用于联合风险分层与事件发生时间估计的深度变分对比学习

Pinar Erbil[1], **Alberto Archetti**[*1]、**Eugenio Lomurno**[1],和 **Matteo Matteucci**[1]

[1]米兰理工大学，*AIRLab*，意大利

摘要

生存分析对于临床决策至关重要，因为它能够帮助临床工作者估计事件发生时间结局、对患者风险特征进行分层，并指导治疗规划。深度学习凭借前所未有的预测能力革新了这一领域，但也面临着性能与可解释性之间的根本性权衡。神经网络虽然准确率很高，但其黑箱特性限制了其在临床中的应用。相比之下，基于深度聚类的方法会将患者划分为可解释的风险组，但通常要以牺牲预测能力为代价。我们提出 *CONVERSE*（用于风险分层与估计的对比变分集成），这是一种深度生存模型，通过将变分自编码器与对比学习相结合，实现可解释的风险分层，从而弥合这一差距。*CONVERSE* 将变分嵌入与多种簇内与簇间对比损失相结合。自步学习逐步纳入从易到难的样本，从而提高训练稳定性。该模型支持簇特定生存头，能够生成准确的集成预测。在四个基准数据集上进行的全面评估表明，与现有深度生存方法相比，*CONVERSE* 的性能具有竞争力或更优，同时能够实现有意义的患者分层。

关键词： 深度生存分析 · 变分自编码器 · 对比学习 · 风险分层 · 聚类 ·可解释性

arXiv:2602.01367v1 [cs.LG] 1 Feb 2026

1 引言

生存分析 [23]在现代医疗保健中发挥着根本性作用，使医务人员能够估计事件发生时间结局并对患者风险特征进行分层。准确的模型对于高效分配资源、制定治疗规划以及支持基于数据的医疗干预至关重要。除了预测性能之外，可解释性也至关重要，因为临床医生在做出直接影响患者结局的安全关键决策时，必须信任模型的判断。

深度学习改变了生存分析，带来了前所未有的预测能力 [24]。这些方法能够捕捉复杂的非线性关系，而不依赖比例风险等严格假设 [7]，并利用传统统计方法难以分析的高维数据。然而，这种性能提升也付出了代价：大多数现代生存模型都如同黑箱，其决策过程晦暗不明。这种缺乏透明度的情况在临床环境中带来了重大挑战，因为了解预测背后的依据与预测结果本身同样重要。

为应对可解释性挑战，人们日益关注将聚类机制融入生存模型的风险分层方法 [6, 8, 17, 20]。这些方法发现患者队列中的潜在亚群体，每个亚群体都具有不同的生存模式。基于分层的方法之所以具有吸引力，在于其固有的可解释性：人们不仅根据患者的个体特征来理解患者，还会结合其所属风险组的动态特征来理解患者。尽管具有这些优势，基于聚类的生存方法在判别性能和校准方面历来落后于纯预测模型，从而限制了其实际应用。

为弥合这一差距，我们提出了 CONVERSE（用于风险分层与估计的对比变分集成），这是一种将变分自编码器 [17] 的表征能力与对比学习 [8]的聚类能力相结合的深度生存模型。基于 DVCSurv 主干网络 [8], , CONVERSE 使用变分自编码器学习数据的多视图表示。随后，通过多种聚类算法对这些表示进行划分，并通过对比学习进行对齐，同时结合簇内损失和簇间损失。由此得到的聚类表示被输入深度生存头集成，通过利用亚群体特异性模式来捕捉群体层面风险。模型通过超参数选择流程进行训练，使架构能够适应每个特定数据集，仅纳入有助于最终模型的组件。我们在四个广泛使用的生存分析数据集上进行了全面评估，验证了 CONVERSE 相较于现有深度生存方法能够取得具有竞争力或更优的性能，同时保持可解释的风险分层。

2 相关工作

生存分析的目标是对生存函数 $S(t|\mathbf{x}) = \Pr(T \geq t|\mathbf{x})$ 进行建模，该函数表示对于具有协变量 $\mathbf{x}$ 的受试者，感兴趣事件在时间 t 之后发生的概率。一个根本性挑战是处理

*通讯作者：alberto.archetti@polimi.it

删失观测，即由于观察期在事件发生前结束或受试者失访，确切的事件时间未知。尽管存在各种删失机制，但右删失在医学应用中最为普遍，并且仍然是本文的重点[23]。

传统生存分析方法可分为非参数、半参数和全参数类别。非参数方法通过对人群亚组进行平均，直接根据数据构建生存函数[11, 12]。半参数模型将人群的非参数基线与基于患者特征的参数化调节相结合[7, 13]。相比之下，全参数模型完全通过有限个参数表示生存函数。在这一框架下，生存函数可以定义在连续或离散时间尺度上。连续时间形式通常采用经典统计分布，并通过非线性方法估计参数[3, 19]。相反，离散时间形式会将时间范围划分为多个区间 τ_1，τ_2，……，τ_T。许多基于神经网络的方法采用这一框架，将生存分析视为跨时间区间的一系列分类问题，并将总生存函数重建为累积乘积[15, 16, 24]。

为增强可解释性，近期方法将聚类机制融入生存模型，以发现具有不同风险特征的潜在亚群。生存聚类分析（SCA）[6]采用贝叶斯非参数方法，利用截断狄利克雷过程自动确定聚类数量，同时对潜在表示施加约束。VaDeSC [17]采用变分自编码器框架，并使用高斯混合先验，学习协变量与生存时间之间的簇特异性关联，再通过威布尔分布混合模型对结局进行建模。深度考克斯混合模型（DCM）[20]通过EM算法和硬聚类分配学习考克斯回归模型的混合，并估计潜在簇内的风险比。近期，DVC-Surv [8]通过孪生自编码器和对比学习引入双视图聚类，以获得稳健的患者表征。尽管这些基于聚类的方法提高了可解释性，但与纯判别模型相比，往往需要以预测性能为代价。

3 方法

CONVERSE 将表示学习、聚类和生存分析整合到一个统一的端到端可训练框架中，如图 1 所示。CONVERSE 基于 DVC-Surv 中提出的双视图聚类范式[8]，，通过模块化设计扩展了这一方法，从而能够灵活配置聚类策略、损失函数形式和生存预测架构。

3.1 CONVERSE架构

该模型由三个核心部分组成：(i) 用于学习患者潜在表征的变分自编码器，(ii) 采用多视图对比学习增强的聚类模块，以及 (iii) 由共享或针对特定聚类的生存预测头构成的集成。

3.1.1 表示学习与自编码器

表示学习组件旨在学习一个低维潜在空间，以捕捉患者协变量的基本结构。该潜在空间为聚类和生存预测奠定基础。与[8], 不同，我们采用变分自编码器（VAEs）[17]，以获得比标准自编码器更灵活且正则化程度更高的表征。给定输入向量 $\mathbf{x}_i$，变分编码器会将其映射为高斯分布的参数$q_\phi(\mathbf{z}_i|\mathbf{x}_i) = \mathcal{N}(\boldsymbol{\mu}_i, \mathrm{diag}(\boldsymbol{\sigma}_i^2))$，其中 $\boldsymbol{\mu}_i$，$\boldsymbol{\sigma}_i \in \mathrm{R}^d$ 是编码器网络生成的均值向量和标准差向量。随后，解码器网络根据潜在表征 $\mathbf{z}_i \in \mathrm{R}^d$ 重构输入。重构损失促使所学习的表征保留原始协变量中的信息：

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

其中，$\hat{\mathbf{x}}_i$ 是重构后的输入。为了规范潜在空间并防止过拟合，我们施加 KL散度约束，促使近似后验保持接近标准高斯先验$p(\mathbf{z}) = \mathcal{N}(\mathbf{0}, \mathbf{I})$：

$$\mathcal{L}_{\mathrm{KLD}} = \frac{1}{N}\sum_{i=1}^{N} \mathrm{KL}\big(q_\phi(\mathbf{z}_i \mid \mathbf{x}_i) \,\|\, p(\mathbf{z})\big). \tag{2}$$

该框架还支持标准（非变分）自编码器，只需省略 KL散度项，此时编码器会直接生成确定性的潜在表征$\mathbf{z}_i$，无需采样。每个数据集采用变分自编码器还是确定性自编码器，由超参数优化过程决定。

为了像 [8], CONVERSE那样捕捉患者数据的互补视图，CONVERSE可选用孪生自编码器。此配置使用两个架构相同但参数独立的编码器，从同一输入$\mathbf{x}_i$生成两个不同的潜在表示$\mathbf{z}_i^{(1)}$,$\mathbf{z}_i^{(2)} \in \mathrm{R}^d$。在孪生配置中，重构损失和KL散度损失在两个视图之间取平均。

3.1.2 聚类与对比学习

学习潜在表示后，CONVERSE 按照 DSACC [9]和DVCSurv [8]中的步骤，通过在潜在空间中进行聚类来发现患者亚群，并通过对比学习优化这些表征。聚类直接在学习到的潜在空间中进行。对于给定的潜在表征（单编码器情形下为$\mathbf{z}_i$，或孪生网络情形下视图v 对应的$\mathbf{z}_i^{(v)}$），K 聚类中心记为$\mathbf{M} = [\mathbf{m}_1, \ldots, \mathbf{m}_K]$，其中$\mathbf{m}_k \in \mathrm{R}^d$。每位患者恰好被分配到一个聚类，并用独热向量$\mathbf{c}_i \in \{0,1\}^K$ 表示，其中$c_{i,k} = 1$ 若患者 i 属于

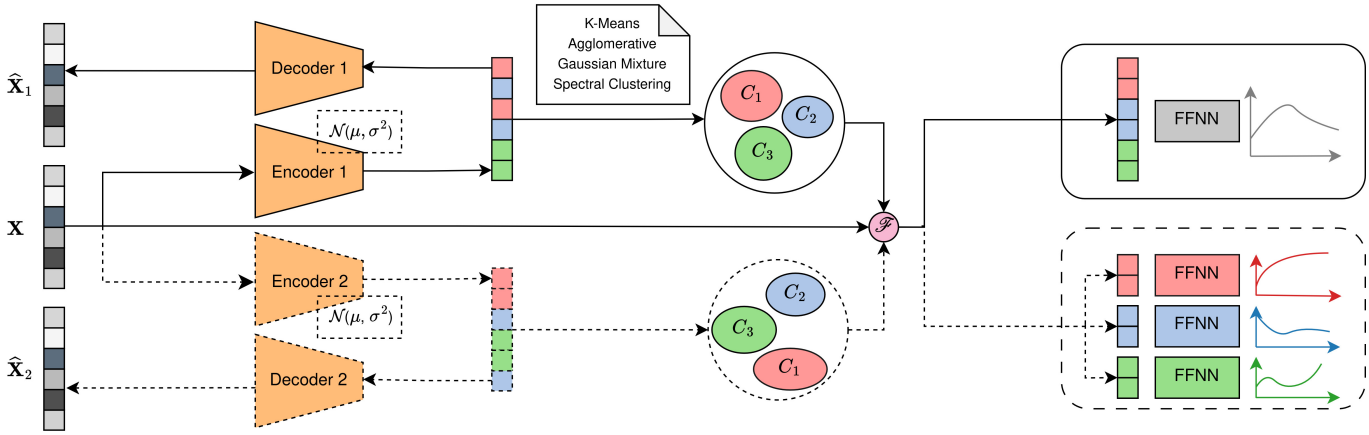

图 1： CONVERSE 的整体架构。模型由三个部分组成：(i) 用于学习潜在表示的变分自编码器，(ii) 在潜在空间中进行对比学习的聚类模块，以及 (iii) 集成生存预测头。虚线表示根据超参数选择而定的可选路径。

聚类 k。聚类损失通过最小化每个潜在表示与其分配到的聚类中心之间的距离来实现，具体如下：

$$\mathcal{L}_{\text{CLUS}} = \frac{1}{N}\sum_{i=1}^{N} \|\mathbf{z}_i - \mathbf{M}\mathbf{c}_i\|_2^2. \tag{3}$$

对于孪生配置，聚类在每个视图中独立执行，损失取平均值：

$$\mathcal{L}_{\text{CLUS}}^{\text{Siam}} = \frac{1}{2N}\sum_{i=1}^{N}\sum_{v=1}^{2} \|\mathbf{z}_i^{(v)} - \mathbf{M}^{(v)}\mathbf{c}_i^{(v)}\|_2^2. \tag{4}$$

与 [8, 9], 不同，我们的框架支持多种聚类算法，包括 K 均值聚类、凝聚层次聚类、高斯混合模型和谱聚类。具体的聚类选择由超参数优化过程决定。

在 [8], 之后，对比学习通过鼓励相似患者聚集在一起、同时分离不相似患者，进一步优化已学习的表征。根据编码器配置，最多采用三种互补的对比目标。第一种目标，即视图内聚类引导对比学习（IVCG），利用未删失患者的信息来改善删失患者的表征。对于每位删失患者（锚点），与同一聚类中的未删失患者构成正样本对，与不同聚类中的未删失患者构成负样本对。这种设计基于以下观察：无论删失状态如何，同一风险组中的患者都应具有相似的潜在表示。使用 InfoNCE损失，可将其形式化为：

$$\mathcal{L}_{\text{IVCG}} = -\frac{1}{N_{\text{cens}}}\sum_{\substack{i=1\\ e_i=0}}^{N}\sum_{j\in\mathcal{P}_i^+} \log\frac{\exp(s(\mathbf{z}_i,\mathbf{z}_j)/\tau)}{\sum_{k=1}^{N}\exp(s(\mathbf{z}_i,\mathbf{z}_k)/\tau)}, \tag{5}$$

其中，$N_{\text{cens}} = \sum_{i=1}^{N}(1-e_i)$ 是删失患者的数量，$\mathcal{P}_i^+ = \{j : e_j = 1,\ c_j = c_i\}$ 是与患者 i 属于同一簇的未删失患者集合，$s(\cdot,\ \cdot)$ 表示余弦相似度，τ 是温度超参数。对于孪生编码器，我们将两个视图的损失相加。

第二个目标是视图间实例级对比学习（IVIW），专为孪生编码器设计，旨在将同一患者的两个潜在表示视为正样本对、将所有其他患者视为负样本对，从而鼓励一致性

$$\mathcal{L}_{\text{IVIW}} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{v=1}^{2}\log\frac{\exp\left(s(\mathbf{z}_i^{(v)},\mathbf{z}_i^{(v')})/\tau\right)}{\sum_{j=1}^{N}\exp\left(s(\mathbf{z}_i^{(v)},\mathbf{z}_j^{(v')})/\tau\right)}. \tag{6}$$

最后，引入第三个目标——视图间逐簇对比学习（IVCW），以确保两个潜在视图之间的一致性。除了聚类算法生成的硬聚类分配外，还采用基于学生t分布的软聚类分配来刻画聚类置信度。其目标是针对每个实例$\mathbf{z}_i^{(v)}$，确定它属于视图 v 中簇 k 的置信度。该置信度的计算方式如下：

$$q_{i,k}^{(v)} = \frac{\left(1 + \left\|\mathbf{z}_i^{(v)} - \mathbf{m}_k^{(v)}\right\|_2^2/\nu\right)^{-\frac{\nu+1}{2}}}{\sum_{k'=1}^{K}\left(1 + \left\|\mathbf{z}_i^{(v)} - \mathbf{m}_{k'}^{(v)}\right\|_2^2/\nu\right)^{-\frac{\nu+1}{2}}}. \tag{7}$$

其中，ν 表示自由度。对于给定的簇 k 和视图 v，将所有患者的置信度得分汇总为$\mathbf{q}_k^{(v)} = [q_{1,k}^{(v)},\ldots,q_{N,k}^{(v)}]^\top \in \mathrm{R}^N$，它表示该视图中的聚类级分布。逐簇对比损失促使不同视图中相对应的簇具有相似的分布：

$$\mathcal{L}_{\text{IVCW}} = -\frac{1}{2K}\sum_{k=1}^{K}\sum_{v=1}^{2}\log\frac{\exp\left(s(\mathbf{q}_k^{(v)},\mathbf{q}_k^{(v')})/\tau\right)}{\sum_{k'=1}^{K}\exp\left(s(\mathbf{q}_k^{(v)},\mathbf{q}_{k'}^{(v')})/\tau\right)}. \tag{8}$$

总对比损失通过可学习权重组合所有适用的目标：

$$\mathcal{L}_{\text{CL}} = \alpha_{\text{IVCG}}\mathcal{L}_{\text{IVCG}} + \alpha_{\text{IVIG}}\mathcal{L}_{\text{IVIG}} + \alpha_{\text{IVCW}}\mathcal{L}_{\text{IVCW}}, \tag{9}$$

其中，使用单编码器时，α_{IVIG} 和 α_{IVCW} 设为零。

3.1.3 集成生存预测

最后一个组件将学得的潜在表征映射为生存预测。我们采用基于多个前馈神经网络的离散时间形式，预测事件在每个时间区间 τ_1，……，τ_T内发生的概率。该组件有两种架构变体：共享头和簇特异性头。在共享头配置中，所有患者均由一个生存网络处理。输入结合潜在表征和原始协变量，以同时保留学得的结构和原始临床信息：

$$\mathbf{h}_i = \begin{cases} [\mathbf{z}_i; \mathbf{x}_i] & \text{if single encoder,} \\ \left[\frac{1}{2}(\mathbf{z}_i^{(1)} + \mathbf{z}_i^{(2)}); \mathbf{x}_i\right] & \text{if dual encoder,} \end{cases} \tag{10}$$

单个前馈神经网络 $g_{\text{surv}}(\cdot)$ 对这一组合表征进行处理，输出离散时间事件概率 $\{\hat{}p_{i,\ t}\}_{t=1}^{T} = g_{\text{surv}}(\mathbf{h}_i)$，其中 $\hat{}p_{i,\ t}$ 表示患者 i 在时间区间t 发生事件的预测概率。

或者，在簇特定头配置中，我们采用由生存分析头 $\{g^{(k)}_{\text{surv}}\}_{k=1}^{K}$组成的集成，其中每个簇 k 都分配给一个独立的头。该设计基于这样一个假设：不同的患者亚群可能表现出本质上不同的生存动态，而专门化模型能够更好地捕捉这些差异。对于被分配到簇 k的患者 i，预测结果为 $\{\hat{}p_{i,\ t}\}_{t=1}^{T} = g^{(k)}_{\text{surv}}(\mathbf{h}_i)$，其中输入 $\mathbf{h}_i$ 按公式（10）构建。这种集成方法使模型能够在维持共享潜在表征的同时，学习簇特定风险模式。

我们使用两个互补目标优化生存头。负对数似然损失（NLL）可确保概率估计得到良好校准：

$$\mathcal{L}_{\text{NLL}} = -\frac{1}{N}\sum_{i=1}^{N}\left[e_i \log(\hat{p}_{i,t_i}) + (1-e_i)\log(\hat{S}_{i,t_i})\right]. \tag{11}$$

为提高判别性能，排序损失促使可比较样本对的生存概率保持正确的顺序。若患者 i 在患者 j 之前发生事件且未被删失（$e_i = 1$ 和 $t_i < t_j$），则两名患者（i，j）可比较。排序损失定义为

$$\mathcal{L}_{\text{RANK}} = \sum_{\substack{i,j \\ e_i=1, t_i<t_j}} \exp\left(\frac{\hat{S}_{i,t_i} - \hat{S}_{j,t_i}}{\sigma}\right), \tag{12}$$

其中 σ 是一个超参数。组合生存损失为 $\mathcal{L}_{\text{SURV}} = \mathcal{L}_{\text{NLL}} + \beta\mathcal{L}_{\text{RANK}}$，其中 β 用于平衡校准与判别性能。

3.2 训练流程

CONVERSE 分三个阶段进行训练：预训练、簇初始化，以及结合自步学习的端到端优化。自步学习（SPL）[14]在优化过程中逐步纳入由易到难的样本，从而提高训练稳定性。易样本是模型置信度高（损失低）的样本，而难样本则位于簇边界附近，类别分配较为模糊。按照 [8], SPL 仅应用于表示学习和聚类组件。在每个训练轮次 e，都会计算一个自适应阈值 λ_e，其计算方式为

$$\lambda_e = \mu(\{\mathcal{L}_i\}_{i=1}^{N}) + \frac{e}{E_{\text{max}}}\sigma(\{\mathcal{L}_i\}_{i=1}^{N}), \tag{13}$$

其中 E_{max} 是总训练轮数，$\mu(\cdot)$ 和 $\sigma(\cdot)$ 分别表示样本级损失的均值和标准差。每个样本级损失由重构项、KL散度项和聚类项构成：

$$\mathcal{L}_i = \alpha_{\text{REC}}\mathcal{L}_{\text{REC},i} + \alpha_{\text{KLD}}\mathcal{L}_{\text{KLD},i} + \alpha_{\text{CLUS}}\mathcal{L}_{\text{CLUS},i}, \tag{14}$$

其中 α_{REC}，α_{KLD}，α_{CLUS} 是权重超参数；使用非变分编码器时，省略 $\mathcal{L}_{\text{KLD},\ i}$。只有满足 $\mathcal{L}_i \le \lambda_e$ 的样本才会参与自步学习（SPL）目标：

$$\mathcal{L}_{\text{SPL}} = \frac{1}{\sum_{i=1}^{N}\mathbb{1}(\mathcal{L}_i \le \lambda_e)}\sum_{i=1}^{N}\mathbb{1}(\mathcal{L}_i \le \lambda_e)\cdot\mathcal{L}_i. \tag{15}$$

随着训练的进行，λ_e 线性增加，逐渐将更难的样本纳入优化过程。

第一阶段通过最小化以下目标，联合预训练自编码器和生存分析头：

$$\min_{\Theta}\left(\alpha_{\text{REC}}\mathcal{L}_{\text{REC}} + \alpha_{\text{KLD}}\mathcal{L}_{\text{KLD}} + \alpha_{\text{SURV}}\mathcal{L}_{\text{SURV}}\right), \tag{16}$$

其中，Θ表示所有网络参数。此阶段用于初始化有意义的表征，然后再引入聚类。

在第二阶段，我们将选定的聚类算法（例如 K-means）应用于学习到的潜在表征，以获得初始聚类中心 $\mathbf{M}$和聚类分配$\mathbf{c}_i$。在后续训练阶段，这些聚类中心将保持固定。

最后，第三阶段通过优化以下目标，以端到端方式训练完整模型，同时启用自定进度学习：

$$\min_{\Theta}\left(\alpha_{\text{SPL}}\mathcal{L}_{\text{SPL}} + \alpha_{\text{CL}}\mathcal{L}_{\text{CL}} + \alpha_{\text{SURV}}\mathcal{L}_{\text{SURV}}\right). \tag{17}$$

每个轮次结束后，我们会更新所有患者的聚类分配$\mathbf{c}_i$，将每位患者重新分配至最近的聚类中心。所有权重超参数（$\alpha., \beta$）、温度参数 τ 以及架构选择，均通过第 4.2 节所述的验证流程进行联合优化。

4 实验

4.1 数据集和基线模型

我们在四个广泛使用的生存分析基准数据集上评估模型。三个数据集聚焦于乳腺癌结局：GBSG（无复发生存期）、METABRIC 和 TCGA_BRCA（总生存期）。第四个数据集 WHAS 用于评估心肌梗死后的生存情况 [13, 21, 22]。为比较性能，我们考察了标准的基于神经网络的方法以及

表1：各数据集的C指数测试结果（↑）。

模型	GBSG	METABRIC	WHAS	TCGA BRCA_
DeepSurv	67.2 ± 4.0	64.6 ± 1.9	77.4 ± 4.7	$\mathbf{74.3 \pm 4.7}$
DeepHit	66.6 ± 3.0	64.8 ± 1.8	77.3 ± 4.1	72.3 ± 4.4
VADESC	67.3 ± 2.8	64.5 ± 2.2	77.3 ± 4.0	72.5 ± 5.8
DCM	65.9 ± 2.4	64.5 ± 2.1	76.4 ± 4.2	70.7 ± 5.1
SCA	66.6 ± 3.2	64.7 ± 2.2	75.2 ± 3.6	67.8 ± 8.4
DVCSurv	$\mathbf{67.8 \pm 2.7}$	65.8 ± 1.7	76.4 ± 4.4	68.2 ± 6.6
CONVERSE	$\mathbf{67.8 \pm 2.5}$	$\mathbf{66.0 \pm 1.8}$	$\mathbf{77.6 \pm 3.8}$	69.3 ± 4.9

表2： IBS测试结果（↓）在各数据集上的表现。

模型	GBSG	METABRIC	WHAS	TCGA BRCA_
DeepSurv	18.7 ± 1.3	19.5 ± 0.8	$\mathbf{15.4 \pm 2.6}$	$\mathbf{10.2 \pm 1.1}$
DeepHit	19.2 ± 0.9	$\mathbf{19.4 \pm 0.7}$	15.6 ± 2.4	10.5 ± 0.7
VADESC	19.3 ± 0.9	$\mathbf{19.4 \pm 0.8}$	16.8 ± 2.3	16.2 ± 2.1
DCM	19.4 ± 1.1	$\mathbf{19.4 \pm 0.8}$	16.7 ± 2.5	11.1 ± 1.0
SCA	31.7 ± 3.3	27.4 ± 1.5	23.9 ± 6.1	27.8 ± 7.6
DVCSurv	20.4 ± 1.1	22.7 ± 0.8	26.8 ± 7.8	13.3 ± 2.2
CONVERSE	$\mathbf{18.6 \pm 1.3}$	21.1 ± 0.8	16.1 ± 2.4	11.5 ± 0.9

基于聚类的方法。前者包括 Deep- Surv [13]和 DeepHit [16], , 作为判别式基线。后者包括 DCM [20], SCA [6],VADESC [17], 和 DVCSurv [8]。

4.2 实验流程与超参数优化

为确保在这些小规模生存数据集上进行可靠的样本外评估，我们对所有模型采用严格的验证流程。我们使用 Optuna [1]，并通过树结构帕尔森估计器采样超参数配置。对于每个配置，我们对原始数据进行五次自助抽样划分，生成五组相互独立的训练集（60%）、验证集（20%）和测试集（20%）。对每组划分分别进行预处理，包括归一化以及对类别变量进行独热编码。我们在五组划分上分别训练每个模型，并根据验证集表现选择最佳超参数配置，同时考虑一致性指数（C指数）和积分布里尔分数（IBS）。最终测试结果报告的是在五组划分上平均验证表现最高的配置所取得的结果。所有模型在离散时间网格上的生存估计均采用线性插值 [2]。

4.3 结果

我们将 CONVERSE 与两类基线方法进行比较：基于神经网络的模型（DeepSurv 和 DeepHit），这类模型优先关注预测性能，而不进行显式的风险分层；以及基于聚类的模型（VADESC、DCM、SCA 和 DVCSurv），这类模型通过聚类整合患者亚组信息。表 1 和表 2 分别报告了 C指数和积分布里尔分数结果，其中最佳性能以粗体显示，次佳性能以下划线标示。所有数值均乘以 100，以提高可读性。

CONVERSE 在三个数据集上取得了最高的 C指数，在四个数据集（GBSG、METABRIC 和 WHAS）中的三个上表现出色，展现了强大的判别性能。与基于聚类的方法相比，CONVERSE 的平均表现稳定提升了 $+0.70\%$。这表明，所提出的变分自编码器集成方法有效提升了用于生存预测的学习表征质量。值得注意的是，CONVERSE 在所有数据集上的表现均与 DVCSurv 持平或更优。与神经网络基线方法相比，CONVERSE 的平均表现相当，但并未胜出。

积分布里尔分数结果揭示了互补性见解。CONVERSE 在 GBSG（18.6）上实现了最佳校准，并在 WHAS 和 TCGA_BRCA 上展现出具有竞争力的性能。与基于聚类的方法相比，CONVERSE 显著改善了校准效果：在积分布里尔分数方面，较基于聚类的方法平均提升 $+1.63\%$，表明集成生存预测头能够有效优化概率估计。同时，在 METABRIC 数据集上，与基于神经网络的模型相比，CONVERSE 的校准效果略有下降。这反映了基于表示的模型所固有的基本权衡：通过聚类划分特征空间可能会在概率估计中引入额外的不确定性，尤其是对于接近簇边界的患者而言。尽管如此，校准差距仍然较小；与以往工作相比，CONVERSE 显著缩小了基于神经网络的方法与基于聚类的方法之间的性能差异。

5 可解释性

为展示 CONVERSE 学习到具有临床意义的患者表征，我们在图 2 中呈现了一组可视化结果，这些结果取自每个数据集表现最佳的超参数对应的模型。该图分为三列，分别突出展示所学习表征的不同方面。第一列展示了所有四个数据集的潜在空间 UMAP 二维投影 [18]，并根据采用 $k = 2$ 的K-均值聚类所得的聚类分配进行着色。我们将聚类数量固定为两个，以探究潜在空间在多大程度上能够自然地将患者区分为较高风险组和较低风险组。在所有数据集中，各聚类在潜在空间中均呈现出良好的分离，表明 CONVERSE 学习到的是结构化且具有判别性的表征，而非任意划分。

第二列展示了每个数据集中两个聚类的 Kaplan-Meier生存曲线。所有四个数据集中的聚类均表现出明显分离，并且对数秩检验 p 值证实差异具有统计学显著性（$p < 0.05$）。总体而言，其中一个聚类表现出较高风险，而另一个聚类则显示出更有利的生存概况（低风险）。这种分离表明，潜在表征捕捉到了具有临床相关性的风险信息。

最后，第三列总结了集群层面的特征重要性。为获得这些估计结果，我们使用集群索引作为标签训练了一个随机森林分类器，并计算基于基尼系数的 [5] 特征重要性，以确定哪些协变量最能区分所学到的亚群。对于乳腺癌数据集，

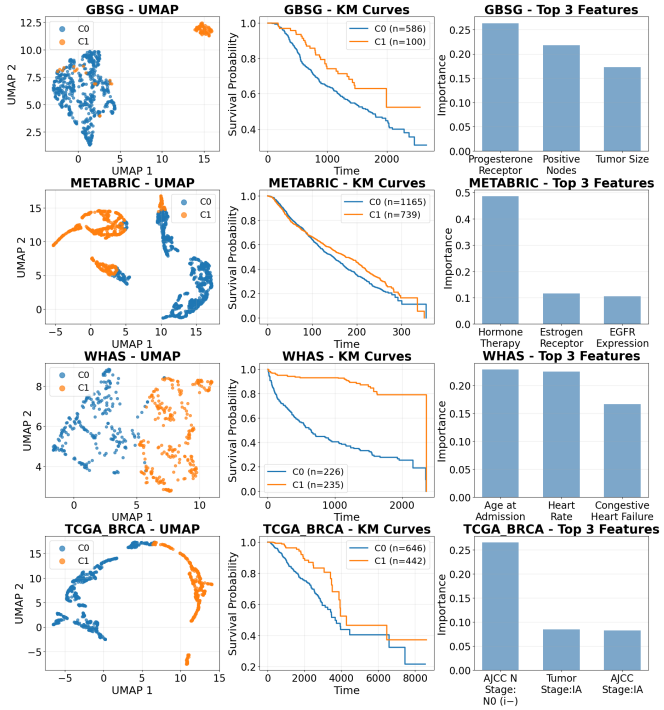

图 2： 训练后的 CON-VERSE 架构的表征可视化。第一列汇集了潜在空间的 UMAP 可视化结果，并标示了基于 K 均值的聚类分配；第二列展示了所得各簇的 Kaplan–Meier 曲线；第三列呈现基于基尼系数的特征重要性，以突出驱动各组划分的协变量。

最重要的特征与公认的预后因素一致。在评估无复发生存期的 GBSG 中，影响最大的特征是孕激素受体状态、阳性淋巴结数量和肿瘤大小——这些都是公认的疾病复发预测因素。 [10]在 METABRIC 中，激素治疗的特征重要性最高，其次是雌激素受体状态和 EGFR 表达，这反映了治疗反应对该队列总生存期的显著影响。在 TCGA_BRCA 中，最重要的特征都与癌症分期有关；这符合临床实际，因为分期仍然是乳腺癌总生存期最有力的预后因素之一。对于关注心肌梗死后生存情况的 WHAS，排名最高的特征是入院年龄、心率和充血性心力衰竭，均属于心血管死亡的主要风险因素。 [4]这些结果表明，CONVERSE 能够识别出具有临床意义的亚群，其特征与各特定结局和数据集的临床知识相符。

6 结论

我们提出了 CONVERSE，这是一种深度生存模型，通过整合变分自编码器、多视图对比学习和簇特定生存头，弥合了判别性能与可解释风险分层之间的差距。在四个数据集上的全面评估表明，与神经网络方法相比，CONVERSE 的表现具有竞争力，同时优于基于聚类的方法。此外，CONVERSE 还提供了具有临床可解释性的患者分层，且各簇对应于公认的预后因素，为基于数据的临床决策提供了实用的解决方案。

致谢

本文得到 FAIR（未来人工智能研究）项目的支持。该项目由 NextGenerationEU 计划资助，属于 PNRR-PE-AI 计划（M4C2，投资 1.3，人工智能方向）。

参考文献

[1] Takuya Akiba、Shotaro Sano、Toshihiko Yanase、Takeru Ohta 和 Masanori Koyama。Optuna：新一代超参数优化框架。载于第*25*届 *ACM SIGKDD* 国际知识发现与数据挖掘会议论文集，第 2623–2631页，2019年。

[2] Alberto Archetti、Francesco Stranieri 和 Matteo Matteucci。弥合差距：通过插值技术改进神经生存模型。《style id='2'>人工智能进展，第1–16页，2024年。

[3] Alberto Archetti、Eugenio Lomurno、Diego Piccinotti 和 Matteo Matteucci。Fpboost：用于生存分析的全参数梯度提升。《style id='2'>人工智能与应用前沿，413：1865–1872，2025年。doi：10.3233/FAIA251019。

[4] Donna K Arnett、Roger S Blumenthal、Michelle A Albert、Andrew B Buroker、Zachary D Goldberger、Ellen J Hahn、Cheryl Dennison Himmelfarb、Amit Khera、Donald Lloyd-Jones、J William McEvoy 等。2019年 ACC/AHA 心血管疾病一级预防指南：执行摘要：美国心脏病学会/美国心脏协会临床实践指南工作组报告。美国心脏病学会杂志，74(10)：1376–1414，2019年。

[5] Leo Breiman、Jerome Friedman、Richard A Olshen 和 Charles J Stone。分类与回归树。Chapman and Hall/CRC，2017。

[6] Paidamoyo Chapfuwa、Chunyuan Li、Nikhil Mehta、Lawrence Carin 和 Ricardo Henao。生存聚类分析。载于 *ACM*健康、推理与学习会议论文集，第 60–68 页，2020 年。

[7] David R Cox。回归模型与生命表。英国皇家统计学会杂志：*B*辑（方法论），34(2):187–202，1972 年。

[8] Chang Cui、Yongqiang Tang 和 Wensheng Zhang。基于双视图聚类的深度对比生存分析。*Electronics*，13(24):4866，2024 年。

[9] Chang Cui、Yongqiang Tang 和 Wensheng Zhang。基于潜在聚类和对比学习的深度生存分析。*IEEE*生物医学与健康信息学杂志，28(5):3090–3101，2024。

[10] 早期乳腺癌试验协作组等。乳腺癌激素受体及其他因素与辅助性他莫昔芬疗效的相关性：随机试验的患者层面荟萃分析。《柳叶刀》，378(9793):771–784，2011。

[11] Hemant Ishwaran、Udaya B Kogalur、Eugene H Blackstone 和 Michael S Lauer。随机生存森林。《应用统计年鉴》，第 841–860 页，2008 年。

[12] Edward L Kaplan 和 Paul Meier。基于不完整观测的非参数估计。《美国统计学会会刊》，53(282):457– 481，1958年。[13] Jared L Katzman、Uri Shaham、Alexander Cloninger、Jonathan Bates、Tingting Jiang 和 Yuval Kluger。DeepSurv：利用 Cox比例风险深度神经网络构建个性化治疗推荐系统。《*BMC*医学研究方法学》，18 (1):24，2018年。[14] M Kumar、Benjamin Packer 和 Daphne Koller。潜变量模型的自适应课程学习。神经信息处理系统进展，23，2010年。[15] Håvard Kvamme 和 Ørnulf Borgan。利用神经网络进行连续时间和离散时间生存预测。《寿命数据分析》，27(4):710–736，2021年。[16] Changhee Lee、William Zame、Jinsung Yoon 和 Mihaela Van Der Schaar。DeepHit：一种用于竞争风险生存分析的深度学习方法。载于*AAAI*人工智能大会论文集，第32卷，2018年。[17] Laura Manduchi、Ričards Marcinkevičs、Michela C Massi、Thomas Weikert、Alexander Sauter、Verena Gotta、Timothy Müller、Flavio Vasella、Marian C Neidert、Marc Pfister 等人。聚类生存数据的深度变分方法。*arXiv*预印本 *arXiv:2106.05763*，2021年。[18] Leland McInnes、John Healy 和 James Melville。UMAP：用于降维的均匀流形近似与投影。*arXiv*预印本 *arXiv:1802.03426*，2018年。[19] Chirag Nagpal、Xinyu Li 和 Artur Dubrawski。深度生存机器：针对含竞争风险删失数据的全参数生存回归与表示学习。*IEEE*生物医学与健康信息学杂志，25(8):3163–3175，2021年。[20]Chirag Nagpal、Steve Yadlowsky、Negar Rostamzadeh 和 Katherine Heller。用于生存回归的深度 Cox 混合模型。载于医疗保健机器学习大会，第674–708页。PMLR，2021年。[21] Jean Ogier du Terrail、Samy-Safwan Ayed、Edwige Cyffers、Felix Grimberg、Chaoyang He、Regis Loeb、Paul Mangold、Tanguy Marchand、Othmane Marfoq、Erum Mushtaq 等人。Flamby：面向真实医疗保健场景中跨机构联邦学习的数据集与基准。神经信息处理系统进展，35:5315–5334，2022年。[22] Sebastian Pölsterl。scikit-survival：一个基于 scikit-learn 构建的事件发生时间分析库。机器学习研究杂志，21(212):1–6，2020年。[23] Ping Wang、Yan Li 和 Chandan K Reddy。生存分析中的机器学习：综述。*ACM*计算调查，51(6):1–36，2019年。[24]Simon Wiegrebe、Philipp Kopper、Raphael Sonabend、Bernd Bischl 和 Andreas Bender。生存分析的深度学习：综述。人工智能评论，57(3):65，2024年。