Review

Multi-modal intermediate integrative methods in neuropsychiatric disorders: A review

Yanlin Wanga,1, Shi Tangb,1, Ruimin Ma, Ibrahim Zamat,c, Yanjie Wei, Yi Pan,a,⇑

aCenter for High Performance Computing, Joint Engineering Research Center for Health Big Data Intelligent Analysis Technology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, Guangdong 518055, China
bLi Chiu Kong Family Sleep Assessment Unit, Department of Psychiatry, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong Special Administrative Region
cUniversity of Chinese Academy of Sciences, Beijing, China

ABSTRACT

The etiology of neuropsychiatric disorders involves complex biological processes at different omics layers, such as genomics, transcriptomics, epigenetics, proteomics, and metabolomics. The advent of high-throughput technology, as well as the availability of large open-source datasets, has ushered in a new era in system biology, necessitating the integration of various types of omics data. The complexity of biological mechanisms, the limitations of integrative strategies, and the heterogeneity of multi-omics data have all presented significant challenges to computational scientists. In comparison to early and late integration, intermediate integration may transform each data type into appropriate intermediate representations using various data transformation techniques, allowing it to capture more complementary information contained in each omics and highlight new interactions across omics layers. Here, we reviewed multi-modal intermediate integrative techniques based on component analysis, matrix factorization, similarity network, multiple kernel learning, Bayesian network, artificial neural networks, and graph transformation, as well as their applications in neuropsychiatric domains. We depicted advancements in these approaches and compared the strengths and weaknesses of each method examined. We believe that our findings will aid researchers in their understanding of the transformation and integration of multi-omics data in neuropsychiatric disorders.

© 2022 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Contents

1. Introduction ................................................................. 6150
2. The characteristics of multi-omics data ............................... 6150
3. The computational tasks ...................................................... 6151
4. Intermediate integration methods ......................................... 6155
   4.1. CA-based integration .............................................. 6155
   4.2. MF-based integration .............................................. 6156
   4.3. sN-based integration ............................................. 6156
   4.4. MKL-based integration ........................................... 6156
   4.5. BN-based integration ............................................ 6157
   4.6. ANN-based integration .......................................... 6157
   4.7. GT-based integration ............................................ 6158
5. Discussion ........................................................................ 6158
6. Conclusion and future perspectives ..................................... 6159
Author contribution .......................................................... 6159
Declaration of Competing Interest .......................................... 6159

⇑ Corresponding authors.
E-mail addresses: yj.wei@siat.ac.cn (Y. Wei), yi.pan@siat.ac.cn (Y. Pan).
1 These authors contributed equally to this study.
1. Introduction

Neuropsychiatric disorders are heterogeneous conditions with a complex molecular etiology, commonly including schizophrenia (SZ), Alzheimer’s disease (AD), Parkinson’s disease (PD), autism spectrum disorders (ASD), bipolar disorder (BD), and major depressive disorder (MDD) [1–3]. In twin studies, the majority of them show high heritability estimates, such as 81% for SZ [4], and 58%–79% for AD [5], yet only a small fraction of this heritability can be recapitulated by genome-wide association analyses (GWAS). Despite considerable efforts, the pathogenesis of neuropsychiatric disorders remains poorly understood [6], owing to the pathophysiology of these disorders being dominated by synergistic interactions between multiple biological entities, such as genetics, proteomics, and neuroimaging. Intra-brain spreading of pathological factors, for example, is commonly observed in specific neuropsychiatric disorders (e.g., α-synuclein (α-Syn) in PD, and amyloid-β (Aβ) and Tau in AD) [7,8]. Many neuropsychiatric disorders are associated with altered brain structure and function [9–11]. Furthermore, phenotypic heterogeneity identified in neuropsychiatric disorders remains unexplained, which could be partly due to restrictive association analysis with a single omics layer.

With the rapid development of neurobiological analytical technologies, the investigation of neuropsychiatric disorders is no longer limited to single omics layer analysis and has expanded to integrative analysis of multi-omics data [12]. Similarly, the computational neuroscience field has shifted to the development of statistical and machine learning models for open-source multi-omics data. Computational methods for neuroimaging [13,14], translational neuroimaging [15], and brain imaging genomics [16,17], for example, have demonstrated potential for joint association analyses, identification of neuro-biomarkers, early disease prediction, data-driven patient stratification, and tailored treatment. Despite the development of various data analysis approaches to assess the flow of information from one omics level to the next, integrating large, heterogeneous, and high-dimensional omics data into appropriate models and techniques remains difficult, particularly in neuropsychiatric disorders [18,19]. As a result, one of the most pressing issues is how to fully exploit the characteristics of these high-throughput data and effectively integrate interactive mechanisms between different omics levels [20,21].

Technological advances have resulted in high-dimensional multi-omics data, allowing for a multiscale understanding of diseases (see Fig. 1) [22]. Whole-genome sequencing, in particular, can identify single-nucleotide polymorphisms (SNPs), copy number variants (CNVs), and structural variations at the genome level [23]. At the intermediate phenotype (or endophenotype) level [16,24,25], magnetic resonance imaging (MRI) (structural, functional, and metabolic) can detect in the brain, white matter integrity, and brain functional activity with high spatial resolution; electroencephalograms (EEG) can record the neural activity of the brain with high temporal resolution; RNA-seq can provide information on gene expression levels and discover alternative splicing, gene fusion, and novel isoforms (transcriptomic level); DNA methylation microarrays can detect CpG, CHH, and CHG sites; miRNA-seq captures expression of noncoding RNA expression levels; protein arrays and mass spectrometers are helpful for detecting the concentration of proteins and metabolites in cerebrospinal fluid (CSF) which can offer insights into the physiological state of the brain; and positron emission tomography/single photon emission computed tomography (PET/SPECT) can allow noninvasive evaluation of functional imaging biomarkers and physiological changes. At the phenotype level, demographic, clinical, and physical measurements can represent a person’s basic information and functional status of patient and help inform treatment decisions and prognosis [26].

Multi-omics data provide complementary perspectives on understanding complex biological systems, but they also pose significant challenges in multi-omics data analysis [27]. As a result, it is critical that we address the massive amounts of data from various sources, which may result in a variety of findings and plausible interpretations. Data integrative analysis, which can combine different sources, scales, and distributions of big data to boost the precision and stability of research results, is one of the most efficient strategies. It is worth noting that data integrative methods are now emerging with the goal of bridging the gap from genotype to phenotype and obtaining in-depth insights into disease biological processes by leveraging distinct (inter-modality) and complementary (inter-modality) information from multi-omics data [28].

Numerous data integration strategies have been developed and are frequently classified into three major approaches: early, intermediate, and late integration [29–31]. Early integration directly combines multi-omics data into a large data matrix, which is then fed into a single model. However, this process produces a more unbalanced, heterogeneous, and high-dimensional matrix, which may cause a number of issues in downstream analysis. In late integration, different modalities are inputted equally into the corresponding model to obtain separated prediction results, which are then integrated to determine the final consensus results. However, this method does not take advantage of the associations between different modalities at the feature level. Unlike early and late integration, intermediate integration can not only effectively extract more informative representations from each modality, but also exploit the interplay of information between intermediate representations from data transformation techniques, thereby addressing the complexities and heterogeneities of original data while retaining the intact information in multi-omics [32,33].

Multi-modal intermediate integrative methods are frequently used to assist researchers in neuropsychiatric fields in understanding how different data types are integrated into distilled information [33]. We systematically summarize the characteristics of various modalities in Section 2, various computational tasks in Section 3, and investigated the applications of various intermediate integrative methods in Section 4. Section 5 discusses the strengths and weaknesses of the investigated methods, as well as potential challenges and future directions, and Section 6 provides our conclusion and additional perspectives.

2. The characteristics of multi-omics data

Understanding the peculiarities of multi-omics data is crucial for the design of omics studies, as it may ensure that the acquired data satisfy the analysis requirements of the integrative approach.
and assist researchers in dealing with various datasets more effectively. In this study, we reviewed and documented the public availability of datasets for neuropsychiatric disorders across seven major omics, namely clinical, neuroimaging, metabolomics, genomics, epigenomics, transcriptomics, and proteomics (see Table 1). Clinical measurements are made up of a series of items that rate participants’ cognitive, emotional, and behavioral performance during the previous 2–3 weeks, which is significant for disease diagnosis. Neuroimaging anatomical features (such as thickness, volume, curvature, and sulcus depth) are frequently derived from brain spatial parcellations (e.g., DK 68-region atlas). The features of regional homogeneity (ReHo), amplitude of low-frequency fluctuations/fractional amplitude of low-frequency fluctuations (ALFF/fALFF), voxel-mirrored homotopic connectivity (VMHC), and degree centrality (DC) are widely used to characterize global functional activity and local spontaneous activity. Furthermore, structural connectivity (SC) refers to the presence of white matter tracts that physically connect brain regions, and functional connectivity (FC) refers to the temporal correlations of neural signals from distinct brain regions [14]. EEG signals are generated by recording cortical position, amplitude, and frequency information at several locations/channels on the scalp. During the experiment, the EEG detects the difference in electrical potential between two sites (generally referred to as active and reference) [34]. Genomic data is often represented as A, T, C, and G sequences. Many hundreds of thousands of specific sites in the genome have genetic variants (e.g., SNPs and CNVs) observed. SNPs are polymorphisms that involve variations of a single base pair and are often recorded as binary variables, whereas CNVs indicate multiple genetic variances between individuals [35]. Transcriptomics features include tens of thousands of numerical gene expression data points acquired through a variety of preprocessing workflows based on sequencing data [36]. Methylation profiling can reveal abnormal methylation markers for disease incidence. Continuous variables are frequently used to encode methylation status (e.g., a CpG site is either methylated or unmethylated in an individual) [37]. Protein sequence, protein domain, and protein–protein interaction (PPI) data are the most common proteomics features. The Basic Alignment Search Tool’s bitescores are utilized for numerical protein sequence representation [38]. Protein domain similarity is assessed using Jaccard scores between protein domains [39]. PPI is defined by its projected biological score and IntAct MI score. PBS has a score ranging from 0 to 10 (the smaller score, the more specific the interaction). The IntAct MI score is based on the manual annotation of every instance of a binary interaction and the cumulative score is normalized between 0 and 1 across the entire IntAct database [38,39]. Abnormal metabolic alterations in CSF and plasma (e.g., glucose, amyloid beta(A), and hyperphosphorylated tau protein (p-tau)) are among the hallmarks of metabolomics [40,41]. Magnetic resonance spectroscopy (MRS) and positron emission computed tomography/single photon computed emission tomography (PET/SPECT) (such as N-acetylaspartate, 18F-fluorodeoxyglucose (18F-FDG), and presynaptic dopamine active transporters (DATs)) can detect metabolites or metabolic tracer retention or reduction [42,43].

### 3. The computational tasks

Another crucial component of planning a multi-omics study is conceptualizing the aim of the omics study and the computational task before analyzing the data [44]. Multi-omics integrative analyses have been applied to a variety of downstream tasks, including patient classification, subtyping, association, and causal inference. Using multi-omics data, classification can differentiate patients from normal populations, and relevant factors can be established as molecular biomarkers, which play an important role in disease...
diagnosis and prevention. Omics-based subtyping is primarily concerned with identifying potential patient subgroups through data-driven methods. Subtyping aims to overcome disease heterogeneity and advance our understanding of neurobiological mechanisms. Multi-omics association analysis can capture interactions across multiple omics types and reveal new information about the relationship between genomic variation and clinical phenotypes [22]. Furthermore, based on prior knowledge, inference analysis can address causal effects between different layers, which is frequently used to represent the hierarchical relationships across multiple omics layers. Overall, an empirical approach is required to investigate the application of various integration models in

---

| Database name   | ND | Clinical | Neuroimaging | Metabolomics | Genomics | Epigenomics | Transcriptomics | Proteomics | Link                          |
|-----------------|----|----------|--------------|--------------|----------|------------|----------------|------------|-------------------------------|
| ADNI            | AD | ✓        | ✓            | ✓            | ✓        | ✓          |                |            | http://adni.loni.usc.edu      |
| PPMI            | PD | ✓        | ✓            | ✓            | ✓        | ✓          |                |            | https://www.ppmi-info.org    |
| ABIDE-I         | ASD| ✓        | ✓            | ✓            | ✓        |            |                |            | https://fcon_1000.projects.nitrc.org/indi/abide/abide_i.html |
| ABIDE-II        | ASD| ✓        | ✓            | ✓            | ✓        |            |                |            | https://fcon_1000.projects.nitrc.org/indi/abide/abide_ii.html |
| ABCD            | NC | ✓        | ✓            | ✓            | ✓        | ✓          |                |            | https://nda.nih.gov/abcd     |
| ADHD-200        | ADHD| ✓        | ✓            | ✓            | ✓        |            |                |            | https://fcon_1000.projects.nitrc.org/indi/adhd200/ |
| FBIRN           | SZ | ✓        | ✓            | ✓            | ✓        |            |                |            | https://fbirndr.nbirn.net:8080/BDR |
| UNM&MCIC & COBRE| SZ | ✓        | ✓            | ✓            | ✓        |            |                |            | https://coins.mrn.org         |
| ROSMAP (AMP-AD) | AD | ✓        | ✓            | ✓            | ✓        | ✓          |                | ✓          | https://adknowledgeportal.synapse.org |
| TADPOLE         | AD | ✓        | ✓            | ✓            | ✓        | ✓          |                | ✓          | https://tadpole.grand-challenge.org/ |
| TGEN            | Human| ✓      | ✓            | ✓            | ✓        |            |                |            | https://www.tgen.org/         |
| HBTRC           | Human| ✓      | ✓            | ✓            | ✓        |            |                |            | https://hbtrc.mclean.harvard.edu/ |
| MEM             | Human| ✓      | ✓            | ✓            | ✓        |            |                |            | https://bio.tools/mem         |
| AHBA            | Human| ✓      | ✓            | ✓            | ✓        |            |                |            | https://human.brain-map.org/  |
| Ahlgenz         | AD | ✓        | ✓            | ✓            | ✓        |            |                |            | https://www.ebi.ac.uk/intact/home |
| IntAct          | Human| ✓      | ✓            | ✓            | ✓        |            |                |            | https://ahlgenz.abs.csiro.au/adni/index.html |
| AIBL            | AD | ✓        | ✓            | ✓            | ✓        |            |                |            | https://www.sanger.ac.uk/Software/Pfam |
| PFAM            | Human| ✓      | ✓            | ✓            | ✓        |            |                |            | https://www.uniprot.org       |
| UniPort         | Human| ✓      | ✓            | ✓            | ✓        |            |                |            | https://www.omim.org/         |

NC normal control, AD alzheimer's disease, HD huntington's disease, SZ schizophrenia, ASD autism spectrum disorder, ADHD attention deficit hyperactivity disorder, PD parkinson disorder, MDD major depression disorder, BD bipolar disorder, SCA spinocerebellar ataxia, cSVD cerebellar small vessel disease.
| Categories | Methods | Author | Application | Task | Omics datatypes | Link |
|------------|---------|--------|-------------|------|----------------|------|
| CA-based   | MCCA    | Sui et al. [45] | SZ | classification | sMRI, fMRI and EEG |  |
|            |         | Sui et al. [46] | AD | association | Cognitive test, sMRI, fMRI, and dMRI |  |
|            |         | Pillai et al. [47] | SZ | association | sMRI, PET and CSF |  |
|            | CaMCCo  | Singanamalli et al. [48] | AD | classification | Cognitive test, sMRI, PET, CSF, APOE |  |
| jICA      |         | Adali et al. [51] | SZ | association | fMRI, sMRI and EEG |  |
| Para-ICA  |         | Vergera et al. [53] | AD | association | fMRI, sMRI and SNPs |  |
| SCCA      | SCA     | Garali et al. [49] | PD | association | sMRI, dMRI, SPECT, and DNA methylation |  |
| JCB-SCCA  |         | Kim et al. [50] | PD | association | sMRI, dMRI, SPECT, and DNA methylation |  |
| MCCA+     | jICA    | Sui et al. [52] | SZ | association | Cognitive test, sMRI, fMRI, and dMRI |  |
| MCCA+     | jICA    | Sui et al. [55] | SZ | association | Cognitive test, sMRI, fMRI, and dMRI |  |
| HGSCCA    |         | Qi et al. [56] | SZ | association | Cognitive test, sMRI, dMRI and fMRI |  |
| MF-based  | RMSMF   | Que et al. [59] | AD | classification | sMRI, PET, and CSF |  |
| jNMF      |         | Wang et al. [58] | AD | association | SNPs, fMRI and DNA methylation |  |
| GSJNMF    |         | Wang et al. [60] | SZ | association | SNPs, fMRI, and DNA methylation |  |
| JCB-SNFM  |         | Wei et al. [61] | AD | association | sMRI, SNPs, and mRNA |  |
| SN-based  | SNF     | Tong et al. [63] | AD | classification | sMRI, PET, CSF, and APOE |  |
|           |         | Stefanik et al. [64] | SZ, ASD, and BD | Clustering | demographic, clinical scores, sMRI |  |
|           |         | Markello et al. [65] | PD | Clustering | clinical scores, SPECT, CSF and sMRI (CT and SV) |  |
|           | SNFM    | Seidritz et al. [68] | AD | Clustering | sMRI (CT, SA, GM, MC, and IC) |  |
|           |         | Morgan et al. [67] | SZ | association | sMRI (CT, SA, GM, GC, MC) and dMRI (FA and MD) |  |
|           |         | Li et al. [69] | MDD | association | sMRI (CT, SA, GM, GC, MC) and dMRI (FA and MD) |  |
| MKL-based | MKL     | Ye et al. [70] | AD | classification | Cognitive test, sMRI, and APOE |  |
|           |         | Zhang et al. [71] | AD | classification | sMRI, PET, and CSF |  |
|           |         | Gupta et al. [72] | AD | classification | sMRI, PET, CSF and APOE |  |
|           |         | Zhou et al. [73] | AD | classification | sMRI, rsMRI, and dMRI |  |
|           |         | Young et al. [74] | AD | classification | sMRI, PET, CSF and APOE |  |
|           |         | Liu et al. [75] | AD | classification | Cognitive test, sMRI, PET, CSF, and APOE |  |
| SimpleMKL |         | Hinrichs et al. [76] | AD | classification | fMRI (Multi-atlas) |  |
| GL-MKL    |         | Liu et al. [77] | AD | classification | sMRI (HIP, HIPR, GM), CSF |  |
| SSR-MKL   |         | Peng et al. [78] | AD | classification | sMRI, PET, and SNPs |  |
| ML-MVC    |         | Zhang et al. [79] | AD | classification | sMRI, PET, and SNPs |  |
| Ensembled MKL model | Shi et al. [80] | AD | classification | sMRI, PET, and CSF |  |
| BN-based  | Discrete BN | Ding et al. [81] | AD | classification | 33 heterogeneous features (demographics items, medicine, cognitive test, sMRI, PET, CSF, and ApoE) |  |
|           |         | Jin et al. [82] | AD | classification | 33 heterogeneous features (demographics items, medicine, cognitive test, sMRI, PET, CSF, and ApoE) |  |
|           |         | Li et al. [83] | AD | inference | sMRI, PET, CSF, APOE and cognitive tests |  |
|           |         | Li et al. [84] | AD | (longitudinal data) | sMRI, PET, CSF, APOE and cognitive tests |  |
| Two-steps BN | Batmanghelich et al. [85] | AD | inference | Phenotype, sMRI, and SNPs |  |

fALFF fractional amplitude of low-frequency fluctuations, GM gray matter, FA fractional anisotropy, SNPs single-nucleotide polymorphisms, CSF cerebrospinal fluid, APOE apolipoprotein E, Aβ amyloid beta, HIPL left hemisphere hippocampus shape, HIPR right hemisphere hippocampus shape.
| Categories | Methods | Author | Application | G(v,E) | Omics datatypes | Link |
|------------|---------|--------|-------------|--------|-----------------|------|
| ANN-based  | Stacked RBM | Li et al. [86] | AD | sMRI, PET, and CSF | – |
|            | Stacked AE/ DAE | Shi et al. [87] | AD | sMRI (GM, 12-month DM, and Aseg) | https://github.com/rasmusbergpalm/DeepLearnToolbox |
|            | Wang et al. [88] | AD | fMRI (Multi-atlas) | Cognitive test, sMRI, and SNPs |
|            | Venugopalan et al. [89] | AD | Cognitive test, sMRI, PET, and CSF |
|            | Suk et al. [91] | AD | sMRI, PET and CSF |
|            | Suk et al. [92] | AD | sMRI (CT, SA, CV, LGI, SD, GH, CURV) and MFN properties (CC, K, and GE) |
|            | mcDNN | Chen et al. [90] | ADHD | sMRI, fMRI, dMRI |
|            | Three-stage DNN | Zhou et al. [93] | AD | sMRI, PET and SNPs |
|            | Two-step DBN | Masoudi et al. [94] | SZ | sMRI, fMRI, dMRI |
| GT-based   | Graph theory | Zheng et al. [97] | AD | sMRI (CT, SA, CV, LGI, SD, GH, CURV) and MFN properties (CC, K, and GE) |
|            | Zheng et al. [98] | AD | sMRI (CT, SA, CV, LGI, SD, GH, CURV) and MFN properties (CC, K, and GE) |
|            | Wang et al. [99] | AD | Cognitive test, sMRI, and PET |
| Matrix      | factorization | Tian et al. [39] | AD and SZ | GO (CC, MF, BP), domain and sequence of protein, genes and Phenotype |
| Graph       | embedding | Parisot et al. [100] | AD and ASD | sample-phenotype |
| GCN         | Arya et al. [101] | ASD | sample-sMRI | sMRI, fMRI (ALFF, fALFF, Autocorrelation, Dcw, Dcb, Entropy, Ecb, Ecw, ReHo, VMHC, LFCdb, LFCdw) |
|            | Wang et al. [102] | ASD | sample-fMRI | fMRI (multi-atlas) |
|            | Zhang et al. [103] | PD | ROI-dMRI (BCG),Region coordinates-Euclidean distances (BGG) | dMRI (different algorithms) |
|            | MWGCN | Wang et al. [104] | AD and ASD | sample-non-imaging |
|            | GraphSAGE | Sugis et al. [38] | AD | Genes, proteins, SNPs-PPIs, epistatic interactions, co-expression interactions |
|            | InceptionGCN | Kaiz et al. [105] | AD and ASD | Demography, cognitive test, fMRI, PET, CSF, and APOE |
|            | TGCN | Yao et al. [106] | ADHD, ASD | Cognitive test, sMRI, SNPs, mRNA, PPIs, GWAS, and PS |
|            | MMTGCN | Yao et al. [107] | ADHD, AD, and cSVD | fMRI(multi-atlas) |
|            | MORONET | Wang et al. [108] | AD | sMRI, mRNA, miRNA, DNA methylation |
|            |            |            |            | mRNA, miRNA, DNA methylation, |

ALFF Amplitude of Low Frequency Fluctuations, fALFF fractional Amplitude of Low Frequency Fluctuations, Dcw Degree centrality weighted, Dcb Degree centrality binarize, Ecb Eigenvector centrality binarize, Ecw Eigenvector centrality weighted, ReHo Regional Homogeneity, VMHC Voxel-Mirrored Homotopic Connectivity, (k) LFCDb Local Functional Connectivity Density binarize, LFCDw Local Functional Connectivity Density weighted, MFN multi-feature-based network, DM deformation magnitude, GM gray matter, SNPs single-nucleotide polymorphisms, CSF cerebrospinal fluid, APOE apolipoprotein E, Aβ amyloid beta, LGI local gyrification index, SD sulcal depth, GH gyrus height, GE global efficiency, K degree, CC clustering coefficient, GWAS genome-wide association studies, PPIs protein–protein interactions, PS positive selection, GO Gene Ontology, CC cellular component, MF molecular function, BP biological process.
neuropsychiatric disorders, as well as the strengths and weaknesses of each method, in order to select an appropriate integration method for the collected dataset and the focused task.

4. Intermediate integration methods

We classified multimodal intermediate integrative strategies into seven categories based on their nature: component analysis (CA)-based, matrix factorization (MF)-based, similarity network (SN)-based, multiple kernel learning (MKL)-based, Bayesian network (BN)-based, artificial neural network (ANN)-based, and graph transformation (GT)-based (see Fig. 2). We cover SNF-based integrative approaches in a distinct section because they are fundamentally GT-based integrative methods and are mostly given in the context of biological patient network mining. We sought to characterize the state of the art in this field for each approach (see Table 2/3) as well as their distinct strengths and shortcomings (see Table 4), with the purpose of identifying how these methods might be applied in future studies to improve the multi-modal intermediate integrative strategy.

4.1. CA-based integration

Blind source separation techniques are a common unsupervised approach that can retrieve low-dimensional components from high-dimensional data under minimum assumptions using statistical properties. Multivariate methods based on blind source separation techniques, such as canonical correlation analysis (CCA), partial least squares (PLS), and parallel independent component analysis, have found widespread application in data fusion (para-ICA). However, these methods can only decompose and relate a pair of linearly connected matrices.

Some studies have looked into multi-modal intermediate integrative approaches that capture cross-modality interactions as well as spatial components between different data types. Sui et al. [45,46] used multi-set CCA (MCCA) to combine data from Multi-modal MRI and (or) EEG to elucidate the abnormalities that underpin SZ patients and covary across multiple modalities in which functional features are associated with clinical symptoms. Pillai et al. [47] compared three integrative approaches for AD classification that combined sMRI, PET, and CSF biomarkers. The MCCA method had the highest accuracy on the ADNI database, followed by the MF-based and MKL-based integrative methods.

Singanamalli et al. proposed the cascaded Multiview canonical correlation (CaMCo) framework to maximize correlations between multiple modalities and class labels in order to optimize classifier performance at each cascade level and fuse a subset of multiple modalities for AD diagnosis [48]. Moreover, Garali and Kim et al. used a sparse CCA (SCCA) algorithm on multimodalities to assess associations and identify imaging genetics biomarkers of spinocerebellar ataxia (SCA) and Parkinson’s disease (PD) [49,50]. Adali et al. [51], on the other hand, used fMRI, sMRI, and EEG data to identify a set of components that report on differences between SZ and controls performing an auditory oddball task using the joint independent component analysis (jICA) and the transposed independent vector analysis (tIVA) separately. As a result, multi-modal CCA with jICA (MCCA + jICA) has become a popular extension of jICA, which is used to combine multi-modal MRI data to distinguish SZ patients from healthy controls (HCs) [52]. Furthermore, Vergara et al. used three-way para-ICA to optimize the independence and linkage of fMRI, sMRI, and EEG data at each cascade level and fuse the highest accuracies on the Multi-Modal MRI integration using prior knowledge as a reference to identify multi-modal biomarker signatures of SZ [55,56].

| Categories | Strengths | Weaknesses |
|------------|-----------|------------|
| CA-based   | Easy to implement | Loss of information |
|            | Capture cross-information | Local optimum |
| MF-based   | Easy to implement | Ignore nonlinear relationship |
|            | Capture sparse and interpretable factors | Prior assumption |
|            | Tolerance to heterogeneity data | Non-negative input data |
| SN-based   | Easy to implement | Loss of information |
|            | Deal with patient network attributes | Local optimum |
| MKL-based  | Dimension-free | Ignore nonlinear relationship |
|            | Nonparametric modeling | Prior assumption |
| BN-based   | Firm probabilistic foundation | Computationally more intensive |
|            | Handle uncertainty | Ignore interactions of datasets |
|            | Robust to noisy data | Independence assumption |
| ANN-based  | Adaptive | Require expert knowledge |
|            | Nonparametric modeling | Data normalization |
|            | Robust to noise | Local optimum |
| GT-based   | Deal with Non-Euclidean data | High computation cost |
|            | Simultaneously learning node and edge attributes | Low interpretability |
|            | representations Local and (or) | Require large scale datasets |
|            | global structural information | Lack the flexibility |

Table 4
The strengths and weaknesses of seven intermediate integration algorithms.

Para-ICA to optimize the independence and linkage of fMRI, sMRI, and EEG data at each cascade level and fuse the highest accuracies on the Multi-Modal MRI integration using prior knowledge as a reference to identify multi-modal biomarker signatures of SZ [55,56].
4.2. MF-based integration

Nonnegative matrix factorization (NMF) is a matrix low-rank approximation method that can extract sparse and interpretable factors from a set of nonnegative data matrices [57,58]. An NMF matrix $X$ can be approximated by the inner product of two non-negative matrices $U$ and $V$:

$$X \approx UV$$

The joint non-NMF (jNMF) is an advanced NMF model for multiple data, as follows:

$$\min_{U,V} \sum_{i=1}^{n} ||X_i - UV_i||_F^2$$

where $\|\cdot\|_F$ is the Frobenius norm, $U \in \mathbb{R}^{m \times r}$ store the basis column vectors and $V \in \mathbb{R}^{n \times r}$ stores the corresponding column coefficient vectors. $r$ is the number of the basis vectors.

Early studies used joint non-NMF (jNMF) and regularized NMF (rNMF) to conduct integrative analysis of multi-modal data in order to extract combinatorial patterns. For example, Que et al. proposed a regularized multi-source matrix factorization (RMSMF) model that can reconstruct multi-modal data (MRI, PET, and CSF) and reduce noise to improve AD classification performance even further. [59]. Moreover, Wang et al. employed a jNMF model, which simultaneously factorized SNP, fMRI, and methylation matrices to identify significant biomarkers associated with SZ [58]. However, they did not take into account specific group structures in the data. As a result, they proposed a group sparse jNMF (GSjNMF) model to address this issue, which incorporates prior knowledge by enforcing group sparse constraints into the corresponding coefficient matrices in the model [60]. Despite the fact that this algorithm added sparse group information to jNMF in order to identify hidden dependent structures between different data modalities, it overlooked brain connection information. Wei et al. proposed joint-connectivity-based sparse NMF (JCB-SNMF) to make full use of connection information, in which connectivity information for each region of the brain and genetic data are added as prior knowledge to identify regions of interest (ROIs), SNPs, and gene-related risks associated with AD patients [61]. In summary, when compared to CA-based integrative methods, the findings of NF-based methods can intuitively capture sparse factors with non-negative constraints. However, both methods may converge to a local optimum and are incapable of extracting nonlinear structures and relationships hidden in multi-modal data.

4.3. sN-based integration

SN-based integration can combine different modalities into a single unified generative modality by leveraging structural attributes and topological information. Similarity network fusion (SNF) is a common example of this fusion method, which frequently generates sample similarity matrices using a scaled exponential similarity kernel:

$$W(i,j) = \exp\left[-\frac{r^2(x_i, x_j)}{\mu_{ij}}\right]$$

where $W(i,j)$ is a measure of similarity between patients $i$ and $j$, $\mu$ is a hyperparameter, and $\epsilon_{ij}$ is used to eliminate the scaling problem.

Following that, all matrices are combined using mean value, iterate fusion, and added weights. To be more specific, a full matrix is first defined by performing a normalization as follows:

$$P(i,j) = \frac{W(i,j)}{\sum_{k \in K} W(k,l)}$$

Then, for each similarity matrix, a sparse kernel is built using the following equation:

$$S_{ij} = \sum_{k \in N_i} \frac{W(i,k)}{\sum_{l \in N_k} W(l,k)} J_{kN_i}$$

where $N_i$ represents a set of neighbors of patients $i$ including itself.

Finally, the key step in SNF is to iteratively update all matrices in the following manner:

$$P_m^{(v)} = S_{m}^{(v)} \times \left[\sum_{k \in N_i} P_m^{(k)}\right]^{-1} \times (S_{m}^{(v)})^T, \quad v = 1, 2, \ldots, m$$

where $P_m^{(v)}$ denotes the learned status matrix of the $v$th data type after $m$ iterations. $S_{m}^{(v)}$ denotes sparse kernel matrix of the $v$th data type. $P_m^{(v)}$ is the status matrix of the $v$th data type after $m$ iterations. $T$ is the transpose operation.

The integrated matrix can be fed into the ML models for prediction, classification, and clustering. For example, Wang proposed SNF to integrate different data types by iteratively updating patient similarity networks [62]. Consequently, some studies applied SNF to fuse different modalities (i.e., demographic, clinical, neuroimaging data, and genetic information) into a unified similarity matrix that can be used for the subtyping and classification of neuropsychiatric disorders (AD, SZ, ASD, BD, PD) [63–65].

To better leverage the Allen Human Brain Atlas (AHBA) microarray expression data, Seidlitz proposed the morphometric similarity network (MSN). This ROI-based data integrative method can characterize brain regions by integrating multi-modal MRI measures into one mean inter-regional correlation value [66]. As a result, Morgan, Seidlitz, and Li used MSN to investigate the expression of disorder-associated genes that correlate with MSN differences in various neuropsychiatric disorders (e.g., SZ, neurodevelopmental disorders, MDD) [67–69].

Because their overall size is based on one common dimension of multi-omics or is patient-based, the number of features has no effect on the workflow of the methods described above. SNF, on the other hand, uses the network structure's similarity information to fuse the obtained multiple similarity matrices while ignoring the significance of the original feature attributes. Furthermore, similarities determined by Euclidean distances may fail to capture complex topological structures.

4.4. MKL-based integration

The MKL is a popular integrative strategy that combines a set of basis kernels by assigning specific kernels and weights to different modalities. Support vector machines (SVMs) are the most widely used kernel-based learning methods for classification. They can maximize the marginal distance in the mapping space of features using the discriminant function $f(x) = w^T \varphi(x) + b$, where $w$ and $b$ are hyperparameters that separate two classes and $\varphi(\cdot)$ is the mapping function. When the Lagrangian dual function is employed, the optimization problem of $w$ can be formulated into:

$$f(x) = \sum_{i=1}^{N} \alpha_i k(\varphi(x_i), \varphi(x_i)) + b$$

where $k(\varphi(x_i), \varphi(x_i))$ is kernel function.

MKL methods can use multiple kernels instead of one specific kernel function:

$$k_p(x_i, x_j) = f_p\left(\left(k_m^p(x_i, x_j)\right)_{m=1}^{p}\right)$$

6156
where the combination function (i.e., a linear or a nonlinear function) \( f_g \) forms a single kernel from \( p \) base kernels using the parameters vector (i.e., weights) \( \eta \).

Kernel integration aims to combine multiple basis kernels by maximizing the similarity between kernels and minimizing kernel weights. Based on these two directions, some optimization methods for MKL have been studied; they mainly yield different kernel selection and/or combination approaches. For instance, Ye et al. combined the feature selection technique and MKL to select and integrate sMRI and demographic and genetic data for AD prediction and discovered useful biomarkers [70]. Another solution to MKL is to perform a grid search and select combined kernel weights that minimize cross-validation error. For example, some studies applied a multitask-based joint feature learning framework to combine the neuroimaging, PET, CSF, and genetic information to classify AD and attention deficit hyperactivity disorder (ADHD) from HCs [71–73].

Moreover, Young applied the Gaussian process (GP) approach that aids the combination of multi-modal data by learning hyper-parameters automatically from training data via type-II maximum likelihood to predict conversion to AD in patients with mild cognitive impairment (MCI) patients [74]. Liu et al. applied a two-step MKL method based on the MKBoost-S2 algorithm to combine six optimal feature subsets and obtained an optimal classifier to perform AD classification [75]. However, this approach is very time-consuming and even intractable when the number of kernels or features is large. Thus, most MKL schemes tend to yield sparse weightings. Hinrichs et al. applied simple MKL to discriminate between different stages of AD using PET, sMRI, other biological measures, and cognitive status measures [76]. Although simple MKL can find sparse kernel combinations, it is susceptible to noise and may discard useful information. Consequently, Liu et al. proposed an MKL framework based on the group Lasso to enhance group sparsity between multi-modal features (CSF and MRI multi-modalities) for AD classification [77]. Contrary to group Lasso (i.e., \( L_2, 1\)-norm penalty), which performs sparse group selection, Peng proposed a novel structured sparsity regularized MKL (SSR-MKL) to acquire a superior multi-modal feature combination by jointly selecting subsets of discriminative features from each modality for AD classification [78].

By combining the different characteristics of kernel functions, we can use the aforementioned MKL methods to benefit from multi-kernel functions and obtain improved mapping performance. Furthermore, one of the key advantages of kernel methods is that their optimizations are not affected by sample size, which is known as dimension-free optimization. These methods, however, typically learn latent representations separately, rarely taking into account the interaction of both feature and modality levels during modeling. To address this issue, Zhang et al. proposed a multi-layer multi-view classification (ML-MVC) method to simultaneously explore the complex input–output correlation and high-order correlation across multiple kernel matrices of different views for AD classification [79]. To achieve a similar goal, Shi et al. proposed an ensemble framework by leveraging the coupled interactions of the feature and modality levels between different modalities (i.e., MRI, PET, and CSF) to improve the performance of AD/MCI diagnosis [80].

4.5. BN-based integration

BN is a probabilistic graphical model that employs directed acyclic graphs (DAGs) to represent variable dependency and conditional probability tables (CPTs) to describe the local probability distribution involving a node and its parents. BNs are well-suited for combining multi-omics data and expert knowledge, exploiting unknown causal relationships, and modeling the latent probability distribution of data sources due to their graphical representation and use of probabilistic theory. If the observations \( X_i \) are statistically independent of one another, the graph structure of BN can represent a joint probability of a domain as:

\[
P_r(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P_r(X_i|\text{Par}_X_i)
\]

where \( \text{Par}_X_i \) denotes the parents of \( X_i \), \( P_r(X_i|\text{Par}_X_i) \) and is a set of CPTs for each \( X_i \).

The current BN-based integrative method has been widely reported in the literature on neuropsychiatric disorders. For example, Ding et al. applied a discrete BN-based integrative framework to efficiently identify key features from coarse-grained data and understand the probabilistic dependencies between multiple AD factors and their changes over time [81]. Jin et al. used a mixed-type BN that can handle both discrete and continuous variables to learn the influential relationships between demographics, sMRI, PET, neuropsychometric tests, and genotypes and provide superior diagnostic and prognostic information [82]. Li et al. applied a latent time join mixed-effects model (LTJMM) to understand the progression of AD using PET imaging, p-tau, A\( _\beta \), and other outcome measures. The proposed model considers inter- and intra-subject variations and the associations between the outcomes [83,84]. Batmanghelich et al. proposed a two-step BN framework that simultaneously captures the associations between genetic markers, brain endophenotype features, and disease phenotypes. This is done to detect genetic variants associated with AD by exploiting image-based features as an intermediate phenotype in a probabilistic model [85]. BN has the advantage of a firm probabilistic foundation. BN-based integrative methods provide a suitable framework for combing heterogeneous data with expert knowledge. They do, however, necessitate the specification of prior distributions as well as computational cost, which may limit their applicability.

4.6. ANN-based integration

Generally, ANN technology in multi-modal data is specific to feature representation learning and deep representation integration. Deep representation learning, as opposed to traditional shallow learning architecture, can learn high-level and effective data representations from multiple modalities. Among the deep representation learning methods developed for multi-modal integration, some common integrative frameworks include denoising sparse autoencoder (DASAE) integration, sparse restricted Boltzmann machine (SRBM) integration, and encoder-decoder models.

Previous studies attempted to construct a hierarchical architecture using an ANN model, which frequently produces better representations, and then integrate them using an MKL model. Li et al., for example, incorporated stability selection and multitask deep learning with dropout into the deep learning framework to identify different progression stages of patients with AD based on MRI, PET, and CSF features [86]. Another solution to ANN data integration is to encode each modality using a stacked autoencoder for potential representation, and the new representation is mapped into a shared layer to obtain a converged representation of features. Shi et al. compared a multi-modal stacked DASAE (MM-S-DASAE) model with four other integrative strategies to improve the identification of AD and MCI [87]. Similarly, Wang et al. applied a stacked denoising autoencoder (DAE) on multi-atlas FCs to learn deep representations separately and used a multilayer perceptron (MLP) with the result voting method to identify ASD/HC [88]. Venugopalan et al. obtained the intermediate features from MRI, gene, and clinical data using a stacked DAE model, and then integrated them into a concatenation layer followed by a classification layer to predict...
the AD stage [89]. In addition, Chen et al. developed a multichannel deep neural network (mcDNN) to identify AHAD using multiscale brain functional connectome data (multi-atlas) [90]. Multiple parallel channels take multiple inputs, fuse them into a signal channel to extract the combined high-level features, and feed them to a classifier. However, in such workflows, single-modal high-level features were learned regardless of the other modalities, suggesting that the synergy between different modalities in feature learning was overlooked.

To reduce the risk of falling into a local optimum, Suk et al. employed a stacked autoencoder (AE) to extract feature representations for MRI, PET, and CSF separately and selected the augmented feature vectors, that is, concatenation of the original low-level features and stacked AE-learned features, using different strategies. Finally, the selected intermediate features were fit to an MKL classifier for AD classification [91,92]. As a result, Zhou et al. proposed a novel three-stage deep neural network (DNN) framework for learning and integrating latent feature representations from MRI, PET, and SNP datasets in order to identify AD and its prodromal status [93]. Similarly, Masoudi et al. proposed a two-step deep belief network (DBN) scheme with a pairwise integration rule to acquire intra-modality and inter-modality information based on multi-modal MRI data for computer-aided SZ/HC diagnosis [94]. One disadvantage of existing methods is that most of them only consider data attributes within each modality while ignoring cross-modality correlations.

4.7. GT-based integration

The graph provides a natural way to represent a population and its interactions [95]. Graph-based representations can convert graph variables (e.g., nodes, edges, and subgraphs) into new representations while preserving their structural and topological properties to the greatest extent possible. The main goal of GT-based integrative algorithms is to capture graph similarity and preserve relevant topological and geometric information between the constituent parts or primitives. GT-based integrative algorithms can be categorized into traditional graph embedding and graph neural networks-based graph embedding methods [96].

In previous studies, traditional graph representations were learned for graph structures from omics data by relying on well-known embedding approaches, such as graph theory, matrix factorization, and random walk techniques. In graph-theoretic techniques, feature representation summarizes node-level attributes in a graph in terms of well-known network features such as degree, closeness, centrality, betweenness, network efficiency, and modularity. Zheng et al. combined multiple morphological features (i.e., MFN) with seven morphological features and other high-level MFN properties (i.e., network properties) to enhance the classification performance of AD and ASD [97,98]. Matrix factorization can reduce data dimensionality while preserving its structure and does so by factorizing graph Laplacian eigenmaps or node proximity matrices. For example, Wang et al. proposed a graph-based transductive learning method that updates the affinity matrix of each observed imaging feature and aligns with clinical data to a common space for AD classification, which can more efficiently integrate different datasets for better explanations of clinical outcomes [99]. Random walks can be used to simulate an imaginary particle that starts at one node and moves randomly to other nodes in order to explore the network topology. After several iterations, a stationary probability distribution is obtained that depicts the seed node's topological properties. For example, Tian et al. [39] employed random walks with a restart algorithm on a phenotype-gene bilayer network, which combines a phenotype similarity network and an integrated gene similarity network constructed using the SNF method, to detect similar nodes and infer causal genes for AD and SZ.

These traditional graph representation approaches have been shown to be effective in learning graph structures across multiple omics datasets. One disadvantage of these algorithms is that they are only used to embed the structure of the graph and not to represent node attributes. It has been demonstrated that graph convolutional networks (GCNs) outperform other traditional methods for many classification tasks when compared to shallow embedding methods. For example, Parisot et al. [100] applied GCN methods to combine MRI data and pairwise interactions using phenotypic data by running node classifications on a population graph for AD classification. Alternatively, Arya et al. utilized relational information from sMRI data as compared to phenotypic data together with multiple FMRI data for ASD classification using GCNs [101]. Wang et al. proposed a multi-atlas graph convolutional network method (MAGCN) based on different brain atlases to identify ASD patients and then combined different feature representations using a stacking ensemble learning method [102]. In general, this group of baseline methods either requires a common network structure or utilizes a constant filter size for node-based classification, which limits the ability to generalize using GCNs and addresses the problem of graph heterogeneity. To address these limitations, Zhang et al. proposed a multi-view GCN model to learn feature representations from six types of brain connectivity features and fuse population graphs with different views, which can capture both the local traits of each graph and the global traits of the shared feature graph [103]. Kazi et al. proposed inception GCN with multi-sized filters, a novel architecture capable of capturing local and global information of heterogeneous graphs and choosing optimal feature representations for disease prediction [104]. As an alternative, Sugis et al. applied the GraphSAGE model by integrating various genes, SNPs, and protein datasets into heterogeneous networks to identify AD-related genes and molecular interactions [38]. However, these studies only focused on heterogeneous networks to learn embeddings from different edge types separately, without considering the biological relationships between different feature sets. To this end, Yao et al. employed a triplet GCN (TGCN) to learn multi-scale graph representations of brain functional connectivity based on multi-scale templates for AD and ASD classifications [105]. To make use of complementary spatial and topological information from a multi-atlas, they consequently proposed a mutual multi-scale triplet graph convolutional network (MMTGCN) for AD, ADHD, and cSVD diagnosis based on functional or structural connectivity networks [106]. For a similar goal, Wang et al. proposed a novel supervised multi-omics integrative framework based on GCN to jointly explore omics-specific learning and cross-omics correlation learning for AD classification [107].

Overall, most existing studies used a common unifying dimension to transform multiple modalities into homogeneous networks, which cannot effectively model high-order associations among samples. Although some studies have attempted to use a heterogeneous network for disease classification [108], building a more sophisticated graph structure from multi-omics datasets remains challenging. Furthermore, because GCN models are susceptible to data variance, designing appropriate GCN models may be crucial for the performance of subsequent analyses.

5. Discussion

Multi-modal intermediate integrative techniques have the ability to transform and incorporate multi-omics data, providing holistic and comprehensive insights into disease biological mechanisms. This is noteworthy because it bridges the gap between multi-omics sources and obtains more distilled informa-
tion to elucidate disease heterogeneity and etiology. Our survey describes the characteristics of different modalities and computational tasks, and provides a snapshot of seven multi-modal intermediate integrative methods and their applications in neuropsychiatric disorders, which can help researchers choose an appropriate method based on the problem at hand. Overall, this review can serve as an accessible entry point into this field for researchers.

Massive effort has gone into developing and applying multi-modal intermediate integrative algorithms to neuropsychiatric disorders for various downstream tasks such as classification, prediction, disease subtyping, association, and causal inference [109]. Although no single best approach has emerged from these integrative methods, various strategies can be adapted to perform an integrative analysis depending on the research purpose and data characteristics [110]. CA-based and MF-based integrative methods are frequently developed under the assumption that different data sets share a common latent space and investigate cross-modality correlations between multimodal data [27]. Structure attributes from different data types can be modelled and integrated into a single unified data format using sN-based integration [62]. MKL-based integration is a versatile method that employs a weighted sum of kernels, with the weight of each kernel optimized non-linearly [111]; BN-based integration provides a rigorous probabilistic graphical framework that allows the combination of various data types and expert knowledge to yield a complete description of the inference pathway and local joint probability distributions [112].

ANN-based and GT-based integrations are deep learning-based methods for discovering very complex interactions between multi-omics data by using non-linear modules and message passing between these modules [113–115]. Notably, GT-based integration offers a powerful solution for learning graph representations by leveraging graph structures and aggregate node information from neighborhoods, reflecting the complex interplay between multiple levels of biological systems [96,116,117]. Although most intermediate integrative methods require that each dataset be transformed into an intermediate form, such as a common component/matrix or a graph, and then used for modeling [70,73,104,105,107], some variants of the model may use ensemble or hybrid frameworks to link and integrate different data types for the best solution [52,54,55,60,61]. Depending on the data characteristics and downstream tasks, various integrative models and variants can be chosen to uncover complex associations between different data types and overcome the substantive issues of heterogeneity and confounding.

Considering the complex interaction between different data types, the heterogeneous multi-modal graph neural network has the potential to model complex disease systems in future applications [114]. However, most studies have either focused on homogeneous networks or learned independent representations for each modality, with only a few studies attempting to construct heterogeneous multi-modal network omics features based on different omics features [105,106]. For example, Yao et al. proposed a heterogeneous graph attention convolution network (HGACN) by leveraging intra-metapath and inter-metapath attention to aggregate and learn useful graph embeddings from heterogeneous graphs with multiple types of nodes and edges for ADHD classification [108]. Nonetheless, constructing a multiscale heterogeneous network from multi-omics datasets that can depict the entire biological process from genotype to phenotype is extremely coveted. Furthermore, the interaction between the graph structure and node attributes must be taken into account, as these characteristics are not independent of human disease processes. Recent research has attempted to incorporate multi-modal graph embedding techniques (i.e., variational autoencoders) into graph convolution networks in order to capture the joint information between graph structure and node attributes [118]. These techniques, however, remain primary and have not yet been applied to the neuropsychiatric domain.

Although deep transformation methods have gained popularity due to their ability to investigate low-level to high-level feature attributes and associations, it is still difficult to explain the relationships between inputs and outputs due to their black-box nature [119,120]. While data transformation and integrative strategies can refer to relevant expert knowledge to improve interpretability [121]. The basic idea behind this approach is to utilize known knowledge to guide the reconstruction of the network or as immediate layers of the model [27]. It is preferable to refer to known expert knowledge of multi-omics data to explore the underlying biological interactions, but neuropsychiatric diseases are complex systems with mixed effects of multi-omics factors, making it difficult to make proper assumptions. On the other hand, these methods cannot identify new biomarkers if some associations have not yet been discovered. In addition, the attention mechanism is a popular interpretation strategy used with deep learning models, especially in GCNs, where attention to different parts of the sequence input on the output can be used as a feature [122]. However, compared with the interpretation models in the ontology domain, their applications in neuropsychiatric disorders are still primitive.

Multi-modal intermediate integrative strategies may provide a more efficient way of analyzing multiple datasets at the same time, but there are several limitations to be aware of. For starters, high-dimensional small sample size datasets in neuropsychiatric domains may have poor stability and generalizability, making clinical application of research results difficult. Despite the fact that researchers have shifted their focus to large and shared data resources from big data consortiums, they should be aware that different big data sites, scales, and formats may end up causing more noise, imbalance, and inconsistency. Second, while various data integrative techniques have been implemented to address the complexity and heterogeneity of multi-omics data, there has been little literature comparing those methods using the same publicly available datasets. Finally, in this review, intermediate integrative methods were divided into seven categories based on their nature and underlying mathematical aspects, despite the fact that some methods, such as sN-based, MKL-based, and GT-based integrative methods, may be intertwined with others.

6. Conclusion and future perspectives

Given that multi-modal integrative strategies may provide an understanding of the full spectral etiology in neuropsychiatric disorders, three aspects of consideration are required for omics studies: the acquisition and analysis of multi-omics data, the advancement of data integrative technologies, and the design of computational tasks. This review systematically summarizes these general experimental factors, allowing researchers to get a head start in this domain. Furthermore, future researchers should focus more on quantifying and validating findings before applying them to clinical practice.

Author contribution

YL.W. contributed to the conception and design of the study. YL.W. and S.T. performed study search and selection. YL.W. drafted the manuscript. S.T., RM.M, and I.Z. participated in the discussion of methods and results. YJ.W. and Y.P. reviewed the manuscript and provided comments. All authors have read and approved the final version of the submitted manuscript.
Declaration of Competing 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.

Acknowledgments
The authors would like to express their gratitude to their tutors and colleagues for their time and assistance. Ibrahim Zamit acknowledges support from the ANSO Scholarship for Young Talents.

Financial support
This study was supported by the Key Research and Development Project of Guangdong Province (Grant No. 2021B0101310002), Shenzhen Science and Technology Program (No. KQTD2020082113106007), China Postdoctoral Science Foundation (No. 2021M703375), National Science Foundation of China (No. 62272449), Strategic Priority CAS Project (No. XDB38050100), the Shenzhen Basic Research Fund (No. RCYX202071411474319), and Youth Innovation Promotion Association (No. Y2021101).

References
[1] Shen L, Thompson PM. Brain imaging genomics: integrated analysis and machine learning. Proc IEEE Eng Med Biol Soc 2020;2010(1):125–62. https://doi.org/10.1109/EMBC.2020.9294722
[2] Forrest MP, Parrell E, Penezes P. Dendritic structural plasticity and neuropsychiatric disease. Nat Rev Neurosci 2018;19(4):215–34. https://doi.org/10.1038/s41583-018-0116-6
[3] Ulfherrmann E, Posthuma D. Emerging methods and resources for biological interpretation of neuropsychiatric polygenic signal. Bioinformatics 2021;38(1):41–53.
[4] Sullivan PF, Kendler KS, Neale MC. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. Arch Gen Psychiatry 2003;60(12):1187–92. https://doi.org/10.1001/archpsyc.60.12.1187
[5] Gatz M et al. Role of genes and environments for explaining Alzheimer’s disease. (in eng). Arch Gen Psychiatry 2006;63(2):168–74. https://doi.org/10.1001/archpsyc.63.2.168
[6] Rees E, Owen MJ. Translating insights from neuropsychiatric genetics and genomics for precision psychiatry. Genome Med 2020;12(1):1–43. https://doi.org/10.1186/s13073-020-00744-6
[7] Stefanis L. Genetic mechanisms of neurodegeneration and neuroprotective strategies. Hum Brain Mapp 2020;41(12):3371–5. https://doi.org/10.1002/hbm.25875
[8] Murphy MP, LeVine 3rd H. Alzheimer’s disease and the amyloid-beta peptide. J Alzheimers Dis 2010;19(1):311–23. https://doi.org/10.3233/jad-2010-1221
[9] Kam GL et al. Development of structure-function coupling in human brain networks during youth. Proc Natl Acad Sci U S A 2020;117(1):771–8. https://doi.org/10.1073/pnas.1912034117
[10] Burhan AM, Marlatt NM, Palaniyappan L, Anazodo UC, Prato FS. Role of hybrid brain imaging in neuropsychiatric disorders. Diagnostics (Basel) 2015;5(6):773–81. https://doi.org/10.3390/diagnostics5060773
[11] Zhang X, Yang Y, Li T, Zhang Y, Wang H, Fujita H. CMC: A consensus multi-omics integrative analysis perspective clustering model for predicting Alzheimer’s disease progression. Med Res Rev 2021;199:1–34. https://doi.org/10.1002/med.21875
[12] Zhai J, Zhang Y, Hu Y, Yang H, Wang L, Wang C, Yang H, Wang H. Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder, Science, vol. 362, no. 6420, Dec 14 2018, doi: 10.1126/science.aat8127.
[13] W. Zhang, T. D. Spector, P. Deloukas, J. T. Bell, and B. E. Engelhardt, Predicting genome-wide DNA methylation using methylation marks, genomic position, and DNA regulatory elements, Genome Biol, vol. 16, no. 1, 14 Jan 2015, doi: 10.1186/s13059-015-0581-9.
[14] Sugi et al., HENA, heterogeneous network-based data set for Alzheimer’s disease, Sci Data, vol. 6, no. 1, 151, Aug 14 2019, doi:10.1038/s41597-019-0152-0.
[15] Z. Tian, M. Guo, C. Wang, L. Xing, L. Wang, and Y. Zhang, Constructing an integrated gene similarity network for the identification of disease genes, J Biomed Semantics, vol. 8, no. Suppl. 1, 32 Sep 20 2017, doi: 10.1186/s13326-017-0141-1.
[16] Verberk IMW et al. Plasma Amyloid as Prescreeener for the Elderly Alzheimer Pathological Changes. Ann Neurol 2018;84(5):548–53. https://doi.org/10.1002/ana.25216
[17] Palmqvist S et al. Cerebrospinal fluid and plasma biomarker trajectories with increasing amyloid deposition in Alzheimer’s disease. EMBO Mol Med 2019;11(1):e11700.
[18] Gong J, Yuan Z, PENG J, Molecular imaging in clinical neuroscience: recent advances in the investigation of CNS diseases. Quantitative imaging in medicine and surgery 2015;5(3):343.
[19] B. K. Puri, MRI and MRS of Neuropsychiatry, MagnRes, 2007.
[20] Mei X, Seldon M, Lusis A. Multi-omics approaches to disease. Genome Biol 2017;18(1):1–15.
[21] Sui J, Hunter R, Yu Q, Segall JM, Calhoun VD. Function-structure associations of the brain: evidence from multimodal connectivity and covariance studies. Neuroimage 2014;102(1):1–11. https://doi.org/10.1016/j.neuroimage.2013.09.044.
