ_j. We then define a loss that enforces a SOM-like neighborhood structure over the centroids as:

$$\begin{aligned}\mathcal{L}_{\text{S-SOM}} &= -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{K} s_{ij} \sum_{e\in N(j)} \log s_{ie} \\ &= \sum_{z=1}^{Z} -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{K} s_{ij} \log s_{in_z(j)} \ .\end{aligned}$$

Intuitively, the objective encourages that if s_{ij} is large, then the s_{ie}'s should also be large and vice versa. Hence, this procedure leads to the same topological neighborhood properties as the classical SOM, while using soft cluster assignments. Interestingly, this loss can also be seen as a sum of Z cross-entropies between the probability distribution of each centroid and the probability distribution of its respective z-th neighbor centroid. Note that $\sum_j s_{in_z(j)} = 1$ because of the union property in the definition above. The complete PSOM clustering loss is then defined as:

$$\mathcal{L}_{\text{PSOM}} = \mathcal{L}_{\text{CAH}} + \beta \mathcal{L}_{\text{S-SOM}} \ , \tag{4}$$

which for $\beta = 0$ becomes equivalent to Cluster Assignment Hardening. We later show empirically that this novel objective does indeed lead to SOM-like behaviour and thus offers a viable way to fit self-organizing maps with probabilistic cluster assignments. Note that the parameter β can be chosen freely as a trade-off between pure clustering performance and spatial coherence, where increasing β improves the smoothness of the learned map (see Fig. 2).

3.3 DPSOM: VAE for representation learning

To increase the expressivity of the PSOM, we apply the clustering in the latent space of a deep representation learning model. In our method, this nonlinear mapping between the input x_i and embedding z_i is realized by a VAE. Instead of directly mapping the input x_i to a latent embedding z_i, the VAE learns a probability distribution $q_\phi(z_i \mid x_i)$ parameterized as a multivariate normal distribution with mean and covariance $(\mu_\phi, \Sigma_\phi) = f_\phi(x_i)$. Similarly, it also learns the probability distribution of the reconstructed output given a sampled latent embedding $p_\theta(x_i \mid z_i)$, where $(\mu_\theta, \Sigma_\theta) = f_\theta(z_i)$. Both f_ϕ and f_θ are neural networks, which are respectively called encoder and decoder. The VAE loss (ELBO) is:

$$\begin{aligned}\mathcal{L}_{\text{VAE}} = \sum_{i=1}^{N} \Big[&- \mathbb{E}_{q_\phi(z|x_i)}(\log p_\theta(x_i \mid z)) \\ &+ D_{KL}(q_\phi(z \mid x_i) \parallel p(z)) \Big] \ ,\end{aligned} \tag{5}$$

where $p(z)$ is an isotropic Gaussian prior over the latent embeddings. The second term can be interpreted as a form of regularization, which encourages the latent space to be compact. For each data point x_i, the latent embedding z_i is sampled from $q_\phi(z \mid x_i)$. Adding the VAE loss to the PSOM loss from the previous subsection, we get the overall loss function of the DPSOM:

$$\mathcal{L}_{\text{DPSOM}} = \gamma \mathcal{L}_{\text{CAH}} + \beta \mathcal{L}_{\text{S-SOM}} + \mathcal{L}_{\text{VAE}} \ , \tag{6}$$

where γ regulates the tradeoff between reconstruction and clustering performances while β is as above. For an in-depth discussion of the choice of these parameters and of our model's robustness to different parameter configurations, we refer to the Appendix. To the best of our knowledge, no previous SOM method attempted to use a VAE to embed the inputs into a latent space. Yet, the VAE is a natural choice, since the compactness of the representations encouraged by the Gaussian prior fits the neighborhood assumptions of the SOM algorithm (see Appendix).

3.4 T-DPSOM: Extension to time series data

To extend our proposed model to time series data, we add a temporal component to the architecture, yielding the Temporal DPSOM (T-DPSOM). Given a set of N time series of length T, $\{x_{i,t}\}_{i=1,\dots,N;t=1,\dots,T}$, the goal is to learn interpretable trajectories on the SOM grid. To do so, the DPSOM could be used directly but it would treat each time step t of the time series independently. To exploit temporal information and enforce smoothness in the trajectories, we design an additional loss term, which is similar to the smoothness loss in the SOM-VAE [11], but is able to act on soft assignments:

$$\mathcal{L}_{\text{smooth}} = -\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} u_{i_t,i_{t+1}} \ , \tag{7}$$

where $u_{i_t,i_{t+1}} = g(z_{i,t}, z_{i,t+1})$ is the similarity between $z_{i,t}$ and $z_{i,t+1}$ using a Student's t-distribution and $z_{i,t}$ refers to the embedding of time series x_i at time index t. It maximizes the similarity between latent embeddings of adjacent time steps, such that large jumps in the latent state between time points are discouraged. This is motivated by the intuition that the true underlying factors of variation in real-world applications usually vary smoothly over time and can be seen as being similar to a Kalman filter prior [28].

One of the main goals in time series modeling is to predict future data points, or alternatively, future embeddings. This can be achieved by adding a long short-term memory network (LSTM) [20] over the latent embeddings of the time series, as shown in Fig 1. Each cell of the LSTM takes as input the latent embedding z_t at time step t, and predicts a probability distribution over the next latent embedding, $p_\omega(z_{t+1} \mid z_t)$. We parameterize this distribution as a multivariate Gaussian distribution where the mean and variance are learnt by the LSTM. The prediction loss is the negative log-likelihood of a sample of the next embedding z_{t+1} under this distribution:

$$\mathcal{L}_{\text{pred}} = -\sum_{i=1}^{N}\sum_{t=1}^{T} \log p_\omega(z_{t+1} \mid z_t) \ . \tag{8}$$

The final loss of the T-DPSOM, which is trainable in a fully end-to-end fashion, is

$$\mathcal{L}_{\text{T-DPSOM}} = \mathcal{L}_{\text{DPSOM}} + \mathcal{L}_{\text{smooth}} + \mathcal{L}_{\text{pred}} \ . \tag{9}$$

Table 2: Clustering performance of DPSOM using 64 clusters arranged in a 8×8 SOM map, compared with baselines. Means and standard errors are computed across 10 runs with different random model initializations. *The DESOM results are taken from [10] which did not provide errors. **The IDEC does not provide pretrained weights for Fashion MNIST.

	MNIST		fMNIST	
	pur	*nmi*	*pur*	*nmi*
K-means	0.845 ± 0.001	0.581 ± 0.001	0.716 ± 0.001	0.514 ± 0.000
VQ-VAE	0.515 ± 0.005	0.354 ± 0.003	0.594 ± 0.003	0.468 ± 0.001
DEC	0.944 ± 0.002	0.682 ± 0.001	0.758 ± 0.002	**0.562 ± 0.001**
IDEC **	0.950 ± 0.001	0.681 ± 0.001	-	-
SOM	0.701 ± 0.005	0.539 ± 0.002	0.667 ± 0.003	0.525 ± 0.001
AE+SOM	0.874 ± 0.004	0.646 ± 0.001	0.706 ± 0.002	0.543 ± 0.001
SOM-VAE	0.868 ± 0.004	0.595 ± 0.004	0.739 ± 0.005	0.520 ± 0.003
DESOM*	0.939 ± N/A	0.657 ± N/A	0.752 ± N/A	0.538 ± N/A
DPSOM (ours)	**0.968 ± 0.001**	**0.701 ± 0.001**	**0.779 ± 0.003**	**0.562 ± 0.001**
DPSOM \ VAE (ablation)	0.813 ± 0.004	0.561 ± 0.002	0.730 ± 0.006	0.530 ± 0.003

This combined objective encourages the learned representations and clusters to preserve similarity between inputs in its topological structure (through the first term), while also preserving smoothness over time (through the second term), and learning representations that are informative about the future of the trajectory (through the third term). It therefore ensures usefulness of the representations for clustering, time series visualization, and time series forecasting. In the following, we will separately evaluate these three use cases empirically.

4 EXPERIMENTS

Firstly, we evaluate the DPSOM and compare its clustering performance to a wide range of state-of-the-art deep and SOM-based clustering methods, on MNIST [29] and Fashion-MNIST [51] data. We then present extensive evidence of the performance of the T-DPSOM on medical time series from the eICU data set [35] on several relevant tasks. Moreover, we discuss how our method increases the spatial coherence of the clusters and we analyze enrichment of organ function parameters and mortality risk in different regions of the SOM map. Lastly, we illustrate how the probabilistic assignments of T-DPSOM enable interpretable visualizations of medical time series with uncertainty estimation. Open source code, with which the most important results and figures can be reproduced, is available at https://github.com/ratschlab/dpsom.

4.1 Model implementation

In implementing our models we focused on retaining a fair comparison with the baselines. Hence, we decided to use a standard network structure, with fully connected layers of dimensions $d - 500 - 500 - 2000 - l$, to implement both the VAEs and the AEs. The latent dimension l is set to 100 for the VAEs, and to 10 for the AEs of the baselines. We set the number of clusters to 64 for the image data and to 256 for the ICU time series for both the baselines and our models. We highlight the fact that our model does not only focus on clustering, but is also a visualization aid. Therefore, we use a much larger amount of clusters than the expected number clusters. We choose our SOM grid to be the 2-dimensional surface of a 3-dimensional torus in all experiments. The neighborhood is chosen to contain one neighboring cluster in each direction (up, down, left, right). For different configurations we refer to the Appendix. For all architectures, no greedy layer-wise pretraining was used to tune the VAE. Instead we simply run the VAE without the clustering loss for a few epochs as initialization. A standard SOM was used to initialize the centroids. Finally, the entire architecture is trained for $100,000$ iterations. To avoid fine-tuning the hyperparameters, β is chosen such that the S-SOM loss has a similar magnitude as the CAH loss and γ follows a simple heuristic which regulates the trade-off between the CAH and the ELBO loss components. Crucially, these simple strategies for choosing the hyperparameters are completely unsupervised and do not rely on ground-truth label information or any expensive optimization. Additionally, the model is shown to be robust under different configurations of β and γ. For an overview of the hyper-parameter setting for both static and temporal data we refer to the Appendix.

4.2 Static image clustering with the DPSOM

As a sanity check on static image data, we evaluate the DPSOM clustering module against various baselines, before proceeding to medical time series. We used two different categories of baselines. The first category contains clustering methods that do not provide any interpretable discrete latent representation. Those include k-means, the well-known DEC model [52], which sequentially applies embedding learning using Stacked Autoencoders (SAE) and the Cluster Assignment Hardening method on the obtained representations, as well as its improved version IDEC [18]. We also include the VQ-VAE [44], which formed the basis for the SOM-VAE model [11]. In the second category, we include state-of-the-art clustering methods based on SOMs. Here, we used a standard SOM, AE+SOM (an architecture composed of an AE and a SOM applied on top of the latent representation, trained sequentially), SOM-VAE [11], and DESOM [10].

Table 2 shows the clustering performance of DPSOM on MNIST and Fashion-MNIST data, compared with the baselines. Purity and

Table 3: Mean NMI and standard error of cluster enrichment for current/future APACHE physiology scores, using a 16×16 SOM map, across 10 runs with different random model initializations.

Model	APACHE-24	APACHE-12	APACHE-6	APACHE-0
SOM	0.0308 ± 0.0013	0.0322 ± 0.0013	0.0334 ± 0.0014	0.0380 ± 0.0016
AE+SOM	0.0278 ± 0.0010	0.0283 ± 0.0009	0.0290 ± 0.0011	0.0321 ± 0.0014
DESOM	0.0847 ± 0.0019	0.0847 ± 0.0019	0.0816 ± 0.0019	0.0868 ± 0.0022
SOM-VAE	0.0824 ± 0.0008	0.0758 ± 0.0007	0.0743 ± 0.0007	0.0803 ± 0.0008
HMM	0.0533 ± 0.0007	0.0463 ± 0.0007	0.0427 ± 0.0007	0.0419 ± 0.0008
DPSOM	0.0919 ± 0.0037	0.0843 ± 0.0031	0.0816 ± 0.0029	0.0875 ± 0.0033
T-DPSOM	**0.1115 ± 0.0006**	**0.10220 ± 0.0005**	**0.0989 ± 0.0004**	**0.1065 ± 0.0005**

Table 4: MSE for predicting the time series of the last 6 hours before ICU dispatch, given the prior time series, across 10 runs with different random model initializations.

Model	LSTM	HMM	SOM-VAE	T-DPSOM
MSE	0.0113 ± 0.0002	0.0146 ± 0.0001	0.0081 ± 0.0001	**0.0049 ± 0.0001**

Normalized Mutual Information are used as evaluation metrics. We observe that our proposed model outperforms the baselines in terms of both metrics on both data sets. Interestingly, the DPSOM not only improves interpretability through the use of the latent PSOM, but also increases the performance compared to DEC/IDEC. Finally, we performed an ablation study to investigate the effect of the VAE in our model. We exchanged the VAE with an AE (DPSOM \ VAE of Table 2) and we observe a significant decrease in performance. This fits our intuition that the Gaussian prior of the VAE encourages a compact latent representation which favors the SOM neighbourhood assumptions.

4.3 eICU dataset preprocessing

The performance of our proposed models is mainly evaluated on the eICU data set [35], which is comprised of multivariate medical time series from the intensive care unit (ICU). Hereby, we use vital sign and lab measurements of ICU patients resampled to a 1-hour based grid using forward filling and filling with population statistics from the training set if no measurements were available prior to the time point. From all available ICU stays, we excluded stays which were shorter than 3 days, longer than 30 days, or which had at least one gap in the continuous vital sign monitoring, which we define as an interval between two heart rate measurements of at least 1 hour. This yielded $N = 10,559$ ICU stays from the eICU database [35]. We included $d_{\text{vitals}} = 14$ vital sign variables and $d_{\text{lab}} = 84$ lab measurement variables, giving an overall data dimension of $d = 98$. To aid reproducibility, included variables are listed in the Appendix. The last 72 hours of these multivariate time series were used for the experiments. As labels, we use a variant of the current dynamic APACHE physiology score (APACHE-0) as well as the worst APACHE score in the next $\{6, 12, 24\}$ hours (APACHE-6/12/24), current lab and vital values and mortality in the next 24 hours. Only those variables from the APACHE score definition which are recorded in the eICU database were taken into account for its definition. The dataset was divided into training, validation, and test sets for both our models and the baselines.

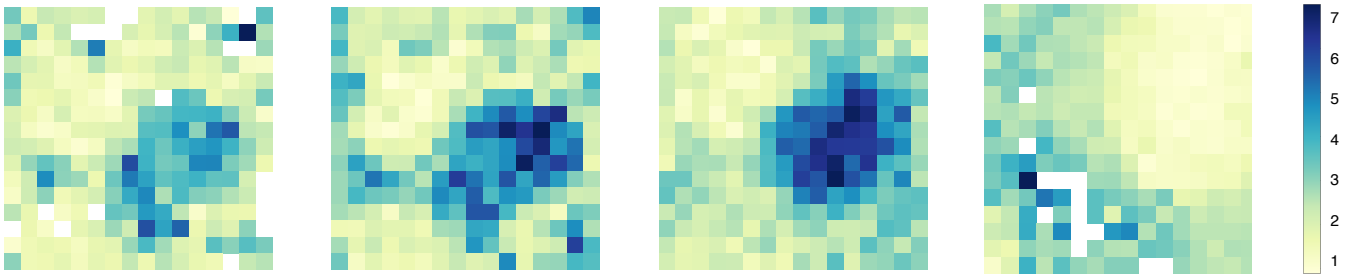
4.4 Clustering and forecasting of patient health states

We compare the clustering performance of our proposed models against the state-of-the-art clustering methods based on SOMs used in Section 4.2 and a Hidden Markov Model. Table 3 shows the cluster enrichment in terms of NMI for four different labels, the current (APACHE-0) and worst future (APACHE-6/12/24) physiology scores in the next 6, 12, 24 hours, respectively. The T-DPSOM clearly achieves superior clustering performance compared to the baselines.

To quantify the performance of T-DPSOM in predicting future trajectories, we predict the final six latent embeddings of each time series, conditioned on all previous time steps. For each predicted embedding, we reconstruct the input using the decoder of the VAE. Finally, we measure the mean squared error (MSE) between the original inputs and the reconstructed inputs for the last six hours of the ICU stay. As baselines, we use an LSTM, an HMM, and the SOM-VAE. The LSTM predicts the future time steps directly in the input space. The HMM predicts by sampling future time steps using its learned transition and emission matrices, starting from the most likely current state. The results (Table 4) indicate that the joint training of clustering and prediction used by T-DPSOM clearly outperforms the baselines. Training times of the T-DPSOM model are reported in the Appendix.

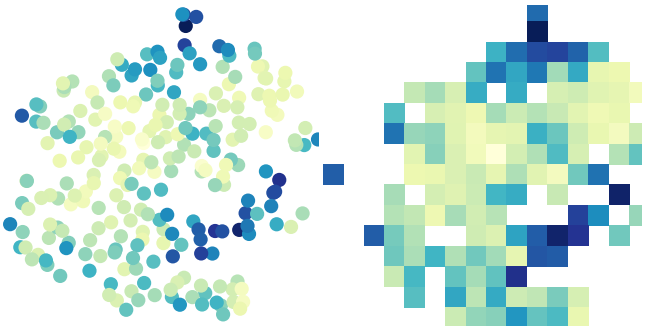
4.5 Spatial coherence through S-SOM loss

The main objective of the PSOM is to enforce a SOM-like structure between the cluster centroids. Thus, we achieve an interpretable 2-D representation of the data in which neighboring centroids should exhibit similar characteristics. To illustrate the topological structure in the latent space and to qualitatively show the effect of the S-SOM loss, we include heatmaps that show the enrichment of cells for the current APACHE physiology score for different values of β in Figure 2(a)-(c). We notice that centroids with a similar APACHE score are more tightly clustered in neighboring centroids for higher values of β.



(a) $I_{\text{moran}} = 0.626$ **(b)** $I_{\text{moran}} = 0.749$ **(c)** $I_{\text{moran}} = 0.764$ **(d)** $I_{\text{moran}} = 0.530$

Figure 2: Visualizations of the SOM grid heat-maps obtained by training T-DPSOMs with (a) $\beta = 0$, (b) $\beta = 10$, (c) $\beta = 100$ and (d) by training SOM-VAE. We see that increasing β increases the correlation between neighboring clusters, which is also shown quantitatively using Moran's index (I_{moran}) and I_d. For all experiments we used the same pretraining weights obtained by running the VAE without the clustering loss for a few epochs and by using a standard SOM to initialize the centroids.



(a) t-SNE **(b)** $I_{\text{moran}} = 0.46$

Figure 3: (a) and (b), t-SNE and gridded t-SNE representations of the latent space when using $\beta = 0$ on eICU data. The colors correspond to the mean APACHE scores.

To assess this property quantitatively, we use Moran's index [34] as a measure of spatial correlation among clusters. Moran's index can be defined as:

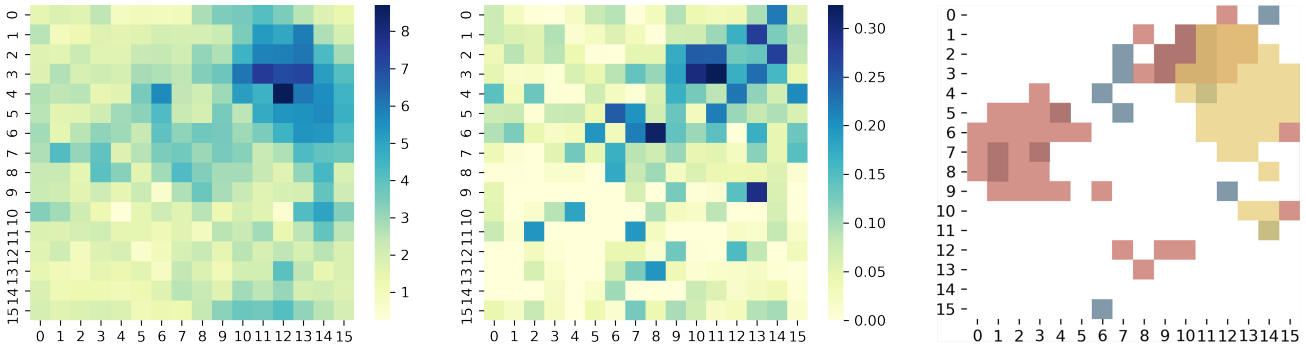
$$I_{\text{moran}} = \frac{N}{W} \frac{\sum_i \sum_j w_{ij} \left(y_i - \bar{y}\right) \left(y_j - \bar{y}\right)}{\sum_i \left(y_i - \bar{y}\right)^2}$$

where N is the number of clusters indexed by i and j, y_i, c_i are continuous cell labels, $\bar{y}$ is the mean of y, $w_{i,j}$ is a matrix of spatial weights and W is the sum of all $w_{i,j}$. We define $w_{i,j} = \exp(-d_{\text{SOM}}(i,j))$, where $d_{\text{SOM}}(i,j)$ is the Manhattan distance between the nodes m_i and m_j in the SOM. For the eICU data-set we use the mean APACHE score as a cell label to express similarities between patient states. We compute Moran's index for different values of the parameter β in Equation 4. We see in Figure 2(a)-(c) that increasing the β coefficient, and thus the relative weight of the S-SOM loss, increases the correlations between neighboring clusters, both visually and quantitatively. We can thus conclude that the S-SOM loss does indeed encourage the latent representations to assume a SOM-like structure, and that the spatial coherence of the clustering can be controlled via β. As a comparison, we computed the Moran's index of the SOM map learned by SOM-VAE in Fig 2(d). It yields a lower Moran's index compared to T-DPSOM, indicating that our model encourages a smoother representation of the latent space than SOM-VAE.

Finally, we compare the spatial coherence of the learned SOM-based map with a *post hoc* visualization of the latent embeddings generated when the soft-SOM loss is disabled ($\beta = 0$). We use the popular visualization method t-SNE as a comparison. In Figure 3a/3b, we show both the original t-SNE plot, and a gridded t-SNE which allows us to compare Moran's index. We see that t-SNE obtains much lower spatial coherence, compared to our method. Moreover, our method enjoys other advantages over t-SNE, which is not trainable end-to-end and does not yield a regular geometric structure (like a grid). Further, it has to be retrained as soon as unseen test data is used and scales poorly with large data-set sizes.

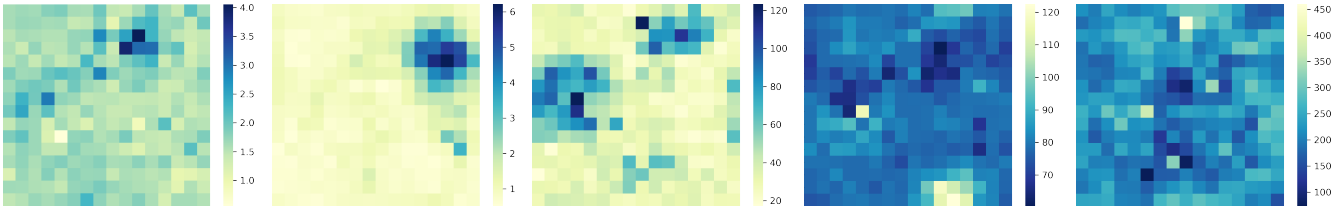
4.6 Cluster enrichment for organ function parameters and mortality risk

In Fig. 4 (a) and (b), heatmaps (colored according to enrichment in the current APACHE score, as well as the mortality risk in the next 24 hours) show compact enrichment structures. The T-DPSOM model succeeds in discovering a meaningful and smooth neighborhood structure with respect to APACHE score enrichment. For mortality, it distinguishes risk profiles with practically zero mortality risk from very high mortality risk in the next 24 hours (reaching up to 30-40%) in different regions of the map, even though it is trained in a purely unsupervised fashion, which is a remarkable result. In Fig. 4 (c), the clusters with the highest 33% (z>0.43) average values per key organ lab value (yellow: Creatinine, red: alanine aminotransferase (ALT), grey: Lactate) are marked by the respective color. The chosen lab values can be seen as key variables for specific organ systems (Creatinine: Kidney, ALT: Liver, Lactate: Circulatory system). For each lab value, the highest values concentrate in a different region on the SOM map. At the same time, clusters with overlapping high values for different lab values exist in all combinations. For example, the cluster (x=14,y=5) has an isolated high kidney lab value; (x=13,y=2), on the other hand, has high kidney and liver values; finally, for the cluster (x=11,y=2), the values for liver, kidney and circulatory lab values are high. The detailed average value distribution per organ system can also be observed



(a) Current APACHE score (b) Mortality risk in the next 24 hours (c) Organ clusters

Figure 4: (a)+(b) Heatmaps of enrichment of the current dynamic APACHE score as well as in mortality risk in the next 24 hours. (c) Regions with the 33% (z>0.43) highest average values per key organ lab value (yellow: Creatinine, red: Alanine aminotransferase (ALT), grey: Lactate). E.g. the patient timepoints assigned to cluster point (x=13,y=2) have on average kidney and liver values in the top 33% of all measurements. For all plots the information is superimposed on the discrete 2D grid learned by the T-DPSOM model.



(a) Lactate [mmol/l] (b) Creatinine [mg/dL] (c) ALT [U/l] (d) MAP [mmHg] (e) P/F ratio [cmH_2O]

Figure 5: Heatmaps of enrichment of key lab vital and lab values, superimposed on the discrete 2D grid learned by the T-DPSOM model. ALT: Alanine aminotransferase, MAP: Mean arterial pressure

in Fig. 5. In this figure, we can also observe that not only regions of high but also of low values exist.

4.7 Visualization of health state trajectories with uncertainty estimation

In Fig. 6, we show how the T-DPSOM could help to distill complex multivariate time series into an intuitive representation. In particular, we show two example trajectories on the SOM grid learned by the T-DPSOM method. The red trajectory represents a patient that died at the end of the ICU stay, while the green one is a control patient who was dispatched alive from the ICU. For a more quantitative assessment, in the Appendix we show more patient trajectories obtained by our model. We observe that the trajectories are located in different parts of the SOM grid, and that their directions of movement fit the intuition when combined with the average physiology scores of each cluster (Fig. 6). One of the advantages of the T-DPSOM over the SOM-VAE algorithm is the use of soft assignments of data points to clusters, which results in the ability to quantify uncertainty in the clustering. For interpreting health states in the ICU, this property is very important [22]. In Fig 6, additionally to the patient trajectories, we show the probability distributions over cluster assignments at different time steps. Our model yields a soft centroid-based probability distribution which evolves over time and which allows estimation of likely discrete health states at any given point in time. For each time step, the distribution of probabilities is plotted using a blue color shading, whereas the overall trajectory is plotted using a solid line.

We see that the assigned probabilities fit well to the intuition that neighboring clusters should be harder to distinguish than more separated ones. Moreover, neighboring clusters with larger assigned probability can sometimes forebode the movement direction of the trajectory, suggesting that the combination of SOM-loss and temporal losses leads to a representation that is smooth in space as well as in time.

5 CONCLUSION

We presented two novel methods for interpretable unsupervised clustering on static and temporal data, DPSOM and T-DPSOM. Both models make use of a VAE and a novel probabilistic clustering method, PSOM, that extends the classical SOM algorithm to include centroid-based probability distributions. They achieve superior clustering performance compared to state-of-the-art deep clustering baselines on image data sets and medical time series. The use of probabilistic assignments of data points to clusters, and the use of a VAE for feature extraction, instead of an AE as used

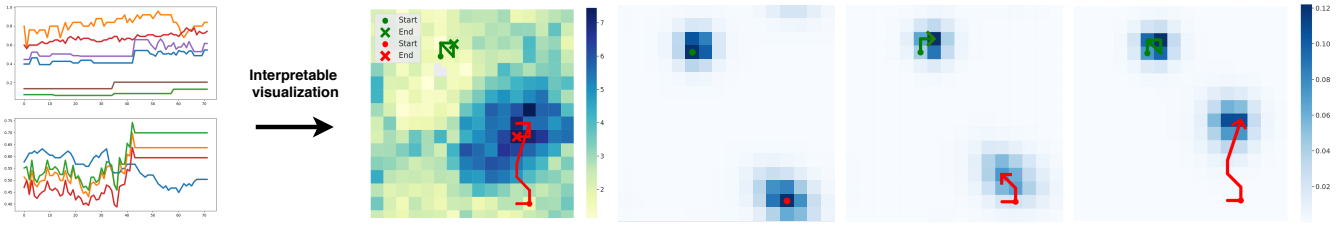


Figure 6: Illustration of two example patient trajectories in the SOM grid of T-DPSOM. One patient died (red), while the other was dispatched alive from the ICU (green). Superimposed is a heatmap which displays mean APACHE score of all time points assigned to each cluster. We observe qualitative differences in the trajectories of the dying and the surviving patient. For each time series we also show the assigned probabilities to the discrete patient health states using a blue color shading.

in previous methods, results in an interpretable model that can quantify uncertainty in the clustering as well as in its predictions of future time series states.

When applied to medical time series, this new methodology improves clustering performance and allows for visualizations of patients' health states. The former can further drive the current developments in personalized medicine by more comprehensive patient phenotypes. This could lead to a better understanding of the pathophysiology of disease and facilitate future therapeutic trials by more homogeneous patient cohorts of likely treatment responders. Intuitive cluster assignment visualizations could improve the grasp of the current patient health state, improving situational awareness, while reducing workload. The next step forward is therefore to apply T-DPSOM to specific patient cohorts (e.g. sepsis) and investigate the derived clusters and trajectories in even greater detail by physicians.

REFERENCES

[1] Elie Aljalbout, Vladimir Golkov, Yawar Siddiqui, and Daniel Cremers. 2018. Clustering with Deep Learning: Taxonomy and New Methods. *CoRR* abs/1801.07648 (2018). arXiv:1801.07648 http://arxiv.org/abs/1801.07648

[2] Pavan K Bhatraju, Leila R Zelnick, Jerald Herting, Ronit Katz, Carmen Mikacenic, Susanna Kosamo, Eric D Morrell, Cassianne Robinson-Cohen, Carolyn S Calfee, Jason D Christie, Kathleen D Liu, Michael A Matthay, William O Hahn, Victoria Dmyterko, Natalie S J Slivinski, Jim A Russell, Keith R Walley, David C Christiani, W Conrad Liles, Jonathan Himmelfarb, and Mark M Wurfel. 2019. Identification of Acute Kidney Injury Subphenotypes with Differing Molecular Signatures and Responses to Vasopressin Therapy. *Am. J. Respir. Crit. Care Med.* 199, 7 (April 2019), 863–872.

[3] Christopher M Bishop, Markus Svensén, and Christopher KI Williams. 1998. GTM: The generative topographic mapping. *Neural computation* 10, 1 (1998), 215–234.

[4] L D Bos, L R Schouten, L A van Vught, M A Wiewel, D S Y Ong, O Cremer, A Artigas, I Martin-Loeches, A J Hoogendijk, T van der Poll, J Horn, N Juffermans, C S Calfee, M J Schultz, and MARS consortium. 2017. Identification and validation of distinct biological phenotypes in patients with acute respiratory distress syndrome by cluster analysis. *Thorax* 72, 10 (Oct. 2017), 876–883.

[5] T. Buchman. 2010. Novel representation of physiologic states during critical illness and recovery. *Critical Care* 14 (2010), 127 – 127.

[6] Carolyn S Calfee, Kevin Delucchi, Polly E Parsons, B Taylor Thompson, Lorraine B Ware, Michael A Matthay, and NHLBI ARDS Network. 2014. Subphenotypes in acute respiratory distress syndrome: latent class analysis of data from two randomised controlled trials. *Lancet Respir Med* 2, 8 (Aug. 2014), 611–620.

[7] Geoffrey J. Chappell and John G. Taylor. 1993. The Temporal Kohonen Map. *Neural Netw.* 6, 3 (March 1993), 441–445. https://doi.org/10.1016/0893-6080(93)90011-K

[8] Shih-Sian Cheng, Hsin-Chia Fu, and Hsin-Min Wang. 2009. Model-based clustering by probabilistic self-organizing maps. *IEEE Transactions on Neural Networks* 20, 5 (2009), 805–826.

[9] Arthur Flexer. 1999. On the Use of Self-Organizing Maps for Clustering and Visualization. In *Principles of Data Mining and Knowledge Discovery*, Jan M. Żytkow and Jan Rauch (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 80–88.

[10] Florent Forest, Mustapha Lebbah, Hanane Azzag, and Jérôme Lacaille. 2019. Deep Embedded SOM: Joint Representation Learning and Self-Organization. In *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2019)*. 1–6.

[11] Vincent Fortuin, Matthias Hüser, Francesco Locatello, Heiko Strathmann, and Gunnar Rätsch. 2018. SOM-VAE: Interpretable Discrete Representation Learning on Time Series. *arXiv preprint arXiv:1806.02199* (2018).

[12] Vincent Fortuin, Gunnar Rätsch, and Stephan Mandt. 2019. Multivariate Time Series Imputation with Variational Autoencoders. *arXiv preprint arXiv:1907.04155* (2019).

[13] Vincent Fortuin, Heiko Strathmann, and Gunnar Rätsch. 2019. Meta-Learning Mean Functions for Gaussian Processes. *arXiv preprint arXiv:1901.08098* (2019).

[14] Jean-Yves Franceschi, Aymeric Dieuleveut, and Martin Jaggi. 2019. Unsupervised Scalable Representation Learning for Multivariate Time Series. *arXiv preprint arXiv:1901.10738* (2019).

[15] Bengt Gårdlund, Natalia O Dmitrieva, Carl F Pieper, Simon Finfer, John C Marshall, and B Taylor Thompson. 2018. Six subphenotypes in septic shock: Latent class analysis of the PROWESS Shock study. *J. Crit. Care* 47 (Oct. 2018), 70–79.

[16] S. Van Gassen, B. Callebaut, Mary J. van Helden, B. Lambrecht, P. Demeester, T. Dhaene, and Y. Saeys. 2015. FlowSOM: Using self-organizing maps for visualization and interpretation of cytometry data. *Cytometry Part A* 87 (2015).

[17] Prasoon Goyal, Zhiting Hu, Xiaodan Liang, Chenyu Wang, and Eric P Xing. 2017. Nonparametric variational auto-encoders for hierarchical representation learning. In *Proceedings of the IEEE International Conference on Computer Vision*. 5094–5102.

[18] Xifeng Guo, Long Gao, Xinwang Liu, and Jianping Yin. 2017. Improved Deep Embedded Clustering with Local Structure Preservation. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*. 1753–1759. https://doi.org/10.24963/ijcai.2017/243

[19] C. Higuera, K. Gardiner, and K. Cios. 2015. Self-Organizing Feature Maps Identify Proteins Critical to Learning in a Mouse Model of Down Syndrome. *PLoS ONE* 10 (2015).

[20] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.

[21] Thomas Hofmann, B. Scholkopf, and Alex Smola. 2008. Kernel methods in machine learning. *Annals of Statistics* 36 (2008), 1171–1220.

[22] Donna L Hudson and Maurice E Cohen. 2010. Uncertainty and complexity in personal health records. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*. IEEE, 6773–6776.

[23] Ian Jolliffe. 2011. *Principal Component Analysis*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1094–1096. https://doi.org/10.1007/978-3-642-04898-2_455

[24] Daniel B Knox, Michael J Lanspa, Kathryn G Kuttler, Simon C Brewer, and Samuel M Brown. 2015. Phenotypic clusters within sepsis-associated multiple organ dysfunction syndrome. *Intensive Care Med.* 41, 5 (May 2015), 814–822.

[25] T. Kohonen. 1990. The self-organizing map. *Proc. IEEE* 78, 9 (Sep. 1990), 1464–1480. https://doi.org/10.1109/5.58325

[26] Teuvo Kohonen. 1995. The Adaptive-Subspace SOM (ASSOM) and its use for the Implementation of Invariant Feature Detection.

[27] Andreas Kopf, Vincent Fortuin, Vignesh Ram Somnath, and Manfred Claassen. 2019. Mixture-of-Experts Variational Autoencoder for clustering and generating from similarity-based representations. *arXiv preprint arXiv:1910.07763* (2019).

[28] Rahul G Krishnan, Uri Shalit, and David Sontag. 2016. Deep Kalman Filters. (2016).

[29] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. 1998. Gradient-based learning applied to document recognition. *Proc. IEEE* 86, 11 (Nov 1998), 2278–2324. https://doi.org/10.1109/5.726791

[30] N. Liu, J. Wang, and Y. Gong. 2015. Deep Self-Organizing Map for visual classification. In *2015 International Joint Conference on Neural Networks (IJCNN)*. 1–6.

https://doi.org/10.1109/IJCNN.2015.7280357
[31] Ezequiel López-Rubio. 2010. Probabilistic self-organizing maps for continuous data. *IEEE Transactions on Neural Networks* 21, 10 (2010), 1543–1554.
[32] T. A. McQueen, A. A. Hopgood, J. A. Tepper, and T. J. Allen. 2004. A Recurrent Self-Organizing Map for Temporal Sequence Processing. In *Applications and Science in Soft Computing*, Ahamad Lotfi and Jonathan M. Garibaldi (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 3–8.
[33] Paulo Afonso Mei, Cleyton de Carvalho Carneiro, Stephen J. Fraser, Li Li Min, and Fabiano Reis. 2015. Analysis of neoplastic lesions in magnetic resonance imaging using self-organizing maps. *Journal of the Neurological Sciences* 359, 1 (2015), 78 – 83. https://doi.org/10.1016/j.jns.2015.10.032
[34] P. A. P. Moran. 1950. NOTES ON CONTINUOUS STOCHASTIC PHENOMENA. *Biometrika* 37, 1-2 (06 1950), 17–23. https://doi.org/10.1093/biomet/37.1-2.17 arXiv:https://academic.oup.com/biomet/article-pdf/37/1-2/17/487420/37-1-2-17.pdf
[35] Tom J Pollard, Alistair EW Johnson, Jesse D Raffa, Leo A Celi, Roger G Mark, and Omar Badawi. 2018. The eICU Collaborative Research Database, a freely available multi-center database for critical care research. *Scientific data* 5 (2018).
[36] Ali Razavi, Aaron van den Oord, and Oriol Vinyals. 2019. Generating Diverse High-Fidelity Images with VQ-VAE-2. *arXiv preprint arXiv:1906.00446* (2019).
[37] Kiran Reddy, Pratik Sinha, Cecilia M O'Kane, Anthony C Gordon, Carolyn S Calfee, and Daniel F McAuley. 2020. Subphenotypes in critical care: translation into clinical practice. *Lancet Respir Med* 8, 6 (June 2020), 631–643.
[38] Margherita Rosnati and Vincent Fortuin. 2019. MGP-AttTCN: An Interpretable Machine Learning Model for the Prediction of Sepsis. *arXiv preprint arXiv:1909.12637* (2019).
[39] P. Tamayo, D. Slonim, J. Mesirov, Q. Zhu, S. Kitareewan, E. Dmitrovsky, E. Lander, and T. Golub. 1999. Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proceedings of the National Academy of Sciences of the United States of America* 96 6 (1999), 2907–12.
[40] W. Teng and Ping-Lin Chang. 2011. Identifying Regions of Interest in Medical Images Using Self-Organizing Maps. *Journal of Medical Systems* 36 (2011), 2761–2768.
[41] S. Tirunagari, N. Poh, K. Aliabadi, D. Windridge, and D. Cooke. 2014. Patient level analytics using self-organising maps: A case study on Type-1 Diabetes self-care survey responses. In *2014 IEEE Symposium on Computational Intelligence and Data Mining (CIDM)*. 304–309. https://doi.org/10.1109/CIDM.2014.7008682
[42] Kazuhiro Tokunaga and Tetsuo Furukawa. 2009. Modular network SOM. *Neural Networks* 22, 1 (2009), 82 – 90. https://doi.org/10.1016/j.neunet.2008.10.006
[43] David Werner Tscholl, Julian Rössler, Sadiq Said, Alexander Kaserer, Donat Rudolf Spahn, and Christoph Beat Nöthiger. 2020. Situation Awareness-Oriented Patient Monitoring with Visual Patient Technology: A Qualitative Review of the Primary Research. *Sensors* 20, 7 (April 2020).
[44] Aäron van den Oord, Oriol Vinyals, and Koray Kavukcuoglu. 2017. Neural Discrete Representation Learning. *CoRR* abs/1711.00937 (2017). arXiv:1711.00937 http://arxiv.org/abs/1711.00937
[45] Laurens van der Maaten. 2009. Learning a Parametric Embedding by Preserving Local Structure. In *Proceedings of the Twelth International Conference on Artificial Intelligence and Statistics (Proceedings of Machine Learning Research, Vol. 5)*, David van Dyk and Max Welling (Eds.). PMLR, Hilton Clearwater Beach Resort, Clearwater Beach, Florida USA, 384–391. http://proceedings.mlr.press/v5/maaten09a.html
[46] Sharad Vikram, Matthew D Hoffman, and Matthew J Johnson. 2018. The loracs prior for vaes: Letting the trees speak for the data. *arXiv preprint arXiv:1810.06891* (2018).
[47] Jean-Louis Vincent. 2017. The coming era of precision medicine for intensive care. *Crit. Care* 21, Suppl 3 (Dec. 2017), 314.
[48] Thomas Voegtlin. 2002. Recursive self-organizing maps. *Neural Networks* 15, 8 (2002), 979 – 991. https://doi.org/10.1016/S0893-6080(02)00072-2
[49] Jennifer G Wilson and Carolyn S Calfee. 2020. ARDS Subphenotypes: Understanding a Heterogeneous Syndrome. *Crit. Care* 24, 1 (March 2020), 102.
[50] S. Wold, K. Esbensen, and P. Geladi. 1987. Principal component analysis. *Chemometrics and Intelligent Laboratory Systems* 2 (1987), 37–52.
[51] Han Xiao, Kashif Rasul, and Roland Vollgraf. 2017. Fashion-MNIST: a Novel Image Dataset for Benchmarking Machine Learning Algorithms. *CoRR* abs/1708.07747 (2017). arXiv:1708.07747 http://arxiv.org/abs/1708.07747
[52] Junyuan Xie, Ross B. Girshick, and Ali Farhadi. 2015. Unsupervised Deep Embedding for Clustering Analysis. *CoRR* abs/1511.06335 (2015). arXiv:1511.06335 http://arxiv.org/abs/1511.06335