MOX-Report No. 01/2026

Multi-view learning and omics integration: a unified perspective with applications to healthcare

Iapaolo V.; Vergani, A.M.; Cavinato, L.; Ieva, F.

MOX, Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

mox-dmat@polimi.it https://mox.polimi.it

-

Multi-view learning and omics integration: a unified perspective with applications to healthcare

Valeria Iapaolo[1] | Andrea Mario Vergani[1,2] | Lara Cavinato[1] | Francesca Ieva[1,2]

[1]MOX lab, Department of Mathematics, Politecnico di Milano, Via Bonardi 9, 20133, Milan, Italy

[2]Human Technopole, Viale Rita Levi-Montalcini 1, 20157, Milan, Italy

Correspondence
Valeria Iapaolo
Email: valeria.iapaolo@mail.polimi.it

Funding information
L. C. and F.I. are funded by the National Plan for Complementary Investments "Advanced Technologies for Human-centred Medicine" (PNC0000003).
All the authors acknowledge the MUR Excellence Department Project 2023-2027 awarded to Dipartimento di Matematica, Politecnico di Milano.

Abstract

Recent technological advances have made it possible to collect diverse biomedical data sources for each individual, ranging from imaging to genetics and digital health records. Integrating such heterogeneous information in a coherent and informative way is a key challenge for modern biomedical data analysis. In this work, we present a unified perspective that bridges the fields of multi-view learning and multi-omics integration, which have traditionally developed in parallel but share the same underlying objective. We organize this vast methodological landscape with respect to learning objectives, providing a structured overview of core paradigms, associated challenges, and emerging directions. Through a case study on UK Biobank data, we highlight the importance of interpretability in biomedical contexts by applying two representative methods, AJIVE and SGCCA, which bridge the multi-omics and multi-view learning streams. The results show that integrative approaches provide more informative and clinically meaningful insights than single-view

analyses, underscoring their practical relevance for biomedical research.

Keywords — multi-view learning, multi-omics, data fusion, interpretability, joint and individual components, cross-modal correlation

1 | INTRODUCTION

In recent years, the exponential growth in data availability has profoundly transformed the landscape of scientific research and clinical practice. The growing heterogeneity in the nature and structure of data sources has opened up unprecedented opportunities to enhance our understanding of complex systems [83]. Traditionally, biomedical analyses have relied on structured data such as demographic variables, diagnostic codes, and laboratory results, sometimes complemented by unstructured information such as medical notes [38]. Today, advances in data acquisition technologies allow for the large-scale collection of complex and high-dimensional information at reduced cost. Examples include high-throughput omics data (e.g., genomics, transcriptomics, proteomics), radiological imaging (MRI, CT), records of medication prescriptions, environmental exposure metrics, and behavioral data from wearable devices or mobile applications [16, 27, 41]. Large-scale initiatives such as The Cancer Genome Atlas (TCGA) [71] have further highlighted the potential of combining diverse data modalities to enable more comprehensive disease modeling and biomarker discovery.

These different modalities provide complementary and often non-redundant perspectives on the biological, clinical, and environmental factors that shape health and disease. Integrating such diverse sources in a principled way is essential to fully exploit their potential, as each modality captures a specific aspect of the underlying system and no single data type provides a complete picture.

However, traditional statistical techniques are generally ill-equipped to address these complexities. Most classical methods rely on strong assumptions such as linearity, independence between variables, and low dimensionality, which rarely hold for modern biomedical data. They struggle with heterogeneity in scale and noise structure, mismatched sample sizes across modalities, and sparse or high-dimensional feature spaces. Moreover, they typically analyze each data source in isolation, making it difficult to capture relationships that emerge only when modalities are considered jointly [24].

Cross-modal interactions are precisely where much of the meaningful biological signal resides: gene expression changes may influence imaging phenotypes; environmental exposure can modulate clinical biomarkers; wearable-based behavioral patterns can predict treatment response only in combination with molecular profiles. Capturing such dependencies enables more accurate prediction, improved patient stratification, and deeper mechanistic insight.

Properly linking data sets can be regarded as introducing a new form of diversity, and this diversity is the basis and driving force of data fusion [32].

Over the past two decades, a broad array of integrative methods has been developed to tackle the challenges associated with combining multiple data modalities [83, 23, 55, 5]. A prominent category of these approaches falls under the umbrella of multi-view learning [73], which focuses on leveraging multiple sources of information—known as views—of the same underlying entities. Originating primarily in the context of machine learning tasks involving web pages, signals, or images, this field has produced a wide range of models capable of jointly analyzing multiple sources of information without resorting to naive concatenation of features [62, 82].

In parallel, the field of multi-omics data integration has gained traction within the biomedical community, with the goal of jointly analyzing different layers of omics data—such as genomics, transcriptomics, proteomics, and epigenomics—to gain deeper insights into biological systems [60].

Despite originating in distinct application domains, multi-view learning and multi-omics share a common goal: integrating complementary sources of information while preserving the specific structure and role of each modality, and avoiding simplistic feature concatenation. However, the corresponding lines of research have largely evolved in parallel, with limited interaction between the two fields. Reviews in multi-omics integration [23, 60, 7, 11, 50, 53, 46, 28] typically do not refer to methods developed within the multi-view learning literature, while surveys on multi-view learning [62, 82, 35, 73, 20, 78, 3, 32, 69] omit the extensive body of work emerging from the multi-omics community.

This separation is difficult to justify, especially given that the underlying methods are not tied to the nature of the data they process. Techniques originally developed for omics integration can be readily applied to non-biomedical scenarios, just as multi-view models from general-purpose machine learning can prove effective in multi-omics applications. To the best of our knowledge, no existing review systematically connects the literature on multi-view learning with that of multi-omics data integration, despite the strong methodological analogies and the potential for mutual enrichment between the two fields.

The aim of this review is precisely to bridge this gap, offering a unified perspective that helps researchers navigate this vast landscape and encouraging a more integrated understanding of approaches that, although currently developed in separate communities, share a common foundation.

In what follows, we will adopt the term multi-view learning to refer broadly to this unified landscape of approaches that aim to extract information from multiple, complementary data views—regardless of whether the context is general-purpose or biomedical.

The remainder of this paper is structured as follows. Section 2 presents a unified overview of the methodological landscape, structured around core principles, key challenges, and a taxonomy of multi-view learning approaches based on learning objectives—including supervised, unsupervised, semi-supervised, and representation learning paradigms. Using data from UK Biobank, Section 3 presents a case study aimed at demonstrating the effectiveness of multi-view integration techniques for extracting clinically meaningful insights from heterogeneous biomedical data. We focus

on two representative methods—AJIVE and SGCCA—drawn from the multi-omics and multi-view learning literature, respectively. These approaches enable a gray-box modeling strategy, striking a balance between predictive power and interpretability. By leveraging the structure across multiple data modalities, both methods identify informative latent components that highlight clinically or biologically relevant variables, and demonstrate how integrative multi-view analysis can outperform single-modality models by offering a more nuanced understanding of complex health phenomena. Finally, Section 4 offers concluding reflections and highlights future research directions.

2 | A UNIFIED PERSPECTIVE ON MULTI-VIEW LEARNING AND MULTI-OMICS

When dealing with a multi-view dataset, several modeling goals may arise. One may aim to construct a low-dimensional joint representation that captures the relationships between modalities, to exploit this representation in downstream tasks such as prediction or stratification, or to directly combine the views for supervised or unsupervised modeling without any prior embedding step. These alternatives reflect different learning objectives and determine the most suitable methodological framework.

To bring structure to this variety, it is helpful to adopt a taxonomy based on the primary learning objective. This perspective allows us to distinguish four major paradigms of multi-view learning: multi-view supervised learning, multi-view unsupervised learning, multi-view semi-supervised learning, and multi-view representation learning [62]. Table 1 summarizes these paradigms by outlining their typical objectives and representative methods.

In parallel, several complementary paradigms have emerged—such as ensemble multi-view learning [74], active multi-view learning [45], and multi-view transfer learning [12]—which reflect different strategies for leveraging multiple modalities in more dynamic or adaptive settings. These paradigms are often built upon or combined with the core categories, expanding the applicability of multi-view learning to more complex or data-constrained scenarios.

Before delving into these paradigms, Section 2.1 introduces the key principles of multi-view learning, outlines the main challenges, and presents the different fusion strategies based on the timing of integration. The subsequent sections explore in greater detail the paradigms listed in Table 1, followed by an overview of the complementary strategies mentioned above.

2.1 | Principles, challenges and fusion strategies

At its core, multi-view learning is guided by two fundamental principles: the complementary principle and the consensus principle [79]. The complementary principle posits that different views should be jointly exploited to achieve a more comprehensive and accurate understanding of the system under study. Recent studies have demonstrated that integrating omics and imaging data can significantly improve predictive performance compared to unimodal approaches. In the context of non–small cell lung cancer, a cross-attention deep learning model combining transcrip-

TABLE 1 Summary taxonomy of multi-view learning paradigms.

Multi-view paradigm	Main objective	Methods
Supervised learning	Predict a specific target using features from multiple views	MFDA [13], SVM-2K [17], multi-view representation learning [35] + standard supervised methods
Unsupervised learning	Discover latent structures or clusters from unlabeled multi-view data	Graph-based approaches (SNF [68], ANF [40], SRF [22]), multi-view representation learning [35] + standard unsupervised methods
Semi-supervised learning	Leverage partial supervision to guide learning on unlabeled data	Co-training [9], co-regularization [56], graph-based methods [61]
Representation learning	Extract informative low-dimensional embeddings that preserve relevant structures across views	Dimensionality reduction techniques (CCA [26], KCCA [33], JIVE [39], AJIVE [19]), probabilistic models (MOFA [2], iCluster [54]), deep learning architectures (AEs [18], GANs [65], GNNs [77]), graph-based models [29]

tomic profiles, clinical data, and CT scans outperformed models based on single modalities [66]. Similarly, in glioma and renal cell carcinoma, the Pathomic Fusion framework showed that fusing histopathological images with genomic features led to enhanced diagnostic and prognostic accuracy [14]. Conversely, the consensus principle aims to promote consistency across views by identifying latent structures that are reflected in multiple modalities, thereby enhancing robustness and interpretability [39].

Although significant methodological advances, multi-view learning continues to face a number of open challenges. These include designing integration strategies that preserve relevant information without redundancy, coping with heterogeneous noise across modalities, managing high-dimensional data with limited sample sizes—as often encountered in biomedical settings—balancing views of differing dimensionality, resolving inconsistencies between views, and ensuring computational scalability [35, 32].

Despite these challenges, the multi-view framework remains attractive even when natural view separation is not explicitly available. In cases where data are collected from clearly distinct sources—such as different sensors or omics layers—the multi-view structure is straightforward. However, it has been empirically demonstrated that artificially creating views by randomly splitting features from a single-view dataset can also enhance performance in downstream tasks [10]. This finding suggests that the benefits of multi-view modeling may extend beyond cases with inherently distinct modalities. Nevertheless, this approach currently lacks theoretical support and offers limited interpretability, highlighting the need for more principled criteria in defining and validating view separations.

Another key question in this domain is *when to fuse* the data, a challenge that arises naturally due to the heterogeneity and varying informativeness of each data type [3]. The decision on the timing of integration can significantly

affect both the performance and interpretability of the resulting models. In fact, depending on whether fusion occurs before, during, or after model training, different integration strategies are adopted, each with distinct theoretical underpinnings and computational implications.

Three principal integration strategies have emerged [49]. Early fusion, or fusion in the data, is an approach in which all view matrices are first concatenated into a single representation, and then standard single-view algorithms are applied to this combined matrix. This enables joint modeling from the outset, but may obscure modality-specific patterns. Late fusion, or fusion in the results, builds separate models for each modality and then combines their outputs, offering flexibility but limiting cross-modal interaction [25, 43]. Both strategies have the advantage of allowing the use of standard machine learning models and software tools. In contrast, intermediate fusion integrates the modalities during the learning process itself. Although this approach often requires the development of dedicated algorithms and cannot rely on off-the-shelf tools, it enables more expressive modeling of cross-view interactions and has the potential to achieve superior performance [83]. For these reasons, the remainder of this paper focuses on intermediate fusion, organizing the review around learning goals and methodological frameworks.

2.2 | Multi-view supervised learning

Multi-view supervised learning addresses scenarios in which multiple data modalities are available and the outcome variable is fully observed. The goal is to exploit complementary information across views to improve predictive performance in tasks such as classification or regression [62]. Unlike semi-supervised settings, this framework relies exclusively on labeled data, which simplifies aspects such as model selection and evaluation.

Several methods have been developed to explicitly model the multi-view structure in supervised tasks. A notable example is the multi-view Fisher discriminant analysis (MFDA) [13], which extends classical discriminant analysis to jointly exploit information from multiple views in both binary and multi-class classification problems. This method seeks discriminative projections across modalities simultaneously, capturing shared class-relevant signals. Similarly, multi-view extensions of support vector machines, such as SVM-2K [17], incorporate cross-view consistency directly into the objective function by enforcing agreement between classifiers trained on each view.

However, in many real-world applications, a more common and pragmatic strategy involves a two-step pipeline: first, a multi-view representation learning method is used to derive informative latent features that summarize the joint structure of the modalities; then, standard statistical models—such as logistic regression, linear discriminant analysis, or Cox proportional hazards models—are applied to these latent representations to perform the final prediction [47, 57, 70]. This hybrid approach benefits from the expressiveness of unsupervised multi-view techniques in capturing shared and complementary patterns, while preserving the interpretability, simplicity, and robustness of well-established supervised models.

2.3 | Multi-view unsupervised learning

Multi-view unsupervised learning, by contrast, seeks to uncover latent structures within multi-modal datasets in the absence of supervision. Prominent applications include clustering and dimensionality reduction, where the objective is to reveal meaningful patterns and structures that emerge from the integrated analysis of all available modalities [8, 20, 79, 64].

Clustering is particularly relevant in biomedical contexts, where researchers are often interested in identifying latent subgroups of patients—such as molecular subtypes of cancer or groups with different response to treatment. Among the various families of clustering algorithms, graph-based approaches have gained significant traction for multi-view data integration.

Graph-based approaches [29] model inter-sample relationships by representing observations as nodes in a graph, with edges encoding pairwise similarity between them. Typically, similarity networks are first constructed independently for each modality, then integrated into a fused graph structure. Representative methods in this family include Similarity Network Fusion (SNF) [68], Affinity Network Fusion (ANF) [40], and Similarity Regression Fusion (SRF) [22]. A key advantage of graph-based approaches is their ability to seamlessly integrate heterogeneous data types, encompassing both categorical and numerical variables, provided that an appropriate similarity measure is defined. Once the fused graph is obtained, spectral clustering is commonly applied to extract group structure from the network.

As in the supervised setting, latent representations learned through multi-view embedding techniques are sometimes used as input for traditional clustering algorithms in a modular, two-step pipeline [49].

Dimensionality reduction represents another important objective within multi-view unsupervised learning. Although unsupervised, it is not tied to a specific task such as clustering or classification. Its primary goal is to learn compact, informative representations that capture the joint structure of multiple views. Given its foundational role across a variety of downstream tasks, we address dimensionality reduction methods in more detail in Section 2.6, within the broader context of multi-view representation learning.

2.4 | Multi-view semi-supervised learning

Multi-view semi-supervised learning operates in settings where labeled data are scarce but large volumes of unlabeled data can be exploited [73]. The goal is to leverage both labeled and unlabeled samples by integrating complementary information from multiple views, improving learning performance when full supervision is not feasible.

A well-established approach within this class is *co-training* [9], which assumes that each view is sufficient on its own to make a prediction, and that views are conditionally independent given the label. In its classical form, two classifiers are trained separately on distinct views using the available labeled data; each classifier then labels the most confident examples among the unlabeled data, and these pseudo-labels are iteratively added to the training set of the other view. This mutual reinforcement process allows the models to gradually expand their labeled set and improve

generalization.

Beyond co-training, various other strategies have been developed, including co-regularization frameworks [56], graph-based approaches, and agreement maximization. Co-regularization methods typically train separate models on each view while introducing regularization terms that penalize discrepancies between their outputs on unlabeled data. This encourages the models to agree where they are confident, without forcing alignment in noisy regions. Graph-based methods construct similarity graphs for each view and propagate label information over a fused or jointly-regularized graph, capturing both view-specific geometry and shared structure [61]. Agreement maximization techniques, on the other hand, explicitly optimize for consistency between predictions across views, often by minimizing divergence measures or enforcing consensus in latent representations [15].

While a large portion of multi-view semi-supervised learning focuses on propagating labels to unlabeled data, another relevant objective is to guide an unsupervised task using partial supervision. For instance, S2GC (Survival Supervised Graph Clustering) [37] incorporates survival time as a weak supervisory signal to guide the clustering of items. The method integrates a Cox model into the similarity graph construction, encouraging the formation of patient groups that share not only similar characteristics but also similar survival times.

More broadly, semi-supervised methods of this kind aim to strike a balance between the exploratory nature of unsupervised learning and the outcome-awareness of supervised approaches. By using outcome information as a guiding signal—rather than a strict target—they help uncover clinically meaningful patterns without fully constraining the discovery process. This makes them particularly suitable for biomedical contexts, where outcome information must inform—but not constrain—the discovery of complex phenotypic subgroups.

At the same time, the presence of supervision, even if weak, can introduce biases that limit the emergence of novel structures unrelated to the outcome. In this sense, supervision becomes a double-edged sword: it improves relevance, but may reduce discovery. Compared to fully supervised models, semi-supervised approaches are more flexible and capable of capturing subgroup structures that do not align neatly with a single predictive goal. However, they are often more challenging to design and tune, as they require balancing exploration and guidance without explicit optimization targets. In contrast, supervised models benefit from well-defined objectives and clearer criteria for model selection and evaluation, which makes them generally easier to implement and validate.

2.5 | Emerging directions in multi-view learning

Beyond the main learning paradigms, several emerging directions have broadened the methodological landscape of multi-view learning. These include ensemble approaches, active learning, and transfer learning, each addressing specific challenges in real-world multi-view scenarios.

Ensemble multi-view learning combines predictions from multiple models trained on different views, often improving robustness and generalization by leveraging view-specific insights. For example, the extension of AdaBoost

to the multi-view setting has been proposed, where weak learners are independently trained on each view and combined through a boosting strategy that adaptively reweights the contributions of each classifier [74].

Active multi-view learning focuses on minimizing annotation costs by selecting the most informative samples to label, while considering the complementary nature of multiple views. One of the earliest approaches adapts the co-training framework by selecting unlabeled examples on which classifiers trained on different views disagree the most. These disagreement-based strategies exploit the assumption that each view is sufficient for learning and that high disagreement signals potential information gain [45].

Multi-view transfer learning, on the other hand, aims to transfer knowledge from one domain or set of modalities to another, enabling learning even in cases with limited labeled data in the target domain. Several strategies have been developed, including co-training for domain adaptation and boosting-based frameworks that leverage multi-view information from source domains to improve performance in target domains [12, 75, 76].

2.6 | Multi-view representation learning

Multi-view representation learning [35] pursues the objective of constructing informative latent representations from multiple sources of information. In contrast to supervised and semi-supervised approaches, representation learning is agnostic to any specific downstream task. Its goal is to embed multi-modal data into a latent space that captures the essential structure of the system under study, thereby facilitating a wide range of subsequent analyses, including clustering, supervised analysis, and exploratory investigations.

Representation learning approaches can be broadly conceptualized as belonging to four principal methodological families: dimensionality reduction approaches, probabilistic models, deep learning-based methods, and graph-based techniques. These categories reflect different modeling paradigms and offer complementary perspectives for capturing multi-view structure. In the following, we examine each of these families in more detail, with particular emphasis on dimensionality reduction techniques, which represent a central focus of our methodological exploration.

2.6.1 | Dimensionality reduction approaches

Dimensionality reduction techniques aim to project high-dimensional multi-modal data into a compact latent space that captures the most informative structures across modalities. This step is critical for enhancing interpretability and computational efficiency, while also mitigating the risk of overfitting. High-dimensional data often contains substantial redundancy, which can obscure relevant signals and degrade the quality of learned representations [20]. By extracting salient patterns and discarding irrelevant variation, dimensionality reduction facilitates more robust and generalizable downstream analyses [42].

Within the broad family of dimensionality reduction methods, we distinguish two main subcategories that are

particularly relevant and widely adopted in multi-view applications: methods based on Canonical Correlation Analysis (CCA) and its extensions, and methods designed to explicitly separate shared and modality-specific components, among which Joint and Individual Variation Explained (JIVE) [39] is one of the most widely used and well-established approaches. While both approaches aim to capture the underlying structure of multi-modal data, the former emphasizes cross-view correlation, and the latter focuses on disentangling shared and individual sources of variation. Notably, while many of these methods are developed within the generic multi-view learning paradigm [26, 21, 52], several have emerged more specifically in the context of multi-omics data integration [39, 67].

Latent space construction via cross-modal correlation

CCA-based methods represent a foundational class of techniques in multi-view learning. The original Canonical Correlation Analysis (CCA) [26] seeks linear projections of two data views such that the corresponding projected variables—known as canonical variates—are maximally correlated. Rather than producing a single latent variable, CCA yields a sequence of canonical variates: each pair represents a direction in the respective data spaces that maximizes correlation, subject to being orthogonal to the previously identified directions. This sequential process, analogous to the extraction of principal components in PCA, constructs a low-dimensional representation that captures multiple axes of cross-view dependence.

Several extensions of CCA have been developed to address its limitations. Kernel CCA (KCCA) [33] introduces non-linear mappings through kernel functions, enabling the capture of more complex relationships between views. Generalized CCA (GCCA) [30] extends classical CCA to handle more than two data sources by maximizing shared correlation across multiple datasets. In contrast, Tensor CCA (TCCA) [31] extends CCA to handle multi-view data structured as high-order tensors—i.e., multidimensional arrays that capture not only features but also internal structure such as spatial or temporal dimensions—by modeling multi-way correlations within and across views through tensor decompositions.

Structured decomposition of shared and modality-specific components

A second important family of dimensionality reduction methods focuses on structured decompositions that separate shared and modality-specific sources of variation. These approaches decompose the input data into three distinct components: a joint structure common to all modalities, individual structures unique to each view, and a residual noise term. This modeling framework provides a valuable tool for dissecting the relative contributions of each data source and has gained particular traction in biomedical applications, where distinguishing between systemic signals and view-specific effects is key to interpretability.

A number of methods have been proposed within this class, which also exhibit methodological intersections with other representation learning paradigms. Prominent examples include Joint and Individual Variation Explained (JIVE) [39], Angle-based Joint and Individual Variation Explained (AJIVE) [19], Bayesian Group Factor Analysis (Bayesian GFA)

[67], Distinct and Common Simultaneous Component Analysis (DISCO-SCA) [52], and Structural Learning and Integrative Decomposition (SLIDE) [21]. These methods differ in modeling assumptions and algorithmic formulation: JIVE adopts a PCA-based iterative optimization framework that decomposes the data into joint, individual, and residual noise matrices through alternating low-rank approximations; AJIVE refines this idea by introducing a rigorous geometric approach for subspace decomposition, enabling a mathematically well-defined solution; Bayesian GFA incorporates structured priors into a Bayesian generative model to achieve similar decomposition; DISCO-SCA draws from sparse component analysis to induce parsimony in shared and individual components; and SLIDE applies structured regularization to enhance interpretability and scalability in large multi-view settings, with the additional capability of identifying partially shared variation by assuming block-sparsity across domains.

2.6.2 | Probabilistic modeling approaches

Probabilistic modeling approaches leverage generative models and Bayesian inference to incorporate domain-specific prior knowledge and to impose explicit assumptions on the statistical properties of the data. These methods provide a rigorous framework for modeling uncertainty, accommodating missing data, and encoding structural assumptions in the learning process. Prominent examples, such as iCluster [54] and MOFA [2], come from the multi-omics setting. iCluster employs a joint latent variable model to integrate multiple data modalities by inferring a shared low-dimensional representation. MOFA (Multi-Omics Factor Analysis), on the other hand, is based on Bayesian group factor analysis and identifies both shared and modality-specific factors underlying variability across views. Although MOFA also distinguishes between shared and unique sources of variation, this separation is not achieved through an explicit matrix decomposition. Instead, it arises from the probabilistic structure of the model, where each latent factor captures patterns of variation with differing relevance across the input modalities. While these approaches are appealing from a modeling standpoint, their complexity and computational demands can limit practical usability. The need for careful model specification, longer runtimes, and advanced statistical expertise may reduce their accessibility in settings that require straightforward and efficient workflows. Moreover, some models rely on strong assumptions—such as data normality—that may not hold in real-world applications, further constraining their applicability.

2.6.3 | Deep learning-based approaches

Deep learning-based approaches employ neural architectures capable of modeling complex, non-linear interactions across multiple modalities [78, 69]. These include a wide range of architectures, such as multi-view Convolutional Neural Networks [59], multi-view Autoencoders [18], multi-view Generative Adversarial Networks [65], multi-view Graph Neural Networks [77], multi-view Restricted Boltzmann Machines [80], Deep CCA [1], and contrastive learning frameworks [81]. These methods are particularly well suited to capturing intricate cross-modal dependencies and

leveraging large-scale datasets. However, they are typically black-box in nature and offer limited interpretability, a notable drawback in domains where transparency and explainability are essential.

2.6.4 | Graph-based approaches

Graph-based approaches [29] model inter-sample relationships by representing observations as nodes in a graph, with edges encoding pairwise similarity between them. As previously discussed in Section 2.3, similarity networks are typically constructed independently for each modality and subsequently integrated into a fused graph that combines complementary information across views. In the context of representation learning, to derive a compact and informative embedding for each observation, spectral methods such as Laplacian Eigenmaps [6] are commonly applied to the fused graph. These techniques project the data into a low-dimensional space that preserves the local geometry of the graph, enabling downstream tasks such as clustering or visualization.

3 | CASE STUDY

One of the most critical challenges in the integration of heterogeneous biomedical data is interpretability. In clinical settings, understanding how and why a model reaches a certain output is not just desirable—it is essential. Beyond predictive accuracy, what ultimately matters is the capacity to extract reliable insights that can support medical decision-making and potentially influence clinical guidelines.

Interpretability allows researchers and clinicians to assess the relevance of specific data modalities, identify the variables that drive associations, and evaluate the consistency of findings with known biological or physiological mechanisms. For example, a latent factor extracted from multi-modal integration that is predominantly explained by a particular ECG-derived feature, or by a region-specific radiomic signature from medical imaging, can shed light on the prognostic or diagnostic relevance of that variable. Such insights, when grounded in a transparent modeling framework, can be used to guide further clinical investigations or be incorporated into evolving standards of care.

However, not all integrative models support this level of transparency. Many state-of-the-art approaches—including deep learning architectures and graph-based methods—achieve impressive results, but do so at the cost of acting as black boxes [51, 36]. Their internal representations are often inaccessible or uninterpretable, making them less suitable for domains in which accountability and mechanistic understanding are key.

In this section, we present a case study aimed at evaluating how interpretable multi-view learning methods can enhance cardiovascular risk prediction. To this end, we use UK Biobank data to assess whether integrative approaches that explicitly preserve interpretability can offer improved predictive performance compared to single-modality models, while also providing clinically meaningful insights into the structure and relevance of the contributing variables.

In light of this, in Section 3.1 we focus on two well-established methods representative of the multi-view learning

and multi-omics integration paradigms that explicitly preserve interpretability while integrating diverse data sources. The results of their application to UK Biobank data are reported in Section 3.2.

3.1 | Methods

We focus on two interpretable methods for multi-view integration: Angle-based Joint and Individual Variation Explained (AJIVE) and Sparse Generalized Canonical Correlation Analysis (SGCCA). These methods belong to two distinct families of approaches and offer complementary insights.

AJIVE [19] falls within the class of structured decomposition techniques, which disentangle joint from modality-specific sources of variation. This modeling framework provides a principled way to quantify the relative contribution of each view, assess redundancy, and identify clinically or biologically relevant components. By decomposing variation into shared and individual subspaces, AJIVE enables a fine-grained understanding of how information is distributed across data modalities.

SGCCA [63], on the other hand, belongs to the family of Canonical Correlation Analysis extensions. These methods aim to identify maximally correlated latent directions across views. SGCCA introduces sparsity constraints to classical CCA, allowing the discovery of interpretable and focused components. These sparse latent directions highlight the variables that contribute most to cross-modal associations, facilitating interpretability and variable selection.

3.1.1 | Angle-based Joint and Individual Variation Explained

Angle-based Joint and Individual Variation Explained (AJIVE) is a recent extension of JIVE that overcomes several of its key limitations. Unlike the original JIVE, which relies on an iterative procedure without guaranteed convergence and potentially unstable rank selection, AJIVE adopts a non-iterative, geometrically principled framework yielding mathematically well-defined and robust decompositions.

Let us consider K data blocks $X_1, \ldots, X_K$, where each matrix $X_k \in \mathbb{R}^{p_k \times n}$ represents a different data view measured on the same set of n subjects. Each data block is assumed to be generated according to the following model:

$$X_k = J_k + I_k + E_k, \quad \text{for } k = 1, \ldots, K,$$

where J_k is a low-rank joint matrix that captures the variation that is common across views, I_k is a low-rank individual component that represents view-specific signal orthogonal to the joint structure, and the noise component E_k accounts for residual unexplained variation.

This decomposition relies on a set of structural assumptions that guarantee identifiability and interpretability. First, all joint components $\{J_k\}_{k=1}^{K}$ are assumed to share the same row space, that is, $\text{row}(J_1) = \cdots = \text{row}(J_K) =$

row(J), where row$(J) \subseteq \mathbb{R}^n$ denotes the common subspace in which the joint variation across all blocks lies. Second, each individual component I_k is orthogonal to the joint space, meaning that row$(I_k) \perp$ row(J), for all $k = 1, \ldots, K$. Finally, the individual components are mutually non-overlapping, in the sense that their row spaces do not share any direction. Formally, this means that $\bigcap_{k=1}^{K}$ row$(I_k) = \{\vec{0}\}$, to guarantee that no variation is simultaneously individual in all views.

The AJIVE procedure begins by approximating each data block with a low-rank matrix, retaining the dominant directions of variation to separate signal from noise via a truncated singular value decomposition. These reduced representations are then analyzed to identify shared variation across blocks by assessing the alignment between their row spaces through principal angles, where smaller angles reflect stronger similarity. To guard against spurious alignments due to noise, thresholds based on perturbation bounds, such as the Wedin bound and the random direction bound, are applied to reliably estimate the joint score space. Finally, each block is projected onto the estimated joint subspace and its orthogonal complement, enabling the decomposition into joint, individual, and residual noise components. As a result, AJIVE provides latent scores for each subject, capturing both their joint and block-specific patterns of variation, which can then be used for downstream analyses.

3.1.2 | Sparse Generalized Canonical Correlation Analysis

Sparse generalized canonical correlation analysis (SGCCA) extends the regularized generalized canonical correlation analysis (RGCCA) framework by introducing an explicit mechanism for variable selection. While RGCCA applies regularization to classical generalized canonical correlation analysis (GCCA) to stabilize the estimation in high-dimensional and collinear settings, SGCCA goes one step further by encouraging sparsity. In this way, it identifies only a relevant subset of variables within each block, improving interpretability and focusing on the most informative features. This is particularly valuable in biomedical contexts, where underlying biological processes are often driven by a limited number of key variables, and interpretability of the resulting latent factors is essential.

Consider K data blocks $X_1, \ldots, X_K$, with each block $X_k \in \mathbb{R}^{n \times p_k}$, where n is the number of subjects and p_k the number of variables in block k. SGCCA searches for weight vectors $\mathbf{a}_k \in \mathbb{R}^{p_k}$, which determine linear combinations of the original variables and define latent components $t_k = X_k \mathbf{a}_k$. These latent components summarize each block's information along directions that maximize the association across the connected blocks, providing a low-dimensional representation of the subjects. The elements of $\mathbf{a}_k$ determine the direction in the variable space along which the data are projected, while the sparsity constraint forces many entries of $\mathbf{a}_k$ to zero. In this way, only a limited set of variables effectively contributes to defining the latent direction, improving both interpretability and stability of the solution.

The inter-block connections are specified through a design matrix $C \in \mathbb{R}^{K \times K}$, where c_{kj} equals 1 if block k and

block j are connected, and 0 otherwise. The SGCCA optimization problem takes the form

$$\max_{\mathbf{a}_1,\ldots,\mathbf{a}_K} \sum_{k,j=1}^{K} c_{kj}\, g\left(\mathrm{cov}\left(X_k\mathbf{a}_k, X_j\mathbf{a}_j\right)\right)$$

subject to

$$\|\mathbf{a}_k\|_2 = 1, \quad \|\mathbf{a}_k\|_1 \leq s_k, \quad k = 1,\ldots,K,$$

where the scheme function $g(\cdot)$ can be selected among the identity $g(x) = x$, the factorial $g(x) = x^2$, or the centroid $g(x) = |x|$ form, depending on the type of association to be emphasized. The hyperparameters s_k govern the degree of sparsity applied to each block, and the optimization is usually performed via an alternating algorithm, updating one $\mathbf{a}_k$ at a time while keeping the others fixed until convergence.

The method allows the extraction of multiple latent components from each block by applying a deflation strategy after estimating the first component, thereby removing its explained variability from the data and recovering subsequent orthogonal directions. Each additional component captures further structure not explained by the previous ones, offering a richer characterization of the relationships across the data blocks. The resulting latent components summarize the shared variation across the data sources in a compact set of features, which can then be used as inputs for downstream analyses, such as predictive models, clustering, or time-to-event analyses, depending on the scientific objective.

3.2 | Results

We considered common predictors of cardiovascular risk, electrocardiogram (ECG)-derived features and measures extracted from cardiac magnetic resonance imaging (referred to as *cardiac measures*) from the UK Biobank, a large scale United Kingdom-based study that followed over 500,000 participants aged 40-69 since 2006. Our objective was to integrate these three data modalities to predict cardiovascular disease occurrence in a population of individuals healthy at baseline, in a survival analysis framework. Cardiovascular risk factors included age, body mass index (BMI), standing height, weight and systolic blood pressure - all measured during the first UKB imaging visit, which constituted the baseline for our study. ECG measures were derived during the visit and comprised ventricular rate, P duration, PP interval, PQ interval, number of QRS complexes, QRS duration, QT interval, QTC interval, RR interval, P axis, R axis, T axis. Finally, the 84 cardiac measures we considered were extracted by [4] and directly available in UKB. Presence of cardiovascular disease, the outcome of our case study, was defined according to Table 2; we excluded from our analyses all the subjects who experienced an endpoint event and/or vascular dementia (ICD-10 code F01)

before the baseline. The dataset used for the analyses includes 18,682 participants, among whom 909 experienced a cardiovascular event during follow-up.

TABLE 2 Definition of cardiovascular disease based on ICD-10 (for hospital diagnoses and causes of death) and ICD-9 (for hospital diagnoses) codes (ICD: International Classification of Diseases).

ICD-10	F01	I20	I21	I22	I23	I24	I25	I50	I60	I61	I62	I63	I64	I65	I66	I67	I68	I69
ICD-9	410	411	412	413	414	428	430	431	432	433	434	436	437	438				

The AJIVE algorithm was applied using an adapted version of the publicly available code from `https://github.com/idc9/mvdr`. Before analysis, the three data views previously described were standardized to zero mean and unit variance to harmonize scales and prevent variables with higher variance from dominating the joint structure estimation.

In the initial dimensionality reduction step, principal components explaining at least 90% of the variance were retained for each view to preserve signal while reducing noise. This resulted in 36 components for cardiac measures, 7 for ECG-derived features, and 4 for clinical variables. This setup led to the identification of 4 joint components shared across the views, along with 33 individual components for cardiac measures, 5 for ECG-derived variables, and 1 for clinical variables. Each subject was thus represented by latent scores summarizing shared and view-specific variation, providing a comprehensive basis for downstream analyses.

Figure 1 summarizes the proportions of explained variance by partitioning each view into joint, individual, and residual components, computed as the variance of each matrix relative to the original data. Overall, the joint structure accounted for a substantial proportion of the variance, particularly for the clinical variables, while retaining meaningful view-specific patterns, especially for cardiac measures.

Among the cardiac measures, the loadings of the first joint component highlight several clinically relevant variables. Notably, *LV circumferential strain global* shows the highest positive loading (0.254), indicating a strong contribution to the shared latent structure. Circumferential strain is a well-established marker of myocardial deformation, revealing subtle impairments in left ventricular mechanics [44, 58]. In contrast, *LV radial strain global* has a large negative loading (–0.251), suggesting an opposite association; radial strain reflects myocardial thickening and provides complementary information on ventricular function [44]. Likewise, *LV ejection fraction* shows a negative loading (–0.217), consistent with its role as a global indicator of systolic performance [34].

For the clinical variables (Figure 2), the joint component shows a strong positive loading for *systolic blood pressure* (0.926) and a negative loading for *body mass index* (–0.244), aligning with evidence that both blood pressure and weight are key cardiovascular risk factors [72, 48]. Their opposite signs suggest the joint component captures a pattern where higher systolic blood pressure co-occurs with lower BMI within the shared variation extracted by AJIVE. Overall, these results confirm the method's ability to recover a clinically meaningful latent structure integrating complementary signals across modalities.

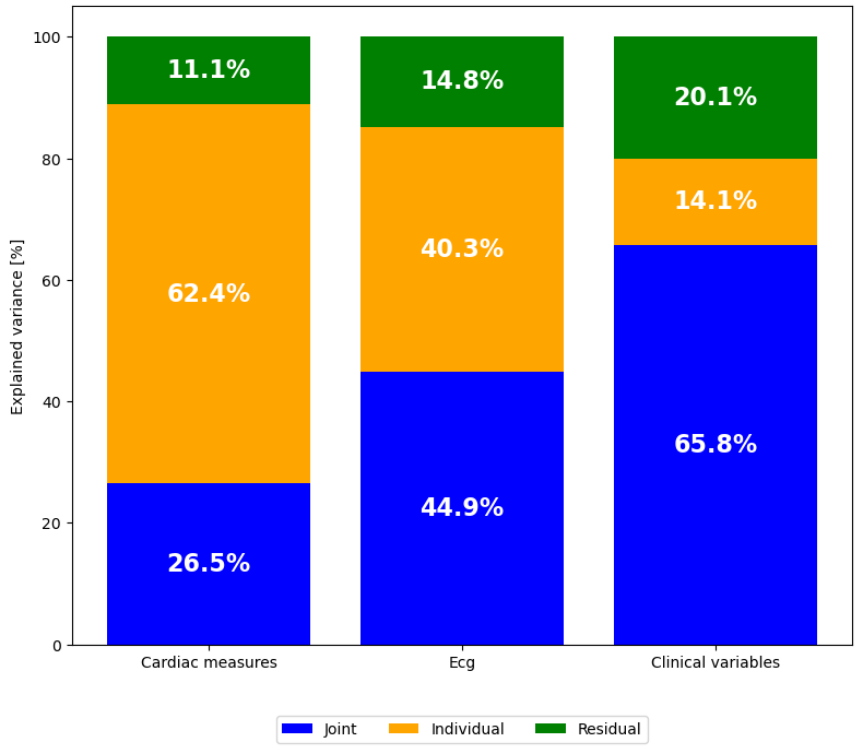

FIGURE 1 Explained variance for each data view (cardiac measures, ECG-derived variables, and clinical variables), partitioned into joint, individual, and residual components obtained from the AJIVE decomposition.

The sparse generalized canonical correlation analysis (SGCCA) was then applied to the same three modalities considered in the previous algorithm, using the `RGCCA` package in R. The variables were rescaled to zero mean and unit variance as done previously, to ensure comparability across blocks.

The SGCCA was implemented using the centroid scheme function $g(x) = |x|$, with sparsity hyperparameters tuned via a 10-fold cross-validation procedure. A Cox penalized regression model with Lasso penalty was adopted to maximize the concordance index for predicting the development of cardiac disease, motivated by the considerable number of components involved. Consistently with the approach used for AJIVE, the number of variables was selected to retain at least 90% of the explained variance, as measured through the Average Variance Explained (AVE) metric.

This procedure led to the selection of 48 variables for the cardiac measures, 7 for the ECG-derived variables, and 4 for the clinical variables. The resulting latent components can be interpreted through their weight vectors, which highlight the most relevant features driving inter-block associations. Among the cardiac measures, the largest absolute weights in the first component were found for *body surface area*, *RV end-diastolic volume*, *LV end-diastolic volume*, *LV myocardial mass*, and *LV stroke volume*, all with negative weights around −0.2 to −0.3, suggesting a coordinated contribution of volumetric and structural metrics. In contrast, *RA ejection fraction* and *average heart rate* displayed

smaller positive weights, indicating a weaker but distinct association.

Regarding the clinical variables (Figure 2), SGCCA assigned the strongest positive weight to *weight* (0.96), with more moderate contributions from *standing height* and *body mass index*, while *systolic blood pressure* and *age* received null weights. Overall, these results suggest that SGCCA emphasizes body size–related metrics as key contributors to shared information across modalities.

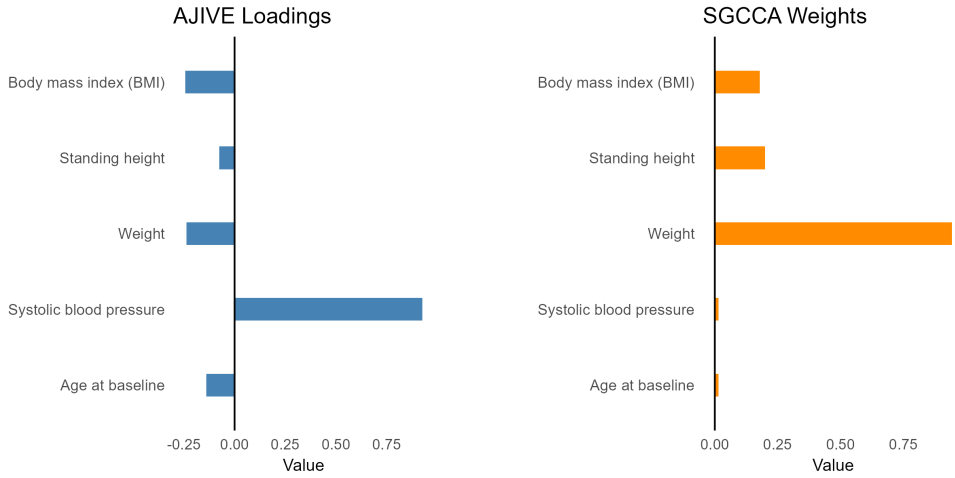

FIGURE 2 Variable contributions to the first joint component (left, loadings from AJIVE) and to the first latent component (right, weights from SGCCA) for the set of clinical variables.

To further evaluate the clinical utility of the latent representations derived through AJIVE and SGCCA, their corresponding scores were included as predictors in a Cox proportional hazards model with Lasso penalization. Figure 3 summarizes the results, comparing the prognostic performance of five sets of scores: the AJIVE-derived scores, the SGCCA-derived latent components, and the scores obtained from applying PCA separately to each of the three modalities, where the number of principal components retained was selected to explain at least 90% of the variance, in order to remain consistent with the previous choices.

The concordance index (C-index), estimated via 10-fold cross-validation, was used to assess discriminative ability, with confidence intervals obtained across folds to quantify uncertainty. This framework enables a direct comparison of integrative latent components versus single-modality features for survival prediction. In terms of point estimates, AJIVE- and SGCCA-based features both reached the highest C-index (0.702), suggesting improved performance over single-view models. Lower mean C-index values were observed for cardiac measures (0.680), clinical variables (0.673), and ECG features (0.590).

While these results indicate comparable predictive performance for the two integrative methods, AJIVE and SGCCA differ in terms of interpretability: the former explicitly separates joint and individual sources of variation, whereas the latter relies on latent components that are less directly interpretable.

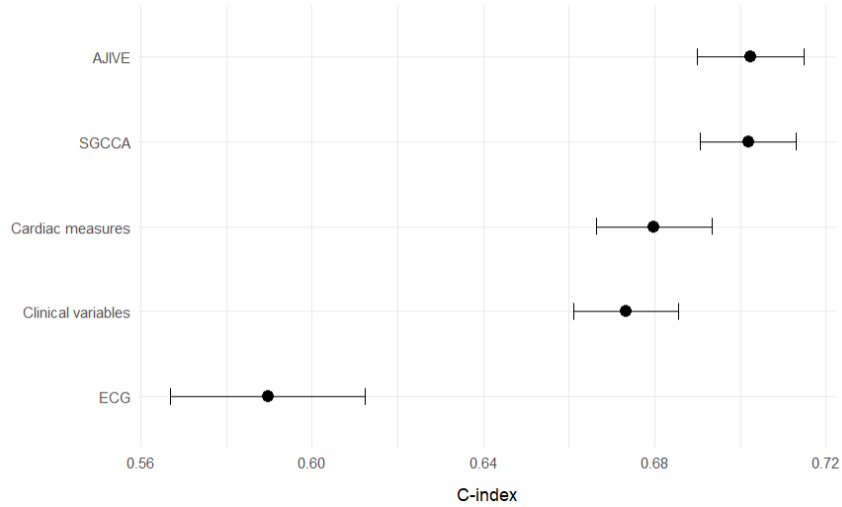

FIGURE 3 Mean C-index and corresponding 95% confidence intervals estimated via 10-fold cross-validation for five different models. The models include latent scores derived from AJIVE, SGCCA, and principal components extracted from each individual modality.

To assess whether multimodal integration leads to significantly improved predictive performance, we conducted Wilcoxon signed-rank tests for paired data on the C-index values obtained across the 10 cross-validation folds. Specifically, we compared the performance of the AJIVE- and SGCCA-based features with each of the single-view models. All comparisons yielded statistically significant results: all p-values were equal to 9.8×10^{-4}, except for the comparison between SGCCA and cardiac measures, which had a p-value of 4.9×10^{-3}. These results confirm that both integrative approaches lead to significantly improved predictive performance compared to each of the single-modality models, supporting the added value of multi-view integration in capturing complementary information across data sources.

These findings support the idea that joint modeling of multiple data modalities can yield more accurate and robust survival predictions compared to single-source approaches. In the biomedical context, this improvement is particularly valuable, as it suggests that methods capable of extracting complementary signals across views—while retaining interpretability—may contribute to more informed clinical decision-making and ultimately enhance patient care.

4 | DISCUSSION

This work set out to offer a unified and structured perspective on the methodological landscape for multi-view data integration. By bridging two historically parallel research communities, we have emphasized their shared conceptual foundations and overlapping goals, grounded in the challenge of integrating heterogeneous data sources in a meaningful way. The proposed taxonomy, organized by learning objective, offers researchers a systematic entry point

into this diverse methodological space—one that not only maps the well-established paradigms, but also highlights lesser-known but powerful learning strategies.

Crucially, the case study demonstrates how methods originating from both fields—such as AJIVE from multi-omics and SGCCA from multi-view learning—can be fruitfully applied to the same real-world dataset with analogous results. Their success confirms that there is no fundamental reason for these communities to remain isolated. On the contrary, drawing from both can expand the range of tools available to researchers and increase the chances of extracting clinically meaningful insights. Notably, we also showed that integrative approaches can lead to significant improvements in predictive performance compared to single-modality models, further reinforcing their practical value.

It is worth noting, however, that although the two methods achieved similar results in terms of overall performance, AJIVE and SGCCA produced quite discordant results in terms of variable selection. One possible explanation is that the components extracted by the two algorithms may simply be ordered differently, giving the impression of inconsistency even when it is not. However, the discrepancies may also be genuine, reflecting differences in the underlying assumptions and sparsity mechanisms of the two algorithms. These two possibilities together point to the importance of applying multiple integrative methods and carefully assessing the robustness of the findings, as variable-level discrepancies may signal sensitivity to method-specific choices or potential instability in the results.

Combining these perspectives helps build models that are not only effective, but also transparent and interpretable. The ability to disentangle shared from modality-specific variation, or to extract sparse correlated components across views, proves essential in biomedical contexts—where each latent factor can be interpreted in terms of measurable signals (e.g., from imaging or ECG) with potential clinical relevance. Such interpretability is not a luxury, but a prerequisite for trust, adoption, and clinical translation, and is expected to play a central role in the next generation of biomedical data analysis.

Acknowledgements

L. Cavinato is funded by the National Plan for NRRP Complementary Investments “Advanced Technologies for Human-centred Medicine” (PNC0000003). All authors acknowledge the MIUR Excellence Department Project 2023-2027 awarded to Dipartimento di Matematica, Politecnico di Milano.
This research has been conducted using the UK Biobank Resource under application number 82779.
The authors would like to thank Emanuele Di Angelantonio for his support in the definition of cardiovascular disease based on hospital diagnoses and causes of death.

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Galen Andrew, Raman Arora, Jeff Bilmes, and Karen Livescu. Deep canonical correlation analysis. In *International conference on machine learning*, pages 1247–1255. PMLR, 2013.

[2] Ricard Argelaguet, Britta Velten, Damien Arnol, Sascha Dietrich, Thorsten Zenz, John C Marioni, Florian Buettner, Wolfgang Huber, and Oliver Stegle. Multi-omics factor analysis—a framework for unsupervised integration of multi-omics data sets. *Molecular systems biology*, 14(6):e8124, 2018.

[3] Pradeep K Atrey, M Anwar Hossain, Abdulmotaleb El Saddik, and Mohan S Kankanhalli. Multimodal fusion for multimedia analysis: a survey. *Multimedia systems*, 16:345–379, 2010.

[4] Wenjia Bai, Hideaki Suzuki, Jian Huang, Catherine Francis, Shuo Wang, Giacomo Tarroni, Florian Guitton, Nay Aung, Kenneth Fung, Steffen E Petersen, et al. A population-based phenome-wide association study of cardiac and aortic structure and function. *Nature medicine*, 26(10):1654–1662, 2020.

[5] Ana R Baião, Zhaoxiang Cai, Rebecca C Poulos, Phillip J Robinson, Roger R Reddel, Qing Zhong, Susana Vinga, and Emanuel Gonçalves. A technical review of multi-omics data integration methods: from classical statistical to deep generative approaches. *Briefings in bioinformatics*, 26(4):bbaf355, 2025.

[6] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation*, 15(6):1373–1396, 2003.

[7] Matteo Bersanelli, Ettore Mosca, Daniel Remondini, Enrico Giampieri, Claudia Sala, Gastone Castellani, and Luciano Milanesi. Methods for the integration of multi-omics data: mathematical aspects. *BMC bioinformatics*, 17:167–177, 2016.

[8] Steffen Bickel and Tobias Scheffer. Multi-view clustering. In *Icdm*, volume 4, pages 19–26, 2004.

[9] Avrim Blum and Tom Mitchell. Combining labeled and unlabeled data with co-training. In *Proceedings of the eleventh annual conference on Computational learning theory*, pages 92–100, 1998.

[10] Ulf Brefeld, Christoph Büscher, and Tobias Scheffer. Multi-view discriminative sequential learning. In *Machine Learning: ECML 2005: 16th European Conference on Machine Learning, Porto, Portugal, October 3-7, 2005. Proceedings 16*, pages 60–71. Springer, 2005.

[11] Chongyang Chen, Jing Wang, Donghui Pan, Xinyu Wang, Yuping Xu, Junjie Yan, Lizhen Wang, Xifei Yang, Min Yang, and Gong-Ping Liu. Applications of multi-omics analysis in human diseases. *MedComm*, 4(4):e315, 2023.

[12] Minmin Chen, Kilian Q Weinberger, and John Blitzer. Co-training for domain adaptation. *Advances in neural information processing systems*, 24, 2011.

[13] Qiaona Chen and Shiliang Sun. Hierarchical multi-view fisher discriminant analysis. In *Neural Information Processing: 16th International Conference, ICONIP 2009, Bangkok, Thailand, December 1-5, 2009, Proceedings, Part II 16*, pages 289–298. Springer, 2009.

[14] Richard J Chen, Ming Y Lu, Jingwen Wang, Drew FK Williamson, Scott J Rodig, Neal I Lindeman, and Faisal Mahmood. Pathomic fusion: an integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis. *IEEE Transactions on Medical Imaging*, 41(4):757–770, 2020.

[15] C Christoudias, Raquel Urtasun, and Trevor Darrell. Multi-view learning in the presence of view disagreement. *arXiv preprint arXiv:1206.3242*, 2012.

[16] Conor John Cremin, Sabyasachi Dash, and Xiaofeng Huang. Big data: Historic advances and emerging trends in biomedical research. *Current Research in Biotechnology*, 4:138–151, 2022.

[17] Jason Farquhar, David Hardoon, Hongying Meng, John Shawe-Taylor, and Sandor Szedmak. Two view learning: Svm-2k, theory and practice. *Advances in neural information processing systems*, 18, 2005.

[18] Fangxiang Feng, Xiaojie Wang, and Ruifan Li. Cross-modal retrieval with correspondence autoencoder. In *Proceedings of the 22nd ACM international conference on Multimedia*, pages 7–16, 2014.

[19] Qing Feng, Meilei Jiang, Jan Hannig, and JS Marron. Angle-based joint and individual variation explained. *Journal of multivariate analysis*, 166:241–265, 2018.

[20] Lele Fu, Pengfei Lin, Athanasios V Vasilakos, and Shiping Wang. An overview of recent multi-view clustering. *Neurocomputing*, 402:148–161, 2020.

[21] Irina Gaynanova and Gen Li. Structural learning and integrative decomposition of multi-view data. *Biometrics*, 75(4):1121–1132, 2019.

[22] Yang Guo, Jianning Zheng, Xuequn Shang, and Zhanhuai Li. A similarity regression fusion model for integrating multi-omics data to identify cancer subtypes. *Genes*, 9(7):314, 2018.

[23] Yehudit Hasin, Marcus Seldin, and Aldons Lusis. Multi-omics approaches to disease. *Genome biology*, 18:1–15, 2017.

[24] Konstantin Hemker, Nikola Simidjievski, and Mateja Jamnik. Healnet: Multimodal fusion for heterogeneous biomedical data. *Advances in Neural Information Processing Systems*, 37:64479–64498, 2024.

[25] Katherine A Hoadley, Christina Yau, Denise M Wolf, Andrew D Cherniack, David Tamborero, Sam Ng, Max DM Leiserson, Beifang Niu, Michael D McLellan, Vladislav Uzunangelov, et al. Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell*, 158(4):929–944, 2014.

[26] HAROLD Hotelling. Relations between two sets of variates. *Biometrika*, 28:321–377, 1936.

[27] Sophie Huhn, Miriam Axt, Hanns-Christian Gunga, Martina Anna Maggioni, Stephen Munga, David Obor, Ali Sié, Valentin Boudo, Aditi Bunker, Rainer Sauerborn, et al. The impact of wearable technologies in health research: scoping review. *JMIR mHealth and uHealth*, 10(1):e34384, 2022.

[28] Mingon Kang, Euiseong Ko, and Tesfaye B Mersha. A roadmap for multi-omics data integration using deep learning. *Briefings in Bioinformatics*, 23(1):bbab454, 2022.

[29] Zhao Kang, Guoxin Shi, Shudong Huang, Wenyu Chen, Xiaorong Pu, Joey Tianyi Zhou, and Zenglin Xu. Multi-graph fusion for multi-view spectral clustering. *Knowledge-Based Systems*, 189:105102, 2020.

[30] Jon R Kettenring. Canonical analysis of several sets of variables. *Biometrika*, 58(3):433–451, 1971.

[31] Tae-Kyun Kim and Roberto Cipolla. Canonical correlation analysis of video volume tensors for action categorization and detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(8):1415–1428, 2008.

[32] Dana Lahat, Tülay Adali, and Christian Jutten. Multimodal data fusion: an overview of methods, challenges, and prospects. *Proceedings of the IEEE*, 103(9):1449–1477, 2015.

[33] Pei Ling Lai and Colin Fyfe. Kernel and nonlinear canonical correlation analysis. *International journal of neural systems*, 10(05):365–377, 2000.

[34] Roberto M. Lang, Luigi P. Badano, Victor Mor-Avi, Jonathan Afilalo, Alison Armstrong, Laurent Ernande, Frank A. Flachskampf, Elyse Foster, Steven A. Goldstein, Tatiana Kuznetsova, Patrizio Lancellotti, Denisa Muraru, Myriam H. Picard, Ernst R. Rietzschel, Lawrence G. Rudski, Kirk T. Spencer, Teresa S. M. Tsang, and Jens-Uwe Voigt. Recommendations for cardiac chamber quantification by echocardiography in adults: An update from the american society of echocardiography and the european association of cardiovascular imaging. *European Heart Journal - Cardiovascular Imaging*, 16(3):233–270, 2015.

[35] Yingming Li, Ming Yang, and Zhongfei Zhang. A survey of multi-view representation learning. *IEEE transactions on knowledge and data engineering*, 31(10):1863–1883, 2018.

[36] Zachary C Lipton. The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue*, 16(3):31–57, 2018.

[37] Cheng Liu, Wenming Cao, Si Wu, Wenjun Shen, Dazhi Jiang, Zhiwen Yu, and Hau-San Wong. Supervised graph clustering for cancer subtyping based on survival analysis and integration of multi-omic tumor data. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 19(2):1193–1202, 2020.

[38] Feifan Liu, Chunhua Weng, and Hong Yu. Natural language processing, electronic health records, and clinical research. *Clinical Research Informatics*, pages 293–310, 2012.

[39] Eric F. Lock, Katherine A. Hoadley, J. S. Marron, and Andrew B. Nobel. Joint and individual variation explained (jive) for integrated analysis of multiple data types. *Annals of Applied Statistics*, 7(1):523–542, 2013.

[40] Tianle Ma and Aidong Zhang. Integrate multi-omic data using affinity network fusion (anf) for cancer patient clustering. In *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 398–403. IEEE, 2017.

[41] Adil Mardinoglu, Hasan Turkez, Minho Shong, Vishnuvardhan Pogunulu Srinivasulu, Jens Nielsen, Bernhard O Palsson, Leroy Hood, and Mathias Uhlen. Longitudinal big biological data in the ai era. *Molecular Systems Biology*, 21(9):1147, 2025.

[42] Chen Meng, Oana A Zeleznik, Gerhard G Thallinger, Bernhard Kuster, Amin M Gholami, and Aedín C Culhane. Dimension reduction techniques for the integrative analysis of multi-omics data. *Briefings in bioinformatics*, 17(4):628–641, 2016.

[43] Stefano Monti, Pablo Tamayo, Jill Mesirov, and Todd Golub. Consensus clustering: a resampling-based method for class discovery and visualization of gene expression microarray data. *Machine learning*, 52(1):91–118, 2003.

[44] Victor Mor-Avi, Roberto M Lang, Luigi P Badano, Marek Belohlavek, Nuno Miguel Cardim, Genevieve Derumeaux, Maurizio Galderisi, Thomas Marwick, Sherif F Nagueh, Partho P Sengupta, et al. Current and evolving echocardiographic techniques for the quantitative evaluation of cardiac mechanics: Ase/eae consensus statement on methodology and indications endorsed by the japanese society of echocardiography. *European Journal of Echocardiography*, 12(3):167–205, 2011.

[45] Ion Muslea, Steven Minton, and Craig A Knoblock. Active learning with multiple views. *Journal of Artificial Intelligence Research*, 27:203–233, 2006.

[46] Milan Picard, Marie-Pier Scott-Boyer, Antoine Bodein, Olivier Périn, and Arnaud Droit. Integration strategies of multi-omics data for machine learning analysis. *Computational and Structural Biotechnology Journal*, 19:3735–3746, 2021.

[47] Erica Ponzi, Magne Thoresen, Therese Haugdahl Nøst, and Kajsa Møllersen. Integrative, multi-omics, analysis of blood samples improves model predictions: applications to cancer. *BMC bioinformatics*, 22:1–17, 2021.

[48] Tiffany M. Powell-Wiley, Paul Poirier, Lauren E. Burke, Jean-Pierre Després, Penny Gordon-Larsen, Carl J. Lavie, Scott A. Lear, Chike Ndumele, Ian J. Neeland, Paul Sanders, Marie-Pierre St-Onge, American Heart Association Council on Lifestyle, Cardiometabolic Health, Council on Cardiovascular, Stroke Nursing, Council on Clinical Cardiology, Council on Hypertension, and Stroke Council. Obesity and cardiovascular disease: A scientific statement from the american heart association. *Circulation*, 143(21):e984–e1010, 2021.

[49] Nimrod Rappoport and Ron Shamir. Multi-omic and multi-view clustering algorithms: review and cancer benchmark. *Nucleic acids research*, 46(20):10546–10562, 2018.

[50] Parminder S Reel, Smarti Reel, Ewan Pearson, Emanuele Trucco, and Emily Jefferson. Using machine learning approaches for multi-omics data analysis: A review. *Biotechnology advances*, 49:107739, 2021.

[51] Cynthia Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature machine intelligence*, 1(5):206–215, 2019.

[52] Martijn Schouteden, Katrijn Van Deun, Sven Pattyn, and Iven Van Mechelen. Sca with rotation to distinguish common and distinctive information in linked data. *Behavior research methods*, 45:822–833, 2013.

[53] Mohamad H Shahrajabian and Wenli Sun. Survey on multi-omics, and multi-omics data analysis, integration and application. *Current Pharmaceutical Analysis*, 19(4):267–281, 2023.

[54] Ronglai Shen, Adam B Olshen, and Marc Ladanyi. Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis. *Bioinformatics*, 25(22):2906–2912, 2009.

[55] Pasquale Sibilio, Enrico De Smaele, Paola Paci, and Federica Conte. Integrating multi-omics data: Methods and applications in human complex diseases. *Biotechnology Reports*, page e00938, 2025.

[56] Vikas Sindhwani and David S Rosenberg. An rkhs for multi-view learning and manifold co-regularization. In *Proceedings of the 25th international conference on Machine learning*, pages 976–983, 2008.

[57] Amrit Singh, Benoit Gautier, Casey P Shannon, Michaël Vacher, Florian Rohart, Scott J Tebbutt, and Kim-Anh Lê Cao. Diablo–an integrative, multi-omics, multivariate method for multi-group classification. *BioRxiv*, page 067611, 2016.

[58] Otto A. Smiseth, Hans Torp, Anders Opdahl, Kristina H. Haugaa, and Stig Urheim. Myocardial strain imaging: how useful is it in clinical decision making? *European Heart Journal*, 37(15):1196–1207, 2016.

[59] Hang Su, Subhransu Maji, Evangelos Kalogerakis, and Erik Learned-Miller. Multi-view convolutional neural networks for 3d shape recognition. In *Proceedings of the IEEE international conference on computer vision*, pages 945–953, 2015.

[60] Indhupriya Subramanian, Srikant Verma, Shiva Kumar, Abhay Jere, and Krishanpal Anamika. Multi-omics data integration, interpretation, and its application. *Bioinformatics and biology insights*, 14:1177932219899051, 2020.

[61] Shiliang Sun. Multi-view laplacian support vector machines. In *International Conference on Advanced Data Mining and Applications*, pages 209–222. Springer, 2011.

[62] Shiliang Sun. A survey of multi-view machine learning. *Neural computing and applications*, 23:2031–2038, 2013.

[63] Arthur Tenenhaus, Cathy Philippe, Vincent Guillemot, Kim-Anh Le Cao, Jacques Grill, and Vincent Frouin. Variable selection for generalized canonical correlation analysis. *Biostatistics*, 15(3):569–583, 2014.

[64] Giulia Tini, Luca Marchetti, Corrado Priami, and Marie-Pier Scott-Boyer. Multi-omics integration—a comparison of unsupervised clustering methodologies. *Briefings in bioinformatics*, 20(4):1269–1279, 2019.

[65] Luan Tran, Xi Yin, and Xiaoming Liu. Disentangled representation learning gan for pose-invariant face recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1415–1424, 2017.

[66] Suraj Verma, Giuseppe Magazzu, Noushin Eftekhari, Thai Lou, Alex Gilhespy, Annalisa Occhipinti, and Claudio Angione. Cross-attention enables deep learning on limited omics-imaging-clinical data of 130 lung cancer patients. *Cell reports methods*, 4(7), 2024.

[67] Seppo Virtanen, Arto Klami, Suleiman Khan, and Samuel Kaski. Bayesian group factor analysis. In *Artificial Intelligence and Statistics*, pages 1269–1277. PMLR, 2012.

[68] Bo Wang, Aziz M Mezlini, Feyyaz Demir, Marc Fiume, Zhuowen Tu, Michael Brudno, Benjamin Haibe-Kains, and Anna Goldenberg. Similarity network fusion for aggregating data types on a genomic scale. *Nature methods*, 11(3):333–337, 2014.

[69] Weiran Wang, Raman Arora, Karen Livescu, and Jeff Bilmes. On deep multi-view representation learning. In *International conference on machine learning*, pages 1083–1092. PMLR, 2015.

[70] Ziwei Wei, Dunsheng Han, Cong Zhang, Shiyu Wang, Jinke Liu, Fan Chao, Zhenyu Song, and Gang Chen. Deep learning-based multi-omics integration robustly predicts relapse in prostate cancer. *Frontiers in oncology*, 12:893424, 2022.

[71] 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.

[72] Paul K. Whelton, Robert M. Carey, Wilbert S. Aronow, Donald E. Casey, Karen J. Collins, Cheryl Dennison Himmelfarb, Samuel M. DePalma, Samuel Gidding, Kenneth A. Jamerson, Daniel W. Jones, Eric J. MacLaughlin, Paul Muntner, Bruce Ovbiagele, Sidney C. Smith, Christopher C. Spencer, Randall S. Stafford, Sandra J. Taler, Ryan J. Thomas, Kim A. Williams, Jeffery D. Williamson, and Jackson T. Wright. 2017 acc/aha/aapa/abc/acpm/ags/apha/ash/aspc/nma/pcna guideline for the prevention, detection, evaluation, and management of high blood pressure in adults. *Journal of the American College of Cardiology*, 71(19):e127–e248, 2018.

[73] Chang Xu, Dacheng Tao, and Chao Xu. A survey on multi-view learning. *arXiv preprint arXiv:1304.5634*, 2013.

[74] Zhijie Xu and Shiliang Sun. An algorithm on multi-view adaboost. In *Neural Information Processing. Theory and Algorithms: 17th International Conference, ICONIP 2010, Sydney, Australia, November 22-25, 2010, Proceedings, Part I 17*, pages 355–362. Springer, 2010.

[75] Zhijie Xu and Shiliang Sun. Multi-view transfer learning with adaboost. In *2011 IEEE 23rd International Conference on Tools with Artificial Intelligence*, pages 399–402. IEEE, 2011.

[76] Zhijie Xu and Shiliang Sun. Multi-source transfer learning with multi-view adaboost. In *Neural Information Processing: 19th International Conference, ICONIP 2012, Doha, Qatar, November 12-15, 2012, Proceedings, Part III 19*, pages 332–339. Springer, 2012.

[77] Fei Xue, Xin Wu, Shaojun Cai, and Junqiu Wang. Learning multi-view camera relocalization with graph neural networks. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11372–11381. IEEE, 2020.

[78] Xiaoqiang Yan, Shizhe Hu, Yiqiao Mao, Yangdong Ye, and Hui Yu. Deep multi-view learning methods: A review. *Neurocomputing*, 448:106–129, 2021.

[79] Yan Yang and Hao Wang. Multi-view clustering: A survey. *Big data mining and analytics*, 1(2):83–107, 2018.

[80] Nan Zhang and Shiliang Sun. Multiview graph restricted boltzmann machines. *IEEE Transactions on Cybernetics*, 52(11):12414–12428, 2021.

[81] Yilin Zhang, Shuo Chen, Rozalina G Mccoy, Chixiang Chen, and Yuzhou Chen. Multi-view k-nearest neighbor graph contrastive learning on multi-modal biomedical data. In *International Conference on Artificial Intelligence in Medicine*, pages 238–248. Springer, 2024.

[82] Jing Zhao, Xijiong Xie, Xin Xu, and Shiliang Sun. Multi-view learning overview: Recent progress and new challenges. *Information Fusion*, 38:43–54, 2017.

[83] Marinka Zitnik, Francis Nguyen, Bo Wang, Jure Leskovec, Anna Goldenberg, and Michael M Hoffman. Machine learning for integrating data in biology and medicine: Principles, practice, and opportunities. *Information Fusion*, 50:71–91, 2019.

MOX Technical Reports, last issues

Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

81/2025 Leimer Saglio, C. B.; Pagani, S.; Antonietti, P. F.
A massively parallel non-overlapping Schwarz preconditioner for PolyDG methods in brain electrophysiology

82/2025 Varetti, E.; Torzoni, M.; Tezzele, M.; Manzoni, A.
Adaptive digital twins for predictive decision-making: Online Bayesian learning of transition dynamics

79/2025 Zacchei, F.; Conti, P.; Frangi, A.; Manzoni, A.
Multi-Fidelity Delayed Acceptance: hierarchical MCMC sampling for Bayesian inverse problems combining multiple solvers through deep neural networks

78/2025 Botti, M.; Mascotto, L.
Trace inequalities for piecewise W^1,p functions over general polytopic meshes

Botti, M.; Mascotto, L.
Trace inequalities for piecewise W^{1,p} functions over general polytopic meshes

77/2025 Bonetti, S.; Botti, M.; Vega, P.
A robust fully-mixed finite element method with skew-symmetry penalization for low-frequency poroelasticity

75/2025 Crippa, B.; Scotti, A.; Villa, A.
A one-dimensional reduced plasma model for the electrical treeing

74/2025 Colombo, S.; Gimenez Zapiola, A.; Ieva, F.; Vantini, S.
Multi-state Modeling of Delay Evolution in Suburban Rail Transports

73/2025 Antonietti, P.F.; Bertoluzza, S.; Credali, F.
The Reduced Basis Multigrid scheme for the Virtual Element Method.

Antonietti, P.F.; Bertoluzza, S.; Credali, F.
The Reduced Basis Multigrid scheme for the Virtual Element Method