

# Navigating the Multiverse: A Hitchhiker's Guide to Selecting Harmonisation Methods for Multimodal Biomedical Data

Murali Aadhitya Magateshvaren Saras<sup>1,2,3</sup>, Mithun K. Mitra<sup>2</sup>,  
Sonika Tyagi<sup>3,4\*</sup>

<sup>1</sup> IITB-Monash Research Academy, Mumbai, Maharashtra, India,  
400076 .

<sup>2</sup> Indian Institute of Technology Bombay, Mumbai, Maharashtra, India,  
400076 .

<sup>3</sup> Central Clinical School, Monash University, Melbourne, Victoria,  
Australia, 3181 .

<sup>4</sup> School of Computational Technologies, RMIT University, Melbourne,  
Victoria, Australia, 3000 .

\*Corresponding author(s). E-mail(s): [sonika.tyagi@monash.edu](mailto:sonika.tyagi@monash.edu);

## Abstract

**Introduction:** The application of machine learning (ML) techniques in classification and prediction tasks has greatly advanced our comprehension of biological systems. There is a notable shift in the trend towards integration methods that specifically target the simultaneous analysis of multiple modes or types of data, showcasing superior results compared to individual analyses. Despite the availability of diverse ML architectures for researchers interested in embracing a multimodal approach, the current literature lacks a comprehensive taxonomy that includes the pros and cons of these methods to guide the entire process. Closing this gap is imperative, necessitating the creation of a robust framework. This framework should not only categorise the diverse ML architectures suitable for multimodal analysis but also offer insights into their respective advantages and limitations. Additionally, such a framework can act as a guide for selecting an appropriate workflow for multimodal analysis. This comprehensive taxonomy would furnish a clear guidance and aid in informed decision-making within the progressively intricate realm of biomedical and clinical data analysis, and is imperative for advancing personalised medicine.

**Objective:** The aims of the work are to comprehensively study and describe the

harmonisation processes that are performed and reported in the literature and present a working guide that would enable planning and selecting an appropriate integrative model.

**Methods:** A systematic review of publications that report the multimodal harmonisation of biomedical and clinical data has been performed.

**Results:** We present harmonisation as a dual process of representation and integration, each with multiple methods and categories. The taxonomy of the various representation and integration methods are classified into six broad categories and detailed with the advantages, disadvantages and examples. A guide flowchart that describes the step-by-step processes that are needed to adopt a multimodal approach is also presented along with examples and references.

**Conclusions:** This review provides a thorough taxonomy of methods for harmonising multimodal data and introduces a foundational 10-step guide for newcomers to implement a multimodal workflow.

**Keywords:** multimodal integration, feature representation, data integration, deep learning, digital health

# 1 Introduction

The growth of biological and healthcare data, in terms of volume, velocity and variety, has been exponential and driven by technological advances in electronics, communication and infrastructure (Laney (2001); Dash et al. (2019)). Concurrently, there has been an increase in data analysis tools to understand and analyse the data. Progress in computational techniques, artificial intelligence (AI) and machine learning (ML) methods have been identified to contribute towards the analysis and interpretation better than traditional analytical methods (Acosta et al. (2022); Li and Ngom (2015)).

Data generated in the context of biological systems can manifest in various forms such as quantitative, qualitative or narrative; each of these has its subtypes, which are collectively referred to as a ‘modality’. These diverse modalities can capture several aspects of a biological system, such as nucleic acid and protein sequences (Neidle (2008)), gene expression (Raghavachari and Garcia-Reyero (2018)) and the biomolecular structure and its activity (Vergoten and Theophanides (2012)). Other modalities include the epigenetic state and methylation information (Paro et al. (2021)) of the genome, metabolites, and anatomic and phenotypic data.

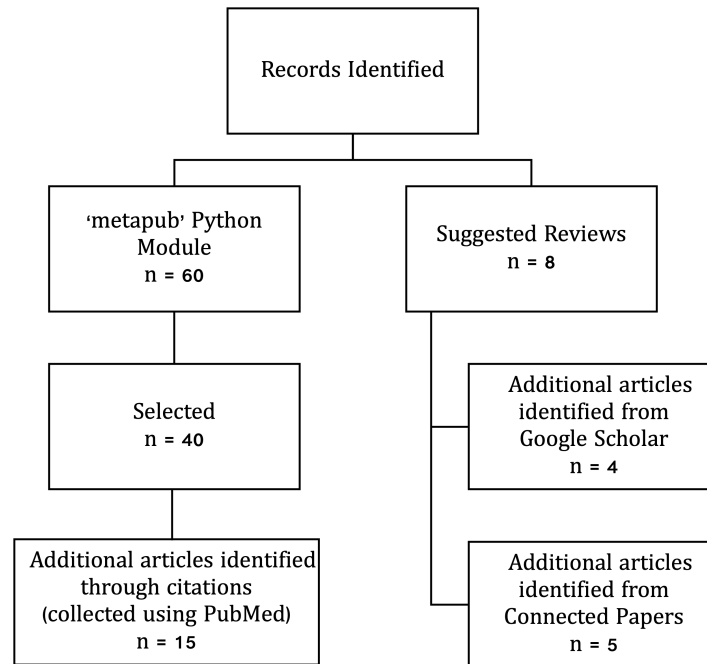
Each data type has driven research towards elucidating the corresponding functional aspects to understand the system. Numerous studies using a single data modality have presented valuable additions to the literature in disease mapping, pathway and network elucidation (Aburajab et al. (2023); Mansuri et al. (2023); Pang et al. (2023)). However, a vast portion of the biological complexity still requires an explanation, which is an ongoing challenge for the research community.

Different modalities capture different aspects of the system. Thus, integrating them provides a comprehensive multi-view understanding of both biological and clinical conditions (Li and Ngom (2015); Nie et al. (2007)). Combining multiple types of omics data or a ‘multiomics’ approach to study biological systems has gained momentum lately due to their demonstrated superiority over single-omics approaches (Chen et al. (2021); Chen and Tyagi (2020); Baltrušaitis et al. (2018); Acosta et al. (2022); Summaira et al. (2021)). Furthermore, healthcare data is integrated with omics datasets to reveal their interconnections, providing a comprehensive 360-degree view of an individual’s condition. (Schiano et al. (2020); Dargazanli et al. (2020)). Such studies have reported results with significant validation and reliability over independent analysis. Thus, the integration of multiple modalities can reveal synergistic effects, where the combined information enhances the overall performance of the model beyond what individual modalities can achieve.

Existing literature primarily focuses on the model architecture and merits of the general ML methods used for analysis (Li and Ngom (2015); Summaira et al. (2021); Sapoval et al. (2022)). However, a gap still persists in delineating between integrated learning and co-learning (harmonisation). Harmonisation aims to elucidate the low-level relationship between features of different modalities (Baltrušaitis et al. (2018)). Often, articles incorporating ‘integration’ as part of their pipeline do not necessarily perform a harmonisation process, instead they focus on the correlation between individual data type analysis (Chen et al. (2021)). The effect of an analysis using multiple modalities is not adequately captured by methods that do not harmonise the features. A co-learning set-up is distinct from an individual analysis since it necessitates fusion

46 of features. Published articles relay information to mitigate data challenges but lack  
47 information on the multimodal process (Mirza et al. (2019); Kline et al. (2022); Acosta  
48 et al. (2022); Zitnik et al. (2019)). A definitive explanation of the methods involved in  
49 a multimodal harmonisation approach is missing. The absence of adequate informa-  
50 tion impedes interested researchers from fully grasping the process and implementing  
51 a workflow. In summary, this literature review addresses the aforementioned gap by  
52 offering insights into data modalities, challenges encountered in data, the processes  
53 involved in a multimodal setup, and a beginner’s guide to multimodal analysis.

## 54 2 Methods



**Fig. 1:** Flowchart of literature screening. The blocks on the left indicate articles searched for representation and integration methods. The blocks on the right describe the targeted search for review articles on biomedical multimodal harmonisation.

55 This systematic review was performed based on the standards of the Preferred  
56 Reporting Items for Systematic Reviews and Meta-analyses (PRISMA 2020) statement  
57 (Page et al. (2021)).

58 Based on existing literature reviews published over the past decade, a general out-  
59 line was followed to select articles that mentioned multimodal learning techniques.  
60 An extensive search of various ML methods focused on biomedical data was initially  
61 gathered using the metapub (<https://github.com/metapub/metapub>) python module

62 based on the following keywords:

63 *'Multimodal', 'Machine learning', 'Integrated learning', 'Multiomics', 'Genomics',*  
64 *'Proteomics', 'Biomedical', 'Healthcare data', 'Biological network', 'Deep learning',*  
65 *'Multitask learning', 'Data fusion', 'Representation', 'Interpretable model', 'Neural*  
66 *network'.*

67 The title and the abstract of the search results reported by the keywords using  
68 the python script was used to select papers for complete reading. A few suggested  
69 articles from co-authors on were used to initiate the search for reviews on biomedical  
70 multimodal integration. Citations from suggested papers were manual searched using  
71 Google Scholar and PubMed. Connected Papers (<https://www.connectedpapers.com>)  
72 was used to identify related articles. We identified less than 10 articles that described  
73 on the topic. Few papers were selected for reading on general multimodal integration  
74 methods. Figure 1 describes the count of articles.

75 The inclusion criteria for this review primarily focused on studies that incorporated  
76 multiple different types of biological and medical data towards a singular analysis  
77 using machine learning algorithms. Reports that did not use biomedical datasets but  
78 employed a multimodal approach for data analysis were also included for review. The  
79 exclusion criteria was marked by the absence of a multimodal approach only. However,  
80 studies that focused on representation methods of different data types were included  
81 for full-text reading. The last date of article search and selection was 20 October 2023.

82 The methods for selected articles were reviewed in detail and information on the  
83 data type, machine learning framework, model advantages, research gaps were col-  
84 lected. They have been classified into groups and presented in tabular format (Tables  
85 1, 4) and the results are discussed in following sections.

## 86 **3 Results**

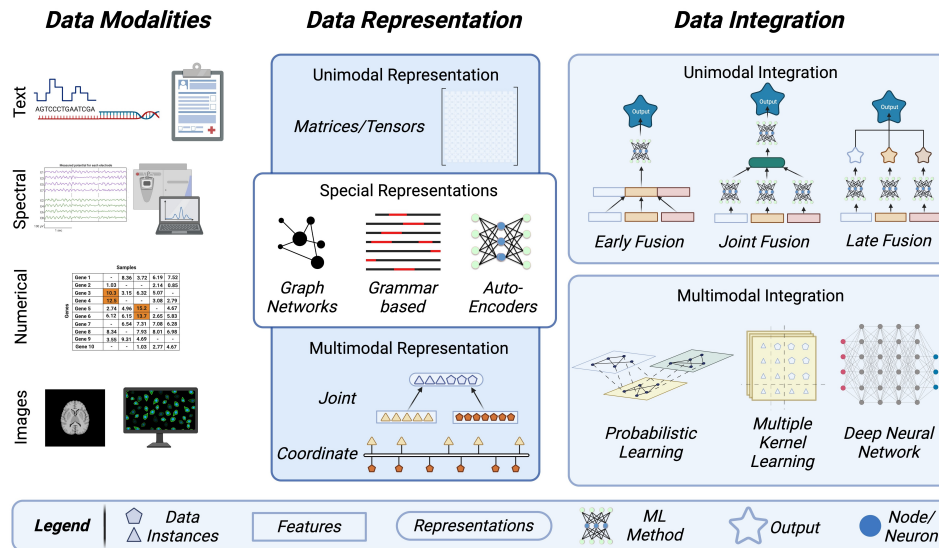
87 In the following sections, we integrate our findings regarding data types, techni-  
88 cal hurdles, and harmonisable methodologies gleaned from the literature we have  
89 examined.

### 90 **3.1 Typical Study Designs in biological and clinical studies**

91 All data generated are usually based on a study or an objective that has a focused  
92 rationale. The richness of information in collected or generated data depends on its  
93 type, determining the range of possible analyses ([Ranganathan and Aggarwal \(2019\)](#)).

94 A static study design acquires data as a 'snapshot', that is, collected at a point  
95 in time. Case series refer to static data collected from positive-group criteria within a  
96 population subgroup, while case-control studies include a negative dataset (controls)  
97 for comparison. On the other hand, cohort studies and randomised controlled trials  
98 sample data over a period of time, capturing the dynamic nature of a biological system,  
99 which allows for a realistic investigation. However, they are resource-intensive methods  
100 that must be maintained regularly, and constant follow-up with the subjects considered  
101 in the study is crucial.

102 Many efforts are being taken to enhance the data collection methods and acces-  
103 sibility across various domains, such as in cancer ( The Cancer Genome Atlas,



**Fig. 2:** An illustration briefly depicting the broad categories of data modalities and the representation and integration methods used in a multimodal harmonisation analysis. The representation methods are split into three groups based on the number and type of datasets. The integration methods are split based on the type of fusion performed. Made with BioRender.

104 <https://www.cancer.gov/ccg/>) and preterm birth (Garbh-Ini). Current literature pre-  
 105 dominantly reports on results based on single, static datasets. Correlation studies use  
 106 multiple datasets to support conclusions through overlapping results (Clarke et al.  
 107 (2017)). Only a few methods take a complementary approach between modalities  
 108 (Welch et al. (2017); Chen et al. (2021)).

### 109 3.2 Common data modalities studied in the literature

110 Data can be collected in forms such as text, numbers, and multimedia. Based on  
 111 the sources, they can be classified as ‘biological’ data and ‘health’ data (Dash et al.  
 112 (2019)). We refer to ‘biological data’ as information from high-throughput experiments  
 113 such as sequencing, expression profiling, microscopic imaging and the vast literature  
 114 corpus for functional annotation. This also includes metadata related to samples,  
 115 experimental design, assay protocols and technologies. ‘Health’ sources refer to data  
 116 primarily collected through healthcare providers in digital forms. This data contains  
 117 an individual’s valuable health and medical history and is stored as time-stamped  
 118 electronic medical records (EMR). EMR data displays significant structural diversity  
 119 since it can be structured data, including vital signs and pathology measurements  
 120 organised in tabular formats, or presented as unstructured data, consisting of clinical  
 121 notes, images, and documents. Table 1 describes the different modalities that stem  
 122 from clinical and biomedical sources.

### 123 **3.2.1 Text Modality**

124 Text as a modality comprises various types, encompassing narrative and sequence  
125 forms of data. Sequence data stemming from biological macromolecules such as DNA,  
126 RNA, and proteins describe and define the relationship between the genotype and the  
127 phenotype of an organism. Differences in sequences among groups differing in demog-  
128 raphy or phenotype are represented as single nucleotide polymorphism (SNP) or small  
129 insertions and deletions of DNA bases and are generated by genome-wide association  
130 studies (GWAS) ([Uffelmann et al. \(2021\)](#)). Information on motifs, interaction net-  
131 works and annotations about biomolecules, drugs and diseases belong to this class.  
132 Healthcare data, such as EMR, contain unstructured clinical notes and prescriptions  
133 manually entered by medical practitioners, which are included in the text category of  
134 datasets ([Lima et al. \(2019\)](#)).

### 135 **3.2.2 Spectral and Signal Modality**

136 Spectral data, typically acquired through mass spectrometry experiments to study  
137 protein molecules and metabolites, provides detailed insights into the structural com-  
138 position, constitution and organisation of the molecules under investigation ([Mansuri  
139 et al. \(2023\)](#); [Mou et al. \(2022\)](#)). ML analysis of spectral data involves features  
140 representing three-dimensional conformations and spatial relationships of molecules,  
141 enabling classification based on functional groups and elements ([Mou et al. \(2022\)](#);  
142 [Sachdev and Gupta \(2019\)](#)). Proteomics, examining proteins through expression, func-  
143 tional relationships, and structural information, includes investigations into protein  
144 folding and structural orientations using methods such as NMR and X-ray crystal-  
145 lography ([Malet-Martino and Holzgrabe \(2011\)](#)). Structural metabolomics stores the  
146 structural data collected from metabolites.

147 Healthcare data in spectral form includes time-dependent Electroencephalogram  
148 (EEG) and electrocardiogram (ECG) analysed with signal processing methods ([Subha  
149 et al. \(2010\)](#); [Abarbanel et al. \(2009\)](#)). Audio data, such as voice notes, undergoes  
150 analysis using appropriate methods after feature extraction ([Camastra and Vinciarelli  
151 \(2015\)](#)).

### 152 **3.2.3 Numerical Modality**

153 A numerical form of biological data can be from any quantifiable assay, broadly called  
154 ‘omics’ data. Transcriptomics represents digital counts of identified expressed tran-  
155 script molecules, available as a two-dimensional (2D) matrix of genes/transcripts and  
156 samples. Similarly, proteomic, lipidomic and metabolomic counts data portray the  
157 expression levels of proteins, lipids or metabolite molecules as a 2D data matrix. This  
158 modal information allows understanding of individual differences regarding genetic  
159 expression and linking related biological pathways. EMR readings document vital and  
160 pathological parameters such as heart rate, weight, height, age, and blood pressure as  
161 numerical time-series data. Frequent time-stamped EMRs enable longitudinal analysis,  
162 capturing changes in the recorded values over time ([Haghverdi et al. \(2016\)](#)).

Type	Data Source	Features of Interest	Reference
Text	DNA, RNA and Protein Sequence	Sequence order and motifs	(Tyagi et al. (2008); Andrikos et al. (2022))
	GWAS Data	Genetic variants	(Chang et al. (2018))
	Clinical Notes	Correlated medical terms and phrases	(Banerjee et al. (2019))
Spectral and signal	X-Ray crystallography, NMR, Mass Spectroscopy	Structural composition and identified functional groups, topology	(Mou et al. (2022))
	Audio signals	Speech to text patterns	(Summaira et al. (2021))
Numerical	Biomolecular profiles (Lipids, Metabolites, Nucleic acids, Proteins)	Expression levels of biomolecules	(Zitnik et al. (2019))
	EMR (Vitals, Lab measurements)	Health factors, trends and trajectories	(Banerjee et al. (2019))
	Interaction Networks (Diseases, drugs, genes, proteins)	Regulatory and Functional relationships	(Lee et al. (2020))
Images	EMR (CT, X-ray, Ultrasound)	Patterns and localisations	(Zhao et al. (2021))
	Cell Imaging	Patterns and localisations	(Schiano et al. (2020))

**Table 1:** The four distinct modalities biological and clinical data investigated in this study are listed and categorised based on their sources and the features targeted for modelling.

### 163 3.2.4 Image Modality

164 This modality encompasses visual information, including images and videos. Videos  
165 are also considered under this modality because each frame can be considered an image  
166 for processing but in a time-dependent manner (Camastra and Vinciarelli (2015)).  
167 Microscopy and cell imaging data are often analysed for morphology studies, protein  
168 localisation and DNA tagging (Fu and Rui (2017)). Manual image analysis methods  
169 include segmentation tasks to identify regions of interest and cell morphology assess-  
170 ment (Kan (2017)). Cell movement and tracking studies creating animated clips from  
171 multiple fluorescence-tagged cell images are a modality of this category. Additionally,  
172 X-rays, CT and MRI images from EMRs supporting non-invasive diagnosis fall in this  
173 category.



### 174 3.3 Common challenges associated with biomedical data

175 Datasets require extensive pre-processing due to incompleteness and imperfections  
176 before analysis (López de Maturana et al. (2019)). Key challenges include high  
177 dimensionality, heterogeneity, missing data, class imbalances, bias and accessibility.

178 Complex high-dimensional data, characterised by large features and file sizes,  
179 require extensive computational resources for understanding the variables (Stephens  
180 et al. (2015)). Addressing the ‘ $p \gg n$ ’ problem, considering the ratio of available sam-  
181 ples ( $n$ ) to features ( $p$ ), is crucial to prevent specific features from being overlooked in  
182 small sample groups (Baltrušaitis et al. (2018); Stephens et al. (2015)).

183 Bias refers to a variety of imbalances found within a dataset and can lead to an  
184 unfair interpretation of results. Bias can manifest in various forms, such as representa-  
185 tion bias or class imbalance, measurement bias (due to incorrect or unrelated values),  
186 aggregation bias (when models are applied to new datasets with a mutually exclusive  
187 relationship to training samples) and evaluation bias (when generic models are used  
188 as benchmarks for targeted datasets) (Suresh and Guttag (2021)). Comparative ana-  
189 lytical methods utilise representative datasets, subsets of the population with samples  
190 from distinct groups like case and control. It is crucial for groups to have samples in  
191 a comparable and an equivalent number to understand the true differences.

192 Irregular clinical data collection processes lead to inconsistent data entries and  
193 contribute to missing data. In datasets, all samples may not provide data for all  
194 possible features, and the resulting matrix could be sparse in a few cases. Importantly,  
195 a missing measurement may carry meaning and should be considered subjectively.

196 Heterogeneity refers to the variety that exists within and across modalities. Within  
197 a modality, the data collected across variables can vary in terms of scale, distribution  
198 and recorded value, such as discrete, continuous, categories and intervals, due to non-  
199 standardised procedures. For example, clinical and genomic data can not be directly  
200 compared and analysed, requiring methods to address heterogeneity.

201 These challenges obstruct the potential in any analysis, but the problem exac-  
202 erbates when multiple datasets are involved in a multimodal set-up (Zhang et al.  
203 (2019)). Multimodal methods are affected by coherence between dataset sets (due to  
204 heterogeneity and missing data), accessibility and computational resources (due to  
205 high-dimensional datasets and bias).

206 Data preprocessing steps prepare a path to check, sort, and select data points so  
207 that informed decisions can be taken to handle samples with anomalies and poor qual-  
208 ity. The lack of data standardisation between multiple collection sites poses a challenge  
209 for seamless data harmonisation (Ramakrishnaiah et al. (2023)) and requires specific  
210 preprocessing for different sources. While imputation methods partially handle missing  
211 data, they are not universal solutions, as approximations may not accurately reflect  
212 the system (Schafer (1997)). Hence, more data-driven approaches may be adopted in  
213 different scenarios (Ramakrishnaiah et al. (2023)). Large datasets can be converted  
214 to latent values to reduce computational load. Representation methods (Section 3.5)  
215 effectively resolve these issues.

216 Biomedical and health data containing personal and sensitive information are  
217 restricted for global access, which limits the extent and scope of analysis. Implement-  
218 ing ethical and legal data practices, both nationally and internationally, is crucial for

219 easing the data sharing process (Tyagi (2023)). These practices establish a structured  
220 and transparent approach to handling data, creating an environment conducive to  
221 sharing valuable information.

### 222 **3.4 Multimodal Data Integration: a systematic investigation of** 223 **data formats and methodologies**

224 Multimodal analytical methods aim to combine information from multiple modalities  
225 towards one or many of the following goals: 1) Explain a biological phenomenon or  
226 phenotype through overlapping results. 2) Account for and impute missing data in one  
227 modality through another linked dataset/modality. 3) Condense the high-dimensional,  
228 sparse and noisy data to a low-dimensional latent representation.

229 The fundamental difference between an unimodal and a multimodal analysis is the  
230 number of different modalities used. The complete dataset can be directly fed into  
231 an ML architecture for an unimodal analysis, but a multimodal approach requires  
232 the fusion of features from multiple datasets. The process of merging and modelling  
233 of features can be classified under ‘representation’ and ‘integration’. The choice of  
234 method varies depending on the task to be achieved and the dataset combinations.  
235 Section 3.5 describes the different ways of feature representation, and section 3.6 briefs  
236 about the fusion methods currently used with examples.

### 237 **3.5 Data Representation**

238 As discussed earlier, biomedical and health data is generated in many forms (Table  
239 1). Data representation methods are crucial as they transform diverse data types  
240 into machine-processable formats such as vectors, matrices, or tensors. Vectors are  
241 one-dimensional representations of numerical values, while matrices and tensors hold  
242 data in 2-dimensional and multidimensional scales. These methods keep track of rela-  
243 tionships between elements of each modality via predefined rules, facilitate feature  
244 extraction by using relevant information, and help in mapping data from one modality  
245 to another. Importantly, the representation methods can be modified to suit the study  
246 conducted (Sapoval et al. (2022)). In this context, we have classified three groups of  
247 data representation approaches.

#### 248 **3.5.1 Unimodal Data Representation**

249 Unimodal data representation involves using a single mode or source of information  
250 to represent features. Each modality qualifies as an unimodal representation when  
251 independently transformed into a numerical format through an encoding or embedding  
252 approach.

253 Encoding involves the conversion of original data into a numerical format, whereas  
254 embedding refers to portraying the original data in a vector space that incorpo-  
255 rates semantic information. The information from biological sequence and text can  
256 be encoded by converting them to a numerical representation based on composition  
257 (tallying frequency of words/monomers), K-mers (segmenting a biological sequence as  
258 a window of ‘k’ letters) and distribution (percentage of occurrence of each monomer  
259 within user-defined ranges) of the sequence (Yang et al. (2020)). K-mers serve as the

260 counterparts to n-grams or tokens in NLP methods, and ongoing efforts are focused  
261 on developing more advanced, data-driven approaches to derive them from sequen-  
262 tial data (Chen et al. (2023)). Many tools have been created to embed text data  
263 into a numerical representation, such as word2vec (Church (2017)) or doc2vec (Lau  
264 and Baldwin (2016)), which preserves the order of information and local neighbouring  
265 relationships.

266 Numerical data, obtained as-is or in other formats is commonly represented as  
267 matrices for analysis purposes. Time-series information is represented as tensors, where  
268 the data is nested within matrices, extending in dimensions to include the tempo-  
269 ral relationship (Zhu et al. (2021)). Similarly, image data is converted to a matrix  
270 representation by splitting a digital pixel into a numerical value between 0-255 for  
271 constituent colours. Further, spectral information from biological mass spectrometry  
272 studies generates coordinate data and is represented as a matrix.

273 Unimodal representation is the fundamental way to proceed with any ML analysis.  
274 The complete set of features obtained through representation methods can vary in  
275 size and dimension depending on the dataset. To alleviate the computational load and  
276 resources during modelling, feature selection and feature reduction methods reduce the  
277 representation into a smaller latent space portraying the complete dataset, which is  
278 used for analysis. There are multiple feature reduction methods, such as the Principal  
279 Component Analysis (PCA), Joint Non-negative Matrix Factorisation (Joint NMF)  
280 and Autoencoders. Wrapper methods (forward, backwards, and stepwise selection),  
281 Filter methods (ANOVA, Pearson correlation, variance thresholding), and embedded  
282 methods (Lasso, Ridge, Decision Tree) are all part of feature selection techniques  
283 (Mirza et al. (2019)).

### 284 3.5.2 Multimodal Representation

285 Multimodal data representation involves using multiple modes or sources of infor-  
286 mation to represent data. A multimodal representation fuses multiple unimodal  
287 representations together onto a shared feature space (joint) or co-represents the  
288 features from the different datasets (coordinate).

289 Each modality is condensed in a joint multimodal representation, and the defin-  
290 ing features selected are concatenated to form a single collective representation. The  
291 ratio of features from each modality contributing to the concatenated representation  
292 is maintained uniformly. This prevents modalities with fewer features from being over-  
293 whelmed by modalities with large dimensions. Zhao et al. describe the application  
294 of joint representation in two publications using image data and clinical information  
295 (Zhao et al. (2020, 2021)). They merge representations of image data (CT scans) and  
296 clinical information in different ratios and predict lymph node (LN) metastasis (Zhao  
297 et al. (2020)). In a subsequent publication focusing on the same diagnosis, they intro-  
298 duce the 3M-CN architecture that utilises a ‘refine layer’ to predict LN metastasis  
299 (Zhao et al. (2021)). The refine layer is a concatenation of key features identified from  
300 clinical information and processed 3D images.

301 Coordinated representations reduce and present the features within each modal-  
302 ity individually but link them towards the same meaning over a common coordinate

303 space. Trajectory inference or pseudotemporal ordering is a method to classify the dif-  
304 ferent stages of the same cell type along an axis representing evolution (Saelens et al.  
305 (2019)). Pseudotime ordering is an excellent example of coordinate representations,  
306 where data from single-cell experiments are projected onto an evolutionary axis (Sae-  
307 lens et al. (2019)). The relationships established with identified patterns and domain  
308 knowledge help associate the features. MATCHER is a tool that has depicted impu-  
309 tation and correlation between modalities using a coordinate representation (Welch  
310 et al. (2017)). The manifold alignment method used in this tool achieved this task by  
311 representing data in low dimensions called a manifold and aligning them in a common  
312 space (alignment) (Wang and Mahadevan (2009)).

313 Multimodal representation methods empower ML architectures to investigate the  
314 interplay between features across diverse modalities. The entities in a biological sys-  
315 tem interact with each other in varied ways. Hence, the ratio of representations in  
316 the shared space as a parameter also affects the results of a multimodal analysis  
317 (Zhao et al. (2020)). Coordinate representations become more difficult than joint  
318 representations when there is no common ground to link the features.

### 319 **3.5.3 Special representations**

320 Special approaches represent data non-conventionally through a generative or a rule-  
321 based approach. These are not mutually exclusive to the previous two categories but  
322 process one or more source modalities differently to generate a representation. Gen-  
323 erative representations learn the underlying patterns and structure of the data and  
324 are capable of generating new instances of data that are similar to the examples they  
325 were trained on. On the other hand, rule-based representations leverage formal rules  
326 and semantics to describe the features within a dataset.

#### 327 **Auto-Encoders (AE)**

328 AE methods compress the entire dataset into a compact set of dimensions through an  
329 ‘encoding’ process, eliminating any non-representative features. A ‘decoding’ process  
330 then reconstructs the original data using the condensed representation, validating the  
331 reduced feature space. The decoding layers are generative of the relationships between  
332 all the variables within the data, and hence, this method is classified under a generative  
333 representation approach.

334 Detlefsen et al. extensively explores AE-based representations, emphasising the  
335 superior results achieved through the non-linear representation method in various tasks  
336 (Detlefsen et al. (2022)). Zhang et al. introduce OmiEmbed as a multitask framework,  
337 utilising the low-dimensional latent space generated by AEs for downstream tasks like  
338 cancer classification and survival prediction (Zhang et al. (2021)). AE representations  
339 also find applications in gene identification and cancer detection using expression  
340 data (Danaee et al. (2017)) and predicting carcinoma primary sites through DNA  
341 methylation data (Leitheiser et al. (2022)).

342 The encoding process in AEs can incorporate any type of model, such as a fully-  
343 connected neural network (FCNN) or convolutional neural network (CNN), to generate  
344 the latent space (Zhang et al. (2021)). Multiple modalities can also be combined at the  
345 input to generate a joint latent space (Huang et al. (2020)). This allows to generate  
346 different variants of the latent space and increases the choices available to work with.

347 The validation by the decoding process makes the latent space devoid of errors or data  
348 misrepresentation. However, the interpretability of AEs is generally low, and reducing  
349 the dimensions of the latent layer further diminishes the model understanding.

### 350 **Graph-based**

351 Graph representations portray relationships between different biological entities as a  
352 network by considering all features as ‘nodes’, and the relationship is depicted using  
353 ‘edges’. The edge values denote characteristics like similarity, interaction, and affinity  
354 between features based on the data. The representation of features as graphs is not  
355 limited to local, adjacent points but links them globally with edges.

356 Intramodal networks describe relationships between identical molecule types (for  
357 example protein-protein), while intermodal networks depict links between distinct  
358 types (Lee et al. (2020)). Clinical information about diseases and drugs can also be  
359 represented as graphs with links depicting common pathologies and targets. Omics  
360 modalities such as genome, lipidome, metabolome, proteome and transcriptome can  
361 be fused with environment and EMR modalities and represented as a Heterogeneous  
362 Multi-layer Network (HMLN) (Himmelstein and Baranzini (2015)).

363 Specific ML architectures are devised to best use a graph network representa-  
364 tion. Graph Convolutional Networks (GCN) are used to learn local graph structures  
365 and scale up based on the number of interactions to represent complex relationships  
366 (Kipf and Welling (2016)). Graph Attention Networks (GAT) incorporate attention  
367 mechanisms to overcome the structural overfitting for higher order GCNs (Veličković  
368 et al. (2017)). Ghorbani et al. present MGCN architecture, which implements graph  
369 representations to consolidate multilevel data (Ghorbani et al. (2019)).

370 Graphs can be generated using data from experimental protocols (e.g. omics),  
371 theory and literature (e.g. disease networks) to represent qualitative and quantitative  
372 information. Appropriate architectures (Section 3.6.2) assist to map and predict links  
373 between nodes using multilevel graph network data. The methods are highly sensitive  
374 to missing and unseen information but excel at discovering links within datasets.  
375 Sparse data matrices are easily translatable into graphs, as they efficiently condense  
376 large dimensional data to relevant nodes.

### 377 **Grammar-based**

378 Grammar-based methods, or semantic methods, rely on a predetermined, ordered set  
379 of ‘vocabulary’ to generate a representation, usually for a text-based modality. High-  
380 level patterns observed in the modality are identified and the complete dataset is  
381 represented based on the discovered patterns through a feature generation procedure.  
382 Additionally, data can either be embedded based on the provided dataset or refer-  
383 enced from the complete knowledge bank. Dictionary-based embedding methods create  
384 embeddings for the complete corpus, and the available data is represented based on  
385 the closest relationship from the complete dictionary (Baltrušaitis et al. (2018)).

386 Tyagi et al. used grammar-based representations to model the syntactic and  
387 semantic rule of RNA folding and used context-free grammars (CFG) to generate  
388 sequences and parse their structures (Tyagi et al. (2008)). Andikos et al. created Kno-  
389 tify (Andrikos et al. (2022)), a tool to predict RNA pseudoknots using CFG. Onokpasa  
390 et al. asserts that CFG representations improve compression ratios of RNA sequences  
391 and structures (Onokpasa et al. (2023)). Grammar-based representations have been

392 used to embed the motif information from sequences with the domain knowledge to  
393 depict the functionally connected regulatory regions (Soylu and Sefer (2023)).

394 Although grammar-based representations are powerful in capturing structured  
395 information, they may face challenges in handling the inherent ambiguity and vari-  
396 ability in real-world data. They require large amounts of data and computing power  
397 to process and generate the rule-based representation (Baltrušaitis et al. (2018)).

398  
399 In a few cases, ML methods do not differentiate between the representation steps  
400 and model training. For instance, dense neural networks and deep-learning architec-  
401 tures do not explicitly have a joint representation stage. They are directly processed  
402 for learning the features of the data and training the model (Section 3.6.2).

**Table 2:** Representation methods detailed with their advantages and disadvantages.

Representation		Advantages	Disadvantages
	Unimodal	Simple; Interpretable; Allows interdependencies	Cannot capture contextual information; Susceptible to noise or biases; High feature sizes
	Multimodal Joint	Combines features to common space; Controls modality size effects; Allows interdependencies; Interpretable; Reduces dimensions;	Requires tailored architecture; Relies on meaningful cross-modal relationships;
	Coordinate	Aligns features to common space; Controls modality size effects; Allows interdependencies; Interpretable; Reduces dimensions; Imputes information; Captures contextual Information	Requires common axis for representation; representation depends on quality and definition of common space
Special	AEs	Creates latent representations; Controls modality size effects; Allows interdependencies; Reduces dimensions; Low susceptibility to noise	Low interpretability; Computationally expensive;
	Graph	Represents qualitative and quantitative; Interpretable; scales with feature size; covers global information	Requires domain-specific knowledge for feature extraction; Complex algorithms needed for irregular structures and dynamic graphs; Susceptible to missing information
	Grammar	Applies for text modality; Captures patterns and semantic information; Reduces dimensions; Interpretable;	Affected by ambiguity and complex language constructs; Computationally expensive; Large datasets needed for processing; Susceptible to missing information



## 403 3.6 Fusion of Data Modalities: Integrating Multifaceted 404 Information

405 Data fusion methods harmonise the different data modalities available towards tasks  
406 such as clustering, regression or classification, utilising the representations from meth-  
407 ods discussed above (Section 3.5). Different modalities can be harmonised using two  
408 broad ML approaches: Unimodal learning and Multimodal learning.

### 409 3.6.1 Unimodal Learning

410 Unimodal learning algorithms elicit information from individual modalities to pre-  
411 dict an output. Here, three categories of unimodal integration exist based on how  
412 the features from different modalities are fused: (i) Early, (ii) Late, and (iii) Joint  
413 integration.

#### 414 Early Fusion

415 Early fusion methods describe ML architectures that concatenate feature representa-  
416 tions from multiple modalities at the input stage for modelling. The data is minimally  
417 processed, primarily to resolve heterogeneity, and samples are removed if imputation  
418 is impossible for missing data. This method disregards prior selection bias and allows  
419 us to investigate all features across modalities. It is time-consuming and computa-  
420 tionally expensive to process the complex combinations of all features from modalities  
421 (Dash et al. (2019)).

422 The benefits of using early integration methods are discussed and reported by Bar-  
423 num et al. (Barnum et al. (2020)). They assert that using immediate fusion techniques  
424 to merge modalities before feeding them into a model works better by integrating  
425 the lowest statistical correlations between input features. Banerjee et al. describe the  
426 PERFORM algorithm, utilising EMR data represented as temporal vectors, to assess  
427 its prediction performance in diagnosing acute pulmonary embolism (PE) (Banerjee  
428 et al. (2019)) with ElasticNet architecture (Zou and Hastie (2005)).

429 In the case of early fusion, class imbalance and differing sample sizes across modal-  
430 ities can affect the contribution of individual datasets, potentially biasing the analysis.  
431 Moreover, as the number of harmonised modalities increases in early fusion methods,  
432 the interpretability of the model drastically reduces. The limited coherence among  
433 data modalities restricts their combined usage, creating challenges in achieving a uni-  
434 fied representation. For example, input data as a combination of metabolomics and  
435 chromatin accessibility data may hinder a unified representation. Chen et al. describe  
436 data agnostic and data specific methods, with choices of modalities that can be used  
437 for coherent analysis (Chen and Tyagi (2020)).

#### 438 Late Fusion

439 Late fusion methods analyse multiple modalities independently using a model that best  
440 fits its representations to a predicted output, and the outputs from each are aggregated  
441 towards a singular result or inference. Late integration is generally performed either  
442 by taking an aggregated average of the predicted probabilities (outputs) from each  
443 modality or passing all the predictions from each modality into a FCNN to process a  
444 final output.

445 Wang et al. presented MOGONET as a tool to classify cancer subtypes using  
446 three modalities: mRNA expression, methylation and miRNA expression (Wang et al.  
447 (2021)). A late fusion architecture was established using GCN to predict an initial  
448 class label and an FCNN to generate a final class label prediction. Luo et al. presented  
449 a modified version of MOGONET, called GRAMINet with GATs instead of GCNs  
450 (Luo et al. (2023)). Both examples pass the same combination of modalities through  
451 different architectures to learn a model for biomedical data classification. However,  
452 this is not universally applicable since the late integration methods provide flexibility  
453 in model selection for different modalities. Huang et al. investigated a multimodal  
454 approach to predict PE and reported the results on seven different architectures,  
455 one early, two joint and four late fusion architectures (Huang et al. (2020)). The  
456 early fusion methods had the highest sensitivity, while the late ElasticNet architecture  
457 outperformed in all other metrics, such as accuracy, AUROC, specificity and positive  
458 predictive value.

459 Ensemble learning may be considered a variant of the late fusion model, where  
460 the outputs of multiple ML models are combined towards a final decision (Zhou  
461 (2012)). The late fusion focuses on combining features or representations after individ-  
462 ual processing, whereas ensemble learning leverages the diversity of multiple models  
463 to improve overall predictive performance.

464 Late fusion methods do not directly allow for the interaction of features from  
465 multiple modalities. This enables to train each modality with independent, unique  
466 models without any interference from other data types. As a result, concerns about  
467 different dataset sizes, heterogeneous measurements, and model compatibility vanish.  
468 The ability of late fusion methods to capture all information within each modality  
469 in an equivalent manner makes it the widely reported harmonisation method in the  
470 literature.

#### 471 **Joint Fusion**

472 Joint fusion methods endeavour to extract a representation of all initial modalities  
473 and model them together to a predicted output. Direct concatenation methods, often  
474 used in early fusion, are not possible between modalities that have different quantities  
475 and may require a heavy preprocessing step. In late integration, the interdependency  
476 between features across modalities is ignored. Joint integration methods provide an  
477 advantage through the interaction of features from different modalities in the training  
478 phase, irrespective of the observed heterogeneity. The heterogeneity is mitigated since  
479 the feature representation and selection procedures reduce and unify the information  
480 numerically.

481 Joint integration methods have been explored using modalities such as CT scans,  
482 EMR, methylation, and expression data to achieve biomedical tasks of classification  
483 towards prognosis and diagnosis (Zhao et al. (2020, 2021); Wang et al. (2021); Huang  
484 et al. (2020)). Zhao et al. investigated the effect of different ratios of EMR features  
485 during an integrated analysis of CT images using the DensePriNet architecture (Zhao  
486 et al. (2021)). Huang et al. propose a joint representation of CT images and clinical  
487 features in a ‘refine layer’ to predict an output that performs better in detecting pul-  
488 monary embolism in comparison to other models (Huang et al. (2020)). MATCHER



489 utilises a joint fusion method to interpolate instances based on the alignment of  
490 multiple modalities to the pseudotime scale (Welch et al. (2017)).

491 Joint methods are more complex than late fusion but easier to interpret than  
492 early fusion methods. Using representations rather than the raw features from each  
493 modality greatly reduces the computational load compared to early fusion. However,  
494 joint harmonisation methods are not data agnostic and may require carefully curated  
495 model designs.

### 496 **3.6.2 Multimodal Learning or Co-learning**

497 Multimodal learning is designed to integrate and model the features from different  
498 modalities more comprehensively than unimodal learning. Joint fusions merge repre-  
499 sentations, but multimodal learning enables co-learning through a direct feed of all  
500 the features interacting at a higher, complex level. They can be classified into three  
501 distinct co-learning methods: probabilistic, multiple kernel learning and deep neural  
502 networks.

#### 503 **Probabilistic Models**

504 Probabilistic methods involve building models that capture the relationships and  
505 dependencies between different modalities using joint or conditional probabilities.  
506 These models are highly interpretable, which allows for the models to integrate expert  
507 knowledge in the fusion approach, granting us the ability to interpret the results better  
508 than other methods. The random walk method uses probabilistic values to simulate a  
509 particle moving between nodes and layers in a network, establishing the relationships  
510 and links between the nodes (Baptista et al. (2022)).

511 MultiXRank module, published by Baptista et al., is an example of the probabilistic  
512 method of integration using a multiplex network of intramodal and intermodal inter-  
513 actions (protein-protein interactions, gene multiplex and disease monoplex networks)  
514 (Baptista et al. (2022)). Pio-Lopez et al. describe a use case of the random walk with  
515 restart architecture, wherein the method predicts long-distance gene-disease interac-  
516 tions using gene interaction network and disease similarity network data (Pio-Lopez  
517 et al. (2021)).

518 Probabilistic methods are applicable to any combination of modalities as long as  
519 they form a multiplex network. They rely heavily on theoretical knowledge to bridge  
520 relationships between elements of multiple domains and hence can be applied to data  
521 from any domain with multiplex and bipartite networks (Pio-Lopez et al. (2021)).

#### 522 **Multiple Kernel Learning**

523 Kernels are linear classifiers that divide the data linearly using lenient boundaries,  
524 and a combined multitude of them assist in classifying non-linear heterogeneous data  
525 (Gönen and Alpaydm (2011)). This method is implemented in support vector machines  
526 (SVM), a popular method to analyse complex data.

527 Liu et al. used SVM to model MRI datasets from multiple sources towards  
528 Alzheimer's disease classification (Liu et al. (2013)). Lanckriet et al. predict the func-  
529 tions of yeast proteins using kernel-based learning (Lanckriet et al. (2003)). Multiple  
530 matrices describing the protein data were used in the algorithm, and results were  
531 reported on the different combinations of kernels used to classify the proteins as per  
532 their functions (Lanckriet et al. (2003)).

533 In MKL, different kernels are applied to each modality, and the combination of  
534 these kernels is learned to optimise the model's overall performance. Kernels can  
535 identify linear boundaries in datasets, making MKL a highly suitable method for clas-  
536 sification tasks (Wilson et al. (2019)). Kernels can be combined in different ways (sum,  
537 product) to generate new kernels. A combination of multiple kernels accounts for a  
538 better classifier than using a single kernel (Hofmann et al. (2008)). This method is  
539 resistant to outliers but is susceptible to missing data (Wilson et al. (2019)).

#### 540 **Deep neural networks (DNNs)**

541 Deep neural integration methods are characterised by a substantial number of neu-  
542 rons and layers constituting neural networks with significant depth and complexity.  
543 DNNs utilise representations of different modalities to reduce the features and pass  
544 them through high-level, intricate architectures, which enable them to uncover hidden  
545 information within the datasets.

546 DNNs are extensively used to understand data at a microscopic level, especially  
547 in the biomedical domain. EMR data can be modelled with omics modalities to  
548 shed light on physical and phenotypic changes and their relationships across time.  
549 Zhu et al. address an ML model to fuse and learn time-series data, with the use of  
550 Stacked Sparse Auto-Encoder (SSAE) and Long Short-Term Memory (LSTM) archi-  
551 tecture (Zhu et al. (2021)). Zhang et al. have reported about OmiEmbed, a multitask  
552 deep-learning framework based on an autoencoder architecture (Zhang et al. (2021)).  
553 AffinityNet, proposed by Ma et al., uses k-nearest neighbours (kNN) attention pooling  
554 where the cluster representations of the data is processed as a GAT (Ma and Zhang  
555 (2019)). The method has asserted good performance for both labelled and unlabelled  
556 datasets.

557 DNNs are computationally expensive to perform due to their dense and complex  
558 architecture. They model data at high degrees of non-linearity, but the process becomes  
559 hard to decipher and elucidate. The meaning of data is lost when modelling and  
560 remains a black box with very low interpretability.

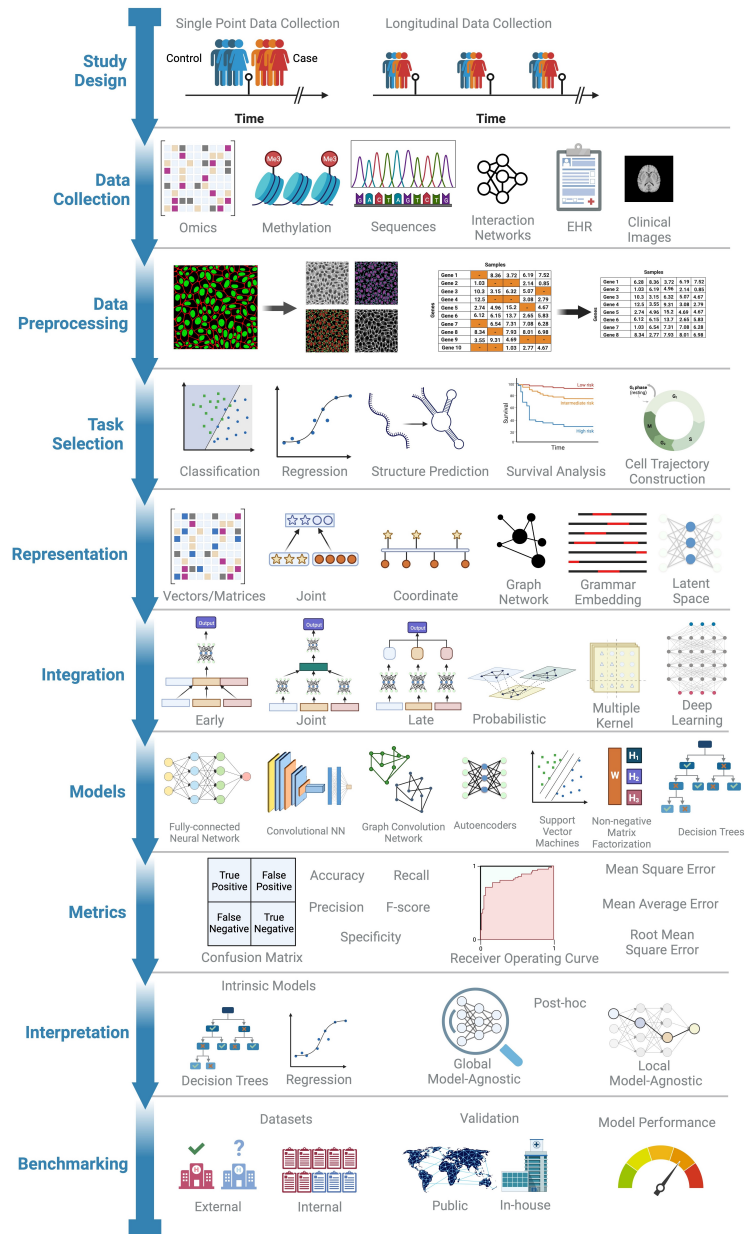
**Table 3:** A description of the integration methods and their advantages and disadvantages for a multimodal set-up.

Integration		Advantages	Disadvantages
Unimodal	Early	Performs lowest level of statistical correlation between features; Disregards selection bias	Computationally complex and expensive; Requires tailored models for effectively modelling and interpreting features
	Joint	Mitigates heterogeneity during modelling; Reduced feature size due to representations eases computational load	Requires tailored models for modelling joint representations
	Late	Independent models for different modalities; Ignores representation bias	Does not allow for feature interaction
Multimodal	Probabilistic	Highly interpretable; Establishes links between entities across different modalities	Requires domain-specific multilayer network information; Susceptible to missing data
	Multiple kernel Learning	Models based on overlapping results; Applicable for non-linear relationships; Resistant to outliers; Interpretable	Susceptible to missing data
	DNNs	Uncovers hidden information without explicit rules; Utilises complex architectures to understand the non-linear relationships between features and modalities	Low interpretability; Computationally complex and expensive

## 4 Guidelines for Model Selection

We propose ten recommendations for initiating a multimodal harmonisation analysis (Figure 3). Before starting out on an analysis, clearly articulate the objectives and aims of the study before initiating the harmonisation analysis. These objectives will guide subsequent data collection, representation, and model selection steps.

- 1. Tailor Study Design to Objectives:** Tailor the study design to the defined objectives, taking into consideration the scale of the study and available resources. Ensure effective study design for sample identification and data collection that aligns with the study's goals.
- 2. Implement Optimised Experimental Protocols:** The data is either already available or generated through new experiments. Employ or select optimised experimental protocols and assays for data collection, ensuring consistency and reliability. These protocols form the foundation for subsequent analysis steps and contribute to the quality of collected data.
- 3. Digitised Data and Global Sharing:** Digitise collected data to facilitate analysis and global data sharing through repositories and databanks. Adopting data and metadata standards enhances data sharing and harmonisation. This step is crucial for collaborative research efforts and ensures data accessibility for future studies.



**Fig. 3:** A ten step guide flowchart that describes the process and order of execution to perform a multimodal integration. The titles on the left of the timeline describe the task order. The illustrations on the right are representative examples of different methods under each category. Made with BioRender.

- 579 4. **Modality Identification and Data Preprocessing:** Classify the collected data  
580 into the broad modalities discussed here and systematically process the data indi-  
581 vidually to create a working subset free of artefacts and low-quality elements. The  
582 analysis and results can vary depending on the level of preprocessing done.
- 583 5. **Task Selection:** Task selection should be guided by the study’s goals, setting the  
584 stage for subsequent processing and analysis. Choose tasks that reflect the aims  
585 and are applicable to the dataset.
- 586 6. **Choose Feature Representation and Integration Methods Wisely:** The  
587 choice of representation or integration methods influences each other. Hence, select  
588 them based on the data type, number of modalities, level of harmonisation, and  
589 modality coherence. Recognise that a one-size-fits-all approach is impractical, and  
590 tailored methods may be needed for different tasks.
- 591 7. **Navigate Model Selection Complexity:** Different ML models can be employed  
592 for the same harmonisation set-up. Employ different models of varying complexity  
593 to assess the data and evaluate their performance using appropriate metrics.
- 594 8. **Model Performance Metrics:** Select model performance metrics corresponding  
595 to the task to compare and choose the optimal model. Provide an explanation of  
596 the metrics used and their relevance to the task.
- 597 9. **Prioritise Interpretable Models:** Prioritise using interpretable models, either  
598 intrinsic or through post-hoc interpretation. Especially in clinical settings, under-  
599 standing how a model arrives at conclusions enhances trust and reliability.
- 600 10. **Validate and Benchmark Models:** Validate models on different datasets and  
601 sources to ensure robustness and generalizability. Benchmark models against state-  
602 of-the-art approaches and external datasets to mitigate aggregation and evaluation  
603 biases.

604 In the end, ensure that these recommendations are adapted to the specific context  
605 and goals of your multimodal harmonisation analysis.

## 606 5 Discussion

### 607 **Lack of Comprehensive Reviews:**

608 The article points out a noticeable gap in the existing literature regarding  
609 comprehensive explanations of workflow and procedures for integrating biomedical  
610 multimodal data. Multiple reviews for machine learning strategies to process multi-  
611 modal data are available, but there is a deficit of articles relating them to biological  
612 and clinical data. A predominant part of research literature presents results with infor-  
613 mation from a single modality. Studies that utilise different biological modalities often  
614 interpret the results of independent analyses together. The concept of co-analysis,  
615 or more aptly, ‘co-learning’ is missed. There is a lack of clarity on how to effec-  
616 tively integrate data from disparate sources at the lowest item level to extract holistic  
617 knowledge.

### 618 **Diverse Taxonomies in Multimodal Analysis:**

619 Biomedical multimodal data from the same sample set is now routinely available  
620 from various research and development activities and healthcare. In the context of  
621 multimodal analysis, there is a distinction in the representation and integration steps

622 compared to unimodal analysis. We highlight the various analysis methods and the  
623 data types available under a limited set of taxonomic categorisation.

624 This classification of data types from biological and clinical sources allows one to  
625 identify methods that will suit the analysis of specific combinations and evaluate the  
626 advantages of each. We describe data harmonisation as a split of representation and  
627 integration methods, each with six distinct categories. Most steps are similar to an  
628 unimodal analysis, and the distinction in a multimodal analysis arises in the represen-  
629 tation and the integration steps. The representation methods proposed emphasise on  
630 the features within the data. The various types of representation methods are key to  
631 uniformly present the multimodal data prior to an analysis. The section on integration  
632 focuses on the various methods to feed the data into ML architectures.

633 The article discusses this difference and provides insights into how to handle  
634 representation and integration methods for multimodal data effectively.

#### 635 **Framework and Model Suggestions for Biomedical Data Combinations:**

636 There is a need for a structured framework or guideline to facilitate the harmoniza-  
637 tion process for multimodal data. The article aims to address this gap by presenting  
638 the first guideline framework towards a data harmonization process and providing a  
639 complete workflow. The recommended procedure consists of 10 steps to plan through  
640 towards a multimodal analysis.

641 To assist those undertaking harmonisation for the first time, we present a guide  
642 matrix showcasing examples from published literature, illustrating different combina-  
643 tions of data modalities. The combinations between the representation and integration  
644 methods are presented as a non-exhaustive list in table 4. Existing studies show that  
645 different choices can yield different results when using the same datasets (Huang et al.  
646 (2020)). The diverse taxonomies outlined in this paper can assist in understanding the  
647 significance of choosing an appropriate integration model for analysis, considering the  
648 concern related to biomedical data and model challenges.

#### 649 **Future Focus for Harmonisable Models:**

650 The article acknowledges the challenges related to data and model selection in  
651 the context of multimodal analysis. It suggests that diverse taxonomies outlined in  
652 the paper can assist in understanding the significance of choosing an appropriate  
653 integration model for analysis, considering these challenges.

654 Data related challenges and model related challenges both arise when implementing  
655 a multimodal analysis. In addition to the challenges described in section 3.3 related to  
656 biomedical data, concerns on data acquisition and maintenance also require attention.  
657 The quality of biomedical data collected needs to be maintained, with appropriate  
658 measures taken for de-identification of the data and global sharing. A vast majority  
659 of the published literature on biomedical multimodal analysis focuses on the model  
660 metrics and parameters scores. However, due focus should be given to the model  
661 interpretability as well. Multimodal analysis with complex architectures may yield  
662 high performance scores, but they cannot be used to understand the biological and  
663 clinical data if the models are not interpretable. Interpretable models are needed to  
664 understand the process, especially with biomedical data to relate to further procedures,  
665 such as diagnosis and intervention strategies .

## 666 6 Conclusion:

667 The article highlights a significant gap in existing literature regarding the integra-  
668 tion of multimodal data, noting a lack of comprehensive explanations and holistic  
669 views in current research. While recognizing the transformative potential of multi-  
670 modal integration, it emphasizes the need for clarity on effectively integrating disparate  
671 data sources to extract comprehensive knowledge. Acknowledging challenges in data  
672 and model selection, the article proposes using diverse taxonomies to aid integra-  
673 tion model selection. Addressing the distinction between unimodal and multimodal  
674 analysis, the article provides insights into representation and integration methods for  
675 multimodal data. Furthermore, it underscores the necessity for a structured framework  
676 to facilitate the harmonization process, presenting the first guideline framework and  
677 workflow. Additionally, it aims to assist researchers new to harmonization by offering  
678 a guide matrix featuring examples from published literature, aiding in the selection of  
679 appropriate integration models.

## 680 Declarations

681 **Funding.** MAMS acknowledges DBT and IITB-Monash research academy for  
682 the PhD sponsorship. ST acknowledges AISRF EMCR fellowship from Australian  
683 Academy of Science. MKM acknowledges support from the IITB Monash research  
684 academy and from Core Research Grant by Science and Engineering Research Board,  
685 India (Grant number: CRG/2022/008142).

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

689 **CRedit Authorship Contribution Statement.** textbfMurali Aadhitya M S:  
690 Data curation, Visualisation, Writing - original draft. **Mithun Mitra:** Supervision,  
691 Validation, Writing - review & editing. **Sonika Tyagi:** Conceptualisation, Data  
692 curation, Supervision, Writing - original draft, Writing - review & editing

693 **Acknowledgements.** We thank Tyrone Chen for his suggestions regarding con-  
694 cepts of the work done.

695 **Ethics approval and consent to participate:** 'Not applicable'

696 **Consent for publication:** 'Not applicable'

697 **Data availability:** 'Not applicable'

698 **Materials availability:** 'Not applicable'

699 **Code availability:** 'Not applicable'

700  
701



**Table 4:** Guide matrix providing examples from literature based on different combinations of modalities and the corresponding tasks, study design, representation, integration, model used, interpretability and reference. \* are all representation methods

Modalities Integrated	Repn	Integration	Study Design	Task	Model Used	Model Interp	Validation Reference
EMR*	Unimodal	MKL	Static	Prediction	SVM	High	Public, Internal (Xu et al. (2010))
Gene Sequence*	Joint	Joint	Time Series	Risk Prediction	EN	Low	In-house, External (Banerjee et al. (2019))
	Grammar	-	Static	RNA Structure Prediction	CFG	High	Public, Internal (Tyagi et al. (2008))
	Grammar	-	Static	RNA Structure Prediction	CFG	High	In-house, Internal (Andrikos et al. (2022))
	Grammar	-	Static	RNA Structure Prediction	CFG	High	Public, Internal (Onokpasa et al. (2023))
Protein Sequence*	AE	-	Static	Clustering	VAE, ResNet, LSTM	Low	Public, Internal (Detlefsen et al. (2022))
Disease network, miRNA Image, EMR	Graph	Probabilistic	Static	Disease Prediction	Katz	High	Public, Internal (Zhang et al. (2017))
	Unimodal, Joint	Early, Joint, Late	Static	Classification	FCNN, CNN	Low	In-house, Internal (Huang et al. (2020))
	Unimodal	Joint	Static	Classification	DDB	Low	In-house, Internal (Zhao et al. (2020))
Methylation, Image	Joint	Joint	Static	Classification, Segmentation	DB, CNN, FCNN	Low	In-house, External (Zhao et al. (2021))
	Unimodal	Early	Static	Multi-variate Regression	GLM	High	In-house, Internal (Schiano et al. (2020))
Metabolomics, Proteomics	Unimodal	Late	Static	Differential Correlation	-	High	In-house, Internal (An et al. (2022))
	Unimodal	Late	Static	Differential Correlation	-	High	In-house, Internal (Wang et al. (2019))
Transcriptomics, DNA-Protein interaction network	Unimodal	Late	Static	Differential Correlation	-	high	Public, Internal (Xu et al. (2019))
	AE	Joint	Static	Classification	AE, FCNN	Low	Public, Internal (Zhang et al. (2019))
Non-Bio	Graph, Joint	Late	Static	Classification	SNF	High	Public, Internal (Wang et al. (2014))
	Graph	Probabilistic	Static	Classification	GCN	Low	Public, Internal (Ghorbani et al. (2019))
Proteomics, EMR	AE	Joint	Static	Classification, Regression	SSAE, LSTM	Low	Public, Internal (Zhu et al. (2021))
	Joint	MKL	Static	Classification	SVM	High	In-house, Internal (Dargazani et al. (2020))

Note: The table is continued in Table 5



**Table 5:** Continuation of Table 4. Guide matrix providing examples from literature based on different combinations of modalities and the corresponding tasks, study design, representation, integration, model used, interpretability and reference. \* are all representation methods

Modalities Integrated	Repn	Integration	Study Design	Task	Model Used	Model Interp	Validation	Reference
Mutation; Drugs	Unimodal	Joint, Late	Static	Drug Prediction	CNN	Low	Public, Internal	(Chang et al. (2018))
Gene Expression, Chromatin Accessibility	Coordinate	Late	Time Series	Cell Trajectory Construction	MA	High	Public, Internal	(Welch et al. (2017))
Networks of Gene; Drug; Disease	Graph	Probabilistic	Static	Link Prediction	RWR		Public, Internal	(Pio-Lopez et al. (2021))
	Graph	Probabilistic	Static	Link Prediction	RWR	High	Public, Internal	(Baptista et al. (2022))
Methylation, mRNA, miRNA	AE	DNN	Static	Classification	AE	Low	Public, Internal	(Franco et al. (2021))
	AE	Joint	Static	Classification, Regression, Survival Analysis	AE, FCNN, CNN	Low	Public, Internal	(Zhang et al. (2021))
	Graph	Joint	Static	Classification	GCN	Low	Public, Internal	(Wang et al. (2021))
	Unimodal, Graph	Late	Static	Classification	GCN	Low	Public, Internal	(Luo et al. (2023))
Lipidomics; Metabolomics; Proteomics; Transcriptomics	Unimodal	Joint	Static	Drug Prediction	PLSDA		Public, Internal	(Chen et al. (2021))
Methylation; Gene Expression; Mutation; Copy Number Alteration; Clinical	AE	Joint, Late	Static	Survival Analysis	FCNN	Low	Public, Internal	(Zhao et al. (2021))
Gene expression; Metabolomics; Proteomics; Cytokine measurements; Cytometric measurements; Microbiome	Unimodal	Late	Time Series	Multi-variate Estimation	EN	Low	In-house, Internal	(Ghaemi et al. (2018))
Interaction networks of drug compound, pharmacologic class, gene of action, pathway, biological process, disease, side effect, symptom, and anatomy	Graph	Probabilistic	Static	Link Prediction	EN	High	Public, Internal	(Himmelstein and Baranzini (2015))

## 702 References

- 703 Abarbanel, A., Evans, J.R., Budzynski, T.H., Budzynski, H.K.: Introduction to Quan-  
704 titative EEG and Neurofeedback: Advanced Theory and Applications. Academic  
705 Press, ??? (2009)
- 706 Acosta, J.N., Falcone, G.J., Rajpurkar, P., Topol, E.J.: Multimodal biomedical ai.  
707 Nature Medicine **28**(9), 1773–1784 (2022)
- 708 Andrikos, C., Makris, E., Kolaitis, A., Rassias, G., Pavlatos, C., Tsanakas, P.: Knotify:  
709 An efficient parallel platform for rna pseudoknot prediction using syntactic pattern  
710 recognition. Methods and Protocols **5**(1), 14 (2022)
- 711 Aburajab, R., Pospiech, M., Alachkar, H.: Profiling the epigenetic landscape of the  
712 antigen receptor repertoire: The missing epi-immunogenomics data. Nature Methods  
713 **20**(4), 477–481 (2023)
- 714 An, R., Yu, H., Wang, Y., Lu, J., Gao, Y., Xie, X., Zhang, J.: Integrative anal-  
715 ysis of plasma metabolomics and proteomics reveals the metabolic landscape of  
716 breast cancer. Cancer & Metabolism **10**(1), 13 (2022) [https://doi.org/10.1186/  
717 s40170-022-00289-6](https://doi.org/10.1186/s40170-022-00289-6)
- 718 Baltrušaitis, T., Ahuja, C., Morency, L.-P.: Multimodal machine learning: A survey  
719 and taxonomy. IEEE transactions on pattern analysis and machine intelligence  
720 **41**(2), 423–443 (2018)
- 721 Baptista, A., Gonzalez, A., Baudot, A.: Universal multilayer network exploration by  
722 random walk with restart. Communications Physics **5**(1), 170 (2022)
- 723 Banerjee, I., Sofela, M., Yang, J., Chen, J.H., Shah, N.H., Ball, R., Mushlin, A.I.,  
724 Desai, M., Bledsoe, J., Amrhein, T., *et al.*: Development and performance of the  
725 pulmonary embolism result forecast model (perform) for computed tomography  
726 clinical decision support. JAMA network open **2**(8), 198719–198719 (2019)
- 727 Barnum, G., Talukder, S., Yue, Y.: On the benefits of early fusion in multimodal  
728 representation learning. arXiv preprint arXiv:2011.07191 (2020)
- 729 Clarke, T.-K., Adams, M.J., Davies, G., Howard, D.M., Hall, L.S., Padmanabhan,  
730 S., Murray, A.D., Smith, B.H., Campbell, A., Hayward, C., *et al.*: Genome-wide  
731 association study of alcohol consumption and genetic overlap with other health-  
732 related traits in uk biobank (n= 112 117). Molecular psychiatry **22**(10), 1376–1384  
733 (2017)
- 734 Church, K.W.: Word2vec. Natural Language Engineering **23**(1), 155–162 (2017)
- 735 Chen, T., Philip, M., Lê Cao, K.-A., Tyagi, S.: A multi-modal data harmonisation  
736 approach for discovery of covid-19 drug targets. Briefings in bioinformatics **22**(6),

- 737 185 (2021)
- 738 Chang, Y., Park, H., Yang, H.-J., Lee, S., Lee, K.-Y., Kim, T.S., Jung, J., Shin,  
739 J.-M.: Cancer drug response profile scan (cdrscan): a deep learning model that  
740 predicts drug effectiveness from cancer genomic signature. *Scientific reports* **8**(1),  
741 8857 (2018)
- 742 Chen, T., Tyagi, S.: Integrative computational epigenomics to build data-driven gene  
743 regulation hypotheses. *GigaScience* **9**(6), 064 (2020)
- 744 Chen, T., Tyagi, N., Chauhan, S., Peleg, A.Y., Tyagi, S.: genomicbert and data-free  
745 deep-learning model evaluation. *bioRxiv*, 2023-05 (2023)
- 746 Camastra, F., Vinciarelli, A.: *Machine Learning for Audio, Image and Video Analysis:*  
747 *Theory and Applications*. Springer, ??? (2015)
- 748 Danaee, P., Ghaeini, R., Hendrix, D.A.: A deep learning approach for cancer detection  
749 and relevant gene identification. In: *Pacific Symposium on Biocomputing 2017*, pp.  
750 219–229 (2017). World Scientific
- 751 Detlefsen, N.S., Hauberg, S., Boomsma, W.: Learning meaningful representations of  
752 protein sequences. *Nature communications* **13**(1), 1914 (2022)
- 753 Dash, S., Shakyawar, S.K., Sharma, M., Kaushik, S.: Big data in healthcare:  
754 management, analysis and future prospects. *Journal of big data* **6**(1), 1–25 (2019)
- 755 Dargazanli, C., Zub, E., Deverdun, J., Decourcelle, M., Bock, F., Labreuche, J.,  
756 Lefèvre, P.-H., Gascou, G., Derraz, I., Riquelme Bareiro, C., Cagnazzo, F., Bonafé,  
757 A., Marin, P., Costalat, V., Marchi, N.: Machine learning analysis of the cerebrovas-  
758 cular thrombi proteome in human ischemic stroke: An exploratory study. *Frontiers*  
759 *in Neurology* **11** (2020) <https://doi.org/10.3389/fneur.2020.575376>
- 760 Fu, J., Rui, Y.: Advances in deep learning approaches for image tagging. *APSIPA*  
761 *Transactions on Signal and Information Processing* **6**, 11 (2017)
- 762 Franco, E.F., Rana, P., Cruz, A., Calderon, V.V., Azevedo, V., Ramos, R.T., Ghosh,  
763 P.: Performance comparison of deep learning autoencoders for cancer subtype  
764 detection using multi-omics data. *Cancers* **13**(9), 2013 (2021)
- 765 Gönen, M., Alpaydm, E.: Multiple kernel learning algorithms. *The Journal of Machine*  
766 *Learning Research* **12**, 2211–2268 (2011)
- 767 Ghorbani, M., Baghshah, M.S., Rabiee, H.R.: Mgcnn: semi-supervised classification in  
768 multi-layer graphs with graph convolutional networks. In: *Proceedings of the 2019*  
769 *IEEE/ACM International Conference on Advances in Social Networks Analysis and*  
770 *Mining*, pp. 208–211 (2019)

- 771 Ghaemi, M.S., DiGiulio, D.B., Contrepois, K., Callahan, B., Ngo, T.T.M., Lee-  
772 McMullen, B., Lehallier, B., Robaczewska, A., Mcilwain, D., Rosenberg-Hasson,  
773 Y., Wong, R.J., Quaintance, C., Culos, A., Stanley, N., Tanada, A., Tsai, A.,  
774 Gaudilliere, D., Ganio, E., Han, X., Ando, K., McNeil, L., Tingle, M., Wise,  
775 P., Maric, I., Sirota, M., Wyss-Coray, T., Winn, V.D., Druzin, M.L., Gibbs,  
776 R., Darmstadt, G.L., Lewis, D.B., Partovi Nia, V., Agard, B., Tibshirani, R.,  
777 Nolan, G., Snyder, M.P., Relman, D.A., Quake, S.R., Shaw, G.M., Stevenson,  
778 D.K., Angst, M.S., Gaudilliere, B., Aghaeepour, N.: Multiomics modeling of the  
779 immunome, transcriptome, microbiome, proteome and metabolome adaptations  
780 during human pregnancy. *Bioinformatics* **35**(1), 95–103 (2018) <https://doi.org/10.1093/bioinformatics/bty537> [https://academic.oup.com/bioinformatics/article-pdf/35/1/95/48962254/bioinformatics\\_35\\_1\\_95\\_s1.pdf](https://academic.oup.com/bioinformatics/article-pdf/35/1/95/48962254/bioinformatics_35_1_95_s1.pdf)  
782
- 783 Garbh-Ini <https://www.garbhnicohort.in>
- 784 Himmelstein, D.S., Baranzini, S.E.: Heterogeneous network edge prediction: a data  
785 integration approach to prioritize disease-associated genes. *PLoS computational*  
786 *biology* **11**(7), 1004259 (2015)
- 787 Haghverdi, L., Büttner, M., Wolf, F.A., Buettner, F., Theis, F.J.: Diffusion pseudotime  
788 robustly reconstructs lineage branching. *Nature methods* **13**(10), 845–848 (2016)
- 789 Huang, S.-C., Kothari, T., Banerjee, I., Chute, C., Ball, R.L., Borus, N., Huang, A.,  
790 Patel, B.N., Rajpurkar, P., Irvin, J., *et al.*: Penet—a scalable deep-learning model  
791 for automated diagnosis of pulmonary embolism using volumetric ct imaging. *NPJ*  
792 *digital medicine* **3**(1), 61 (2020)
- 793 Huang, S.-C., Pareek, A., Zamanian, R., Banerjee, I., Lungren, M.P.: Multimodal  
794 fusion with deep neural networks for leveraging ct imaging and electronic health  
795 record: a case-study in pulmonary embolism detection. *Scientific reports* **10**(1),  
796 22147 (2020)
- 797 Hofmann, T., Schölkopf, B., Smola, A.J.: Kernel methods in machine learning (2008)
- 798 Kan, A.: Machine learning applications in cell image analysis. *Immunology and cell*  
799 *biology* **95**(6), 525–530 (2017)
- 800 Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional  
801 networks. arXiv preprint arXiv:1609.02907 (2016)
- 802 Kline, A., Wang, H., Li, Y., Dennis, S., Hutch, M., Xu, Z., Wang, F., Cheng, F., Luo,  
803 Y.: Multimodal machine learning in precision health: A scoping review. *npj Digital*  
804 *Medicine* **5**(1), 171 (2022)
- 805 Laney, D.: 3d data management: Controlling data volume, velocity and variety. META  
806 Group Research Note (2001)

- 807 Lau, J.H., Baldwin, T.: An empirical evaluation of doc2vec with practical insights into  
808 document embedding generation. arXiv preprint arXiv:1607.05368 (2016)
- 809 Leitheiser, M., Capper, D., Seegerer, P., Lehmann, A., Schüller, U., Müller, K.-R.,  
810 Klauschen, F., Jurmeister, P., Bockmayr, M.: Machine learning models predict the  
811 primary sites of head and neck squamous cell carcinoma metastases based on dna  
812 methylation. *The Journal of Pathology* **256**(4), 378–387 (2022)
- 813 Lanckriet, G.R., Deng, M., Cristianini, N., Jordan, M.I., Noble, W.S.: Kernel-  
814 based data fusion and its application to protein function prediction in yeast. In:  
815 *Biocomputing 2004*, pp. 300–311. World Scientific, ??? (2003)
- 816 Maturana, E., Alonso, L., Alarcón, P., Martín-Antoniano, I.A., Pineda, S., Piorno, L.,  
817 Calle, M.L., Malats, N.: Challenges in the integration of omics and non-omics data.  
818 *Genes* **10**(3), 238 (2019)
- 819 Luo, H., Liang, H., Jia, M., Li, Y., Yao, X., Cong, S.: Multi-omics integration for dis-  
820 ease prediction via multi-level graph attention network and adaptive fusion. *bioRxiv*,  
821 2023–03 (2023)
- 822 Li, Y., Ngom, A.: Data integration in machine learning. In: 2015 IEEE International  
823 Conference on Bioinformatics and Biomedicine (BIBM), pp. 1665–1671 (2015).  
824 IEEE
- 825 Lima, D.M., Rodrigues-Jr, J.F., Traina, A.J., Pires, F.A., Gutierrez, M.A.: Transform-  
826 ing two decades of epr data to omop cdm for clinical research. *Stud Health Technol*  
827 *Inform* **264**(August), 233–7 (2019)
- 828 Lee, B., Zhang, S., Poleksic, A., Xie, L.: Heterogeneous multi-layered network model  
829 for omics data integration and analysis. *Frontiers in genetics* **10**, 1381 (2020)
- 830 Liu, F., Zhou, L., Shen, C., Yin, J.: Multiple kernel learning in the primal for mul-  
831 timodal alzheimer’s disease classification. *IEEE journal of biomedical and health*  
832 *informatics* **18**(3), 984–990 (2013)
- 833 Malet-Martino, M., Holzgrabe, U.: Nmr techniques in biomedical and pharmaceutical  
834 analysis. *Journal of pharmaceutical and biomedical analysis* **55**(1), 1–15 (2011)
- 835 Mou, M., Pan, Z., Lu, M., Sun, H., Wang, Y., Luo, Y., Zhu, F.: Application of machine  
836 learning in spatial proteomics. *Journal of Chemical Information and Modeling*  
837 **62**(23), 5875–5895 (2022)
- 838 Mansuri, M.S., Williams, K., Nairn, A.C.: Uncovering biology by single-cell pro-  
839 teomics. *Communications Biology* **6**(1), 381 (2023)
- 840 Mirza, B., Wang, W., Wang, J., Choi, H., Chung, N.C., Ping, P.: Machine learning  
841 and integrative analysis of biomedical big data. *Genes* **10**(2), 87 (2019)

- 842 Ma, T., Zhang, A.: Affinitynet: semi-supervised few-shot learning for disease type  
843 prediction. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol.  
844 33, pp. 1069–1076 (2019)
- 845 Neidle, S.: 2 - the building-blocks of dna and rna. In: Neidle, S. (ed.)  
846 Principles of Nucleic Acid Structure, pp. 20–37. Academic Press,  
847 New York (2008). <https://doi.org/10.1016/B978-012369507-9.50003-0>  
848 <https://www.sciencedirect.com/science/article/pii/B9780123695079500030>
- 849 Nie, L., Wu, G., Culley, D.E., Scholten, J.C., Zhang, W.: Integrative analysis of  
850 transcriptomic and proteomic data: challenges, solutions and applications. Critical  
851 reviews in biotechnology **27**(2), 63–75 (2007)
- 852 Onokpasa, E., Wild, S., Wong, P.W.: Rna secondary structures: from ab initio pre-  
853 diction to better compression, and back. In: 2023 Data Compression Conference  
854 (DCC), pp. 278–287 (2023). IEEE
- 855 Paro, R., Grossniklaus, U., Santoro, R., Wutz, A.: Introduction to Epigenetics.  
856 Springer, ??? (2021)
- 857 Pio-Lopez, L., Valdeolivas, A., Tichit, L., Remy, É., Baudot, A.: Multiverse: a multi-  
858 plex and multiplex-heterogeneous network embedding approach. Scientific reports  
859 **11**(1), 8794 (2021)
- 860 Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow,  
861 C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., et al.: The prisma 2020  
862 statement: an updated guideline for reporting systematic reviews. *Bmj* **372** (2021)
- 863 Pang, B., Weerd, J.H., Hamoen, F.L., Snyder, M.P.: Identification of non-coding  
864 silencer elements and their regulation of gene expression. *Nature Reviews Molecular  
865 Cell Biology* **24**(6), 383–395 (2023)
- 866 Ranganathan, P., Aggarwal, R.: Study designs: Part 3-analytical observational studies.  
867 *Perspectives in clinical research* **10**(2), 91 (2019)
- 868 Raghavachari, N., Garcia-Reyero, N.: Gene Expression Analysis. Springer, ??? (2018)
- 869 Ramakrishnaiah, Y., Macesic, N., Webb, G., Peleg, A.Y., Tyagi, S.: Ehr-qc: A  
870 streamlined pipeline for automated electronic health records standardisation and  
871 preprocessing to predict clinical outcomes. *Journal of Biomedical Informatics*,  
872 104509 (2023)
- 873 Sapoval, N., Aghazadeh, A., Nute, M.G., Antunes, D.A., Balaji, A., Baraniuk, R., Bar-  
874 beran, C., Dannenfelser, R., Dun, C., Edrisi, M., et al.: Current progress and open  
875 challenges for applying deep learning across the biosciences. *Nature Communications*  
876 **13**(1), 1728 (2022)

- 877 Schiano, C., Benincasa, G., Infante, T., Franzese, M., Castaldo, R., Fiorito, C., Man-  
878 sueto, G., Grimaldi, V., Della Valle, G., Fatone, G., Soricelli, A., Nicoletti, G.F.,  
879 Ruocco, A., Mauro, C., Salvatore, M., Napoli, C.: Integrated analysis of dna methy-  
880 lation profile of hla-g gene and imaging in coronary heart disease: Pilot study. PLOS  
881 ONE **15**(8), 1–16 (2020) <https://doi.org/10.1371/journal.pone.0236951>
- 882 Schafer, J.L.: Analysis of Incomplete Multivariate Data. CRC press, ??? (1997)
- 883 Saelens, W., Cannoodt, R., Todorov, H., Saeys, Y.: A comparison of single-cell  
884 trajectory inference methods. Nature biotechnology **37**(5), 547–554 (2019)
- 885 Sachdev, K., Gupta, M.K.: A comprehensive review of feature based methods for drug  
886 target interaction prediction. Journal of biomedical informatics **93**, 103159 (2019)
- 887 Suresh, H., Gutttag, J.: A framework for understanding sources of harm throughout  
888 the machine learning life cycle. In: Equity and Access in Algorithms, Mechanisms,  
889 and Optimization, pp. 1–9 (2021)
- 890 Subha, D.P., Joseph, P.K., Acharya U, R., Lim, C.M.: Eeg signal analysis: a survey.  
891 Journal of medical systems **34**, 195–212 (2010)
- 892 Stephens, Z.D., Lee, S.Y., Faghri, F., Campbell, R.H., Zhai, C., Efron, M.J., Iyer,  
893 R., Schatz, M.C., Sinha, S., Robinson, G.E.: Big data: astronomical or genetical?  
894 PLoS biology **13**(7), 1002195 (2015)
- 895 Summaira, J., Li, X., Shoib, A.M., Li, S., Abdul, J.: Recent advances and trends in  
896 multimodal deep learning: a review. arXiv preprint arXiv:2105.11087 (2021)
- 897 Soyly, N.N., Sefer, E.: Bert2ome: Prediction of 2'-o-methylation modifications from  
898 rna sequence by transformer architecture based on bert. IEEE/ACM Transactions  
899 on Computational Biology and Bioinformatics (2023)
- 900 Tyagi, S., Vaz, C., Gupta, V., Bhatia, R., Maheshwari, S., Srinivasan, A., Bhat-  
901 tacharya, A.: Cid-mirna: a web server for prediction of novel mirna precursors  
902 in human genome. Biochemical and biophysical research communications **372**(4),  
903 831–834 (2008)
- 904 Tyagi, S.: Technical issues in implementing ai in healthcare. In: Translational Appli-  
905 cation of Artificial Intelligence in Healthcare, pp. 60–70. Chapman and Hall/CRC,  
906 ??? (2023)
- 907 Uffelmann, E., Huang, Q.Q., Munung, N.S., De Vries, J., Okada, Y., Martin, A.R.,  
908 Martin, H.C., Lappalainen, T., Posthuma, D.: Genome-wide association studies.  
909 Nature Reviews Methods Primers **1**(1), 59 (2021)
- 910 Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y.: Graph  
911 attention networks. arXiv preprint arXiv:1710.10903 (2017)

- 912 Vergoten, G., Theophanides, T.: Biomolecular Structure and Dynamics vol. 342.  
913 Springer, ??? (2012)
- 914 Welch, J.D., Hartemink, A.J., Prins, J.F.: Matcher: manifold alignment reveals cor-  
915 respondence between single cell transcriptome and epigenome dynamics. *Genome*  
916 *biology* **18**(1), 1–19 (2017)
- 917 Wilson, C.M., Li, K., Yu, X., Kuan, P.-F., Wang, X.: Multiple-kernel learning for  
918 genomic data mining and prediction. *BMC bioinformatics* **20**, 1–7 (2019)
- 919 Wang, C., Mahadevan, S.: A general framework for manifold alignment. In: 2009 AAAI  
920 Fall Symposium Series (2009)
- 921 Wang, B., Mezlini, A.M., Demir, F., Fiume, M., Tu, Z., Brudno, M., Haibe-Kains, B.,  
922 Goldenberg, A.: Similarity network fusion for aggregating data types on a genomic  
923 scale. *Nature Methods* **11**(3), 333–337 (2014) <https://doi.org/10.1038/nmeth.2810>
- 924 Wang, T., Shao, W., Huang, Z., Tang, H., Zhang, J., Ding, Z., Huang, K.: Mogenet  
925 integrates multi-omics data using graph convolutional networks allowing patient  
926 classification and biomarker identification. *Nature Communications* **12**(1), 3445  
927 (2021)
- 928 Wang, B., Wang, P., Parobchak, N., Treff, N., Tao, X., Wang, J., Rosen, T.: Integrated  
929 rna-seq and chip-seq analysis reveals a feed-forward loop regulating h3k9ac and key  
930 labor drivers in human placenta. *Placenta* **76**, 40–50 (2019)
- 931 Xu, Z., Jin, R., Yang, H., King, I., Lyu, M.R.: Simple and efficient multiple kernel  
932 learning by group lasso. In: Proceedings of the 27th International Conference on  
933 Machine Learning (ICML-10), pp. 1175–1182 (2010). Citeseer
- 934 Xu, W., Xu, M., Wang, L., Zhou, W., Xiang, R., Shi, Y., Zhang, Y., Piao, Y.:  
935 Integrative analysis of dna methylation and gene expression identified cervical  
936 cancer-specific diagnostic biomarkers. *Signal transduction and targeted therapy*  
937 **4**(1), 55 (2019)
- 938 Yang, S., Wang, Y., Lin, Y., Shao, D., He, K., Huang, L.: Lncmirnet: predicting lncrna-  
939 mirna interaction based on deep learning of ribonucleic acid sequences. *Molecules*  
940 **25**(19), 4372 (2020)
- 941 Zhao, L., Dong, Q., Luo, C., Wu, Y., Bu, D., Qi, X., Luo, Y., Zhao, Y.: Deepomix: A  
942 scalable and interpretable multi-omics deep learning framework and application in  
943 cancer survival analysis. *Computational and structural biotechnology journal* **19**,  
944 2719–2725 (2021)
- 945 Zou, H., Hastie, T.: Regularization and variable selection via the elastic net. *Journal*  
946 *of the Royal Statistical Society Series B: Statistical Methodology* **67**(2), 301–320  
947 (2005)



- 948 Zhou, Z.-H.: Ensemble Methods: Foundations and Algorithms. CRC press, ??? (2012)
- 949 Zitnik, M., Nguyen, F., Wang, B., Leskovec, J., Goldenberg, A., Hoffman, M.M.:  
950 Machine learning for integrating data in biology and medicine: Principles, practice,  
951 and opportunities. *Information Fusion* **50**, 71–91 (2019)
- 952 Zhao, X., Wang, X., Xia, W., Li, Q., Zhou, L., Li, Q., Zhang, R., Cai, J., Jian, J.,  
953 Fan, L., *et al.*: A cross-modal 3d deep learning for accurate lymph node metastasis  
954 prediction in clinical stage t1 lung adenocarcinoma. *Lung Cancer* **145**, 10–17 (2020)
- 955 Zhao, X., Wang, X., Xia, W., Zhang, R., Jian, J., Zhang, J., Zhu, Y., Tang, Y.,  
956 Li, Z., Liu, S., *et al.*: 3d multi-scale, multi-task, and multi-label deep learning for  
957 prediction of lymph node metastasis in t1 lung adenocarcinoma patients' ct images.  
958 *Computerized Medical Imaging and Graphics* **93**, 101987 (2021)
- 959 Zhang, X., Xing, Y., Sun, K., Guo, Y.: Omiembed: a unified multi-task deep learning  
960 framework for multi-omics data. *Cancers* **13**(12), 3047 (2021)
- 961 Zhang, X., Zou, Q., Rodriguez-Paton, A., Zeng, X.: Meta-path methods for prioritizing  
962 candidate disease mirnas. *IEEE/ACM transactions on computational biology and*  
963 *bioinformatics* **16**(1), 283–291 (2017)
- 964 Zhang, X., Zhang, J., Sun, K., Yang, X., Dai, C., Guo, Y.: Integrated multi-omics  
965 analysis using variational autoencoders: application to pan-cancer classification. In:  
966 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM),  
967 pp. 765–769 (2019). IEEE
- 968 Zhang, S.-F., Zhai, J.-H., Xie, B.-J., Zhan, Y., Wang, X.: Multimodal representa-  
969 tion learning: advances, trends and challenges. In: 2019 International Conference on  
970 Machine Learning and Cybernetics (ICMLC), pp. 1–6 (2019). IEEE
- 971 Zhu, Q., Zhang, S., Zhang, Y., Yu, C., Dang, M., Zhang, L.: Multimodal time series  
972 data fusion based on ssae and lstm. In: 2021 IEEE Wireless Communications and  
973 Networking Conference (WCNC), pp. 1–7 (2021). IEEE