MODIS: MULTI-OMICS DATA INTEGRATION FOR SMALL AND UNPAIRED DATASETS

Daniel Lepe-Soltero
Aix Marseille Univ & INSERM
MMG
Marseille, France
daniel.lepe-soltero@univ-amu.fr

Thierry Artières
Aix Marseille Univ & Ecole Centrale de Marseille
LIS CNRS UMR 7020
Marseille, France
thierry.artieres@lis-lab.fr

Anaïs Baudot
Aix Marseille Univ & INSERM
MMG, CNRS
Marseille, France
anais.baudot@univ-amu.fr

Paul Villoutreix
Aix Marseille Univ & INSERM
MMG, Turing Centre for Living Systems
Marseille, France
paul.villoutreix@univ-amu.fr

ABSTRACT

A key challenge today lies in the ability to efficiently handle multi-omics data since such multimodal data may provide a more comprehensive overview of the underlying processes in a system. Yet it comes with challenges: multi-omics data are most often unpaired and only partially labeled, moreover only small amounts of data are available in some situation such as rare diseases. We propose MODIS which stands for Multi-Omics Data Integration for Small and unpaired datasets, a semi supervised approach to account for these particular settings. MODIS learns a probabilistic coupling of heterogeneous data modalities and learns a shared latent space where modalities are aligned. We rely on artificial data to build controlled experiments to explore how much supervision is needed for an accurate alignment of modalities, and how our approach enables dealing with new conditions for which few data are available. The code is available at https://github.com/VILLOUTREIXLab/MODIS.

MEANINGFULNESS STATEMENT

We propose to integrate multiple representations of a given biological system, into a common, meaningful, latent representation. This common representation is built by training adversarially multiple coupled auto-encoders. We take advantage of a large reference dataset to build this representation which will be used to align smaller multi-omics dataset. This approach finds its applications in the study of rare diseases where samples are in low number, unpaired and possibly with one or more missing modalities.

1 INTRODUCTION

The recent explosion in multi-modal and particularly multi-omics data holds immense promise for understanding complex biological mechanisms and ultimately help address diagnostic challenges. Rare diseases in particular, defined as diseases affecting less than one person in 2000, are characterized by diagnostic deadlocks due to the lack of understanding of the specific processes and could benefit from multi-omics approaches Banerjee et al. (2023). To take full advantage of these multi-omics datasets, it is useful to derive a joint representation in a lower dimensional space, a task named multi-omics data integration. This joint representation helps disease identification by classifying or clustering samples, leading subsequently to biomarker discovery Cantini et al. (2021); Hirst et al. (2024). Moreover, since it is not always possible to acquire samples in multiple omics simultaneously, and it has been proposed to use the joint representation as a basis for generation missing modalities Yang et al. (2021). Various strategies have been proposed for multi-omics integration, ranging from vertical approaches that combine layers of omics data for paired samples to

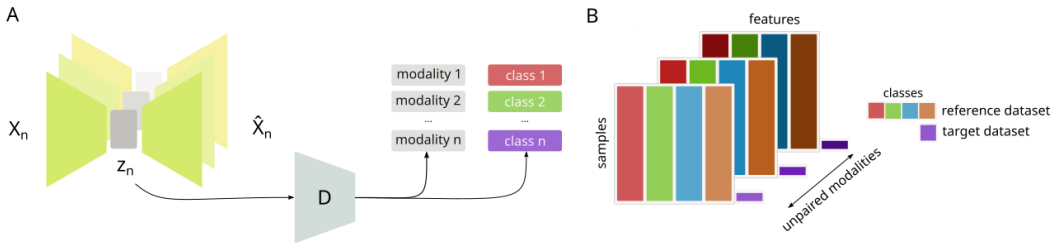


Figure 1: **Schematic of our MODIS model and our approach to addressing the challenges in rare diseases**. **A**, The embeddings generated by each VAE, customized for a specific modality, are integrated and aligned in a shared latent space using adversarial learning in a fully supervised or semi-supervised setting. **B**, A large reference dataset is combined with a small target dataset into an unbalanced (paired or unpaired) multi-omics dataset leveraging the knowledge from the reference to improve the alignment of the target dataset.

diagonal and mosaic strategies that enable the integration of diverse combinations of unpaired and paired samples across datasets Argelaguet et al. (2021). We are interested here in the applicability of data integration for disease identification (a classification task), for difficult cases such as rare diseases and more generally scarce situations, where the number of samples is low, the training samples are not paired across modalities, and where some modalities might even be missing for some class. Most of the approaches for multi-omics data integration Argelaguet et al. (2021) require large datasets for their training and are thus not straightfully applicable to this difficult setting. To extend their applicability to rare diseases or other types of diseases with scarce data, we propose to take advantage of large datasets of multi-omics data such as the TCGA database Weinstein et al. (2013) to help with the integration.

From a machine learning perspective, we consider that the training dataset includes data from multiples modalities (the various -omics) and for many classes (the various diseases or cancer or conditions from the available databases). Yet this training dataset is not complete in multiple ways. First, all the samples are not observed in all modalities, they are unpaired. Moreover a particular modality may be missing for a particular class in the training set, meaning that there are no example of this modality in the training samples for this disease. Finally a number of samples may miss a class label, since labeling requires expensive expert annotation. The goal of this study is to propose solutions, within this context of multi-omics data integration, for the learning and the recognition of disease with small unpaired dataset, a situation exemplified by rare diseases. To be able to cope with this particular situation requires a set of desired features for the model to be designed. First it should be able to predict the class of a sample which is observed in a single modality, whatever it is, as it is often the case that only one modality is observed. This requires to be able to translate or link in some way the samples observed in one modality to samples observed in another modality, keeping in mind that the data are unpaired across modalities. Second, the method should be able to take advantage of unlabeled data (wrt class). Third, the classifier should be learnable for a set of classes with eventually small training data, hence it should be able to deal with an imbalanced classification problem. To answer these three challenges, we propose an approach named MODIS for Multi-Omics Data Integration for Small and unpaired datasets.

2 State of the art

Several approaches have been dedicated to dealing with multiview or multimodal data with the goal of performing a prediction task, e.g., classification, from a number of very different and complementary modalities whose heterogeneity prevents any simple combination (see for example Han et al. (2021)). More generic approaches have been proposed that combine modalities, such as audio and text, in the CLAP model Elizalde et al. (2023), to learn rich representations that may be used for (any) downstream task. Another line of works have focused on extending data analysis classics such as canonical correlation analysis (CCA) to explore relationship between different modalitites Andrew et al. (2013). Besides, a number of works have been proposed to learn to infer a modality from another one, usually in a supervised setting, in a general machine learning setting Ngiam et al.

(2011) and for multiomics data Yang et al. (2021). A common backbone of these methods consists in a set of auto-encoders, one for each modality, that are learned in such a way that their latent space are aligned, thus enabling prediction of a modality from another one Ngiam et al. (2011); Yang et al. (2021) by encoding with a modality's encoder and decoding with another modality's decoder. Learning to align the latent representation spaces of multiple modalitie's autoencoders can be done using paired data (all modalities of training samples are observed) if these data are available Tang et al. (2023) or through distribution constraints Yang et al. (2021); He et al. (2024), extending the seminal work from Makhzani et al. (2015), in the opposite case of unpaired data where only one (or not all) modality is available for every training sample.

In computational biology and bioinformatics, the question of integrating multiple modalities for classification or clustering has been addressed by various authors in the past few years. In the case of paired samples across modalities, several linear methods such as joint matrix factorization have been used, as reviewed in Cantini et al. (2021), and extended with transfer learning Hirst et al. (2024). More recently, non-linear methods such as multiple auto-encoders with a shared latent space have been proposed, for paired samples (also known as vertical integration Argelaguet et al. (2021)) Tang et al. (2023), unpaired samples (diagonal integration) Yang et al. (2021); Samaran et al. (2024), or when considering a combinaison of paired or unpaired samples (mosaic integration) He et al. (2024).

A few approaches allow to deal with imbalanced classification problems Huang et al. (2016) or to handle learning classes with very few training samples (few shot learning Wang et al. (2020)). However, the question of combining multi-omics data integration with the question of class imbalance remains largely unexplored.

3 METHODS

As stated in the introduction, we propose an approach, MODIS, whose aim is to perform multi-omics data integration with small and unpaired datasets. We propose to train coupled autoencoders (Fig. 1A) on both a large reference dataset and a small target dataset (Fig. 1B), the latter corresponding to rare diseases. This approach aims at leveraging the structure learned from the larger dataset to improve the alignment and translation of modalities and label classification in the smaller dataset. Additionally, since we were particularly interested in datasets with multiple **biological conditions** which are modeled as multiple classes in the data, we incorporated a discriminative loss and a regularization, following Tang et al. (2023). These modifications help preserve the underlying data structure while ensuring meaningful alignment across modalities.

MODIS is composed of multiple coupled variational auto-encoders (VAEs), as many as the number of modalities. To link the samples from different modalities, we want to align the latent representation space of each of the modality encoders. To do so, we enforce the latent representation of a sample, computed by one of the modality encoders, to be *modality-free*, following Makhzani et al. (2015). Practically we use an auxiliary discriminator that operates in the latent space which is trained adversarially on the modality. More precisely, this classifier is trained to recognize the modality of a sample whose latent representation is input to the classifier, while the encoders of each of the modalities are trained to fool this classifier. Doing so, one aligns the latent space of all modality autoencoders so that it becomes a shared latent space. This is a key aspect of our method since it enables both the possibility of reconstructing a modality from another one, by chaining the encoder of a modality with the decoder of another modality, and the possibility of learning and using a single classifier whatever the observed modality, since it operates in the shared latent space, see Figure 1.

Since the training involves an adversarial discriminator, the optimization iteratively alternates two steps. In the first step, only the discriminator's weights are updated, to improve modality and class identification from the latent space using the following loss

$$L_D = L_{M,D} + L_C + L_{\text{clustering}} \tag{1}$$

where $L_{M,D}$ is the cross-entropy loss for modality prediction (the D denotes that this is the variant of this loss that is used to learn the classifier D), L_C is the cross-entropy loss for class prediction and, following Tang et al. (2023); Kampffmeyer et al. (2019), we used a clustering loss, noted $L_{\text{clustering}}$, that ensures the separability and the compactness of the class clusters.

In the second step, the discriminator's weights are frozen and all other parameters (all the VAEs' weights) are updated using a combined loss. It includes all VAEs' standalone losses in a first part (a reconstruction loss L_{recon} and a Kullback-Leibler divergence loss D_{KL} for every VAE), a classification (adversarial) loss for modality prediction, the label classification clustering loss as above.

$$L_{VAE} = \left(\sum_{i=1}^{m} \left(L_{\text{recon}}^{(i)} + \beta D_{KL}^{(i)} \right) \right) + (L_{M,VAE} + L_C + L_{\text{clustering}}) \tag{2}$$

where m is the number of modality, β is an hyperparameter (set to $\beta = 5 \times 10^{-4}$ in our experiments), $L_{M,VAE}$ is the *adversarial* cross-entropy loss of the discriminator with respect to the modalities (a variant of $L_{M,D}$ for training the encoders, as is standard in adversarial learning Makhzani et al. (2015)).

The proposed approach may naturally be learned in a semi-supervised way with supervised data whose class label is known and unsupervised data whose class label is unknown. Moreover the learning of the latent space from all data from all classes enables an accurate learning of this latent space whose dimension is significantly lower than original dimensions of data independently of the modality (Fig.5). Training a classifier in this latent space is then easier and is likely to require much less training data than in the original data space. It is then a possible approach for handling situations with small datasets, such as rare diseases, as we will investigate.

4 EXPERIMENTS

4.1 DATASETS

To test our approach, we generated realistic datasets by simulation using the **InterSIM** R package Chalise et al. (2016). The three -omics represented are DNA methylation (367 features), gene expression (131 features), protein abundance (160 features). We started by creating a large paired multi-omics datasets with 11,500 samples (each sample being a triplet of -omics) and distributed among 5 balanced classes (each class representing $\sim$ 20 % of the samples). We then derived an unpaired dataset by keeping only one representative -omic modality per sample while preserving the balanced distribution among the 5 classes. We finally used the `train_test_split` function of sklearn with a ratio of 0.2 and balanced classes to create the train/test datasets. We adapted this dataset to the various experimental settings (semi-supervised, class imbalance, missing modality). To evaluate the class prediction accuracy of MODIS, we used the classical accuracy (ACC), the inverted class frequency accuracy (W-ACC) and the Jaccard Index (JI) Pedregosa et al. (2011). To evaluate the reconstruction and the translation between modalities, we used the Mean Square Error. In the case of the translation, which corresponds to the composition of an encoder for one modality and a decoder for another modality, since we originally generated paired samples, given a sample in one -omic, we have the ground truth in each of the other -omics. The translation error is computed between the predicted -omic and the ground truth.

4.2 RESULTS

Latent space alignment MODIS consists of coupled auto-encoders sharing a common latent representation such that the modalities of origin of the samples are indistinguishable while the classes from which each sample belongs to are easily identifiable.

To establish a baseline for the quality of the latent space alignment, we first trained our architecture on a fully supervised (class labels available for all samples) and class-balanced dataset (Fig. 2, first column). In this ideal setting, the model achieved 100% class prediction accuracy on test data using the trained class discriminator, demonstrating its capacity to learn a well-structured latent representation when provided with full supervision. Moreover, the reconstruction error and translation error were down to 0.01 (Fig. 2.C), confirming that 1) the latent representation of each modality is sufficient for the faithful reconstruction of the data and 2) the latent representations between modalities are sufficiently well aligned that the latent encoding of one modality leads to faithful reconstruction in another modality. Of course one cannot achieve any meaningful alignment between modalities in the absence of class labels, i.e. no assignment constraint between the clusters of the various modalities, which is shown in the right column of the figure. The model did recover the clustered structure

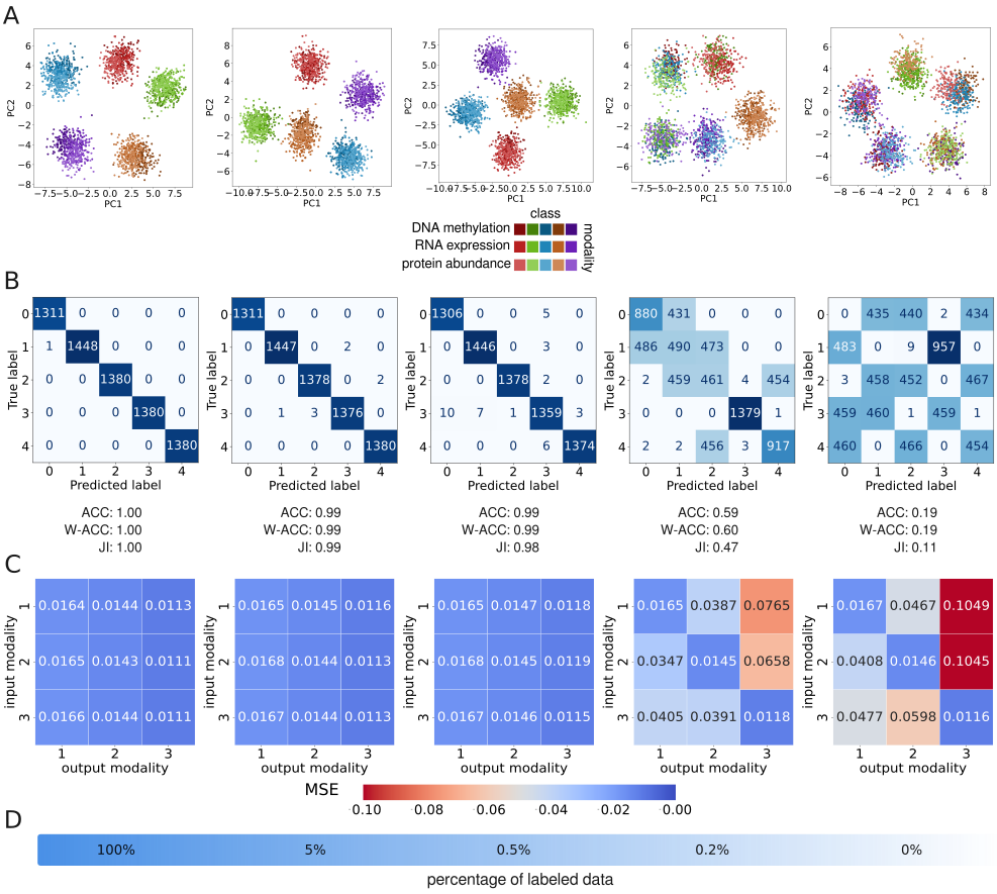


Figure 2: **Latent space alignment as a function of available class labels**. All panels display visualizations of five different models, each trained on varying amounts of labeled data. **A**, PCA projections of 800 samples per modality in the joint latent space. **B**, Confusion matrices of the discriminator's predictions. **C**, Heatmaps with the mean squared error (MSE) for modality reconstructions and translations. Each column of results correspond to the results gained with the percentage of (class) labeled examples in panel D. **D**, Percentages of class labels used in each case.

of the data, but as expected, struggled to match the right clusters between the various modalities, leading to random class prediction (class prediction accuracy of 19% across five classes Fig. 2.B fifth column). And similarly, as expected, the reconstruction error is as low as in the fully supervised case, with value around 0.01 for each of the three modality, but the translation error is much higher, e.g. with value 0.1 from modality 1 to modality 3. This confirms the misalignment of the modalities in the latent space. Overall, these results highlight the importance of explicit guidance in ensuring a well-aligned latent space across modalities.

We explored an intermediate, semi-supervised, setting in which only a subset of the training samples were annotated with their class labels. Our results show that even with minimal supervision, where only 5% or as little as 0.5% of the training data were annotated with their class label, our architecture was able to recover a nearly perfect class prediction accuracy on test data (Fig. 2.B, second and third column) and a reconstruction and translation error almost identical to the fully supervised baseline (Fig. 2.C, second and third column). However, when the fraction of annotated data was further reduced to 0.2% of the training data, the model's performance deteriorated with a class prediction accuracy decreasing to 59% and a translation error increasing to 0.0765, indicating a threshold below which supervision is no longer effective in guiding latent space alignment (Fig. 2.B and C, fourth column).

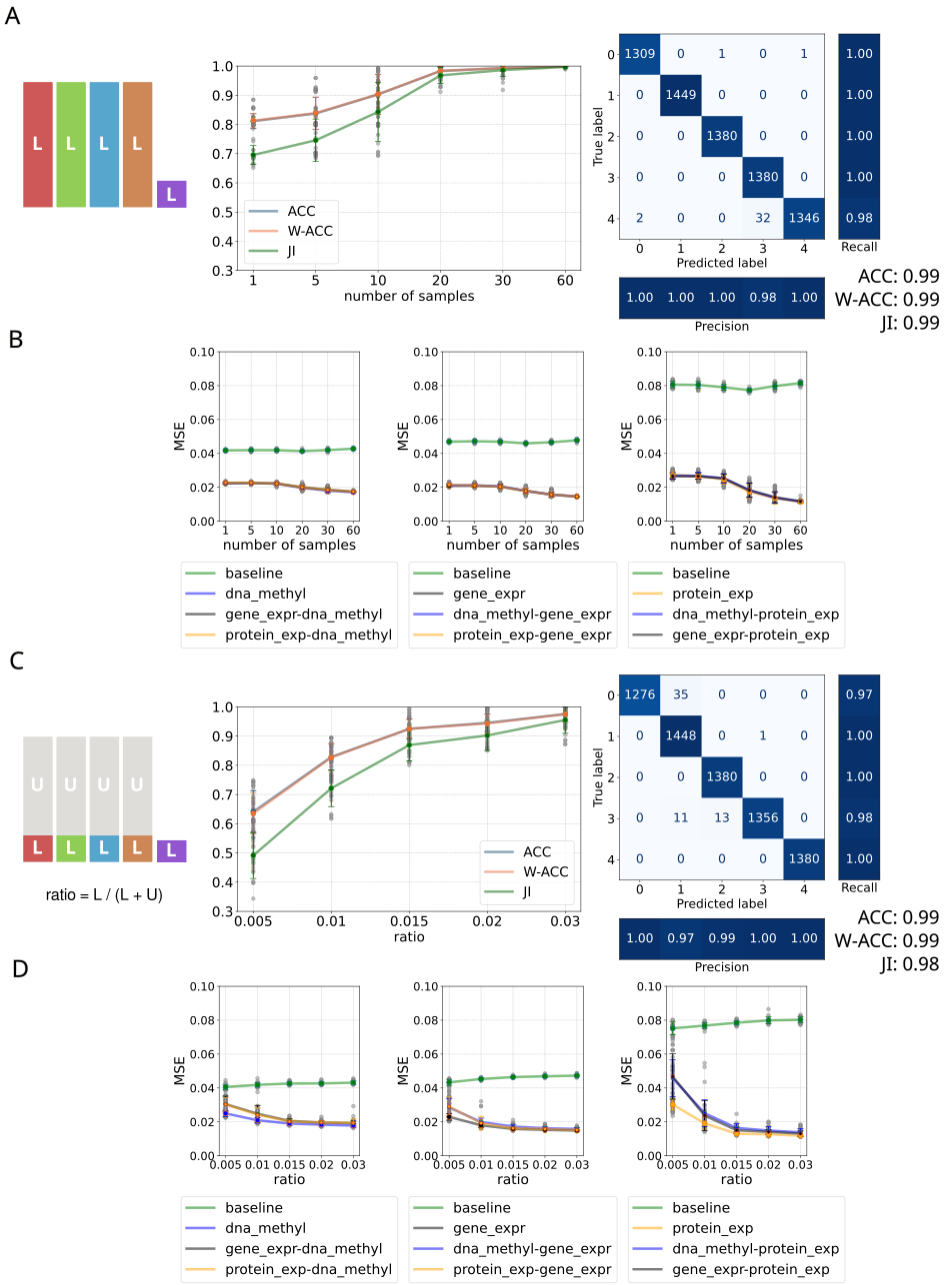


Figure 3: **Class imbalance between a reference dataset and a target dataset in fully supervised and semi-supervised settings**. **A and C**, First columns, depiction of the dataset distribution in the fully supervised and semi-supervised settings, L stands for labeled and U for unlabeled data respectively. The target dataset was fixed at 20 samples in the semi-supervised setting. Second columns, scatter plots with the accuracy (ACC), the inverted class frequency accuracy (W-ACC), and the Jaccard index (JI) metrics for different numbers of samples per modality in the target dataset and different ratios of labeled data in the reference dataset, respectively. Third column, confusion matrices illustrating the predictions of the discriminator for a target dataset with 20 samples per modality, and a ratio of 0.02 of annotated data with class labels, respectively. **B and D**, Reconstruction and translation errors for each omic in the fully supervised and semi-supervised settings, respectively. The baseline consists of the mean squared error (MSE) for all pairs between each sample and all reconstructions in a given omic.

Handling small datasets with transfer learning In the context of rare diseases, training datasets are often small, making it challenging to develop robust predictive models Banerjee et al. (2023). With MODIS, we propose to address this aspect by training our model on both a large reference dataset containing multiple classes of unpaired samples and a small target dataset containing one class of unpaired samples Fig. 1.B.

To evaluate the effect of class imbalance, we first considered a setting in which the reference dataset contained a large number of samples across multiple classes, while the target dataset consisted of only a few samples in a specific class (Fig. 3A). We evaluated the impact of sample size on classification performance by analyzing the evolution of class prediction accuracy as a function of the number of samples in the target dataset. Our results indicate that, even with a low number of samples, the model achieves good class prediction accuracy. More precisely, with 5 samples per modality in the target dataset (0.2% of the whole training dataset), MODIS achieves more that 0.8 class prediction accuracy (Fig. 3A). Similary, with 20 samples per modality in the target dataset (0.8% of the whole training dataset), MODIS achieves more than 0.98 class prediction accuracy. Naturally, the class prediction accuracy grows with the increase in the number of samples in the target dataset. This suggests that the knowledge transfer from the large reference dataset plays a crucial role in stabilizing the representation of the target class, allowing for accurate predictions despite the small dataset size.

Next, we investigated the minimal level of class supervision required in this highly unbalanced setting (Fig. 3C). In this case, we fixed the number of samples in the target dataset to 20 per modality and systematically varied the proportion of class-labeled versus unlabeled samples in the reference dataset. Our results reveal that the percentage of annotated data in the reference dataset influences the class prediction accuracy of both the reference and the target dataset. Notably, we found that the class prediction accuracy could exceed 0.95 with 2% of the reference dataset samples being annotated, and increases with the proportion of labeled samples. This finding underscores the efficiency of our semi-supervised approach, demonstrating that even a minimal amount of supervision in the reference dataset is sufficient to enable accurate classification in the rare disease setting. Overall, these results highlight the potential of leveraging large, well-annotated datasets to enhance model performance in low-data scenarios, a crucial advantage for rare disease research where labeled data is often scarce.

Assessing the impact of missing modality for a given class Beyond the issue of class imbalance, we also investigated the impact of missing modalities in one or more classes on the performance of our model. To systematically evaluate this challenge, we designed a series of increasingly difficult settings and assessed the model's ability to maintain latent space alignment and accurate class prediction despite missing data.

In the first case (Fig. 4.A), we removed all samples from one modality in a single class to observe whether the model could still accurately classify samples from the missing modality. Remarkably, despite the complete absence of data for that modality-class pair, the model achieved a class prediction accuracy of 97%. This suggests that our approach successfully leverages information from the available modalities and classes to infer meaningful representations, allowing for accurate predictions even in the presence of missing data. To further explore the impact of introducing minimal data in the missing modality, we gradually increased the number of samples available in that modality-class pair (Fig. 4.B). We found that adding just a single sample was sufficient to restore the class prediction accuracy to a near-perfect 99%, indicating that even minimal supervision can significantly improve model performance in such cases.

Next, we tested a more challenging scenario in which two modalities were missing in two different classes (Fig. 4.C). As expected, this setting led to a reduction in overall model accuracy, with a global class prediction accuracy of 85%. Notably, while the model still performed well across the dataset, the accuracy dropped for some specific classes and modalities. Importantly, these classes and modalities are not necessarily the ones in which data was completely missing. To determine whether the introduction of even a small number of samples could mitigate this effect, we incrementally added samples to the missing modalities (Fig. 4.D). Once again, we observed a substantial improvement, with the overall accuracy increasing to 97% after the addition of just one sample per missing modality-class pair. These findings emphasize the robustness of our approach in handling incomplete multi-modal datasets while also highlighting the critical role of even minimal supervi-

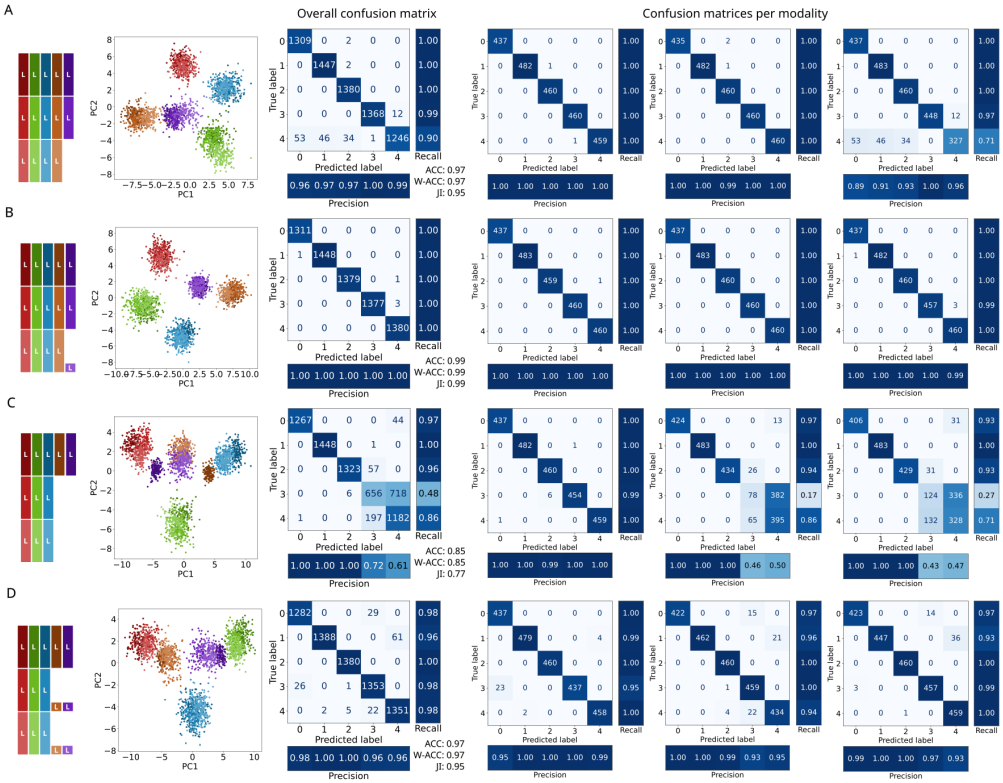


Figure 4: **Effect of missing or partially missing modalities on latent space alignment and class prediction accuracy**. Panels **A**, **B**, **C**, and **D** display the same visualizations in different settings. The first column displays the setting. L stands for labeled data. The second column presents the PCA projection of the joint latent space. The third column shows the overall confusion matrix of the discriminator's predictions. The fourth, fifth, and sixth columns display the confusion matrices for predictions of DNA methylation, gene expression, and protein abundance, respectively. Confusion matrices in **B** and **D** were obtained from datasets with only one sample in the shrunken class-modality pairs.

sion in restoring prediction accuracy. This capability is particularly relevant in real-world biomedical applications, where data collection across multiple modalities is often incomplete or unbalanced across different classes.

5 Conclusion

In summary, MODIS effectively addresses the challenges of multi-modal data integration and alignment in scenarios with limited annotated samples and small dataset size. Its ability to predict classes, generate missing data, and to align latent representations highlights its potential for a broad range of applications for the study of disease with scarce datasets such as rare diseases.

Acknowledgments

This research was supported by l'Agence Nationale de la Recherche (ANR), project ANR21-CE45-0001-01

References

Galen Andrew, Raman Arora, Jeff A. Bilmes, and Karen Livescu. Deep canonical correlation analysis. In *Proceedings of the 30th International Conference on Machine Learning, ICML 2013,*

Atlanta, GA, USA, 16-21 June 2013, volume 28 of *JMLR Workshop and Conference Proceedings*, pp. 1247–1255. JMLR.org, 2013. URL http://proceedings.mlr.press/v28/andrew13.html.

Ricard Argelaguet, Anna SE Cuomo, Oliver Stegle, and John C Marioni. Computational principles and challenges in single-cell data integration. *Nature biotechnology*, 39(10):1202–1215, 2021.

Jineta Banerjee, Jaclyn N Taroni, Robert J Allaway, Deepashree Venkatesh Prasad, Justin Guinney, and Casey Greene. Machine learning in rare disease. *Nature Methods*, 20(6):803–814, 2023.

Laura Cantini, Pooya Zakeri, Celine Hernandez, Aurelien Naldi, Denis Thieffry, Elisabeth Remy, and Anaïs Baudot. Benchmarking joint multi-omics dimensionality reduction approaches for the study of cancer. *Nature communications*, 12(1):124, 2021.

Prabhakar Chalise, Rama Raghavan, and Brooke L Fridley. Intersim: Simulation tool for multiple integrative 'omic datasets'. *Computer methods and programs in biomedicine*, 128:69–74, 2016.

Benjamin Elizalde, Soham Deshmukh, Mahmoud Al Ismail, and Huaming Wang. CLAP learning audio concepts from natural language supervision. In *IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP 2023, Rhodes Island, Greece, June 4-10, 2023*, pp. 1–5. IEEE, 2023. doi: 10.1109/ICASSP49357.2023.10095889. URL https://doi.org/10.1109/ICASSP49357.2023.10095889.

Zongbo Han, Changqing Zhang, Huazhu Fu, and Joey Tianyi Zhou. Trusted multi-view classification. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021. URL https://openreview.net/forum?id=OOsR8BzCnl5.

Zhen He, Shuofeng Hu, Yaowen Chen, Sijing An, Jiahao Zhou, Runyan Liu, Junfeng Shi, Jing Wang, Guohua Dong, Jinhui Shi, et al. Mosaic integration and knowledge transfer of single-cell multimodal data with midas. *Nature Biotechnology*, pp. 1–12, 2024.

David Hirst, Morgane Térézol, Laura Cantini, Paul Villoutreix, Matthieu Vignes, and Anaïs Baudot. Motl: enhancing multi-omics matrix factorization with transfer learning. *bioRxiv*, pp. 2024–03, 2024.

Chen Huang, Yining Li, Chen Change Loy, and Xiaoou Tang. Learning deep representation for imbalanced classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 5375–5384, 2016.

Michael Kampffmeyer, Sigurd Løkse, Filippo M Bianchi, Lorenzo Livi, Arnt-Børre Salberg, and Robert Jenssen. Deep divergence-based approach to clustering. *Neural Networks*, 113:91–101, 2019.

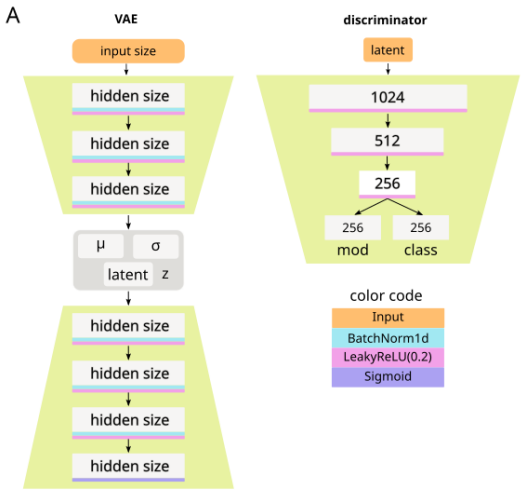
Alireza Makhzani, Jonathon Shlens, Navdeep Jaitly, and Ian J. Goodfellow. Adversarial autoencoders. *CoRR*, abs/1511.05644, 2015. URL http://arxiv.org/abs/1511.05644.

Jiquan Ngiam, Aditya Khosla, Mingyu Kim, Juhan Nam, Honglak Lee, and Andrew Y. Ng. Multimodal deep learning. In Lise Getoor and Tobias Scheffer (eds.), *Proceedings of the 28th International Conference on Machine Learning, ICML 2011, Bellevue, Washington, USA, June 28 - July 2, 2011*, pp. 689–696. Omnipress, 2011. URL https://icml.cc/2011/papers/399_icmlpaper.pdf.

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830, 2011.

Jules Samaran, Gabriel Peyré, and Laura Cantini. scconfluence: single-cell diagonal integration with regularized inverse optimal transport on weakly connected features. *Nature Communications*, 15 (1):7762, 2024.

Xin Tang, Jiawei Zhang, Yichun He, Xinhe Zhang, Zuwan Lin, Sebastian Partarrieu, Emma Bou Hanna, Zhaolin Ren, Hao Shen, Yuhong Yang, et al. Explainable multi-task learning for multimodality biological data analysis. *Nature communications*, 14(1):2546, 2023.



B

	DNA methylation	mRNA expression	protein abundance
input size	367	131	160
hidden size	512	256	256
latent size	64	64	64

Figure 5: **Architecture of the MODIS model**. **A**, General structure of a VAE. The specific size of each layer depends on the modality. The discriminator is designed based on the one used in an AC-GAN, which simultaneously predicts modality and class labels. **B**, Layer sizes for each modality-specific VAE in the InterSIM dataset.

Yaqing Wang, Quanming Yao, James T Kwok, and Lionel M Ni. Generalizing from a few examples: A survey on few-shot learning. *ACM computing surveys (csur)*, 53(3):1–34, 2020.

John N Weinstein, Eric A Collisson, Gordon B Mills, Kenna R Shaw, Brad A Ozenberger, Kyle Ellrott, Ilya Shmulevich, Chris Sander, and Joshua M Stuart. The cancer genome atlas pan-cancer analysis project. *Nature genetics*, 45(10):1113–1120, 2013.

Karren Dai Yang, Anastasiya Belyaeva, Saradha Venkatachalapathy, Karthik Damodaran, Abigail Katcoff, Adityanarayanan Radhakrishnan, GV Shivashankar, and Caroline Uhler. Multi-domain translation between single-cell imaging and sequencing data using autoencoders. *Nature communications*, 12(1):31, 2021.

A Appendix

A.1 Network architecture

The architecture of MODIS is shown on Fig. 5. In summary, it is defined by

- Meth VAE parameters: 1,531,375
- Gene expression VAE parameters: 382,723
- Protein VAE parameters: 397,600
- Discriminator parameters: 724,744
- Total model parameters: 3,036,442