OXFORD

Briefings in Bioinformatics, 2025, **26(2)**, bbaf088
https://doi.org/10.1093/bib/bbaf088
Case Study

A novel integrative multimodal classifier to enhance the diagnosis of Parkinson's disease

Xiaoyan Zhou[1], Luca Parisi[2], Wentao Huang[1], Yihan Zhang[1], Xiaoqun Huang[1], Mansour Youseffi[3], Farideh Javid[4], Renfei Ma[1,*]

[1]Department of Biology, Shenzhen MSU-BIT University, Longcheng Street, Shenzhen 518115, Guangdong, China
[2]Department of Computer Science, Tutorantis, 5 South Charlotte Street, Edinburgh EH2 4AN, United Kingdom
[3]Department of Engineering and Informatics, University of Bradford, Richmond Road, Bradford BD7 1DP, United Kingdom
[4]Department of Pharmacy, University of Huddersfield, Queensgate, Huddersfield HD1 3DH, United Kingdom

*Corresponding author. Faculty of Biology, Shenzhen MSU-BIT University, Longcheng Street, Shenzhen 518115, Guangdong, China.
E-mail: marenfei@smbu.edu.cn

Abstract

Parkinson's disease (PD) is a complex, progressive neurodegenerative disorder with high heterogeneity, making early diagnosis difficult. Early detection and intervention are crucial for slowing PD progression. Understanding PD's diverse pathways and mechanisms is key to advancing knowledge. Recent advances in noninvasive imaging and multi-omics technologies have provided valuable insights into PD's underlying causes and biological processes. However, integrating these diverse data sources remains challenging, especially when deriving meaningful low-level features that can serve as diagnostic indicators. This study developed and validated a novel integrative, multimodal predictive model for detecting PD based on features derived from multimodal data, including hematological information, proteomics, RNA sequencing, metabolomics, and dopamine transporter scan imaging, sourced from the Parkinson's Progression Markers Initiative. Several model architectures were investigated and evaluated, including support vector machine, eXtreme Gradient Boosting, fully connected neural networks with concatenation and joint modeling (FCNN_C and FCNN_JM), and a multimodal encoder-based model with multi-head cross-attention (MMT_CA). The MMT_CA model demonstrated superior predictive performance, achieving a balanced classification accuracy of 97.7%, thus highlighting its ability to capture and leverage cross-modality inter-dependencies to aid predictive analytics. Furthermore, feature importance analysis using SHapley Additive exPlanations not only identified crucial diagnostic biomarkers to inform the predictive models in this study but also holds potential for future research aimed at integrated functional analyses of PD from a multi-omics perspective, ultimately revealing targets required for precision medicine approaches to aid treatment of PD aimed at slowing down its progression.

Keywords: Parkinson's disease; multimodal data; cross-attention; deep learning; classification; transformer

Introduction

Parkinson's disease (PD) is the second most prevalent neurodegenerative disorder after Alzheimer's disease (AD), characterized primarily by motor symptoms such as bradykinesia, hand tremor, freezing gait, muscle stiffness, and balance impairments [1, 2]. In addition, some individuals experienced non-motor symptoms, including speech disorders like dysarthria [3], which contributed to the complexity of the disease [4]. The defining feature of PD is the progressive loss of dopamine-producing neurons in the brain [5], which underpins the current therapeutic approaches of administering dopamine precursors like levodopa to address dopamine deficiency [6, 7]. Given the progressive nature of PD and the variability in symptoms among patients, early diagnosis is essential for initiating timely treatments to slow disease progression and improve patients' quality of life [8].

Traditional diagnostic methods relied heavily on clinical and cognitive observations [9], such as speech analysis [10] and spiral drawing assessments [11, 12]. However, these approaches often overlook molecular-level changes and the underlying etiologies that underlie PD. The prevalence of PD varies significantly across different regions, largely influenced by genetic and environmental factors [4, 9]. Recent studies on circular RNA [13] and noncoding RNA [14] linked PD pathogenesis to specific metabolic processes [15]. Furthermore, the hematological long noncoding RNA-messenger RNA (mRNA) co-expression network identified immune-related biomarkers, likely playing a key role in the immune response associated with PD [16]. Another significant factor in the development of neurodegenerative diseases was pathological protein aggregates. For instance, the accumulation of misfolded α-synuclein was a central feature in PD pathology, contributing to its progression [17]. Advances in single-cell sequencing further elucidated the association between specific cell populations and PD through single-cell RNA sequencing [18]. Consequently, with the development of multi-omics analysis techniques encompassing imaging, genomics, transcriptomics, proteomics, and metabolomics, more studies have begun to explore PD's molecular and cellular mechanisms. As traditional methods fail to fully capture the intricate interplay of various biological factors contributing to PD, leading to delayed

Received: October 22, 2024. **Revised:** December 30, 2024. **Accepted:** February 18, 2025

or missed diagnoses, a multimodal classification approach addresses these challenges by integrating diverse data types captured across various modalities, such as RNA sequencing (RNA-seq), proteomics, metabolomics, hematology, and imaging [e.g. dopamine transporter scan (DaTscan)]. Each modality provides unique, complementary insights into different biological layers of the disease, from molecular alterations to physiological dysfunctions. The integrative approach enables a holistic understanding of PD by leveraging the complementary information from these modalities, improving diagnostic accuracy and early detection/treatment to slow down the progressive nature of this neurodegenerative disorder.

Despite these advancements, integrating multi-omics data into coherent functional pathways for studying PD pathogenesis remained challenging. As the volume of available data grew, machine learning (ML) and deep learning methods emerged as powerful tools for PD classification and diagnosis due to their ability to process large datasets and improve diagnostic accuracy [19–21]. However, relatively few studies leveraged the full range of available biological data, with many focusing on only a few data modalities [22]. For instance, [23] combined clinical data with DaTscan imaging for PD classification. Similarly, Höglinger *et al.* [24] proposed a biologically based PD classification system incorporating pathological α-synuclein in tissues or cerebrospinal fluid (CSF), neuroimaging, and pathogenic gene variants. Furthermore, few analytical frameworks existed in the literature for comprehensively integrating multiple data modalities in PD research.

Multi-omics data provide a more comprehensive understanding of the molecular mechanisms behind specific biological processes or complex diseases than models that use single-omics data [25, 26]. Deep learning-based multi-omics fusion methods have been widely used in cancer studies [27–29]. However, as mentioned in the Introduction, multi-omics data integration remains relatively underexplored in PD-related research. Each modality captures unique aspects of the PD, but together, they form a more comprehensive picture. For example, RNA-seq provides insights into gene expression levels, offering a snapshot of transcriptional activity. However, not all mRNA transcripts are translated into proteins, so mRNA profiling does not capture regulatory processes or posttranscriptional modifications that might affect the amount of active protein [30], as proteins undergo posttranslational modifications (PTMs) that impact their function. Proteomics bridges this gap by capturing the protein-level expression and modifications [31], enabling a deeper understanding of how gene expression translates into functional changes. Proteins, especially enzymes, directly regulate metabolic pathways, influencing metabolite concentrations. Metabolism plays an instructive role in cell signaling or gene expression, as the stoichiometries of PTMs caused by a metabolite typically depend on its cellular concentration [32]. Similarly, metabolic changes observed in PD often have systemic effects that can be reflected in blood markers captured by hematological data. Therefore, although each data modality captures different PD biological or clinical aspects, complementary information can be obtained from other modalities. A fused method using a combination of multi-omics data enables a more comprehensive study of complex biological processes and highlights the interrelationship of relevant data modalities [27].

In the literature, three types of multimodal data fusion, namely, early fusion, intermediate fusion, and late fusion, were proposed for integrating various data sources, as summarized by [33]. Standard techniques like data concatenation often fail to capture interactions between data modalities effectively. Given the limitations of standard techniques, a similarity network fusion approach was introduced to integrate multiple data types [34]. Transformer models, widely applied in imaging and natural language processing (NLP) [35], were also explored in PD research, although primarily focused on audio or imaging data [36, 37]. With their powerful cross-attention mechanism [38], transformer models might have the potential to capture intrinsic interactions between multiple biological modalities, thereby enhancing PD classification.

In this study, hematological information, proteomics, RNA-seq, metabolomics, and DaTscan imaging data were obtained from the Parkinson's Progression Markers Initiative (PPMI) to investigate PD classification. Differential expression analysis was performed on the RNA-seq data to identify genes with significantly altered expression between PD patients and healthy controls. After data cleaning, features from each modality were used as input for various ML or deep learning models, including support vector machines (SVMs), eXtreme Gradient Boosting (XGBoost), fully connected neural networks (FCNNs), and multimodal transformers with cross-attention (MMT_CA). Finally, SHapley Additive exPlanations (SHAP) values and learnable modality scaling weights were employed to explore the contributions of each feature and modality to model predictions, enhancing the interpretability of the predictive models.

Materials

Dataset

The dataset used in this study was obtained with approval from the PPMI database (https://www.ppmi-info.org/) as of 24 June 2024. RNA-seq counts were analyzed to identify differentially expressed genes, and transcripts per million (TPM) values were then normalized for these genes. Blood test parameters available for most participants were retained and normalized according to their respective reference ranges. Although the PPMI database includes CSF and plasma proteomics across various projects, proteomics data from Project 115, which featured CSF tests on participants with PD, Prodromal, and Healthy Control statuses, were chosen to ensure a good balance between sample size and data availability. Additionally, metabolomics data and DaTscan imaging results were extracted. In total, 306 participants (192 with PD and 114 healthy controls) were retained for further analysis. For participants with multiple clinical visits, the average value for each data type was used, and all data were normalized prior to model training. The overview of the data modalities used in the predictive models of this study can be found in the Supplementary Table S1.

Methods

Schematic of the proposed method

Figure 1 depicts a schematic of this study that includes data collection and preprocessing, ML- and deep learning-based PD classification model development, and model evaluation.

In the literature, SVM was widely used in PD detection and other ML-based classification models [39–41]. Known for its exceptional predictive and computational performance, XGBoost [42] was also commonly used in neurodegenerative diseases classification [43, 44]. Moreover, XGBoost is optimized for large datasets, making it suitable for high-dimensional data. Therefore, ML-based SVM and XGBoost were chosen in this study for PD classification with multiple features. The grid search technique was leveraged for hyperparameter optimization in both SVM and

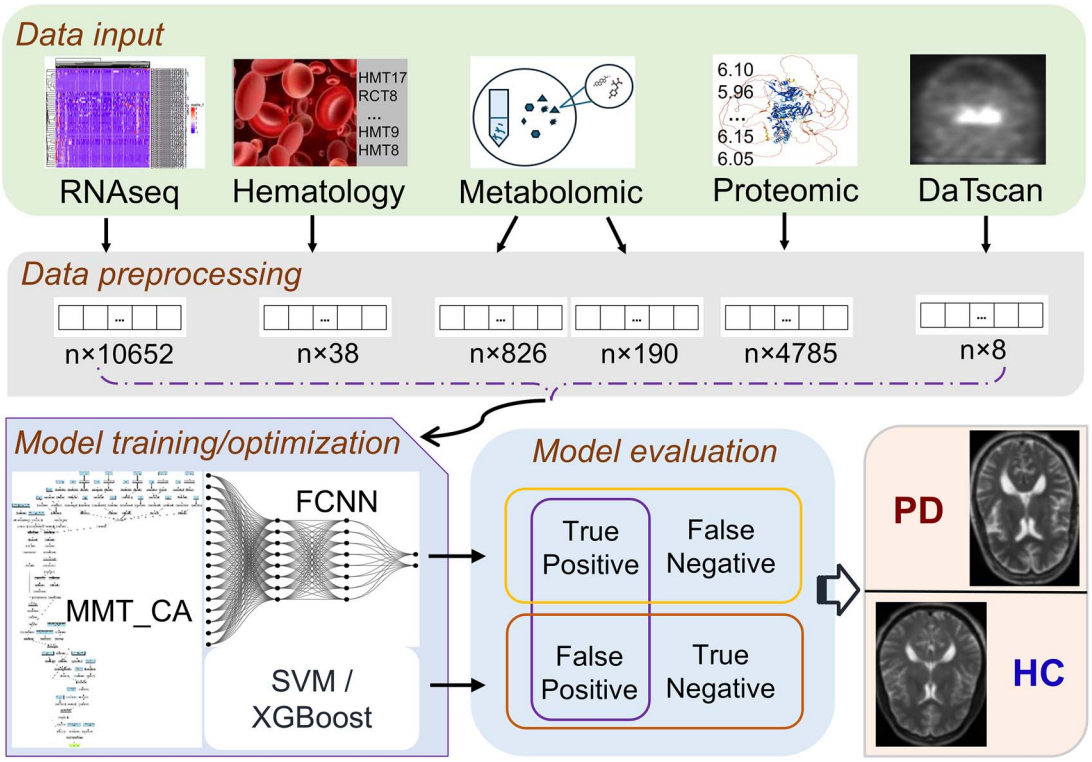

Figure 1. The predictive modeling workflow for PD classification incorporated RNA-seq, hematological information, metabolomics, proteomic, and DaTscan imaging data, exploring various model architectures. HC, healthy control.

Table 1. Grid search parameters for SVM and XGBoost models

Model	Hyperparameter	Values
SVM	C	[0.1, 1, 10]
	kernel	["linear," "rbf," "poly"]
	gamma	["scale," "auto"]
XGBoost	n_estimators	[50, 100, 150]
	learning_rate	[0.01, 0.1, 0.2]
	max_depth	[3, 5, 7]
	subsample	[0.8, 1.0]

XGBoost models in the model training process. Details of the grid search parameters for these models can be found in Table 1.

In addition to the above two ML-based methods, deep learning-driven frameworks were leveraged in this study. Since five data types were considered, different data fusion techniques and model architectures were explored for multimodal deep learning models.

As mentioned in the Introduction section, different multimodal data fusion techniques were proposed in the literature. In this study, concatenation and joint modeling methods were used. The concatenation method combined hematological information, protein, RNA-seq, metabolomics part 1, metabolomics part 2, and DaTscan imaging features into a single feature matrix, which was then fed into FCNN. In contrast, the joint modeling approach processed each feature type separately through the initial layers of the FCNN and merged them in a joint layer [45]. The models were referred to as FCNN_C and FCNN_JM, respectively.

As highlighted in a recent review of transformer-based models in neuroscience, neurology, and psychiatry [35], transformer-based models had primarily been applied to imaging data [46, 47] and signal data [48]. However, transformer-based models incorporating cross-attention mechanisms demonstrated strong performance in various other classification tasks [49]. While some studies focused on PD diagnosis using transformer-based models on specific modalities like imaging and audio data [50] or genomics data [51], there is still limited research on integrating comprehensive biological data modalities for PD classification using transformer-based model architecture. The cross-attention mechanism [52] was particularly useful for capturing multimodal features. Thus, this study employed an MMT_CA model architecture to address this gap in PD classification. Further details of the MMT_CA model are provided in the following section.

Multi-head cross-attention

Each data modality in this study, namely, hematological information, protein, RNA-seq, metabolomics, and DaTscan imaging, had its dedicated cross-attention module. These modules allowed each modality to attend to and integrate information from the others. Furthermore, multi-head attention enabled the model to focus on information from multiple feature representation sub-spaces simultaneously, capturing diverse patterns at various scales and across different positions [38].

Given the query matrix Q, key matrix K, and value matrix V, the multi-head attention can be expressed as [38]

$$MultiHead(Q, K, V) = Concat(head_1, \ldots, head_h)W^O, \quad (1)$$

where

$$\begin{aligned} head_i &= Attention\left(QW_i^Q, KW_i^K, VW_i^V\right) \\ &= softmax\left(\frac{QW_i^Q(KW_i^K)^T}{\sqrt{d_k}}\right)VW_i^V, \end{aligned} \quad (2)$$

where $W_i^Q, W_i^K \in \mathbb{R}^{d_{model} \times d_k}$, $W_i^V \in \mathbb{R}^{d_{model} \times d_v}$, and $W^O \in \mathbb{R}^{hd_v \times d_{model}}$.

Table 2. Key general model parameters, optimization parameters, and regularization parameter of the MMT_CA model

Parameter	Description	Value
Transformer_dim (D)	Embedding dimension for transformer layers	512
Num_heads (H)	Number of attention heads in multi-head attention	8
Dropout	Dropout rate applied in attention and feed-forward networks	0.1
Learning_rate	Learning rate for the optimizer	1.00e-04
Optimizer	Optimization algorithm used	Adam
Dropout_rate	Dropout rate to prevent overfitting	0.1

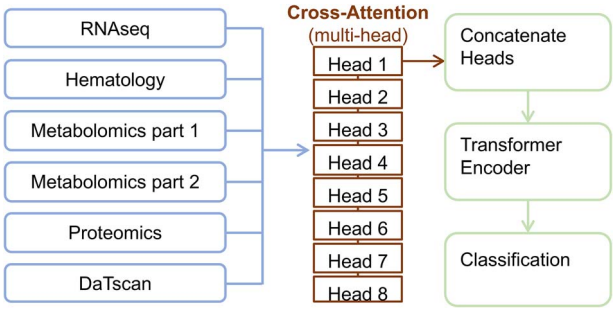

Figure 2. Features from each modality were input into a cross-attention module, where outputs from all eight heads were concatenated into a 512-dimensional vector, which was then passed through a transformer encoder for PD classification.

In this study, the number of attention heads is set to 8, with each head having dimensions $d_k = d_v = d_{model}/h$, following standard practice in the literature [38].

After processing each modality through its respective cross-attention module, the MMT_CA model aggregated the enriched embeddings and fed them into a transformer encoder to capture higher level interactions.

A schematic diagram illustrating the flow within the MMT-CA model is shown in Fig. 2. Briefly, the features of each modality were fed into their respective cross-attention modules, each consisting of eight heads processing the data in parallel. The outputs from these heads were then concatenated and passed through a transformer encoder for further processing before being used for PD classification. A detailed introduction of the multi-head cross-attention mechanism can be found in the study of [38]. Briefly, six modalities were leveraged in this study, as related to hematological information, proteomics, RNA-seq, metabolomics, and the DaTscan imaging, and each modality was passed through a separate cross-attention layer, which would interact with the other modalities. For instance, the cross-attention mechanism allowed the proteomic modality to interact with the other five modalities and vice versa. Since eight heads were used in this study, the attention heads help the model learn eight aspects of the relationships between these modalities.

The key model parameters, along with the optimization and regularization parameters of the MMT_CA model, are summarized in Table 2. During training, learnable scalar weights were assigned to each modality to scale its feature vector before integration. The weights were initially set to 1 and updated through back-propagation. These learnable weights provide insights into the contribution of each modality to the model's predictive performance.

Model evaluation

The model evaluation and comparison relied on several commonly used metrics, including balanced accuracy, weighted precision, weighted recall, weighted F1-score, and the Area Under the Receiver Operating Characteristic Curve (ROC-AUC). These metrics, in their raw form, are defined as follows and were then balanced and weighted accordingly to consider imbalances in the data, which are typical of medical diagnostics data:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{3}$$

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

$$F1\_score = \frac{2 \times precision \times recall}{precision + recall}, \tag{6}$$

where TP, TN, FP, and FN, respectively, denote true positives, true negatives, false positives, and false negatives.

Results

RNA-seq analysis

With RNA-seq analysis, the DESeq2 dispersion plot displaying the relationship between gene mean expression and variance and the enhanced Volcano plot highlighting the relationship between statistical significance and the magnitude of fold change (FC) for genes are shown in Fig. 3(A) and (B), respectively. The dispersion plot shows a clear downward trend. Genes with an adjusted P-value <0.05 were considered statistically significant. In total, there were 10 652 genes retained, and the TPM values of these significant genes were normalized and passed as features for model training. A more detailed description of this RNA-seq data source can be found in two studies mentioned by the PPMI [53, 54].

Performance comparison between different model architectures

A comparison of the results obtained from leveraging different model architectures in this study is provided in Table 3. The corresponding evaluation metrics were assessed comprehensively across all available data via five-fold cross-validation. The FCNN_C model, which utilizes a simple data concatenation approach, yielded poor performance in PD classification. In contrast, the joint modeling architecture (FCNN_JM), involving separate FCNNs for hematological information, proteomics, RNA-seq, metabolomics, and DaTscan imaging data, followed by a joint fusion layer, significantly enhanced predictive performance, increasing the accuracy from 65.0% to 87.5%. While both ML-based models achieved relatively high predictive accuracy, the XGBoost model outperformed the SVM model, achieving an accuracy of 96.7%, as compared with 83.0%. By applying grid search-based hyperparameter optimization, the optimal performance

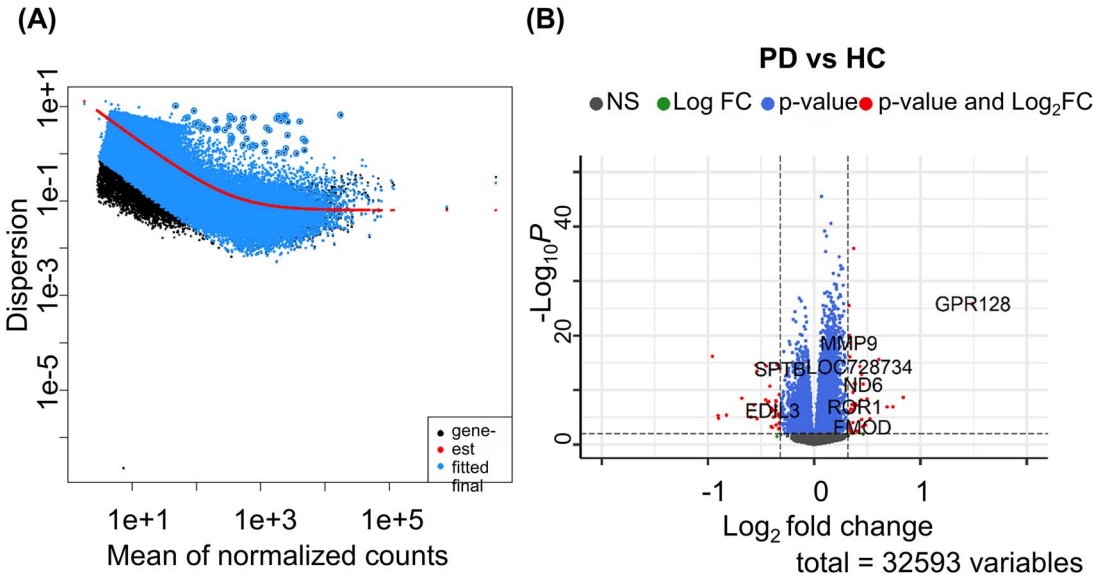

Figure 3. RNA-seq analysis results. (A) DESeq2 dispersion plot displaying the relationship between gene mean expression and variance. (B) Enhanced volcano plot highlighting differentially expressed genes, with representative genes labeled. The fold-change cutoff (FCcutoff) was set to 0.32, and the significance threshold (pCutoff) was set at 10e-3.

Table 3. Comparison of mean predictive performance across different model architectures using accuracy, weighted average precision, recall, F1 score, and ROC-AUC. The metrics were scaled from 0 to 1. Thus, they can be easily converted into percentages by multiplying them by 100. Acc, accuracy

Models	Acc	Precision	Recall	F1 score	AUC
SVM	0.830 ± 0.031	0.834 ± 0.028	0.830 ± 0.031	0.828 ± 0.032	0.914 ± 0.024
XGBoost	0.967 ± 0.023	0.967 ± 0.023	0.967 ± 0.023	0.967 ± 0.023	0.988 ± 0.015
FCNN_C	0.650 ± 0.044	0.662 ± 0.044	0.650 ± 0.044	0.650 ± 0.042	0.696 ± 0.059
FCNN_JM	0.875 ± 0.042	0.887 ± 0.033	0.875 ± 0.042	0.874 ± 0.043	0.924 ± 0.044
MMT_CA	**0.977 ± 0.017**	**0.977 ± 0.017**	**0.977 ± 0.017**	**0.977 ± 0.017**	**0.994 ± 0.008**

of the SVM model was attained with parameters set to $C = 0.1$, *gamma* = *scale* and *kernel* = *linear*, and the XGBoost model yielded its best performance with *n_estimators* = 50, *learning_rate* = 0.2, *max_depth* = 3, and *subsample* = 0.8. Noteworthily, the MMT_CA model yielded slightly higher classification performance across multiple evaluation metrics, including accuracy, weighted F1-score, weighted precision, weighted recall, and ROC-AUC, than the XGBoost model, highlighting the superior capability of the MMT_CA model in capturing cross-modality interactions for improving PD detection.

To gain deeper insights into the stability and generalization of these models' performance, ROC-AUC curves for each fold of the five-fold cross-validation in the SVM model were plotted, as illustrated in Fig. 4(A). The low variability across folds indicates that the model performs consistently. Furthermore, the ROC curves and corresponding AUC values for various models are presented in Fig. 4(B) to facilitate a straightforward comparison across different model architectures. It can be observed that the MMT_CA and XGBoost models demonstrated comparable performance, with both models showing a significant improvement over the FCNN_JM and SVM models.

Feature importance

For enhancing the interpretability of each type of data in the detection of PD, the SHAP method was leveraged to assess feature importance, as illustrated in Fig. 5, which shows the top 20 features identified by SHAP and their impact on the model output.

Figure 5(A) illustrates that the top eight features, namely, *Lowput_expected, Mean_putamen, Datscan_putamen_L, Datscan_putamen_R, Mean_striatum, Mean_caudate, Datscan_caudate_L, and Datscan_caudate_R*, which had the most significant impact on the prediction using the SVM model, were derived from DaTscan imaging results. In contrast, the top features in the XGBoost model were more diverse, as shown in Fig. 5(B). Although *Lowput_expected* remained the most influential feature in the XGBoost output, five of the next six top features were proteins, while one feature originated from RNA-seq data. Among the top 20 features, many were from the proteomics and RNA-seq data modalities, along with features from the DaTscan imaging modality. A full description of these top features presented in Fig. 5 can be found in the Supplementary Table S2. Consistently with the feature importance rankings from both the SVM and XGBoost models, the learnable weights from the MMT_CA model also assigned a higher modality weight to DaTscan imaging data as compared with other modalities, as presented in Table 4.

Conclusion and Discussion

To enhance PD classification, hematological information, proteomics, RNA-seq, metabolomics, and DaTscan imaging data were retrieved from the PPMI database. Various model architectures

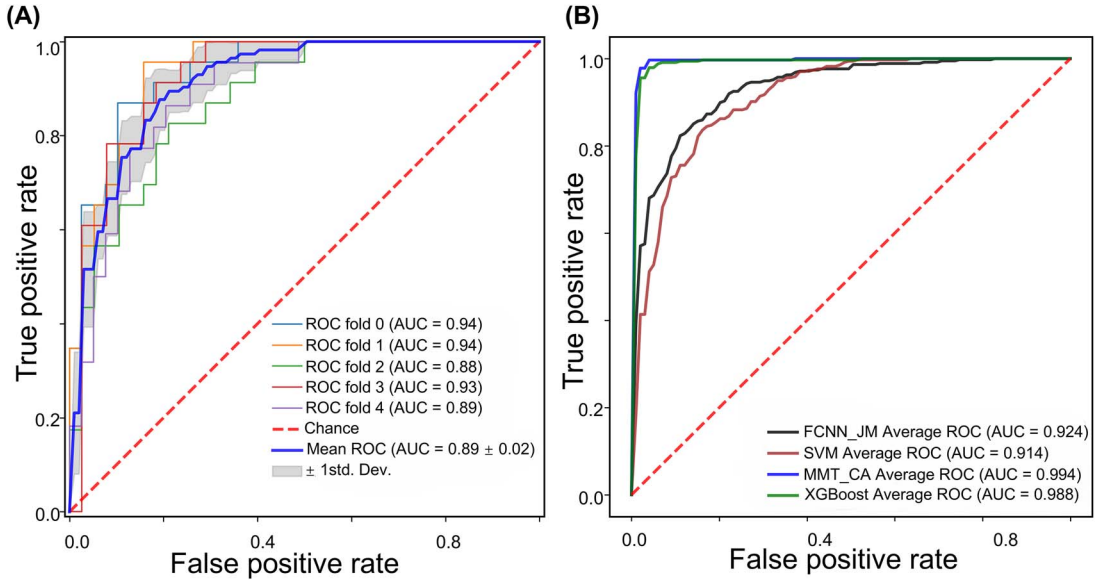

Figure 4. Classification performance comparisons between models based on ROC curves. (A) Performance comparison of the SVM model, showing the ROC curves for each fold in a five-fold cross-validation. In this process, the dataset was divided into five equally-sized subsets, or "folds." The model was trained on four folds and validated on the remaining fold. The ROC curves represent the model's classification performance on the validation set for each fold. (B) Performance comparison of SVM, XGBoost, FCNN_JM, and MMT_CA models, each evaluated using five-fold cross-validation.

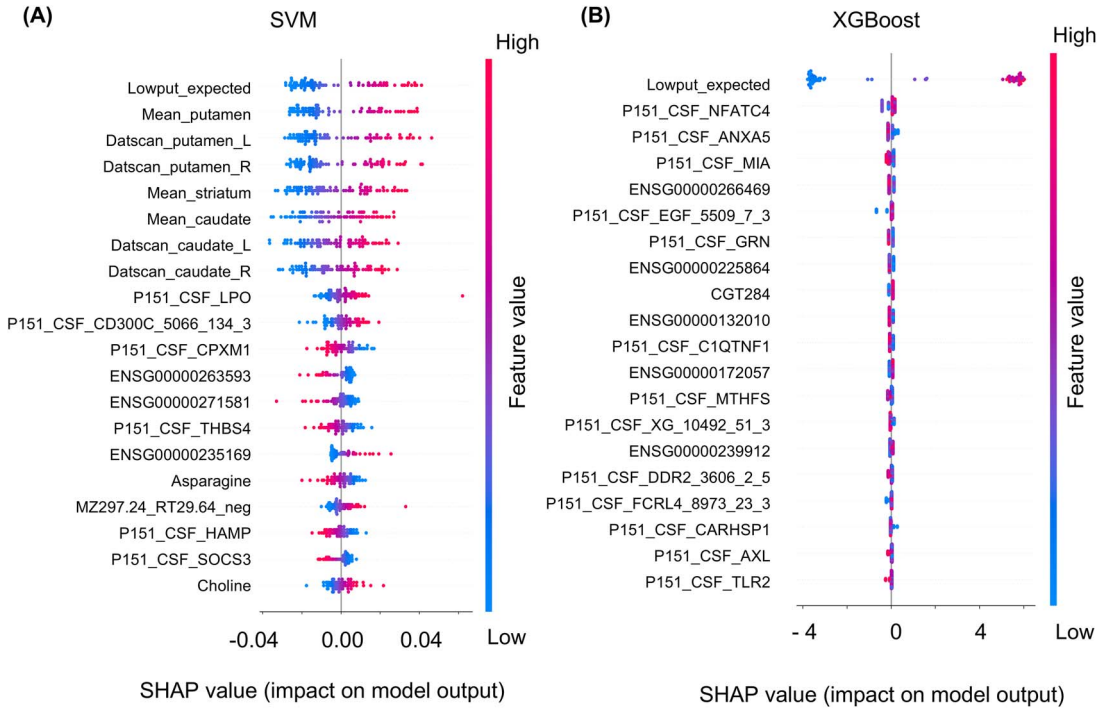

Figure 5. Features explained by SHAP values. (A) SHAP values show the relative impact of each feature on the SVM model output. (B) SHAP values show the relative impact of each feature on the XGBoost model output.

were investigated and evaluated, including the ML-based SVM and XGBoost models and the deep learning-based FCNN_C, FCNN_JM, and MMT_CA models. The results of this study demonstrated that both XGBoost and MMT_CA models achieved high predictive performance compared with the other state-of-the-art models. Multimodal data enabled a more comprehensive approach to discriminating PD from healthy controls, reducing the potential bias and practical diagnostic limitation of relying on a single data modality. The key novelty of this study lies in its innovative approach to enhancing PD prediction by integrating multi-omics data collected via different but complementary modalities that span both structural/morphological and physiological/functional techniques, which is relatively uncharted territory in PD-related diagnostic research. While conventional ML-driven algorithms, such as SVM, XGBoost, and FCNN_C, have been widely adopted to aid various predictive tasks in the literature, from classification to regression, this study highlights the unique potential of MMT_CA models, typically used for aiding imaging analyses or NLP, in capturing complex interactions of multi-omics data captured via different but complementary biological data modalities.

Table 4. Average modality weights derived from the MMT_CA model

Modality	Modality weights
Hematology	0.9996
Protein	0.9998
RNA-seq	0.9991
Image	**1.0010**
Metabolomics part1	0.9989
Metabolomics part2	0.9988

Integrative analysis and modeling of multi-omics data can provide a more holistic view of the underlying genetic and biological mechanisms driving PD, providing the foundation for further analysis to identify critical targets for enabling early treatments for PD.

Based on the taxonomy of early, intermediate, and late fusion outlined in [33], the FCNN_C model of this study represents an early fusion technique. In contrast, FCNN_JM and MMT_CA models represent an intermediate fusion technique. Performance metrics in Table 3 revealed that a simple concatenation of features failed to integrate the data modalities effectively, as reflected by the relatively low accuracy of 65% from five-fold cross-validation. However, when the joint modeling technique was applied, the classification performance improved significantly, increasing accuracy from 65% to 87.5%. These results demonstrated that handling hematological information, proteomics, RNA-seq, metabolomics, and DaTscan imaging data separately in the initial layers before merging them in a joint layer more accurately captured the inter-dependencies between these data modalities. This finding is consistent with prior research [34], which showed that standard techniques like feature concatenation yielded a lower predictive performance in capturing interactions among different data modalities as compared with advanced data fusion-driven methods, such as similarity network-based fusion.

Although late fusion exists, it typically processes each modality independently and combines the outputs or predictions at the final decision layer, usually via equal/weighted averaging and meta-learning [33, 55]. However, multi-omic data modalities might provide supplementary information to each other, and biological interactions can exist between them. For instance, apparent changes in the levels of the endocytic proteins clathrin and endophilin, as well as Ras-associated binding proteins, were identified in leucine-rich repeat kinase 2 (LRRK2)-G2019S and LRRK2-R1441C mutation carriers [56], and mutations in LRRK2 have been regarded as the primary cause of autosomal dominant PD. When investigating the potential therapeutic effect of Acanthopanax senticosus extracts in treating Fu *et al.* [57] revealed in their experiments that the expression of key proteins was affected by the regulation of the level of metabolites. Additionally, considering the interactions between multi-omics data modalities demonstrated an advantage in finding novel molecular and pathway alterations associated with neurodegenerative diseases, such as AD pathology and cognitive impairment [58]. Therefore, late fusion is less suitable for this study, as it fails to capture the complex yet important interdependencies between modalities.

Among the ML-based models, XGBoost demonstrated a higher performance than the SVM, likely due to its architecturally higher ability to capture complex patterns in the data. Noteworthily, the MMT_CA model outperformed XGBoost slightly, achieving an accuracy of 97.7% as compared with 96.7%, and was significantly better performing than the SVM, FCNN_C, and FCNN_JM models. These findings support the higher predictive performance of transformer-based models with cross-attention mechanisms in comprehensively integrating and leveraging features derived from multimodal biological data for aiding predictive analytics. An accuracy of 96.88% was achieved in [59] in identifying PD and health control for 101 participants from PPMI data, using magnetic resonance imaging and clinical features and a stacking ensemble learning method. With DaTscan single photon emission computed tomography, Challa *et al.* [60] achieved an accuracy of 97.159% for 402 PD from the PPMI database. The high accuracies of 97.7% and 96.7%, respectively, achieved with the MMT_CA model and XGBoost model, demonstrate a comparable performance of models in this study with the state-of-art models in the literature. However, it should be noted that this comparison is not strictly comparable, as the samples are different, although they are all from the PPMI database.

The following 13 proteins were used to perform functional enrichment analysis, which were also among the top 20 features in the XGBoost model, including Nuclear Factor of Activated T-cells Cytoplasmic 4 (NFATC4), Annexin A5 (ANXA5), Melanoma Inhibitory Activity (MIA), Epidermal Growth Factor (EGF), Granulin (GRN), C1q and Tumor Necrosis Factor Related Protein 1 (C1QTNF1), 5,10-Methenyltetrahydrofolate Synthase (MTHFS), XG Blood Group Antigen (XG), Discoidin Domain Receptor Tyrosine Kinase 2 (DDR2), Fc Receptor-Like 4 (FCRL4), Cytokine-Responsive Gene Regulator Protein 1 (CARHSP1), AXL Receptor Tyrosine Kinase (AXL), and Toll-Like Receptor 2 (TLR2). After the enrichment analysis using the gprofiler2 package in R, nine Gene Ontology (GO) terms from Biological Process and Molecular Function were identified, as listed in the Supplementary Table S3. Among them, GO:0051897, GO:0051896, and GO:0043491 are related to the Phosphatidylinositol 3-kinases/protein kinase B (PI3K/AKT) signaling pathway, which has been found to show close relations with PD in extensive studies in the literature [61–63]. For instance, the PI3K/AKT pathway could influence oxidative stress by modulating downstream molecular targets like the mammalian target of rapamycin (mTOR), and a decrease in mTOR activity may cause neurodegeneration [62]. Additionally, the PI3K/AKT/mTOR signaling pathway is critical in Tau protein phosphorylation, while Tau hyperphosphorylation is related to the pathogenesis of dementia disorders [61, 63]. Proteins are crucial in disease mechanisms if changes in their levels or function are linked to the onset or progression of the disease. Therapeutic discovery can be facilitated by screening small molecules or antisense oligonucleotides that can alter the protein's function. A PD-related biomarker is identified when changes in specific key proteins associated with the disease, as typically detected in blood or CSF before the disease progression [64].

As shown in Fig. 5, features from DaTscan imaging had the most significant influence on model outputs in both the SVM and XGBoost models. Similarly, in the MMT_CA model, DaTscan imaging had the highest modality weights, indicating its substantial contribution to enhancing PD detection. These consistent results suggest that DaTscan imaging could be a reliable tool in clinical settings for accurately discriminating PD from healthy controls. Nevertheless, some studies argued that caution should still be applied in interpreting the role of DaTscan [22]. Furthermore, the significant role of features related to proteomics and RNA-seq, as highlighted by the XGBoost model, highlights the importance of multi-omics data in PD diagnostic research. Several CSF protein

measures ranked among the top contributors to the models' predictive performance, emphasizing the relevance of CSF biomarkers in aiding the diagnosis of PD. Protein PTMs regulate various signaling pathways and biological functions, and their relevance to PD has been demonstrated. For instance, increased LRRK2 activity was linked to PD [65]. Future studies could integrate abnormal proteins implicated in PD with regulatory networks and functional associations from PTM databases, such as dbPTM [66], to further investigate the pathogenesis of PD. This deeper investigation would explore PD's functional pathway, e.g. by examining pathophysiological phosphorylation events to identify key kinases that may offer additional targets for aiding the development of personalized therapies.

As multiple data modalities were used in this study, data availability of these modalities restricts the number of participants available. This can be demonstrated by the fact that there are around 4000 volunteers in the PPMI database, but only 306 participants were included in this study to balance biological data completeness with sample size. As a result, the limited number might fail to account for the effects of demographic differences, hindering the generalizability of the results. In the future, when more data modalities become available as PPMI projects go on, such an integrative analysis will be more generalizable. Additionally, as modalities have differing amounts of features or quality, the model may rely more heavily on the dominant modality, the DaTscan modality in this case, potentially hindering the statistical significance of the contribution of other modalities. However, other modalities, such as proteomics and metabolomics, are crucial to reveal PD-related pathophysiological mechanisms. In the future, weighted fusion strategies can be further optimized to ensure the contribution of each modality is captured more coherently, although weights supported by biological evidence may be challenging to derive but potentially more significant. Another limitation was that the classification models in this study focused solely on differentiating PD from healthy controls. As mentioned in the Introduction, PD is characterized by various symptoms, including tremor and non-tremor sub-types regulated by distinct functional pathways. A more nuanced PD sub-type classification model incorporating biological and clinical data could be more informative in identifying specific biomarkers for PD sub-types and advancing precision medicine. Moreover, as more features were included and the model architecture with cross-attention to model the relationships between modalities became more sophisticated, the increased complexity inevitably led to higher computational costs. As feature importance has been analyzed with SHAP in ML-based SVM and XGBoost models, feature selection by removing features with low SHAP importance scores can also be a good strategy to balance the accuracy and model complexity. In the future, further dimensionality reduction strategies, such as principal component analysis and t-distributed stochastic neighbor embedding, can help reduce the number of features and potentially transform them into more powerful predictors, thus mitigating the computational cost and complexity of these models.

Additionally, the early diagnosis of PD is challenging, and the research on the etiology and biomarkers has a significant medical value for early diagnosis. As functional pathways revealed by transcriptomics and proteomics would change with the age of PD patients, and these changes are in agreement with disease progression [67], integrative multi-omics analysis can provide the foundation for revealing more in-depth personalized molecular network changes in PD in follow-up studies, contributing to personalized diagnosis and treatment of PD. Moreover, dysregulation in specific pathways, such as inflammation and dopamine-related pathways, are closely related to PD progression [67]; these dysregulation pathways are often driven by kinase activity [39]. As a result, kinase inhibitors targeting related dysregulation pathways can be helpful in treating PD. Take the enriched PI3K/AKT pathway in this study as an example; kinase inhibitors targeting PI3K/AKT/mTOR signaling pathway have drawn extensive attention and efforts to treat many human diseases, including PD [68]. Therefore, pathway analysis, combined with protein–protein interactions and network analysis in the future, can discover more disrupted signaling pathways in different sub-types of PD, particularly those involving kinases, and these kinases could become novel therapeutic targets for PD treatment.

In conclusion, this study leveraged comprehensive biological data related to PD to build classification models. Among the various model architectures investigated and evaluated, the proposed MMT_CA model achieved the highest classification performance, demonstrating its higher effectiveness in capturing the interdependencies in the underlying features extracted from different biological data modalities. This model architecture holds potential to be extended into PD sub-type classification with more data available in the future.

Key Points

- This study integrated multimodal data, including hematological information, proteomics, RNA-seq, metabolomics, and DaTscan imaging, to enhance the detection of PD.
- A transformer-based model with cross-attention can accurately capture cross-modality interactions and has higher predictive performance in detecting PD compared with state-of-the-art ML-driven algorithms.
- SHAP feature importance from SVM and XGBoost models, along with the learnable modality weights from the MMT_CA model, highlights the crucial role of diagnostic biomarkers extracted from various data modalities in aiding the detection of PD.

Author contributions

All authors contributed and approved the final manuscript. Xiaoyan Zhou (Conceptualization, Formal Analysis, Validation, Writingoriginal draft), Luca Parisi (Conceptualization, Validation, Writingreview & editing), Wentao Huang (Formal Analysis, Writingreview & editing,), Yihan Zhang (Formal Analysis, Writingreview & editing), Xiaoqun Huang (Formal Analysis, Writingreview & editing), Mansour Youseffi (Writingreview & editing), Farideh Javid (Formal Analysis), and Renfei Ma (Conceptualization, Formal Analysis, Validation, Writingreview & editing)

Supplementary data

Supplementary data is available at *Briefings in Bioinformatics* online.

Conflict of interest: No competing interest is declared.

Funding

This work was supported by the National Natural Science Foundation of China (32300540) and the Science, Technology

and Innovation Commission of Shenzhen Municipality (Grant No. RCBS20221008093338092).

Data availability

The dataset leveraged in this study was obtained with approval from the Parkinson's Progression Markers Initiative (PPMI) (https://www.ppmi-info.org/) as of 24 June 2024. The Python code supporting the methodologies and findings of this article is available on GitHub at https://github.com/tom-209/PD_classification.

References

1. Váradi C. Clinical features of Parkinson's disease: the evolution of critical symptoms. *Biology* 2020;**9**:103. https://doi.org/10.3390/biology9050103
2. Sveinbjornsdottir S. The clinical symptoms of Parkinson's disease. *J Neurochem* 2016;**139**:318–24. https://doi.org/10.1111/jnc.13691
3. Pfeiffer RF. Non-motor symptoms in Parkinson's disease. *Parkinsonism Relat Disord* 2016;**22**:S119–22. https://doi.org/10.1016/j.parkreldis.2015.09.004
4. McGregor MM, Nelson AB. Circuit mechanisms of Parkinson's disease. *Neuron* 2019;**101**:1042–56. https://doi.org/10.1016/j.neuron.2019.03.004
5. Zaman V, Shields DC, Shams R. *et al.* Cellular and molecular pathophysiology in the progression of Parkinson's disease. *Metab Brain Dis* 2021;**36**:815–27. https://doi.org/10.1007/s11011-021-00689-5
6. Cacabelos R. Parkinson's disease: from pathogenesis to pharmacogenomics. *Int J Mol Sci* 2017;**18**:551. https://doi.org/10.3390/ijms18030551
7. Stoker TB, Barker RA. Recent developments in the treatment of Parkinson's disease. *F1000Research* 2020;**9**:9. https://doi.org/10.12688/f1000research.25634.1
8. Smedinga M, Darweesh SK, Bloem BR. *et al.* Towards early disease modification of Parkinson's disease: a review of lessons learned in the Alzheimer field. *J Neurol* 2021;**268**:724–33. https://doi.org/10.1007/s00415-020-10162-5
9. Ball N, Teo WP, Chandra S. *et al.* Parkinson's disease and the environment. *Front Neurol* 2019;**10**:421551. https://doi.org/10.3389/fneur.2019.00218
10. Parisi L, Ma R, Zaernia A. *et al.* M-ark-support vector machine for early detection of Parkinson's disease from speech signals. *Int J Math Comput Simul* 2021;**15**:34–41. https://doi.org/10.46300/9102.2021.15.7
11. Parisi L, Ma R, Zaernia A. *et al.* Hyper-sinh-convolutional neural network for early detection of Parkinson's disease from spiral drawings. *WSEAS Trans Comput Res* 2021;**9**:1–7. https://doi.org/10.37394/232018.2021.9.1
12. Parisi L, Neagu D, Ma R. *et al.* Quantum ReLU activation for convolutional neural networks to improve diagnosis of Parkinson's disease and COVID-19. *Expert Syst Appl* 2022;**187**:115892. https://doi.org/10.1016/j.eswa.2021.115892
13. Jia E, Zhou Y, Liu Z. *et al.* Transcriptomic profiling of circular RNA in different brain regions of Parkinson's disease in a mouse model. *Int J Mol Sci* 2020;**21**:3006. https://doi.org/10.3390/ijms21083006
14. Zhang H, Yao L, Zheng Z. *et al.* The role of non-coding RNAs in the pathogenesis of Parkinson's disease: recent advancement. *Pharmaceuticals* 2022;**15**:811. https://doi.org/10.3390/ph15070811
15. Zhang J, Chen R, Shi F. *et al.* Genome-wide data mining to construct a competing endogenous RNA network and reveal the pivotal therapeutic targets of Parkinson's disease. *J Cell Mol Med* 2021;**25**:5912–23. https://doi.org/10.1111/jcmm.16190
16. Lei K, Zhang L, He Y. *et al.* Immune-associated biomarkers for early diagnosis of Parkinson's disease based on hematological lncRNA–mRNA co-expression. *Biosci Rep* 2020;**40**:BSR20202921. https://doi.org/10.1042/BSR20202921
17. Kawahata I, Finkelstein DI, Fukunaga K. Pathogenic impact of α-Synuclein phosphorylation and its kinases in α-Synucleinopathies. *Int J Mol Sci* 2022;**23**:6216. https://doi.org/10.3390/ijms23116216
18. Ma SX, Lim SB. Single-cell RNA sequencing in Parkinson's disease. *Biomedicine* 2021;**9**:368. https://doi.org/10.3390/biomedicines9040368
19. Sivaranjini S, Sujatha C. Deep learning based diagnosis of Parkinson's disease using convolutional neural network. *Multimed Tools Appl* 2020;**79**:15467–79. https://doi.org/10.1007/s11042-019-7469-8
20. Burgos N, Bottani S, Faouzi J. *et al.* Deep learning for brain disorders: from data processing to disease treatment. *Brief Bioinform* 2021;**22**:1560–76. https://doi.org/10.1093/bib/bbaa310
21. Aşuroğlu T, Oğul H. A deep learning approach for Parkinson's disease severity assessment. *Health Technology* 2022;**12**:943–53. https://doi.org/10.1007/s12553-022-00698-z
22. Gerraty RT, Provost A, Li L. *et al.* Machine learning within the Parkinson's progression markers initiative: review of the current state of affairs. *Front Aging Neurosci* 2023;**15**:1076657. https://doi.org/10.3389/fnagi.2023.1076657
23. Hosseinzadeh M, Gorji A, Fathi Jouzdani A. *et al.* Prediction of cognitive decline in Parkinson's disease using clinical and DAT SPECT imaging features, and hybrid machine learning systems. *Diagnostics* 2023;**13**:1691. https://doi.org/10.3390/diagnostics13101691
24. Höglinger GU, Adler CH, Berg D. *et al.* A biological classification of Parkinson's disease: the SynNeurGe research diagnostic criteria. *Lancet Neurol* 2024;**23**:191–204. https://doi.org/10.1016/S1474-4422(23)00404-0
25. Gomez-Cabrero D, Abugessaisa I, Maier D. *et al.* Data integration in the era of omics: Current and future challenges. *BMC Syst Biol* 2014;**8**:I1–10. https://doi.org/10.1186/1752-0509-8-S2-I1
26. Subramanian I, Verma S, Kumar S. *et al.* Multi-omics data integration, interpretation, and its application. *Bioinf Biol Insights* 2020;**14**:117793221989905. https://doi.org/10.1177/1177932219899051
27. Leng D, Zheng L, Wen Y. *et al.* A benchmark study of deep learning-based multi-omics data fusion methods for cancer. *Genome Biol* 2022;**23**:171. https://doi.org/10.1186/s13059-022-02739-2
28. Yang Y, Tian S, Qiu Y. *et al.* MDICC: Novel method for multi-omics data integration and cancer subtype identification. *Brief Bioinform* 2022;**23**:1–13. https://doi.org/10.1093/bib/bbac132
29. He X, Liu X, Zuo F. *et al.* Artificial intelligence-based multi-omics analysis fuels cancer precision medicine. *Semin Cancer Biol* 2023;**88**:187–200. https://doi.org/10.1016/j.semcancer.2022.12.009
30. Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet* 2012;**13**:227–32. https://doi.org/10.1038/nrg3185
31. Kumar D, Bansal G, Narang A. *et al.* Integrating transcriptome and proteome profiling: strategies and applications. *Proteomics* 2016;**16**:2533–44. https://doi.org/10.1002/pmic.201600140
32. Figlia G, Willnow P, Teleman AA. Metabolites regulate cell signaling and growth via covalent modification of proteins. *Dev Cell* 2020;**54**:156–70. https://doi.org/10.1016/j.devcel.2020.06.036

33. Stahlschmidt SR, Ulfenborg B, Synnergren J. Multimodal deep learning for biomedical data fusion: a review. *Brief Bioinform* 2022;**23**:1–15. https://doi.org/10.1093/bib/bbab569
34. Markello RD, Shafiei G, Tremblay C. *et al.* Multimodal phenotypic axes of Parkinson's disease. *npj Parkinson's Disease* 2021;**7**:6. https://doi.org/10.1038/s41531-020-00144-9
35. Cong S, Wang H, Zhou Y. *et al.* Comprehensive review of transformer-based models in neuroscience, neurology, and psychiatry. *Brain-X* 2024;**2**:e57. https://doi.org/10.1002/brx2.57
36. Lv C, Fan L, Li H. *et al.* Leveraging multimodal deep learning framework and a comprehensive audio-visual dataset to advance Parkinson's detection. *Biomed Signal Process Control* 2024;**95**:106480. https://doi.org/10.1016/j.bspc.2024.106480
37. Mohaghegh M, Gascon J. Identifying Parkinson's disease using multimodal approach and deep learning. In: *2021 6th International Conference on Innovative Technology in Intelligent System and Industrial Applications (CITISIA)*, pp. 1–6. Sydney, Australia: IEEE, 2021.
38. Vaswani A. Attention is all you need. *Advances in Neural Information Processing Systems* 2017;**30**.
39. Ma R, Li S, Li W. *et al.* KinasePhos 3.0: Redesign and expansion of the prediction on kinase-specific phosphorylation sites. *Genomics Proteomics Bioinf* 2023;**21**:228–41. https://doi.org/10.1016/j.gpb.2022.06.004
40. Tuan TA, Pham TB, Kim JY. *et al.* Alzheimer's diagnosis using deep learning in segmenting and classifying 3D brain MR images. *Int J Neurosci* 2022;**132**:689–98. https://doi.org/10.1080/00207454.2020.1835900
41. Soumaya Z, Taoufiq BD, Benayad N. *et al.* The detection of Parkinson disease using the genetic algorithm and SVM classifier. *Appl Acoust* 2021;**171**:107528. https://doi.org/10.1016/j.apacoust.2020.107528
42. Chen T, Guestrin C. Xgboost: A scalable tree boosting system. In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (KDD '16)*, pp. 785–94. San Francisco, USA: ACM, 2016.
43. Yi F, Yang H, Chen D. *et al.* XGBoost-SHAP-based interpretable diagnostic framework for Alzheimer's disease. *BMC Med Inform Decis Mak* 2023;**23**:137. https://doi.org/10.1186/s12911-023-02238-9
44. Lee SB, Kim YJ, Hwang S. *et al.* Predicting Parkinson's disease using gradient boosting decision tree models with electroencephalography signals. *Parkinsonism Relat Disord* 2022;**95**:77–85. https://doi.org/10.1016/j.parkreldis.2022.01.011
45. Ma R, Li S, Parisi L. *et al.* Holistic similarity-based prediction of phosphorylation sites for understudied kinases. *Brief Bioinform* 2023;**24**:1–12. https://doi.org/10.1093/bib/bbac624
46. Sahu SK, Chowdhury AS. AG-PDCnet: An attention guided Parkinson's disease classification network with MRI, DTI and clinical assessment data. In: *International Conference on Computer Vision and Image Processing*, pp. 176–87. Jammu, India: Springer, 2023.
47. Cai H, Gao Y, Liu M. Graph transformer geometric learning of brain networks using multimodal MR images for brain age estimation. *IEEE Trans Med Imaging* 2022;**42**:456–66. https://doi.org/10.1109/TMI.2022.3222093
48. De S, Sayyad A, Kotian H. *et al.* ParViT: A modified vision transformer architecture for Parkinson's disease identification using EEG signals. In: *2024 International Conference on Smart Systems for applications in Electrical Sciences (ICSSES)*, pp. 1–6. Tumakuru, India: IEEE, 2024.
49. Hammad A, Moretti S, Nojiri M. Multi-scale cross-attention transformer encoder for event classification. *J High Energy Phys* 2024;**2024**:1–31. https://doi.org/10.1007/JHEP03(2024)144
50. Hemmerling D, Wodzinski M, Orozco-Arroyave JR. *et al.* Vision transformer for Parkinson's disease classification using multilingual sustained vowel recordings. In: *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 1–4. Sydney, Australia: IEEE, 2023.
51. Reyes DM, Kim M, Chaoh H. *et al.* Genomics transformer for diagnosing Parkinson's disease. In: *2022 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, pp. 01–4. Ioannina, Greece: IEEE, 2022.
52. Li H, Wu XJ. CrossFuse: A novel cross attention mechanism based infrared and visible image fusion approach. *Inf Fusion* 2024;**103**:102147. https://doi.org/10.1016/j.inffus.2023.102147
53. Craig DW, Hutchins E, Violich I. *et al.* RNA sequencing of whole blood reveals early alterations in immune cells and gene expression in Parkinson's disease. *Nat Aging.* 2021;**1**:734–47. https://doi.org/10.1038/s43587-021-00088-6
54. Kern F, Fehlmann T, Violich I. *et al.* Deep sequencing of sncRNAs reveals hallmarks and regulatory modules of the transcriptome during Parkinson's disease progression. *Nat Aging* 2021;**1**:309–22. https://doi.org/10.1038/s43587-021-00042-6
55. Boulahia SY, Amamra A, Madi MR. *et al.* Early, intermediate and late fusion strategies for robust deep learning-based multimodal action recognition. *Mach Vision Appl* 2021;**32**:121. https://doi.org/10.1007/s00138-021-01249-8
56. Connor-Robson N, Booth H, Martin JG. *et al.* An integrated transcriptomics and proteomics analysis reveals functional endocytic dysregulation caused by mutations in LRRK2. *Neurobiol Dis* 2019;**127**:512–26. https://doi.org/10.1016/j.nbd.2019.04.005
57. Fu J, Gao X, Lu Y. *et al.* Integrated proteomics and metabolomics reveals metabolism disorders in the α-syn mice and potential therapeutic effect of Acanthopanax senticosus extracts. *J Ethnopharmacol* 2024;**318**:116878. https://doi.org/10.1016/j.jep.2023.116878
58. Clark C, Dayon L, Masoodi M. *et al.* An integrative multi-omics approach reveals new central nervous system pathway alterations in Alzheimer's disease. *Alzheimer's Res Ther* 2021;**13**:1–19. https://doi.org/10.1186/s13195-021-00814-7
59. Yang Y, Wei L, Hu Y. *et al.* Classification of Parkinson's disease based on multi-modal features and stacking ensemble learning. *J Neurosci Methods* 2021;**350**:109019. https://doi.org/10.1016/j.jneumeth.2020.109019
60. Challa KNR, Pagolu VS, Panda G. *et al.* An improved approach for prediction of Parkinson's disease using machine learning techniques. In: *2016 international conference on signal processing, communication, power and embedded system (SCOPES)*, pp. 1446–51. Paralakhemundi, India: IEEE, 2016.
61. Alonso AC, Zaidi T, Novak M. *et al.* Hyperphosphorylation induces self-assembly of τ into tangles of paired helical filaments/straight filaments. *Proc Natl Acad Sci* 2001;**98**:6923–8. https://doi.org/10.1073/pnas.121119298
62. Long HZ, Cheng Y, Zhou ZW. *et al.* PI3K/AKT signal pathway: A target of natural products in the prevention and treatment of Alzheimer's disease and Parkinson's disease. *Front Pharmacol* 2021;**12**:648636. https://doi.org/10.3389/fphar.2021.648636
63. Boozari M, Hosseinzadeh H. Crocin molecular signaling pathways at a glance: a comprehensive review. *Phytother Res* 2022;**36**:3859–84. https://doi.org/10.1002/ptr.7583
64. O'Connor LM, O'Connor BA, Lim SB. *et al.* Integrative multi-omics and systems bioinformatics in translational neuroscience: a

data mining perspective. *J Pharm Anal* 2023;**13**:836–50. https://doi.org/10.1016/j.jpha.2023.06.011

65. Alessi DR, Sammler E. LRRK2 kinase in Parkinson's disease. *Science* 2018;**360**:36–7. https://doi.org/10.1126/science.aar5683
66. Li Z, Li S, Luo M. *et al.* dbPTM in 2022: An updated database for exploring regulatory networks and functional associations of protein post-translational modifications. *Nucleic Acids Res* 2022;**50**:D471–9. https://doi.org/10.1093/nar/gkab1017
67. Tng TJ, Wong BWY, Sim EH. *et al.* Systematic analysis of multi-omics data reveals component-specific blood-based biomarkers for Parkinson's disease. *Transl Med Commun* 2024;**9**:12. https://doi.org/10.1186/s41231-024-00169-9
68. Huang J, Chen L, Wu J. *et al.* Targeting the PI3K/AKT/mTOR signaling pathway in the treatment of human diseases: current status, trends, and solutions. *J Med Chem* 2022;**65**:16033–61. https://doi.org/10.1021/acs.jmedchem.2c01070

Briefings in Bioinformatics, 2025, **26(2)**, bbaf088
https://doi.org/10.1093/bib/bbaf088
Case Study