POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Executive Summary of the Thesis

CONVERSE: Discovering Patient Groups in Survival Data via Deep Latent Clustering

Laurea Magistrale in Computer Engineering - Ingegneria Informatica

Author: Pinar Erbil

Advisor: Prof. Matteo Matteucci

Co-advisors: Alberto Archetti, Eugenio Lomurno

Academic year: 2025-2026

1. Introduction

Survival analysis is fundamental to clinical decision-making, enabling the modelling of time-to-event outcomes and patient risk stratification. While recent deep learning–based survival models achieve strong predictive performance, they often lack interpretability, which limits their use in medical settings. To improve transparency, clustering-based survival approaches aim to identify latent patient subgroups with distinct risk profiles, providing intuitive interpretability; however, they frequently underperform compared to purely discriminative models and struggle to balance subgroup discovery with predictive accuracy.

We propose **CONVERSE** (CONtrastive Variational Ensemble for Risk Stratification and Estimation), a deep survival framework that jointly learns latent representations, patient clusters, and survival risk. CONVERSE combines variational representation learning with contrastive clustering and ensemble survival heads, and employs self-paced training to improve robustness. Experiments on multiple datasets demonstrate that CONVERSE achieves competitive predictive performance while producing meaningful risk stratification.

2. Related Work

Survival analysis aims to model the survival function $S(t|\mathbf{x}) = \Pr(T \geq t|\mathbf{x})$, which captures the probability that the event time exceeds t given $\mathbf{x}$. A central challenge is handling censored observations, particularly right-censoring, and remains the focus of this work.

Among classical survival methods, the Cox proportional hazards model remains a widely used semi-parametric formulation that models covariate effects without assuming a specific baseline hazard. In contrast, many modern deep learning methods adopt discrete-time formulations in which the time axis is partitioned into intervals and event probabilities are estimated within each bin. This perspective frames survival prediction as a sequence of classification tasks [3].

Recent work has focused on improving interpretability by incorporating representation learning and clustering into survival models. DVCSurv introduces dual-view latent representations learned via Siamese encoders and contrastive objectives to discover patient subgroups [1]. While such methods enhance interpretability, they often involve trade-offs between predictive performance and clustering quality.

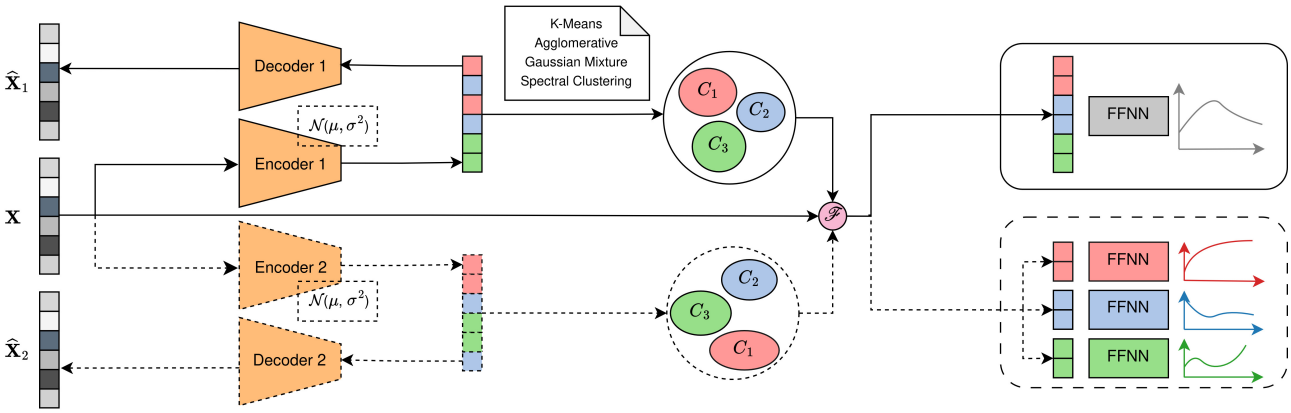

Figure 1: Overall architecture of CONVERSE. Dashed lines indicate optional pathways depending on hyperparameter selection

3. Proposed Method

CONVERSE is an end-to-end framework that jointly performs representation learning, clustering, and survival analysis, as illustrated in Figure 1. Inspired by dual-view clustering-based survival models [1], CONVERSE extends prior work through a modular design that supports flexible encoder architectures, clustering strategies, and survival heads. The model consists of three components: (i) latent representation learning via autoencoders, (ii) clustering with contrastive refinement in latent space, and (iii) discrete-time survival prediction.

3.1. Representation Learning

CONVERSE learns a low-dimensional latent representation $\mathbf{z}_i$ from patient covariates $\mathbf{x}_i$ using either deterministic or variational autoencoders. For variational encoders, a Gaussian posterior $q_\phi(\mathbf{z}_i \mid \mathbf{x}_i) = \mathcal{N}(\boldsymbol{\mu}_i, \mathrm{diag}(\boldsymbol{\sigma}_i^2))$ is learned by minimizing the reconstruction loss

$$\mathcal{L}_{\mathrm{REC}} = \|\hat{\mathbf{x}}_i - \mathbf{x}_i\|_2^2 \tag{1}$$

together with the KL divergence regularizer

$$\mathcal{L}_{\mathrm{KLD}} = \mathrm{KL}(q_\phi(\mathbf{z}_i \mid \mathbf{x}_i) \,\|\, \mathcal{N}(\mathbf{0}, \mathbf{I})) \tag{2}$$

When using deterministic autoencoders, the KL term is omitted and the encoder directly outputs a latent representation.

To capture complementary views of the data, CONVERSE optionally employs a Siamese configuration with two encoders of identical architecture but independent parameters, producing two latent representations $\mathbf{z}_i^{(1)}$ and $\mathbf{z}_i^{(2)}$. If a Siamese configuration is used, the losses are simply averaged across views.

3.2. Contrastive Learning and Clustering

Patient subgroups are identified by clustering latent representations. Given latent vectors $\mathbf{z}_i$ and cluster centres $\{\mathbf{m}_k\}_{k=1}^K$, we define the clustering loss as

$$\mathcal{L}_{\mathrm{CLUS}} = \frac{1}{N} \sum_{i=1}^{N} \|\mathbf{z}_i - \mathbf{m}_{c_i}\|_2^2, \tag{3}$$

where c_i denotes the assigned cluster. The framework supports multiple clustering methods, including K-means, GMMs, spectral, and agglomerative clustering.

To refine the latent space, CONVERSE incorporates contrastive learning. Let $\mathcal{C} = \{i \mid e_i = 0\}$ denote the set of censored patients and $\mathcal{U} = \{i \mid e_i = 1\}$ the set of uncensored patients. For a censored patient i, we define $\mathcal{P}_i^+ = \{j \in \mathcal{U} \mid c_j = c_i\}$ as the set of uncensored patients assigned to the same cluster. The Intra-View Cluster-Guided objective (IVCG) leverages uncensored patients to improve representations of censored ones by forming positive pairs within clusters and negative pairs across clusters. Using the InfoNCE loss, this is defined as

$$\mathcal{L}_{\mathrm{IVCG}} = \frac{-1}{|\mathcal{C}|} \sum_{i \in \mathcal{C}} \sum_{j \in \mathcal{P}_i^+} \log \frac{e^{s(\mathbf{z}_i, \mathbf{z}_j)/\tau}}{\sum_{k=1}^{N} e^{s(\mathbf{z}_i, \mathbf{z}_k)/\tau}} \tag{4}$$

where $s(\cdot, \cdot)$ denotes cosine similarity, and τ is a temperature parameter.

For Siamese encoders, two additional objectives enforce inter-view consistency. The Inter-View Instance-Wise loss (IVIW) treats the two latent

views of the same patient as positive pairs:

$$\mathcal{L}_{\text{IVIW}} = \frac{-1}{N}\sum_{i=1}^{N}\sum_{v=1}^{2}\log\frac{e^{s(\mathbf{z}_i^{(v)},\mathbf{z}_i^{(v')})/\tau}}{\sum_{j=1}^{N}e^{s(\mathbf{z}_i^{(v)},\mathbf{z}_j^{(v')})/\tau}} \quad (5)$$

The Inter-View Cluster-Wise loss (IVCW) aligns clusters across views using soft assignments computed via Student's t-distribution:

$$q_{i,k}^{(v)} = \frac{\left(1+\|\mathbf{z}_i^{(v)}-\mathbf{m}_k^{(v)}\|_2^2/\nu\right)^{-\frac{\nu+1}{2}}}{\sum_{k'=1}^{K}\left(1+\|\mathbf{z}_i^{(v)}-\mathbf{m}_{k'}^{(v)}\|_2^2/\nu\right)^{-\frac{\nu+1}{2}}} \quad (6)$$

with cluster-wise contrastive loss

$$\mathcal{L}_{\text{IVCW}} = \frac{-1}{2K}\sum_{k=1}^{K}\sum_{v=1}^{2}\log\frac{e^{s(\mathbf{q}_k^{(v)},\mathbf{q}_k^{(v')})/\tau}}{\sum_{k'=1}^{K}e^{s(\mathbf{q}_k^{(v)},\mathbf{q}_{k'}^{(v')})/\tau}} \quad (7)$$

The total contrastive loss is the sum of all: $\mathcal{L}_{\text{CL}} = \alpha_{\text{IVCG}}\mathcal{L}_{\text{IVCG}} + \alpha_{\text{IVIW}}\mathcal{L}_{\text{IVIW}} + \alpha_{\text{IVCW}}\mathcal{L}_{\text{IVCW}}$, where α_{IVIW} and α_{IVCW} are set to zero for single-encoder models.

3.3. Survival Prediction

The final component maps latent representations to survival predictions using a discrete-time formulation. An ensemble of feed-forward networks predicts the event probability in each time bin $\tau_1,\ldots,\tau_T$, following a DeepHit-based [3] discrete-time formulation. We consider two architectural variants: a shared survival head and cluster-specific heads. In both cases, the survival input combines latent features with raw covariates:

$$\mathbf{h}_i = \begin{cases} [\mathbf{z}_i;\mathbf{x}_i] & \text{if single encoder,} \\ \left[\frac{1}{2}(\mathbf{z}_i^{(1)}+\mathbf{z}_i^{(2)});\mathbf{x}_i\right] & \text{if dual encoder} \end{cases} \quad (8)$$

In the shared-head configuration, a single network $g_{\text{surv}}(\cdot)$ produces discrete-time event probabilities $\{\hat{p}_{i,t}\}_{t=1}^{T} = g_{\text{surv}}(\mathbf{h}_i)$, where $\hat{p}_{i,t}$ represents the predicted probability that the event occurs at time bin t for patient i. Alternatively, the cluster-specific-heads variant assigns each cluster k its own survival head $g_{\text{surv}}^{(k)}(\cdot)$, yielding $\{\hat{p}_{i,t}\}_{t=1}^{T} = g_{\text{surv}}^{(k)}(\mathbf{h}_i)$ for patients assigned to cluster k. Survival heads are trained using a combination of calibration and ranking objectives. The negative log-likelihood (NLL) loss is defined as

$$\mathcal{L}_{\text{NLL}} = \frac{-1}{N}\sum_{i=1}^{N}\left[e_i\log(\hat{p}_{i,t_i}) + \bar{e}_i\log(\hat{S}_{i,t_i})\right] \quad (9)$$

where $\bar{e}_i = 1-e_i$ denotes the censoring indicator and $\hat{S}_{i,t} = \sum_{k>t}\hat{p}_{i,k}$ denotes the discrete survival probability. To encourage correct risk ordering, we employ a ranking loss over comparable pairs (i,j) with $e_i = 1$ and $t_i < t_j$:

$$\mathcal{L}_{\text{RANK}} = \sum_{\substack{i,j \\ e_i=1,\, t_i<t_j}}\exp\left(\frac{\hat{S}_{i,t_i}-\hat{S}_{j,t_i}}{\sigma}\right) \quad (10)$$

where σ is a temperature parameter. The exponential term acts as a smooth pairwise ranking surrogate, penalizing incorrect risk ordering. The final survival objective is $\mathcal{L}_{\text{SURV}} = \alpha_{\text{NLL}}\mathcal{L}_{\text{NLL}} + \alpha_{\text{RANK}}\mathcal{L}_{\text{RANK}}$.

3.4. Training Procedure

CONVERSE is trained in three stages: pre-training, cluster initialization, and end-to-end refinement with self-paced learning (SPL) [2]. SPL progressively incorporates instances from easy to hard, improving training stability, and is applied only to representation learning and clustering components following [1]. At epoch e, the SPL threshold is defined as

$$\lambda_e = \mu(\{\mathcal{L}_i\}_{i=1}^{N}) + \frac{e}{E_{\max}}\sigma(\{\mathcal{L}_i\}_{i=1}^{N}), \quad (11)$$

where $E_{\max}$ is the total number of epochs, and $\mu(\cdot)$ and $\sigma(\cdot)$ denote the mean and standard deviation of the instance-level losses. Each instance loss is given by $\mathcal{L}_i = \alpha_{\text{REC}}\mathcal{L}_{\text{REC},i} + \alpha_{\text{KLD}}\mathcal{L}_{\text{KLD},i} + \alpha_{\text{CLUS}}\mathcal{L}_{\text{CLUS},i}$ where $\mathcal{L}_{\text{KLD},i}$ is omitted for non-variational encoders. Only instances with $\mathcal{L}_i \le \lambda_e$ contribute to the SPL objective:

$$\mathcal{L}_{\text{SPL}} = \frac{1}{\sum_{i=1}^{N}\mathbb{1}(\mathcal{L}_i \le \lambda_e)}\sum_{i=1}^{N}\mathbb{1}(\mathcal{L}_i \le \lambda_e)\mathcal{L}_i \quad (12)$$

In the first stage, autoencoder(s) and survival head(s) are jointly pre-trained by minimizing $\min_\Theta(\alpha_{\text{REC}}\mathcal{L}_{\text{REC}} + \alpha_{\text{KLD}}\mathcal{L}_{\text{KLD}} + \mathcal{L}_{\text{SURV}})$ to obtain meaningful latent representations. In the second stage, a clustering algorithm (e.g., K-means) is applied to the latent space to initialize cluster centres $\mathbf{M}$ and assignments $\mathbf{c}_i$; cluster centres $\mathbf{M}$ are fixed thereafter. Finally, the full model is trained end-to-end with SPL by optimizing $\min_\Theta(\mathcal{L}_{\text{SPL}} + \mathcal{L}_{\text{CL}} + \mathcal{L}_{\text{SURV}})$. Cluster assignments are updated after each epoch by nearest-centre reassignment. All hyperparameters and architectural choices are selected via validation.

Table 1: Combined C-index (higher is better) and IBS (lower is better) test results across datasets, scaled by 100. Best value is shown in bold and second best is underlined. Mean values are reported.

Datasets	AIDS		BREAST		GBSG		Metabric		PBC		TCGA		Veterans		WHAS	
	CI	IBS	CI	IBS	CI	IBS	CI	IBS	CI	IBS	CI	IBS	CI	IBS	CI	IBS
Statistical Models																
CoxPH	73.0	**6.7**	55.3	17.9	66.9	18.9	63.8	19.6	75.5	**16.7**	73.0	10.0	66.8	16.4	76.1	15.8
AFT	70.9	7.1	66.0	21.1	67.2	18.8	65.2	19.7	74.8	17.5	73.2	10.5	67.3	17.6	75.5	16.9
PC-Hazard	68.6	6.8	63.2	23.0	65.5	19.3	64.3	19.5	72.2	18.9	72.6	10.0	67.3	20.6	73.1	18.9
Log-Hazard	71.2	6.8	69.8	16.0	67.2	18.9	64.3	19.6	76.0	17.0	**74.7**	**9.9**	69.3	18.3	77.3	15.9
Machine Learning Models																
RSF	72.3	**6.7**	68.7	**15.7**	68.5	19.0	64.2	19.6	75.7	**16.7**	70.0	10.2	67.6	16.5	77.8	15.5
XGBCox	**74.6**	**6.7**	69.7	16.8	**68.7**	**18.3**	64.3	19.6	75.5	16.9	74.6	**9.9**	69.3	**16.1**	**79.6**	**14.8**
DeepSurv	73.6	**6.7**	66.5	16.6	67.3	18.7	64.6	19.5	75.8	17.2	74.3	10.2	68.5	16.5	77.4	15.4
DeepHit	71.8	6.7	**66.7**	16.5	66.6	19.2	64.8	**19.4**	74.9	17.3	72.3	10.5	70.0	16.8	77.3	15.6
Clustering-Based Models																
SCA	71.1	34.0	69.4	39.5	66.6	31.7	64.7	27.4	75.9	28.2	67.8	27.8	66.5	24.9	75.2	23.9
VaDeSC	70.2	6.9	**70.8**	17.0	67.3	19.3	64.5	**19.4**	**76.1**	19.1	72.5	16.3	**70.1**	16.6	77.3	16.8
DCM	71.2	7.1	61.8	17.4	65.9	19.4	64.5	**19.4**	73.2	19.6	70.7	11.1	68.8	17.5	76.4	16.7
DVCSurv	72.8	**6.7**	65.5	18.9	67.8	20.5	65.8	22.7	75.1	18.6	68.2	13.3	68.9	29.6	76.4	26.8
CONVERSE	73.0	**6.7**	63.7	17.1	68.0	18.5	**65.9**	21.2	75.7	17.9	73.1	10.5	68.9	19.3	77.9	15.3

4. Experimental Setup

We evaluate CONVERSE on eight real-world survival datasets spanning oncology and clinical outcomes, covering a wide range of cohort sizes, feature dimensionalities, and censoring rates. The datasets include both low- and high-dimensional settings, with sample sizes ranging from small clinical cohorts to larger population studies and varying censoring, reflecting diverse real-world survival scenarios.

We compare CONVERSE against a diverse set of baseline models, including classical statistical approaches, tree-based methods, deep learning models, and clustering-based survival methods. These baselines span a broad range of modelling assumptions, from interpretable parametric formulations to flexible neural and subgroup-aware approaches, providing a comprehensive benchmark for evaluating both predictive performance and risk stratification quality.

Hyperparameters for CONVERSE are optimized using Optuna with a Tree-structured Parzen Estimator due to the size and interdependence of the search space. For each configuration, models are trained and validated across five bootstrap splits, and selection is based on validation performance using the concordance index and integrated Brier score.

The final results correspond to the best-performing configuration averaged across splits. The same tuning and evaluation protocol is applied to all baseline models.

5. Results and Discussion

5.1. Performance Analysis

We evaluate CONVERSE against three categories of baselines: (i) classical statistical survival models, (ii) machine learning-based models, and (iii) clustering-based survival approaches. All results are reported in Table 1.

Across all datasets, CONVERSE demonstrates consistently competitive discrimination (C-index). Compared to classical statistical models, CONVERSE achieves the best C-index on half of the datasets and ranks second on most others, particularly improving performance on GBSG, Metabric, and WHAS. Against neural network–based models, CONVERSE remains competitive and frequently ranks among the top three methods, despite not being solely optimized for predictive accuracy. This indicates that incorporating structured latent clustering does not compromise discriminative ability too much. Most importantly, within the category of clustering-based survival models, CONVERSE consistently matches or outperforms existing methods. It achieves the highest C-index on 5 out of 8 datasets and improves upon DVCSurv, the base model upon which CONVERSE is built, on all datasets except BREAST, demonstrating the effectiveness of the proposed variational clustering framework for risk stratification. In terms of calibration (IBS), CONVERSE achieves performance comparable to classical statistical models and attains the best

IBS on AIDS, GBSG, and WHAS. However, neural network–based approaches such as XGBCox and RSF often achieve lower IBS values overall, indicating stronger probabilistic calibration when prediction accuracy is the primary objective. Within clustering-based models, CONVERSE achieves the lowest IBS on 5 out of 8 datasets and the second-lowest on two additional datasets, establishing state-of-the-art calibration performance in its category. Although it does not consistently surpass purely predictive neural models in IBS, it provides a favourable balance between discrimination, calibration, and interpretability. Overall, the results demonstrate that CONVERSE successfully bridges the gap between predictive performance and structured interpretability, delivering state-of-the-art results among clustering-based survival models.

5.2. Interpretability

Beyond predictive performance, a central objective of CONVERSE is to learn clinically meaningful representations. We analyse the native clusters produced by the model and, when necessary, the structure of the learned latent space through post-processing, which is purely for interpretability analysis and does not affect the trained model. Across datasets, two behaviours emerge: fragmented native clustering despite structured latent representations (BREAST, GBSG, TCGA, Veterans); clear and directly interpretable native clustering (AIDS, Metabric, PBC, WHAS). We illustrate these behaviours using two representative cohorts.

5.2.1 Structured Latent Space with Fragmented Clustering

GBSG dataset highlights a key property of CONVERSE: even when the native clustering is overly granular, the learned latent space remains highly informative. The original model partitions the cohort into several small clusters that are difficult to interpret clinically. However, applying K-means ($k = 3$) directly to the learned latent representations reveals three clearly separated survival groups with statistically distinct Kaplan–Meier trajectories. These clusters correspond to clinically coherent risk strata. The low-risk group (K2, green) is characterized by higher prevalence of hormone therapy and post-menopausal status, both associated with improved prognosis. In contrast, the high-risk group (K0, red) shows absence of hormone therapy and higher tumour grade, consistent with its poor survival outcomes. The intermediate cluster (K1, blue) exhibits mixed characteristics aligned with its survival trajectory. This result demonstrates that the latent space encodes meaningful clinical structure, even when the native clustering mechanism requires refinement. Importantly, the strong C-index and IBS performance on GBSG confirm that these representations are not only interpretable but also predictive. It is important to emphasize that, unlike GBSG, the remaining datasets did not require re-clustering of the latent space. In those cases, simply merging closely related CONVERSE clusters was sufficient to obtain more clinically meaningful survival groups. The primary issue was not the absence of structure, but rather an over-granulation of clusters, where the model identified finer sub-partitions within otherwise similar survival strata.

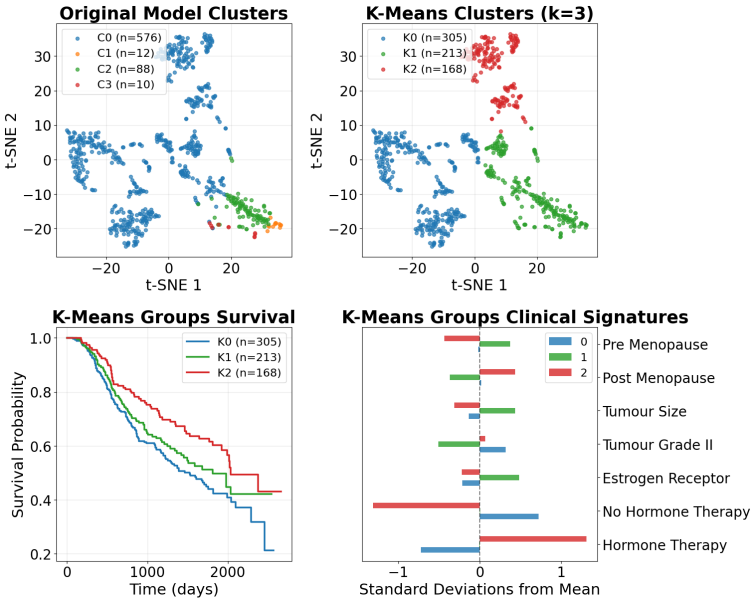

Figure 2: Post-processing analysis of CONVERSE latent representations on GBSG. Top: t-SNE visualization of original clusters (left) and K-means clusters (k = 3) (right). Bottom: Kaplan–Meier survival curves for K-means groups (left) and standardized clinical signatures across clusters (right).

5.2.2 Directly Interpretable Clustering

The PBC dataset illustrates a scenario in which CONVERSE natively identifies clinically distinct subgroups without requiring post-processing. The model isolates a high-risk clus-

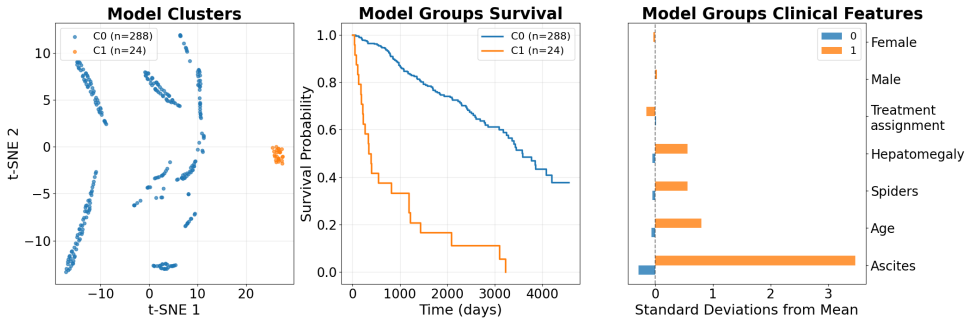

Figure 3: Post-processing analysis of CONVERSE latent representations on PBC. Left: t-SNE visualization of model-defined clusters. Middle: Kaplan–Meier survival curves for model groups. Right: Standardized clinical feature profiles across groups.

ter with markedly poor survival and a low-risk group with substantially better outcomes.

The high-risk subgroup (C1, orange) is characterized by features associated with advanced liver disease, including hepatomegaly, ascites, and spider angiomas, which are established indicators of severe cirrhosis. The clean separation of Kaplan–Meier curves confirms that the model captures pronounced survival-driven heterogeneity inherent in this cohort. This case highlights that when strong survival structure is present in the data, CONVERSE directly translates it into interpretable risk stratification.

Overall, these findings demonstrate that CONVERSE does not merely optimize predictive performance; it learns structured, clinically interpretable embeddings that encode real survival-driven relationships between patients.

5.3. Computational Considerations

Lastly, we assess the computational requirements of CONVERSE. Due to its joint optimization of latent representations and cluster assignments, the model incurs one of the highest training costs among the evaluated approaches, particularly when compared to purely discriminative baselines such as the fast tree-based method XGBCox. Empirical scaling analysis shows that training time increases super-linearly with the number of samples and scaling with respect to feature dimensionality is sublinear, indicating that cohort size is the primary driver of runtime.

6. Conclusions

We introduced CONVERSE, a deep survival framework that balances strong discriminative performance with interpretable risk stratification by combining variational representation learning, multi-view contrastive objectives, and cluster-specific survival heads. Extensive experiments on eight benchmark datasets show that CONVERSE matches the performance of neural survival models while consistently outperforming existing clustering-based approaches. Moreover, CONVERSE yields clinically meaningful patient subgroups aligned with established prognostic factors, supporting transparent and data-driven clinical decision-making. Future work will focus on improving cluster robustness and reducing reliance on post-processing, narrowing the performance gap between clustering-aware and purely predictive survival models, enhancing computational efficiency, and developing clinically deployable pipelines with automated subgroup explanations.

Publication. A preprint of this work is available at `https://arxiv.org/abs/2602.01367` and the public code is available at `https://github.com/erbilpinar/CONVERSE`.

References

[1] Chang Cui, Yongqiang Tang, and Wensheng Zhang. Deep contrastive survival analysis with dual-view clustering. *Electronics*, 13(24):4866, 2024.

[2] M Kumar, Benjamin Packer, and Daphne Koller. Self-paced learning for latent variable models. *Advances in neural information processing systems*, 23, 2010.

[3] Changhee Lee, William Zame, Jinsung Yoon, and Mihaela Van Der Schaar. Deephit: A deep learning approach to survival analysis with competing risks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.