MIND: Multimodal Integration with Neighbourhood-aware Distributions

Hanwen Xing
Nuffield Department of
Women's & Reproductive Health
University of Oxford
United Kingdom
hanwen.xing@wrh.ox.ac.uk

Christopher Yau
Nuffield Department of
Women's & Reproductive Health
University of Oxford
Health Data Research UK
United Kingdom
christopher.yau@wrh.ox.ac.uk

Abstract

Multimodal data integration is a powerful tool for combining different data modalities to improve predictive and classification performance when multiple data modalities are available. In biology, multi-omics profiling has become a powerful tool for biomedical applications such as cancer patient stratification and clustering. However, the characterisation and integration of multi-omics data remain challenging because of missingness and inherent heterogeneity. Methods such as imputation and sample exclusion often rely on strong assumptions that could potentially lead to information loss or distortion. To address these limitations, we propose MIND (Multimodal Integration with Neighbourhood-aware Distributions) that learns patient-specific embeddings from incomplete multi-omics data based on a multimodal Variational Autoencoder with a data-driven prior that injects neighbourhood structure of the observed dataset encoded as affinity matrices into the prior of embeddings through exponential tilting. Our proposed method handles high missing rate and unbalanced missingness pattern well, and is robust in the presence of data with a low signal-to-noise ratio. Compared with existing data integration methods, the proposed method achieves better performance on a range of supervised and unsupervised downstream tasks on both synthetic and real data. MIND can also be applied to other multimodal learning domains such as neuroscience, healthcare, and sensor fusion.

1 Introduction

Multi-omics data integration has become essential for understanding complex biological systems and advancing precision medicine (Shin et al., 2017; Subramanian et al., 2020; Steyaert et al., 2023). Modern high-throughput technologies enable simultaneous profiling of genomics, transcriptomics, proteomics, and epigenomics from the same samples, each providing complementary views of cellular function (Cao et al., 2018; Regner et al., 2021; Fu et al., 2024). However, integrating these heterogeneous, high-dimensional datasets presents significant computational challenges, particularly when dealing with the pervasive missing data patterns characteristic of real-world multi-omics studies.

Existing integration methods face fundamental limitations in handling incomplete data. Network-based approaches often require overlapping observations or rely on ad hoc imputation or graph fusion strategies (Rappoport and Shamir, 2019; Xu et al., 2021; Ma et al., 2025). Matrix factorisation methods assume linear relationships and struggle with complex missing patterns (Shen et al., 2012; Yang and Michailidis, 2016). Although variational autoencoders (VAE) can naturally accommodate missing data (Gayoso et al., 2021), aggregating information from multimodal data with missing

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: .

values using VAE requires careful design of both the training scheme (Wu and Goodman, 2018) and the modelling architecture (Ballard et al., 2025; Beaude et al., 2025), which can be computationally intensive. Furthermore, unlike network-based approaches, current VAE models do not exploit the neighbourhood structures within multi-omics datasets explicitly. As a result, they may not be able to preserve the intrinsic clustering structure present in biological data.

We present Multimodal Integration with Neighbourhood-aware Distributions (MIND), a lightweight and conceptually general multimodal Variational Autoencoder (VAE) framework. Unlike existing methods that either rely on modality-specific heuristics or complex interaction architectures, MIND directly incorporates data-driven neighbourhood priors into the latent space. This design principle enables the model to maintain interpretable and biologically meaningful representations in biomedical domains, while also generalising to any multimodal learning setting where incomplete data and heterogeneous noise are pervasive. MIND utilises t-SNE-derived affinity matrices to construct a data-driven prior that preserves neighbourhood structures from individual omics modalities, ensuring biologically meaningful patient-level representations while maintaining VAE's probabilistic modelling framework. Additionally, MIND employs cross-modal regularisation to encourages consistency and information sharing between modality-specific encodings from the same patient, thereby stabilising the aggregation step while maintaining flexibility for arbitrary missing patterns.

When applied to multi-omics, we demonstrate MIND's superior performance across synthetic datasets and real-world applications including The Cancer Genome Atlas (Weinstein et al., 2013), the Childhood Cancer Multi-omics Atlas (Sun et al., 2023), and the Cancer Cell Line Encyclopedia (Ghandi et al., 2019). Our method consistently outperforms or achieves performance on par with existing approaches on cancer classification and data reconstruction tasks, providing a robust framework for multi-omics integration in diverse biological contexts.

2 Method

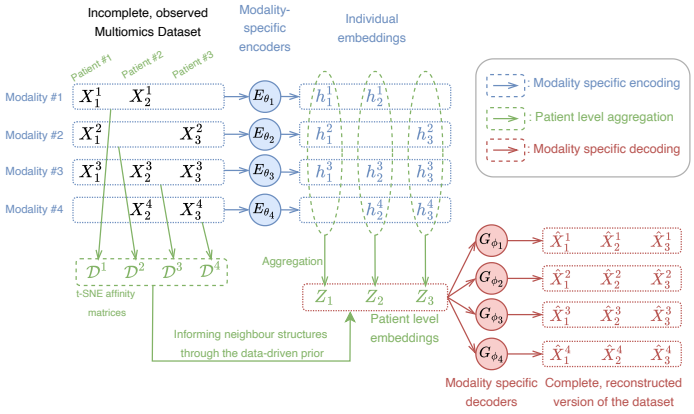

Figure 1: **Schematic illustration of MIND**. We use an incomplete multiomics dataset with $M = 4$ modalities and $N = 3$ patients as an example. Available X_n^ms for each m are first mapped to individual embeddings h_n^ms. Individual embeddings h_n^m associated with the same patient n are then aggregated into a patient level embedding Z_n. Each X_n^m is reconstructed or predicted by passing the patient level embedding Z_n to the modality-specific decoder G_{ϕ_m}. For each modality, a t-SNE affinity matrix $\mathcal{D}^m$ encoding neighbour structure of the relevant patients is computed. These matrices guide the neighbour structure of patient level Z_ns through the prior.

Our proposed method can be interpreted as a multimodal variational autoencoder. Before we give the details of the proposed model, we first fix the notations. Let M be the total number of modalities. Let N be the total number of individual patients. Let $[N] = \{1, ..., N\}$, $[M] = \{1, ..., M\}$. Let X_n^m be the data for the mth modality of the nth patient. For each patient n, denote $A_n = \{m : m \in [M], X_n^m \text{ is available}\}$ the index set of modalities in which the data of the nth patient are available. Similarly, define $B^m = \{n : n \in [N], X_n^m \text{ is available}\}$ be the index set of available patients for each modality. Denote $\mathbf{X}^m = \{X_n^m\}_{n \in B^m}$ the observed data from the mth modality. Denote $\mathbf{X}_n = \{X_n^m\}_{m \in A_n}$ the collection of observed data associated with the nth patient. Denote $\mathbf{X}_{obs} = \{\mathbf{X}^m\}_{m=1}^M = \{\mathbf{X}_n\}_{n=1}^N$ the full multi-omics dataset. In this paper, we assume $X_n^m \in \mathbb{R}^{d_m}$ for all $n \in [N], m \in [M]$ for sake of clarity. Extension to e.g. count data is straightforward.

2.1 Encoder architecture

For each modality $m \in [M]$, let $E_{\theta_m} : \mathbb{R}^{d_m} \to \mathbb{R}^{2d_l}$ be a modality-specific encoder governed by the parameter vector θ_m that maps X_n^m, $n \in [N]$ to the latent space. Denote $h_n^m = E_{\theta_m}(X_n^m)$ and μ_n^m, s_n^m the first and last d^l entries of h_n^m, respectively. Denote $\boldsymbol{\theta} = \{\theta_m\}_{m=1}^M$ the collection of encoder parameters. Since our goal is to learn patient level embeddings Z_n , we need to further aggregate the modality-specific $\{\mu_n^m, s_n^m\}_{m \in A_n}$ encoded by E_{θ_m}s, and pass the information to Z_n. In this paper, we choose to bridge patient level Z_n and modality-specific $\{\mu_n^m, s_n^m\}_{m \in A_n}$ for each patient n by setting the VAE posterior

$$q_{\boldsymbol{\theta}}(Z_n|\mathbf{X}_n) = \mathcal{N}\left(Z_n; \frac{1}{|A_n|}\sum_{m \in A_n} \mu_n^m,\ \text{diag}\left(\exp\left(\frac{1}{|A_n|}\sum_{m \in A_n} s_n^m\right)\right)\right), \tag{1}$$

where $\text{diag}(z)$ represents a diagonal matrix with diagonal elements equal to z and off-diagonal elements 0, and $\exp(z)$ denotes element-wise exponential. In principle, one could consider more sophisticated aggregation functions such as Set Transformer (Lee et al., 2019) instead of simple averaging. However, we found that it gave satisfactory results in all numerical examples. We therefore do not investigate other alternatives in this paper for simplicity.

2.2 Prior distribution on embeddings

So far we have discussed encoder architecture and posteriors of the embeddings. In this section, we discuss the choice of prior on Z_ns. In this paper, we use a partially data-driven prior on Z_ns to inject the neighbouring structure information into the embeddings. Our choice of prior is inspired by the variational formulation of t-SNE (Maaten and Hinton, 2008) given in Ravuri et al. (2023). Let $\mathbf{Z} = \{Z_i\}_{i=1}^N$ and $\mathcal{D}(\mathbf{Z}) \in \mathbb{R}^{N \times N}$ be the pairwise similarity matrix of embeddings $\mathbf{Z}$ where each entry $\mathcal{D}_{i,j}(\mathbf{Z}) = \frac{1}{1+||Z_i - Z_j||_2^2}$. For each modality $m \in [M]$, let $\mathcal{D}^m \in \mathbb{R}^{|B^m| \times |B^m|}$ be the t-SNE's sparse data affinity matrix obtained by applying t-SNE to $\mathbf{X}^m$. Let $\mathcal{D}^m(\mathbf{Z}) \in \mathbb{R}^{|B^m| \times |B^m|}$ be a sub-matrix of $\mathcal{D}(\mathbf{Z})$ whose rows and columns are selected to match $\mathcal{D}^m$. Let $\bar{\mathcal{D}}^m(\mathbf{Z}), \bar{\mathcal{D}}^m$ be the normalised and vectorised versions of $\mathcal{D}^m(\mathbf{Z}), \mathcal{D}^m$, respectively. Denote I_p a $p \times p$ identity matrix. We define the exponential-tilted prior on $\mathbf{Z}$ as

$$p(\mathbf{Z}) \propto \prod_{n=1}^N \mathcal{N}\left(Z_i; \mathbf{0}, I_{d_l}\right) \times \exp\left(-\sum_{m=1}^M KL\left(\bar{\mathcal{D}}^m || \bar{\mathcal{D}}^m(\mathbf{Z})\right)\right). \tag{2}$$

Note that $\text{KL}\left(\bar{\mathcal{D}}^m || \bar{\mathcal{D}}^m(\mathbf{Z})\right)$, the KL divergence between two categorical distributions specified by probability vectors $\bar{\mathcal{D}}^m, \bar{\mathcal{D}}^m(\mathbf{Z})$, respectively, is nonnegative by definition. This ensures that $p(\mathbf{Z})$ is still a valid probability distribution whose probability density is known up to a multiplicative constant. Compared to i.i.d. isotropic Gaussian priors, our choice of $p(\mathbf{Z})$ additionally encourages different subsets of $\mathbf{Z}$ to cluster in a way similar to the neighbouring structures of $\mathbf{X}^m$s, which are encoded as t-SNE's data affinity matrices. See Fig 1 for a schematic illustration of the proposed method.

2.3 Training and inference

Let $G_{\phi_m} : \mathbb{R}^{d_l} \to \mathbb{R}^{d_m}$ be the modality-specific decoder governed by the parameter vector ϕ_m for each $m \in [M]$. Let $\boldsymbol{\phi} = \{\phi_m\}_{m=1}^M$ be the collection of decoder parameters. In addition to the standard evidence lower bound (Kingma and Welling, 2014) objective of a VAE model, we also incorporate a regularisation term on the encoder outputs $\{h_n^m\}_{m \in A_n}$ for all $n \in [N]$ taking the form

$$R(\boldsymbol{\theta}; \mathbf{X}_{obs}) = \frac{1}{d_l}\sum_{n=1}^N \sum_{m \in A_n} \sum_{m' < m} ||h_n^m - h_n^{m'}||_2^2. \tag{3}$$

Intuitively speaking, $R(\boldsymbol{\theta}; \mathbf{X}_{obs})$ penalises pairwise distance between $\{h_n^m\}_{m \in A_n}$ for each patient n. Recall that $\{h_n^m\}_{m \in A_n}$ are generated using different modality-specific encoders. We encourage $\{h_n^m\}_{m \in A_n}$ to be close to each other as the data used to generate these quantities are from the same patient, and we want the modality-specific encoders E_{θ_m} to be aware of this cross-modality connection. Details of the training and inference procedure can be found in Appendix A.

3 Numerical experiments

We applied our proposed method to three multiomics datasets: The Cancer Genome Atlas (TCGA) (Weinstein et al., 2013), the Childhood Cancer Model Atlas (CCMA) (Sun et al., 2023) and the Cancer Cell Line Encyclopedia (CCLE) (Ghandi et al., 2019). Details regarding the datasets can be found in Appendix B. We compare MIND with three current state-of-the-art models: the VAE-based JASMINE (Ballard et al., 2025) and MOVE (Allesøe et al., 2023), and the network-based IntegraO (Ma et al., 2025). We also include MSNE (Xu et al., 2021) as a further baseline. We set the dimension of embeddings $d_l = 64$ for all methods. A Python implementation of MIND and codes for reproducing all experiments can be found on https://anonymous.4open.science/r/RAND-A1F7.

We investigate two supervised tasks, cancer type classification and multiomics data reconstruction. For cancer type classification, we trained XGboost classifiers (Chen and Guestrin, 2016) to predict cancer types using embeddings generated by different methods. Classification accuracies estimated by 5-fold CV are reported in Table 1, showing that MIND and JASMINE derived embeddings gave strongest classification results.

Dataset/Method	MIND	IntegraO	JASMINE	MOVE	MSNE
TCGA	<u>0.974</u>	0.948	**0.975**	0.964	0.891
CCMA	**0.793**	<u>0.761</u>	0.704	0.669	0.548
CCLE	**0.659**	0.547	<u>0.610</u>	0.497	0.367

Table 1: **Cancer type classification**. Classification accuracy estimated using 5-fold CV. For each cancer type, the best result is highlighted in **boldface**. Second best is <u>underlined</u>.

We then compared the reconstruction performance of MIND with JASMINE and MOVE. Recall that IntegraO and MSNE do not have this feature. Here, for each modality of each dataset, we first randomly mask 10% of its data subject to the constraint that, for every dataset, each patient must be present in at least one modality of the resulting masked multiomics dataset. We then train the models using the masked datasets, reconstruct the masked data using the learned embeddings, and compare the reconstructed values with the masked observed values. Since the output embeddings of JASMINE are obtained by concatenating M modality-specific embeddings and a global embedding, each of its modality-specific embedding only has length $\lceil \frac{d_l}{M+1} \rceil$, which could be too stringent to accurately reconstruct data from each modality. We therefore additionally fit an augmented JASMINE such that each individual embedding has length $d_l = 64$ (i.e. the final embedding has length $d_l(M+1)$). We refer to the larger model as JASMINE$_{aug}$. For each modality of each dataset, we computed the Pearson correlation between the reconstructed and masked observed data. We report the Pearson correlations in Table 2. This shows that correlations between MIND reconstructions of the masked data were more accurate than those provided by both variants of JASMINE and MOVE.

Dataset	Modality / Method	MIND	JASMINE	JASMINE$_{aug}$	MOVE
TCGA	mRNA	**0.797**	0.398	0.554	<u>0.794</u>
	DNA methyl	<u>0.778</u>	0.285	0.398	**0.799**
	CNV	**0.559**	0.110	0.149	<u>0.531</u>
	RPPA	**0.514**	0.126	0.336	<u>0.437</u>
	miRNA	**0.741**	0.358	0.702	<u>0.725</u>
CCMA	mRNA	**0.550**	0.064	0.106	<u>0.261</u>
	DNA methyl	**0.648**	0.198	0.431	<u>0.450</u>
	CNV	**0.836**	0.461	0.548	<u>0.653</u>
CCLE	RNA	**0.572**	0.239	0.280	<u>0.568</u>
	DNA methyl	**0.447**	0.163	0.219	<u>0.405</u>
	CNV	<u>0.174</u>	0.042	0.072	**0.176**
	miRNA	**0.244**	0.064	0.073	<u>0.225</u>
	RPPA	<u>0.379</u>	0.108	0.210	**0.401**
	Metabolomics	**0.338**	0.121	0.227	<u>0.289</u>

Table 2: **Reconstruction Accuracy**. Pearson correlations between reconstructed and observed values of the individual modalities. For each modality, the best result is highlighted in **boldface**. Second best is <u>underlined</u>.

4 Discussion

Our proposed data integration approach, MIND, adopts a multimodal VAE architecture to accommodate incomplete multimodal data sets. Using multi-omics, we demonstrate state-of-the-art performance in experiments that show that MIND consistently outperforms or achieves performance on par with existing VAE- and network-based state-of-the-art methods on down-streaming tasks such as cancer type classification. This suggests that MIND can better extract biologically meaningful infor-

mation from incomplete multiomics data. Furthermore, we also demonstrate that MIND consistently outperforms competing methods in terms of predicting unseen data from learnt patient embeddings. This further confirms that MIND is capable of accurately identifying and capturing characteristics of different patients groups in diverse biological contexts.

References

Allesøe, R. L., Lundgaard, A. T., Hernández Medina, R., Aguayo-Orozco, A., Johansen, J., Nissen, J. N., Brorsson, C., Mazzoni, G., Niu, L., Biel, J. H., et al. (2023). Discovery of drug–omics associations in type 2 diabetes with generative deep-learning models. *Nature biotechnology*, 41(3):399–408.

Ballard, J. L., Dai, Z., Shen, L., and Long, Q. (2025). Jasmine: A powerful representation learning method for enhanced analysis of incomplete multi-omics data. *bioRxiv*, pages 2025–06.

Beaude, A., Augé, F., Zehraoui, F., and Hanczar, B. (2025). Crossattomics: multiomics data integration with cross-attention. *Bioinformatics*, 41(6):btaf302.

Cao, J., Cusanovich, D. A., Ramani, V., Aghamirzaie, D., Pliner, H. A., Hill, A. J., Daza, R. M., McFaline-Figueroa, J. L., Packer, J. S., Christiansen, L., et al. (2018). Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science*, 361(6409):1380–1385.

Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794.

Fu, Z., Jiang, S., Sun, Y., Zheng, S., Zong, L., and Li, P. (2024). Cut&tag: a powerful epigenetic tool for chromatin profiling. *Epigenetics*, 19(1):2293411.

Gayoso, A., Steier, Z., Lopez, R., Regier, J., Nazor, K. L., Streets, A., and Yosef, N. (2021). Joint probabilistic modeling of single-cell multi-omic data with totalvi. *Nature methods*, 18(3):272–282.

Ghandi, M., Huang, F. W., Jané-Valbuena, J., Kryukov, G. V., Lo, C. C., McDonald III, E. R., Barretina, J., Gelfand, E. T., Bielski, C. M., Li, H., et al. (2019). Next-generation characterization of the cancer cell line encyclopedia. *Nature*, 569(7757):503–508.

Kingma, D. P. and Welling, M. (2014). Auto-encoding variational bayes. In Bengio, Y. and LeCun, Y., editors, *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*.

Lee, J., Lee, Y., Kim, J., Kosiorek, A., Choi, S., and Teh, Y. W. (2019). Set transformer: A framework for attention-based permutation-invariant neural networks. In *International conference on machine learning*, pages 3744–3753. PMLR.

Ma, S., Zeng, A. G., Haibe-Kains, B., Goldenberg, A., Dick, J. E., and Wang, B. (2025). Moving towards genome-wide data integration for patient stratification with integrate any omics. *Nature Machine Intelligence*, 7(1):29–42.

Maaten, L. v. d. and Hinton, G. (2008). Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605.

Rappoport, N. and Shamir, R. (2019). Nemo: cancer subtyping by integration of partial multi-omic data. *Bioinformatics*, 35(18):3348–3356.

Ravuri, A., Vargas, F., Lalchand, V., and Lawrence, N. D. (2023). Dimensionality reduction as probabilistic inference. *arXiv preprint arXiv:2304.07658*.

Regner, M. J., Wisniewska, K., Garcia-Recio, S., Thennavan, A., Mendez-Giraldez, R., Malladi, V. S., Hawkins, G., Parker, J. S., Perou, C. M., Bae-Jump, V. L., et al. (2021). A multi-omic single-cell landscape of human gynecologic malignancies. *Molecular cell*, 81(23):4924–4941.

Shen, R., Mo, Q., Schultz, N., Seshan, V. E., Olshen, A. B., Huse, J., Ladanyi, M., and Sander, C. (2012). Integrative subtype discovery in glioblastoma using icluster. *PloS one*, 7(4):e35236.

Shin, S. H., Bode, A. M., and Dong, Z. (2017). Precision medicine: the foundation of future cancer therapeutics. *Npj precision oncology*, 1(1):12.

Steyaert, S., Pizurica, M., Nagaraj, D., Khandelwal, P., Hernandez-Boussard, T., Gentles, A. J., and Gevaert, O. (2023). Multimodal data fusion for cancer biomarker discovery with deep learning. *Nature machine intelligence*, 5(4):351–362.

Subramanian, I., Verma, S., Kumar, S., Jere, A., and Anamika, K. (2020). Multi-omics data integration, interpretation, and its application. *Bioinformatics and biology insights*, 14:1177932219899051.

Sun, C. X., Daniel, P., Bradshaw, G., Shi, H., Loi, M., Chew, N., Parackal, S., Tsui, V., Liang, Y., Koptyra, M., et al. (2023). Generation and multi-dimensional profiling of a childhood cancer cell line atlas defines new therapeutic opportunities. *Cancer cell*, 41(4):660–677.

Weinstein, J. N., Collisson, E. A., Mills, G. B., Shaw, K. R., Ozenberger, B. A., Ellrott, K., Shmulevich, I., Sander, C., and Stuart, J. M. (2013). The cancer genome atlas pan-cancer analysis project. *Nature genetics*, 45(10):1113–1120.

Wu, M. and Goodman, N. (2018). Multimodal generative models for scalable weakly-supervised learning. *Advances in neural information processing systems*, 31.

Xu, H., Gao, L., Huang, M., and Duan, R. (2021). A network embedding based method for partial multi-omics integration in cancer subtyping. *Methods*, 192:67–76.

Yang, Z. and Michailidis, G. (2016). A non-negative matrix factorization method for detecting modules in heterogeneous omics multi-modal data. *Bioinformatics*, 32(1):1–8.

A Details of the training and inference procedure

We discuss here the training and inference procedures of the proposed model. Let $G_{\phi_m} : \mathbb{R}^{d_l} \to \mathbb{R}^{d_m}$ be the modality-specific decoder governed by the parameter vector ϕ_m for each $m \in [M]$. Let $\boldsymbol{\phi} = \{\phi_m\}_{m=1}^M$ be the collection of decoder parameters. The training procedure of the proposed model is similar to a VAE (Kingma and Welling, 2014): Denote $q_{\boldsymbol{\theta}}(\mathbf{Z}|\mathbf{X}_{obs}) = \prod_{n=1}^N q_{\boldsymbol{\theta}}(Z_n|\mathbf{X}_n)$ the joint posterior distribution of $\mathbf{Z}$. Suppose $p_{\boldsymbol{\phi}}(X_n^m|Z_n) = \mathcal{N}(X_n^m; G_{\phi_m}(Z_n), I_{d_m})$, the evidence lower bound of the proposed model takes the form

$$\mathrm{ELBO}(\boldsymbol{\theta}, \boldsymbol{\phi}; \mathbf{X}_{obs}) = -\frac{1}{2}\sum_{n=1}^{N}\sum_{m\in A_n} E_{Z_n\sim q_{\boldsymbol{\theta}}(\cdot|\mathbf{X}_n)}\left(||X_n^m - G_{\theta_m}(Z_n)||_2^2\right) - KL\left(q_{\boldsymbol{\theta}}(\mathbf{Z}|\mathbf{X}_{obs})||p(\mathbf{Z})\right), \tag{4}$$

where $E_{Z_n\sim q_{\boldsymbol{\theta}}(\cdot|\mathbf{X}_n)}$ means taking expectation w.r.t. $Z_n \sim q(_{\boldsymbol{\theta}}\cdot|\mathbf{X}_n)$. The first and second term of $\mathrm{ELBO}(\boldsymbol{\theta}, \boldsymbol{\phi}; \mathbf{X}_{obs})$ are the log likelihood and the KL divergence between the posterior and prior, respectively. Here we assume X_n^m follows a Gaussian distribution. Extension to other likelihoods is straightforward.

In addition to the standard ELBO of a VAE model, we also incorporate a regularisation term on the encoder outputs $\{h_n^m\}_{m\in A_n}$ for all $n \in [N]$ taking the form

$$R(\boldsymbol{\theta}; \mathbf{X}_{obs}) = \frac{1}{d_l}\sum_{n=1}^{N}\sum_{m\in A_n}\sum_{m'<m} ||h_n^m - h_n^{m'}||_2^2. \tag{5}$$

Intuitively speaking, $R(\boldsymbol{\theta}; \mathbf{X}_{obs})$ penalises pairwise distance between $\{h_n^m\}_{m\in A_n}$ for each patient n. Recall that $\{h_n^m\}_{m\in A_n}$ are generated using different modality-specific encoders. We encourage $\{h_n^m\}_{m\in A_n}$ to be close to each other as the data used to generate these quantities are from the same patient, and we want the modality-specific encoders E_{θ_m} to be aware of this cross-modality connection.

The resulting loss function of the proposed model is

$$L(\boldsymbol{\theta}, \boldsymbol{\phi}; \mathbf{X}_{obs}) = -\mathrm{ELBO}(\boldsymbol{\theta}, \boldsymbol{\phi}; \mathbf{X}_{obs}) + \alpha R(\boldsymbol{\theta}; \mathbf{X}_{obs}), \tag{6}$$

where $\alpha > 0$ is a hyperparameter controlling the strength of the regulariser $R(\boldsymbol{\theta}; \mathbf{X}_{obs})$. We set $\alpha = 0.05$ in all numerical examples. The VAE parameters $\{\boldsymbol{\theta}, \boldsymbol{\phi}\}$ are trained using reparameterisation trick (Kingma and Welling, 2014) and gradient descent methods.

A.1 Scalability of the t-SNE informed prior

The prior we proposed in Sec 2.2 does not factorise. As a result, we are not able to directly apply standard minibatch stochastic gradient descent. We use a Gibbs sampling style training scheme to address this issue: Denote BC the index set of a minibatch of patients. Denote $\mathbf{Z}_{BC}$ the corresponding embeddings and $\mathbf{Z}_{-BC} = \mathbf{Z} \setminus \mathbf{Z}_{BC}$ its complement. For each minibatch stochastic gradient descent step, we replace $KL\left(q_{\boldsymbol{\theta}}(\mathbf{Z}|\mathbf{X}_{obs})||p(\mathbf{Z})\right)$ in Eq (4) by the conditional version $KL\left(q_{\boldsymbol{\theta}}(\mathbf{Z}_{BC}|\mathbf{Z}_{-BC}, \mathbf{X}_{obs})||p(\mathbf{Z}_{BC}|\mathbf{Z}_{-BC})\right)$. Since the posterior $q(\mathbf{Z}|\mathbf{X}_{obs})$ is factorisable by design, we have

$$q_{\boldsymbol{\theta}}(\mathbf{Z}_{BC}|\mathbf{Z}_{-BC}, \mathbf{X}_{obs}) = q_{\boldsymbol{\theta}}(\mathbf{Z}_{BC}|\mathbf{X}_{obs}) = \prod_{n \in BC} q_{\boldsymbol{\theta}}(Z_n|\mathbf{X}_n). \tag{7}$$

The second term $p(\mathbf{Z}_{BC}|\mathbf{Z}_{-BC})$ is equivalent to $p(\{\mathbf{Z}_{BC}, \text{sg}(\mathbf{Z}_{-BC})\}$ in the training process, where sg is the stop-gradient operator (i.e. we treat $\mathbf{Z}_{-BC}$ as fixed by detaching it from the computation graph). Computing the corresponding minibatch log likelihood and regularisation is straightforward. The final minibatch loss is obtained by combing the individual minibatch terms the same way as in Eq (6).

A.2 Prediction and imputation

Suppose $\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\phi}}$ are approximate minimisers of $L(\boldsymbol{\theta}, \boldsymbol{\phi}; \mathbf{X}_{obs})$. Then each patient embedding Z_n is represented by the posterior $q_{\hat{\boldsymbol{\theta}}}(Z_n|\mathbf{X}_n)$. We compute the reconstructed or predicted data $\hat{X}_n^m$ for any $n \in [N], m \in [M]$ by passing either the mean or a sample from $q_{\hat{\boldsymbol{\theta}}}(Z_n|\mathbf{X}_n)$ to the corresponding trained decoder $G_{\hat{\phi}_m}$. See Fig 1 for a schematic illustration of the proposed method.

B Data preprocessing

TCGA

For the TCGA example, we leveraged TCGA multi-omics data sets across 17 types of cancer: Head and Neck squamous cell carcinoma (HNSC), Lung squamous cell carcinoma (LUSC), Liver hepatocellular carcinoma (LIHC), Cervical and endocervical cancers (CESC), Lung adenocarcinoma (LUAD), Kidney renal clear cell carcinoma (KIRC), Breast invasive carcinoma (BRCA), Brain Lower Grade Glioma (LGG), Ovarian serous cystadenocarcinoma (OV), Skin Cutaneous Melanoma (SKCM), Thyroid carcinoma (THCA), Bladder urothelial carcinoma (BLCA), Glioblastoma multiforme (GBM), Stomach adenocarcinoma (STAD), Uterine Corpus Endometrial Carcinoma (UCEC), Colorectal adenocarcinoma (COADREAD), and Colon adenocarcinoma (COAD). We obtain mRNA expression, DNA methylation, MicroRNA, protein expression data and relavent patient information from the Broad Institute's Firehose source data. Copy number variation are obtained from cBioportal. We adpot the data preprocessing steps used in Ma et al. (2025): we first stack the 17 data sets, then remove columns with more than 20% missing values for each of the modalities. We then impute missing values in the resulting datasets using k-nearest neighbour. We stress that the removal and imputation steps are applied only to the missing entries in the data vectors of patients already available in a modality, and does not apply to the patients that are not present in a modality (i.e. if a patient is present in a modality, and the corresponding data vector contains missing values, the preprocessing steps will impute those missing values. However, if a patient is not present in a modality, its corresponding data remain completely unknown). For copy number variation, mRNA expression, and DNA methylation data, we select the top 2,000 most variable features. Data from all modalities are then standardised so that each feature dimension has mean 0 and standard deviation 1. Datasets for each individual cancer type are preprocessed in the same fashion.

CCMA

The CCMA dataset consists of three modalities: RNA-sequencing, DNA methylation, and copy number variation. The CCMA dataset and relevant patient information are obtained from the web portal. Missing values are imputed using k-nearest neighbour. For RNA-sequence and DNA methylation data, we select the top 2000 most variable features. All data are then standardised so that each feature dimension has mean 0 and standard deviation 1.

CCLE

The CCLE dataset consists of six modalities: RNA-sequencing, DNA methylation, copy number variation, MicroRNA, protein expression data and metabolomics data. The CCLE dataset and relevant patient information are obtained from the web portal. For RNA-sequence and DNA methylation data, we select the top 2000 most variable features. All data are then standardised so that each feature dimension has mean 0 and standard deviation 1.

NeurIPS Paper Checklist

The checklist is designed to encourage best practices for responsible machine learning research, addressing issues of reproducibility, transparency, research ethics, and societal impact. Do not remove the checklist: **The papers not including the checklist will be desk rejected.** The checklist should follow the references and follow the (optional) supplemental material. The checklist does NOT count towards the page limit.

Please read the checklist guidelines carefully for information on how to answer these questions. For each question in the checklist:

- You should answer [Yes] , [No] , or [NA] .
- [NA] means either that the question is Not Applicable for that particular paper or the relevant information is Not Available.
- Please provide a short (1–2 sentence) justification right after your answer (even for NA).

The checklist answers are an integral part of your paper submission. They are visible to the reviewers, area chairs, senior area chairs, and ethics reviewers. You will be asked to also include it (after eventual revisions) with the final version of your paper, and its final version will be published with the paper.

The reviewers of your paper will be asked to use the checklist as one of the factors in their evaluation. While "[Yes] " is generally preferable to "[No] ", it is perfectly acceptable to answer "[No] " provided a proper justification is given (e.g., "error bars are not reported because it would be too computationally expensive" or "we were unable to find the license for the dataset we used"). In general, answering "[No] " or "[NA] " is not grounds for rejection. While the questions are phrased in a binary way, we acknowledge that the true answer is often more nuanced, so please just use your best judgment and write a justification to elaborate. All supporting evidence can appear either in the main paper or the supplemental material, provided in appendix. If you answer [Yes] to a question, in the justification please point to the section(s) where related material for the question can be found.

IMPORTANT, please:

- **Delete this instruction block, but keep the section heading "NeurIPS Paper Checklist"**,
- **Keep the checklist subsection headings, questions/answers and guidelines below.**
- **Do not modify the questions and only use the provided macros for your answers**.

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: See Sec 2.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: See Appendix on discussion on batched training.

Guidelines:

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [NA]

   Justification: No proof included.

   Guidelines:

   - The answer NA means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: See Sec 3 and Appendix.

   Guidelines:

   - The answer NA means that the paper does not include experiments.

- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: See Sec 3 and Appendix.

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).

- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: See Sec 3.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [No]

   Justification: Did not include repeated runs.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.
   - It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
   - For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
   - If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: See Sec 3 and Appendix.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.

- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: See our open sourced Github page

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [NA]

    Justification: Does not apply.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: [NA]

    Guidelines:

    - The answer NA means that the paper poses no such risks.

- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [NA]

    Justification: Does not apply.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [No]

    Justification: We used only publicly available datasets.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: Does not apply

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: Does not apply.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

    Answer: [No]

    Justification: No LLM was used.

    Guidelines:

    - The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.