POEMS: Product of Experts for Interpretable Multi-omic Integration using Sparse Decoding

Mihriban Kocak Balik[1] **Pekka Marttinen**[1] **Negar Safinianaini**[1]
[1]Department of Computer Science, Aalto University, Espoo, Finland
mihribannurkocak@gmail.com, {pekka.marttinen, negar.safinianaini}@aalto.fi

Abstract

Integrating different molecular layers, i.e., multi-omics data, is crucial for unraveling the complexity of diseases; yet, most deep generative models either prioritize predictive performance at the expense of interpretability or enforce interpretability by linearizing the decoder, thereby weakening the network's nonlinear expressiveness. To overcome this trade-off, we introduce POEMS: **P**roduct **O**f **E**xperts for Interpretable **M**ulti-omics Integration using **S**parse Decoding, an unsupervised probabilistic framework that preserves predictive performance while providing interpretability. POEMS provides interpretability *without linearizing* any part of the network by 1) *mapping* features to latent factors using sparse connections, which directly translates to biomarker discovery, 2) allowing for *cross-omic associations* through a shared latent space using product of experts model, and 3) reporting *contributions of each omic* by a gating network that adaptively computes their influence in the representation learning. Additionally, we present an efficient sparse decoder. In a cancer subtyping case study, POEMS achieves competitive clustering and classification performance while offering our novel set of interpretations, demonstrating that biomarker-based insight and predictive accuracy can coexist in multi-omics representation learning.

1 Introduction

Integrating signals from different molecular layers, i.e, omics layers (e.g., gene expression, DNA methylation, miRNA), is a key step toward understanding the complexity of biological systems and diseases. Multi-omics data provide complementary perspectives, but their high dimensionality, heterogeneity, and noise make analysis particularly challenging [4, 2]. Deep generative models such as variational autoencoders (VAEs) [6, 14] have become popular for representation learning in this setting. Multiple extensions have explored the combination of information across different modalities. In the context of bioinformatics, Minoura et al. [9] proposed scMM, a Mixture-of-Experts (MoE) based VAE for single-cell multi-omics integration, and Chen et al. [24] developed MOCSS, an autoencoder combined with contrastive learning to allow for shared and specific representation learning for multi-omics cancer subtyping. Other approaches utilize dimensionality reduction, classical clustering algorithms, and contrastive learning to perform multi-omics integration [13, 22, 23]. These approaches do not address interpretability. Those frameworks that address interpretability are typically restricted to single-omic representation learning [3]. Moreover, even in this setting, they often enforce interpretability by linearizing the decoder network, thereby limiting the expressive power of the network [20, 25, 18]. The trade-off between predictive performance and interpretability in these methods limits their broader utility in machine learning, where it is increasingly critical to understand not only whether a model performs well, but also *why it learns a particular structure in the latent space and how this relates to the observed features* [11, 1, 19]. In the context of generative models, Moran et al. [10] introduced the Sparse VAE, enforcing sparsity in feature-to-factor mapping

EurIPS 2025 Workshop: Multimodal Representation Learning for Healthcare.

(mapping features to latent factors) to promote identifiable and interpretable latent variables. However, it was designed for unimodal, not multi modal data integration.

To address the above limitations, we introduce POEMS: **P**roduct **O**f **E**xperts for Interpretable **M**ulti-omics Integration using **S**parse Decoding, a probabilistic unsupervised representation learning framework with a novel set of interpretation tools. We achieve interpretability and strong predictive performance without linearizing the network or compromising accuracy. POEMS introduces: **1) Sparse feature-to-factor mappings** [10] for interpretable associations between latent factors and omics features, **2) Product-of-Experts (PoE) posterior** [5] that integrates modality-specific posteriors into a closed-form joint distribution, and **3) Gating mechanism** that adaptively weighs the contribution of each omic modality, offering a new dimension of interpretability. To scale Sparse VAEs to high-dimensional omics data, we implemented a vectorized decoder that accelerates training.

Our novel interpretability arises from: 1) *sparse feature-to-factor* mappings that enable biomarker discovery, 2) *cross-omic associations* captured via a shared latent space using a PoE model, and 3) *adaptive per-omic contribution* estimation through a gating network. To the best of our knowledge, POEMS is the first framework to unify interpretable sparse feature–factor mappings with scalable multi-omics generative modeling in an unsupervised learning setting.

Evaluating POEMS on the cancer subtyping task using breast and kidney cancer data, i.e., BRCA and KIRC [12], indicates that POEMS achieves competitive cancer subtyping performance compared to the state-of-the-art methods while providing interpretable insights. These findings may contribute to the design of targeted treatments and, moreover, demonstrate that high predictive performance and interpretability can be achieved simultaneously in multi-omics unsupervised representation learning.

2 Method

We propose **POEMS: P**roduct **O**f **E**xperts for Interpretable **M**ulti-omics Integration using **S**parse Decoding [1], a probabilistic and interpretable multi-omics representation learning tool. POEMS constructs a *shared latent space* via a Product-of-Experts (PoE) and learns *per-omic sparse feature-to-factor* mappings from this joint representation, see Fig 1. This shared latent representation connects different omics modalities, enabling cross-omic interpretability. In addition, a data-dependent gating mechanism assigns modality weights that regulate *each omic's contribution* to the joint posterior. These components yield a unified interpretable latent space with sparse associations between latent dimensions and omic features. To address the feature-wise decoding bottleneck of SparseVAE and make it practical for high-dimensional omics, we implement a *vectorized* decoder, see Appendix C.1.

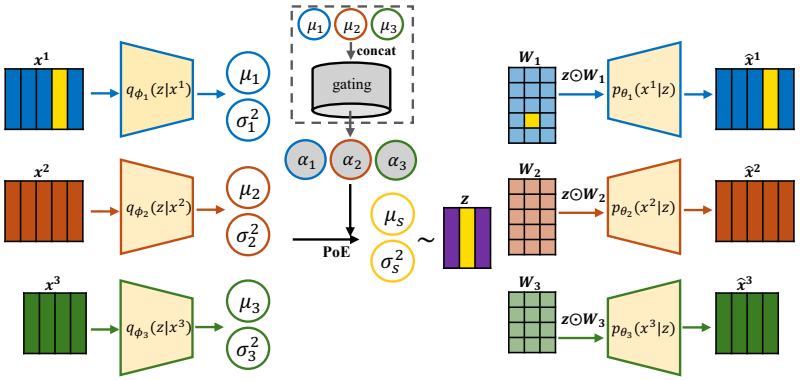


Figure 1: Schematic overview of **POEMS**. Each omics, $\mathbf{x}^i$, is encoded into a Gaussian posterior with mean $\boldsymbol{\mu}_i$ and variance $\boldsymbol{\sigma}_i^2$. The posteriors are fused via a Product-of-Experts, with gating weights $\boldsymbol{\alpha}_v$ controlling each omic's contribution to the shared latent $\mathbf{z}$. Before reconstruction, $\mathbf{z}$ is modulated by the sparse feature-to-factor matrix $\mathbf{W} \in \mathbb{R}^{D\times K}$, ensuring each feature depends on a limited subset of latent dimensions. These masked versions are then passed through modality-specific decoders for reconstruction. The yellow highlight shows the interpretable link between the 4th feature of $\mathbf{x}^1$ and the 2nd latent factor, $\mathbf{W}_1^{4,2}$.

[1] For implementation, see the **link.**

Shared latent space Each omic $v \in \{1, \ldots, V\}$ has an encoder $q_{\phi_v}(\mathbf{z} \mid \mathbf{x}^v)$ that outputs a Gaussian $\mathcal{N}(\boldsymbol{\mu}_v, \boldsymbol{\sigma}_v^2)$. We fuse them with a Product-of-Experts (PoE) [5] to obtain a single inference distribution $q_\phi(\mathbf{z} \mid \mathbf{x}^{1:V}) \propto \prod_{v=1}^{V} q_{\phi_v}(\mathbf{z} \mid \mathbf{x}^v) \propto \mathcal{N}(\boldsymbol{\mu}_s, \boldsymbol{\sigma}_s^2)$, which yields precision-weighted closed forms. To improve robustness to heterogeneous modalities and enhance interpretability, a *gating network* predicts normalized weights $\boldsymbol{\alpha}_v \in [0, 1]$ and rescales the precisions, as shown in Equation (1). The $\boldsymbol{\alpha}_v$ provides relative modality contributions and mitigates domination by overconfident experts.

$$\boldsymbol{\sigma}_s^2 = \Big(\sum_{v=1}^{V} \boldsymbol{\alpha}_v \boldsymbol{\tau}_v \Big)^{-1}, \qquad \boldsymbol{\mu}_s = \frac{\sum_{v=1}^{V} \boldsymbol{\alpha}_v \boldsymbol{\tau}_v \, \boldsymbol{\mu}_v}{\sum_{v=1}^{V} \boldsymbol{\alpha}_v \boldsymbol{\tau}_v}, \quad \boldsymbol{\tau}_v = \boldsymbol{\sigma}_v^{-2}. \tag{1}$$

Sparse feature-to-factor mapping In conventional deep generative models, every latent factor is assumed to contribute to every observed feature. However, in many real-world domains such as genomics, this assumption is unrealistic: only a small subset of latent factors typically influences each feature. Sparse deep generative models address this by introducing a *feature-to-factor mapping* that selectively links factors to features. Building on this idea, SparseVAE [10] enforces sparsity in these mappings, enabling interpretable associations between latent dimensions and meaningful subsets of features. POEMS adopts the SparseVAE approach, enforcing sparsity in the feature-to-factor mappings ($\mathbf{W}_v$) via a Spike-and-Slab Lasso [15] prior, yielding localized loadings *connected across omics* through the shared latent variable $\mathbf{z}$.

Per-omic sparse decoding Given the shared latent $\mathbf{z} \in \mathbb{R}^K$, each omic v is reconstructed via a SparseVAE-style decoder, i.e., $p_{\theta_v}(\mathbf{x}^v \,|\, \mathbf{z})$, equipped with its own sparse feature-to-factor mapping $\mathbf{W}_v \in \mathbb{R}^{D_v \times K}$. For feature j in omic v, the decoder conditions on a masked latent input $\tilde{\mathbf{z}}_{v,j} = \mathbf{z} \odot (\mathbf{W}_v)_j$, where $(\mathbf{W}_v)_j$ denotes j-th row of $\mathbf{W}_v$, and $\odot$ denotes element-wise multiplication between the latent vector and the corresponding feature–factor row. Thus, for omic v, $\mathbf{z} \odot \mathbf{W}_v$ generates all D_v masked versions of the shared latent vector simultaneously, which are passed to their corresponding decoder components for feature-wise reconstruction.

Objective function (ELBO) POEMS optimizes a single shared posterior $q_\phi(\mathbf{z} \mid \mathbf{x}^{1:V})$. We denote $q_\phi(\mathbf{z} \mid \mathbf{x}^{1:V})$ as $Q_{\phi,1:V}$ and $p_{\theta_\nu}(\mathbf{x}^\nu|\mathbf{z})$ as $P_{\theta,\nu}$ in the ELBO, as defined in Equation (2). That is, there is one KL term for the shared PoE posterior and a sum of per-omic reconstruction terms; the sparsity prior over each $\mathbf{W}_v$ enters via the same MAP/EM-style block used in single-omic SparseVAE (mask penalty and Beta–Bernoulli updates), applied independently for each omic v.

$$\sum_{v=1}^{V} \left\{ \underbrace{\mathbb{E}_{Q_{\phi,1:V}}[\log P_{\theta,\nu}]}_{\text{reconstruction}} + \underbrace{\mathbb{E}_{\boldsymbol{\Gamma}_v|\mathbf{W}_v,\,\eta_v}[\log p(\mathbf{W}_v \mid \boldsymbol{\Gamma}_v)\, p(\boldsymbol{\Gamma}_v \mid \eta_v)\, p(\eta_v)]}_{\text{sparsity prior over } \mathbf{W}_v} \right\} - \underbrace{\mathrm{KL}(Q_{\phi,1:V} \,\|\, p(\mathbf{z}))}_{\text{shared KL}}. \tag{2}$$

3 Experiments

We evaluate POEMS on breast and kidney cancer multi-omics datasets, i.e., BRCA and KIRC from The Cancer Genome Atlas (TCGA) [12], using mRNA expression, DNA methylation, and miRNA expression modalities, with details provided in Appendix A.1.POEMS is compared with representative multi-omics baselines; see Appendix A.3. POEMS achieves competitive predictive performance across datasets. On BRCA, it attains the highest clustering and classification scores, demonstrating that combining a shared PoE-based latent space with sparsity yields discriminative representations. On KIRC, the deterministic baseline performs slightly better in clustering due to the smaller sample size, while all models reach near-perfect supervised accuracy. As this paper focuses on interpretability, detailed results of predictive performance are provided in Appendix B.2. We examine the interpretability of POEMS on the BRCA, as we have omic labels for this dataset.

Biomarker detection The sparse feature-to-factor mappings learned by POEMS enable biological interpretation through feature-level inspection. Figure 2 shows the top 10 most influential features for each omic modality across latent dimensions. For example, latent factor 7 shows strong activation in DNA methylation and miRNA, highlighting relevant genes and regulatory patterns. Since the latent factors live in the same space among omics, one can also infer associations between omics. This finding is particularly informative, as predictive performance in multi-omics models is often dominated by mRNA, as shown in the Modality contributions section.

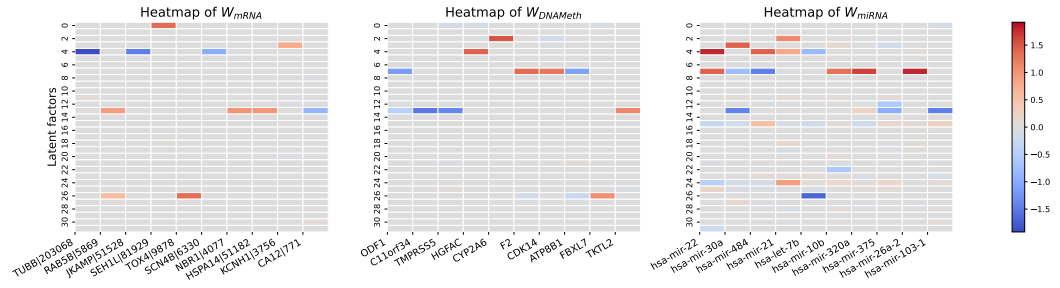


Figure 2: Top 10 activated features per omic, showing modality-specific feature-to-factor associations.

Cancer subtype correlations To assess biological structure in the latent space, we compare subtype correlation matrices computed from input features and learned latent representations. As shown in Figure 3, the latent space reveals clearer subtype distinctions. Because some subtypes share characteristics such as hormone-receptor expressions (ER, PR, HER2), residual correlations are expected (see Appendix B.3). Overall, POEMS successfully separates the subtypes while capturing biologically plausible correlations, yielding latent factors that align with clinical subtype structure.

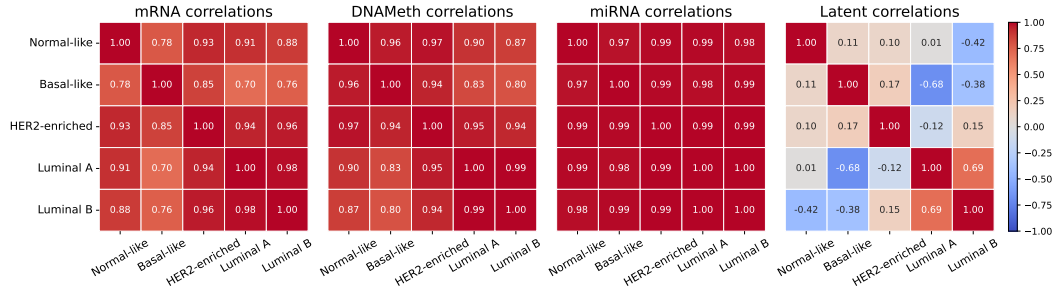


Figure 3: Subtype correlation maps with respect to input features and latent factors.

Modality contributions The gating network in POEMS assigns modality-specific weights quantifying each omic's contribution to the shared posterior. Figure 4 indicates that mRNA dominates overall, while DNA methylation and miRNA provide complementary signals, showing that the model adaptively balances modalities based on their informativeness.

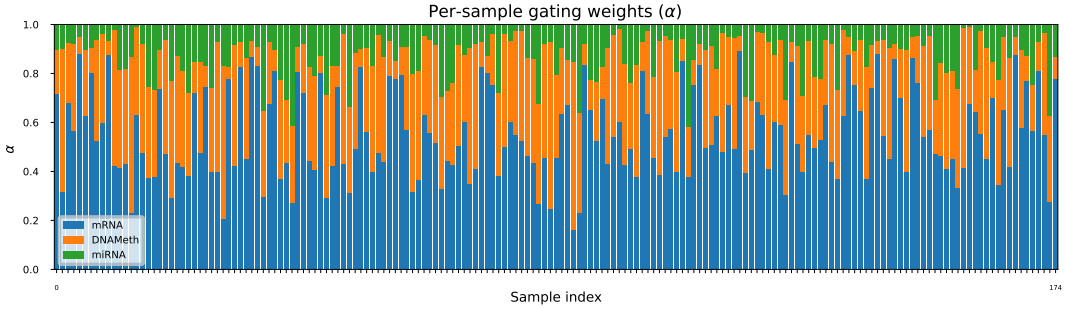


Figure 4: Per-sample gating weights (α) indicating each omic's contribution.

4 Conclusion

We presented POEMS, a sparse interpretable deep generative framework for unsupervised multi-omics integration that 1) maps features to latent factors via sparse connections for biomarker discovery, 2) captures cross-omic associations through a shared latent space using a Product-of-Experts posterior, and 3) quantifies omic-specific contributions via a gating mechanism. POEMS offers feature- and modality-level interpretability while maintaining strong cancer subtyping performance. Experiments on limited breast and kidney data reveal biologically coherent latent structures, capturing cross-omic links and modality contributions. To account for more structure in latent space, e.g., [17], and robustness towards hyperparameter tuning, future work can be to refine the latent space and its invariance to hyperparameter choices [16]. Overall, POEMS demonstrates that interpretability and predictive power can coexist in deep multi-omics unsupervised learning.

Acknowledgments

The results presented here are, in whole or in part, based upon data generated by the TCGA Research Network: `https://www.cancer.gov/tcga`. We acknowledge the computational resources provided by the Aalto University School of Science "Science-IT" project, which supported the experiments conducted in this work. We thank the anonymous reviewers for their constructive feedback, which strengthened this paper.

References

[1] A.B. Arrieta and et al. Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information fusion*, 58:82–115, 2020.

[2] V. Baiao, Z. Cai, R.C. Poulos, P.J. Robinson, R.R. Reddel, Q. Zhong, S. Vinga, and E. Goncalves. A technical review of multi-omics data integration methods: from classical statistical to deep generative approaches. *Brief Bioinform*, 2025.

[3] Y. Choi, Li R., and G. Gerald Quon. sivae: interpretable deep generative models for single-cell transcriptomes. *Genome Biology*, 2023.

[4] Y. Hasin, M. Seldin, and A. Lusis. Multi-omics approaches to disease. *Genome biology*, 18:1–15, 2017.

[5] G.E. Hinton. Products of experts. In *Proceedings of the 9th International Conference on Artificial Neural Networks (ICANN'99)*, volume 1999, pp. 1–6. IEE, 1999.

[6] D.P. Kingma and M. Welling. Auto-encoding variational bayes, 2013.

[7] H.W. Kuhn. The Hungarian Method for the Assignment Problem. *Naval Research Logistics Quarterly*, 2(1–2):83–97, March 1955.

[8] Harvey N. Mayrovitz. *Breast Cancer*. Exon Publications, Brisbane (AU), 2022.

[9] K. Minoura, K. Abe, H. Nam, H. Nishikawa, and T. Shimamura. A mixture-of-experts deep generative model for integrated analysis of single-cell multiomics data. *Cell reports methods*, 1(5), 2021.

[10] G.E. Moran, D. Sridhar, Y. Wang, and D.M. Blei. Identifiable deep generative models via sparse decoding. *TMLR*, 2022.

[11] W.J. Murdoch, C. Singh, K. Kumbier, R. Abbasi-Asl, and B. Yu. Definitions, methods, and applications in interpretable machine learning. *Proceedings of the National Academy of Sciences*, 116(44):22071–22080, 2019.

[12] National Cancer Institute. The cancer genome atlas (tcga). `https://www.cancer.gov/tcga`, 2025. Accessed: 2025-11-15.

[13] N. Rappoport, R. Shamir, and R. Schwartz. Nemo: cancer subtyping by integration of partial multi-omic data. *Bioinformatics*, 2019.

[14] D.J. Rezende, S. Mohamed, and D. Wierstra. Stochastic backpropagation and approximate inference in deep generative models. In *International conference on machine learning*, pp. 1278–1286. PMLR, 2014.

[15] V. Rovckova and E.I. George. The spike-and-slab lasso. *Journal of the American Statistical Association*, 113(521):431–444, 2018.

[16] H. Roy, M. Miani, C.H. Ek, P. Hennig, M. Pfortner, L. Tatzel, and S. Hauberg. Reparameterization invariance in approximate bayesian inference. *NeurIPS*, 2024.

[17] N. Safinianaini, N. Valimaki, R. Bresson, A. Gorbonos, K. Rajamaki, L.A. Aaltonen, and P. Marttinen. Omidient: Multiomics integration for cancer by dirichlet auto-encoder networks. Jul 2025.

[18] L. Seninge, I. Anastopoulos, H. Ding, and J. Stuart. Vega is an interpretable generative model for inferring biological network activity in single-cell transcriptomics. *Nat Commun*, 2021.

[19] D. Sidak, J. Schwarzerova, W. Weckwerth, and S. Waldherr. Interpretable machine learning methods for predictions in systems biology from omics data. *Frontiers in Molecular Biosciences*, 9:926623, 2022.

[20] V. Svensson, A. Gayoso, N. Yosef, and L. Pachter. Interpretable factor models of single-cell rna-seq via variational autoencoders. *Bioinformatics*, 2020.

[21] p. Turova and et al. The breast cancer classifier refines molecular breast cancer classification to delineate the her2-low subtype. *Nature npj breast cancer*, 2025.

[22] W. Wang, X. Zhang, and D. Dai. Defusion: a denoised network regularization framework for multi-omics integration. *Briefings in Bioinformatics*, 2021.

[23] M. Yang, Y. Yang, C. Xie, M. Ni, J. Liu, H. Yang, F. Mu, and J. Wang. Contrastive learning enables rapid mapping to multimodal single-cell atlas of multimillion scale. *Nature Machine Intelligence*, 2022.

[24] C. Yuxin, W. Yuqi, X. Chenyang, C. Xinjian, H. Song, B. Xiaochen, and Z. Zhongnan. Mocss: Multi-omics data clustering and cancer subtyping via shared and specific representation learning. *Iscience*, 26(8), 2023.

[25] Y. Zhao, H. Cai, Z. Zhang, J. Tang, and Y. Li. Learning interpretable cellular and gene signature embeddings from single-cell transcriptomic data. *Nat Commun*, 2021.

A Experimental setup

A.1 Datasets description

Table 1: Summary of multi-omics data for BRCA (breast cancer) and KIRC (kidney cancer) datasets

Dataset	mRNA expression	DNA methylation	miRNA expression	Samples	Subtypes
BRCA	1000	1000	503	875	5
KIRC	58316	22928	1879	289	2

The evaluation of all models is conducted using two publicly available multi-omics datasets obtained from The Cancer Genome Atlas (TCGA) project [12]. A summary of the dataset statistics is provided in Table 1. The first dataset, BRCA, represents breast cancer samples, while the second, KIRC, corresponds to kidney renal carcinoma. The preprocessed BRCA dataset was obtained from the MOCSS [24]. Each dataset comprises three omic layers: mRNA expression, DNA methylation, and miRNA expression profiles. The BRCA dataset consists of 875 samples encompassing 1,000 mRNA genes, 1,000 DNA methylation sites, and 503 miRNA features, distributed across five molecular subtypes. In contrast, the KIRC dataset contains 289 samples with 58,316 mRNA genes, 22,928 methylation sites, and 1,879 miRNA features, spanning two subtypes. Together, these datasets offer complementary evaluation conditions such that BRCA providing a larger and more diverse cohort suitable for examining clustering and interpretability, and KIRC serving as a smaller-scale benchmark to assess the robustness of model performance in low-sample scenarios.

A.2 Evaluation metrics

To thoroughly evaluate the learned representations, we employ both unsupervised and supervised evaluation schemes. The unsupervised metrics assess how well the latent space captures the intrinsic subtype structure, while the supervised evaluation measures how well the learned embeddings separate subtypes in a supervised setting.

For the unsupervised evaluation, we perform K-means clustering on the latent embeddings and compute two complementary metrics: Normalized Mutual Information (NMI) and Accuracy (ACC). These metrics quantify the alignment between the predicted cluster assignments $\hat{\mathbf{y}}$ and the ground-truth subtype labels $\mathbf{y}$.

– **Normalized Mutual Information (NMI)** quantifies the mutual dependence between two assignments, as defined in Equation (3):

$$\mathrm{NMI}(\mathbf{y}, \hat{\mathbf{y}}) = \frac{2 \cdot I(\mathbf{y}; \hat{\mathbf{y}})}{H(\mathbf{y}) + H(\hat{\mathbf{y}})}, \tag{3}$$

where $I(\cdot;\cdot)$ denotes mutual information and $H(\cdot)$ denotes entropy. NMI values range from 0 (no mutual information) to 1 (perfect alignment).

– **Accuracy (ACC)** measures the proportion of correctly clustered samples after optimal label matching via the Hungarian algorithm [7], as shown in Equation (4):

$$\mathrm{ACC} = \frac{1}{N} \sum_{i=1}^{N} \delta(y_i, \pi(\hat{y}_i)), \tag{4}$$

where $\delta(\cdot,\cdot)$ is the Kronecker delta and π denotes the optimal permutation mapping between cluster labels and ground truth labels.

For the supervised evaluation, we train a k-Nearest Neighbors (KNN) classifier using the latent representations and report its classification accuracy (ACC). This metric provides an additional perspective on how well the embeddings separate subtypes when label information is used, complementing the unsupervised clustering results.

A.3 Baseline methods

To ensure a fair and comprehensive comparison, we evaluate five model variants that differ in their architectural components and training objectives but share the common goal of learning representations across multiple omic modalities:

– **MOCSS-AE**: The baseline MOCSS framework [24], which employs modality-specific autoencoders (AEs) to extract both shared and specific latent representations across omics.

– **MOCSS-VAE**: Our modified version of MOCSS in which the autoencoders that are learning omic-specific representations are replaced by variational autoencoders (VAEs), enabling probabilistic latent inference.

– **MOCSS-SparseVAE**: Our variant extends MOCSS-VAE by replacing VAEs with Sparse VAEs [10] for omic-specific representation learning.

– **POEM**: Our variant of POEMS, replacing SparseVAEs with VAEs.

– **POEMS**: Our main model proposed in this study, combining the PoE posterior with Sparse VAEs to jointly achieve interpretable, sparse feature–factor mappings and robust multi-omic integration.

A.4 Training configurations

All models were trained under a unified experimental configuration to ensure consistency and fairness in comparison. The dataset was partitioned into training, validation, and test sets following an 80%–20% split for training and testing, and an additional 20% of the training portion was reserved for validation. Each model was optimized for 5,000 epochs with a latent dimensionality of 32, while employing early stopping to mitigate overfitting. The MOCSS-AE model (MOCSS [24]) followed the training procedure outlined in its original implementation, utilizing the Adam optimizer. In contrast, all remaining models adopted the AdamW optimizer. Hyperparameters (batch size (BS), learning rate (LR), and weight decay (WD)) were tuned separately for each model using a shared search grid, and the optimal configuration was determined based on the minimum total validation loss.

All model trainings were performed using a fixed random seed of 21 to ensure reproducibility. For the evaluation phase, K-means clustering and k-NN classification were each repeated using five different random seeds $0, 12, 21, 42, 1234$, and the reported subtyping results in Tables 2 and 3 correspond to the mean and standard deviation across these runs.

All experiments were executed on a high-performance computing (HPC) environment using a SLURM workload scheduler. Each job was run on a single CPU node with 16GB of memory and a 10-hour wall-time limit. The experiments were managed through an automated SLURM array job script that executed independent tasks corresponding to combinations of hyperparameters. The script automatically launched training runs for each configuration and stored separate output and error

logs for all jobs. The exact per-model training times and total compute consumption were not systematically recorded, as computational efficiency was not the primary focus of this study. The experiments were designed to evaluate interpretability and methodological performance rather than computational scalability. Consequently, while all models were executed under identical hardware and resource constraints to ensure fair comparison, detailed runtime profiling was not required for the conclusions drawn in this work. It should be noted that the overall research project required more computational resources than the experiments directly reported in the paper. This includes additional exploratory and trial runs conducted during model development and debugging phases. These preliminary experiments were essential for refining the final methodology but are not included in the presented results.

B Complementary results

B.1 Interpretability

To support the interpretability analysis presented in Section 3, we provide additional qualitative results obtained on the test set from the BRCA dataset. These complementary visualizations show that POEMS learns biologically interpretable and structured feature–factor relationships across modalities.

Figure 5 illustrates the aggregated activation strengths of the top 10 features identified within each omic (mRNA, DNAMeth, and miRNA). It highlights the dominant features associated with each latent factor, supporting the identification of shared or distinct pathways across omics. Figure 6 expands on this analysis by presenting the top contributing features per latent dimension. This mapping provides a direct interpretation of which biological variables drive particular latent directions in the representation space.

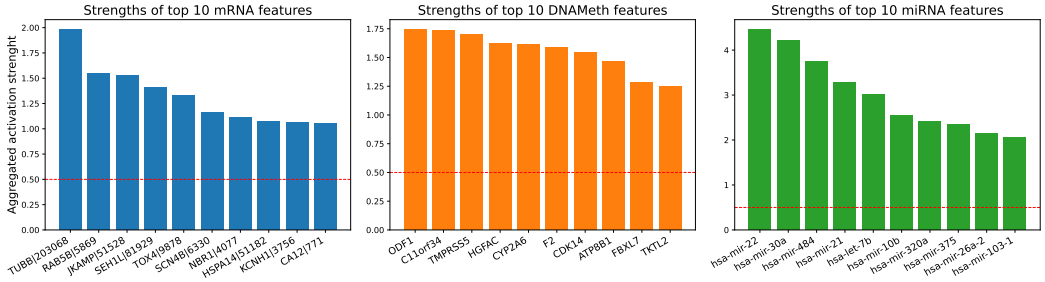


Figure 5: Aggregated activation strengths of the top 10 features across mRNA, DNAMeth, and miRNA modalities, derived from each omic's feature–factor mapping matrix $\mathbf{W}$. For each feature, the aggregated strength is computed as the sum of absolute loading values across all latent dimensions in $\mathbf{W}$. The x-axis labels are the original omic feature names.

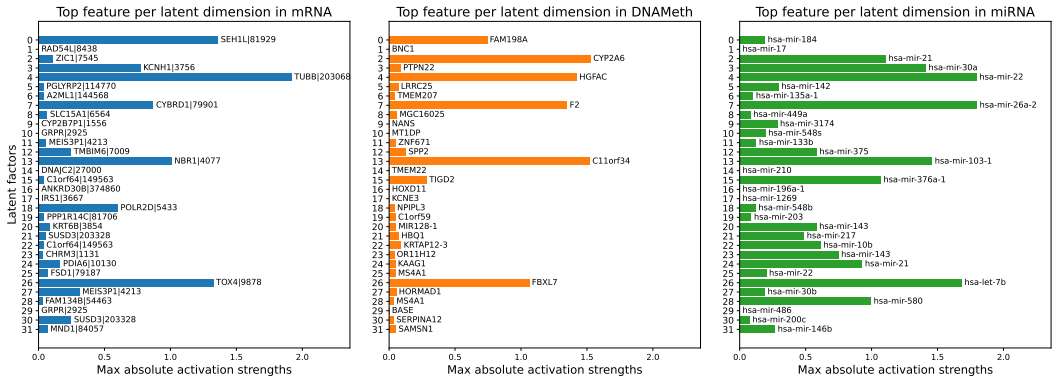


Figure 6: Top contributing features per latent dimension across mRNA, DNAMeth, and miRNA modalities. Each bar represents the maximum absolute activation strength of a feature within a given latent dimension, derived from the corresponding feature–factor mapping matrix $\mathbf{W}$ of each omic. The feature names on the bars correspond to the most dominant input features associated with each latent factor.

To visualize sparsity and activation structure within the learned weight matrices, Figure 7 displays binary activation maps of the absolute loadings $|\mathbf{W}|$ for all three omics. The resulting sparse and localized activation patterns confirm that the spike-and-slab prior effectively enforces interpretability at the feature level.

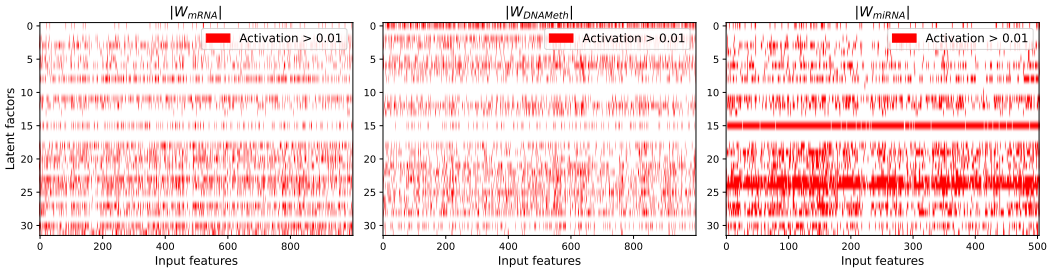


Figure 7: Binary activation maps of the absolute feature–factor loadings $|\mathbf{W}|$ corresponding to each omic modalities mRNA, DNAMeth, and miRNA. Each row corresponds to a latent factor, and each column to an input feature. Red regions indicate active feature–factor connections with absolute activation strength greater than 0.01.

Figure 8 presents the heatmap of latent embeddings, where samples are sorted according to their cluster assignments. The block-like structures along the vertical axis indicate subtype-specific activation patterns, suggesting that the latent space encodes biologically coherent sample groupings consistent with known BRCA subtypes. Finally, Figure 9 compares two-dimensional projections of the learned latent representations using t-SNE and UMAP. Each point corresponds to a single BRCA sample colored by its molecular subtype. Both visualizations reveal meaningful clustering in the latent space, with samples of the same subtype forming separable groups. These results further confirm that the learned representations are both interpretable and discriminative with respect to relevant subtypes.

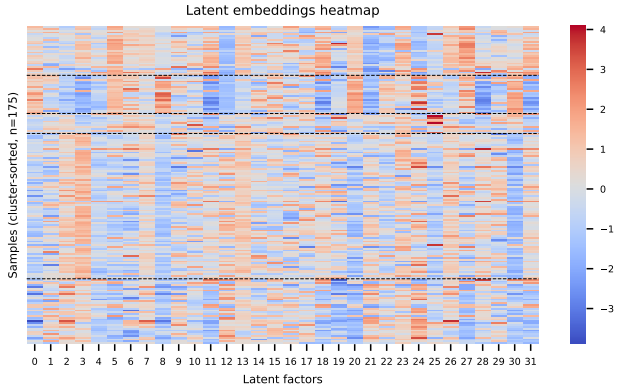
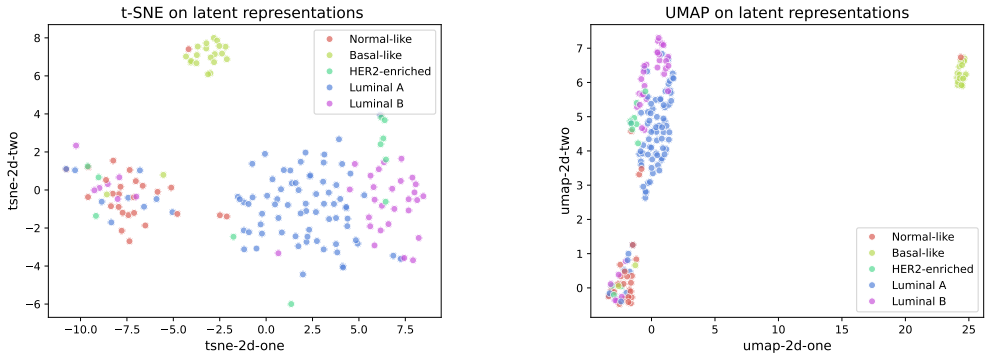


Figure 8: Heatmap of the learned latent embeddings, with samples sorted by cluster assignment. Each row represents a sample, and each column a latent factor. The dashed horizontal lines mark the boundaries between sample clusters, revealing structured variation in the latent space that reflects subtype-specific embedding patterns.

B.2 Subtyping performance

The following quantitative results complement and support the interpretability analysis presented in Section 3. They demonstrate that the proposed models, particularly those integrating sparsity and multi-omic fusion, not only yield latent representations that are more structured and interpretable but also achieve strong subtyping performance. These findings reinforce the conclusions drawn from the qualitative analyses, highlighting the consistency between the model's interpretability and its discriminative capability.

As shown in Table 2, **POEMS** achieves the best overall performance across all three evaluation metrics on the BRCA dataset. It attains the highest K-means clustering accuracy (0.63), NMI (0.45), and KNN classification accuracy (0.78), demonstrating that combining a PoE integration with sparsity decoding enhances both discriminative power and clustering coherence. The improvements over



(a) t-SNE projection of the latent representations.

(b) UMAP projection of the latent representations.

Figure 9: Two-dimensional visualizations of the learned latent representations using t-SNE and UMAP. Each point corresponds to a sample colored by its breast cancer subtype (Normal-like, Basal-like, HER2-enriched, Luminal A, and Luminal B). Both projections reveal meaningful structure in the latent space, where samples of the same subtype form coherent clusters.

Table 2: Subtyping performance on the BRCA dataset using K-means clustering and KNN classification applied on the resulting latent representations of the test set. Bold indicates best results.

Model	BS	LR	WD	ACC_{kmeans}	NMI_{kmeans}	ACC_{knn}
MOCSS-AE	256	3e–4	7e–4	0.62 ($\pm$0.07)	0.41 ($\pm$0.04)	0.68 ($\pm$0.04)
MOCSS-VAE	512	5e–4	5e–4	0.57 ($\pm$0.06)	0.43 ($\pm$0.03)	0.73 ($\pm$0.02)
MOCSS-SparseVAE	512	7e–4	5e–4	0.58 ($\pm$0.05)	0.42 ($\pm$0.01)	0.71 ($\pm$0.02)
POEM	512	7e–4	3e–4	0.54 ($\pm$0.02)	0.38 ($\pm$0.03)	0.72 ($\pm$0.05)
POEMS	512	9e–4	1e–4	**0.63** ($\pm$0.05)	**0.45** ($\pm$0.04)	**0.78** ($\pm$0.04)

MOCSS-AE and **POEM** indicate that probabilistic inference and sparse decoding jointly contribute to the quality of the learned representations.

Table 3: Subtyping performance on the KIRC dataset using K-means clustering and KNN classification applied on the resulting latent representations of the test set. Bold indicates best results.

Model	BS	LR	WD	ACC_{kmeans}	NMI_{kmeans}	ACC_{knn}
MOCSS-AE	32	3e–4	3e–4	**0.93** ($\pm$0.03)	**0.60** ($\pm$0.10)	0.99 ($\pm$0.03)
MOCSS-VAE	32	5e–4	7e–4	0.91 ($\pm$0.00)	0.54 ($\pm$0.00)	0.99 ($\pm$0.03)
MOCSS-SparseVAE	32	3e–4	7e–4	0.91 ($\pm$0.00)	0.54 ($\pm$0.00)	**1.00** ($\pm$0.00)
POEM	32	9e–4	9e–4	0.69 ($\pm$0.04)	0.21 ($\pm$0.07)	**1.00** ($\pm$0.00)
POEMS	32	3e–4	1e–4	0.90 ($\pm$0.00)	0.50 ($\pm$0.01)	**1.00** ($\pm$0.00)

On the other hand, Table 3 shows that on the KIRC dataset, **MOCSS-AE** outperforms all VAE-based variants in clustering metrics, achieving the highest K-means accuracy (0.93) and NMI (0.60). This suggests that deterministic architectures may be better suited for low-sample scenarios such as KIRC, where the stochasticity introduced by VAEs can lead to over-regularization. Nonetheless, all models-including **POEMS**-achieve near-perfect KNN accuracy, indicating that their latent representations remain highly discriminative when evaluated in a supervised manner.

B.3 Subtyping correlations

In this section, we present a detailed analysis of the breast cancer subtypes based on the latent correlations, guided by the established molecular and clinical characteristics of breast cancer [8, 21].

HER2-enriched: In Fig. 3, HER2-enriched subtype is not highly correlated with Luminals A or B nor correlated with Normal-like subtype, which is shown to resemble the Luminals [21]. This observation can be confirmed by the lack of HER2 expression in the Luminals [8]. The lack of high correlation with Basal-like subtype can be due to the opposite presence of ER and PR expressions.

Normal-like: In Fig. 3, we see a correlation between Normal-like subtype and Luminals A and B. This observation is partly aligned with previously shown result on the correlation between Luminal B and Normal-like subtypes [21].

Basal-like: In Fig. 3, we observe correlation between Basal-like subtype and Luminals A and B. This can be explained by their similarity in ER and PR expressions [8].

Luminal A: In Fig. 3, we can see correlation with Basal-like, which can be explained by their common ER and PR expressions [8]. Luminal A correlation with Luminal B in this figure can be explained by the common ER expression and the fact that both subtypes are primary and main breast cancer subtypes [8].

Luminal B: The correlations shown in Fig. 3, i.e., correlation of Luminal B with Luminal A, Basal-like and Normal-like subtypes are explained in the above and aforementioned analysis.

C Implementation details

C.1 Optimization of the Sparse VAE decoder

As declared for the original Sparse VAE [10], a key computational limitation of the model lies in its decoder design, which performs feature-wise reconstruction using separate masked latent vectors for each input dimension. This results in a computational complexity that scales linearly with the number of features, posing a bottleneck for high-dimensional omics data.

To address this inefficiency, we implemented a fully vectorized version of the Sparse VAE decoder that leverages tensor broadcasting and batched operations to parallelize computations across all features simultaneously. This removes the need for sequential feature-wise decoding, significantly improving runtime efficiency. In our experiments, this optimization reduced the average training time per epoch from approximately 6.5 seconds to 1.7 seconds under identical conditions, yielding more than a threefold speedup.

This optimization substantially enhances the scalability of Sparse VAE while preserving its interpretability benefits. Consequently, the model becomes feasible for large-scale omics applications and serves as a more practical foundation for our multi-omics extensions presented in the paper.