Multi-Modal Parkinson's Disease Classification

Shreyas Anand
School of Computer Science and Engineering
Vellore Institute of Technology
Vellore, India
shreyas.anand2021@vitstudent.ac.in

Anuradha J
School of Computer Science and Engineering
Vellore Institute of Technology
Vellore, India
januradha@vit.ac.in

Abstract*— Parkinson's disease (PD) is a neurological condition that worsens over time and affects motor abilities, leading to stiffness, bradykinesia, and tremors. Diagnosing it early and accurately is difficult due to its complicated patterns, which frequently overlap with other neurological diseases. Advances in machine learning and artificial intelligence have created new opportunities for more accurate PD identification. To increase the precision of PD categorization, this study suggests a multi-modal diagnostic paradigm incorporating various data sources, including MRI scan images and tabular data having UPDRS (Unified Parkinson's Disease Rating Scale) scores. Utilizing multi-modal data fusion, this method employs the Vision Transformer model (ViT) for image analysis and a Gradient Descent model for tabular data. An increasingly popular application of multi-modal technology is image search, where users can retrieve similar images by analyzing visual content and features. With the advancements in computer vision and deep learning, such searches have become highly effective and precise. The study was carried out using the Parkinson's Progression Markers Initiative (PPMI) dataset, which included midbrain T1 and T2-weighted MRI images and UPDRS scores. The ViT model achieved high sensitivity and specificity, and the multi-modal integration greatly increased prediction accuracy. In the PPMI dataset, the total model is targeted to attain an area under the curve (AUC) of 90% and show strong validation in other datasets. The findings suggest that, in comparison to single-modality techniques, combining data from multiple modalities—such as clinical and imaging data—offers more thorough and reliable diagnostic tools. This methodology can help with early PD diagnosis and improve individualized treatment regimens. Subsequent efforts will focus on expanding the research to additional modalities, including speech and gait analysis.

Keywords— Multi-modal image search, Vision Transformer network, Deep learning, Visual content analysis, Feature extraction, Image retrieval, Multi-modal fusion, UPDRS

I. INTRODUCTION

One of the most prevalent neurodegenerative diseases, Parkinson's disease (PD) affects millions of people globally. Parkinson's disease (PD) is typified by increasing motor symptoms such as bradykinesia, tremors, and rigidity. Although the precise origin of Parkinson's disease (PD) is yet unknown, a complex interaction between aging-related, environmental, and genetic variables is thought to be at play.

The disease is a progressive neurological disorder that gradually develops over time and tends to affect the central nervous system. It's often characterized by the degeneration of dopamine-producing cells in a specific region of the brain known as the substantia nigra. The common symptoms of PD are motor symptoms including tremors, slow movement, rigidity, and difficulty walking. Tremors can affect the hands, arms, legs, jaw, or head. Non-motor symptoms may include muscle stiffness and cramps, Impaired balance and coordination, loss of smell, small handwriting, bladder or bowel problems, depression, or anxiety. Most current diagnostic techniques are clinical and depend on observing these motor symptoms, which usually appear after a considerable amount of neuronal death.

Recently, there have been encouraging approaches to solve the diagnostic difficulties related to Parkinson's disease (PD) thanks to the convergence of big data, biomedicine, and machine learning (ML). The diagnosis of Parkinson's disease (PD) could be completely transformed by multimodal techniques, which combine information from several sources, including behavioral measurements, clinical imaging (Fig 1), and genetic profiles. In contrast to conventional single-modality approaches, Multimodal frameworks capture several facets of the disease's pathogenesis and offer a holistic understanding of the condition. These models can identify early-stage symptoms and subtle biomarkers of Parkinson's disease (PD) that would otherwise go unnoticed by merging several data streams.

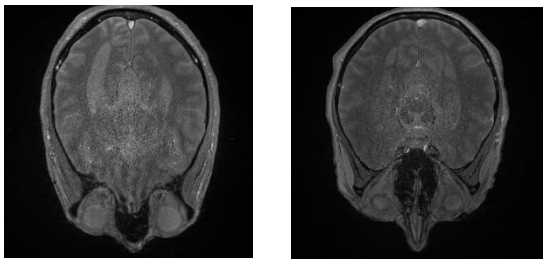


Fig. 1. Brain MRI images of a normal Brain and a brain Parkinson's Disease

The limits of conventional PD diagnostic techniques and the increasing demand for early detection measures are the main driving forces behind this research. As Parkinson's disease progresses, considerable neuronal damage frequently occurs before neuroprotective treatments can offer a discernible benefit, by the time clinical symptoms manifest. Furthermore, it has been shown that single-modality diagnostic techniques, like genetic testing or neuroimaging, are insufficient to fully capture the complexity of the illness. The difficulty of integrating several data types—such as genetic, imaging, and clinical record data—is both philosophically and technically fascinating from the standpoint of machine learning. It is now feasible to examine these intricate, high-dimensional datasets in a meaningful way thanks to developments in processing power, the availability of massive biobank datasets, and the emergence of automated machine learning methods. The chance to develop a diagnostic framework that offers

biological insights in addition to increasing prediction accuracy is what drives this study in particular. Novel biomarkers can be found, genetic pathways unearthed, and the comprehension of the molecular mechanisms behind Parkinson's disease (PD) be improved by evaluating multimodal data. Ultimately, this research could transform the current diagnostic landscape, providing a more precise, early, and personalized diagnosis for individuals at risk of developing PD.

The objective of this study is to create a novel, multimodal machine-learning framework for Parkinson's disease early detection and categorization. The project's scope includes clinical assessments specifically UPDRS and neuroimaging (MRI brain scans). The objective is to develop a holistic diagnostic model with superior accuracy, sensitivity, and specificity compared to conventional single-modality approaches.

The Unified Parkinson's Disease Rating Scale (UPDRS) is a comprehensive tool widely used to assess and track the progression of Parkinson's disease (PD). It evaluates various aspects of the disease, including motor and non-motor symptoms. The UPDRS consists of four parts: (I) Mentation, Behavior, and Mood, (II) Activities of Daily Living (ADL), (III) Motor Examination, and (IV) Complications of Therapy. Each section is scored separately, with higher scores indicating greater impairment. The scale helps clinicians gauge the severity of symptoms and monitor the effectiveness of treatments over time. In the context of machine learning, UPDRS scores often serve as valuable features for models designed to detect or predict Parkinson's disease progression. By incorporating these scores, algorithms can better capture the complexity of the disease, enabling early detection and potentially improving patient outcomes.

Publicly accessible datasets from programs like the Parkinson's Progression Markers Initiative (PPMI) were used for this study. These datasets are perfect for training and evaluating machine learning models because they offer an abundance of multimodal data, such as genetic, imaging, and clinical information. The objective of this study is to create a novel, multimodal machine-learning framework for Parkinson's disease early detection and categorization. In summary, by utilizing the strength of multimodal data and machine learning, this research aims to push the limits of existing PD diagnostic techniques.

II. RELATED WORKS

Recent advancements in non-intrusive patient monitoring have revolutionized the identification and management of Parkinson's disease (PD). The integration of advanced technologies, particularly in machine learning and data analytics, has significantly enhanced the efficacy of healthcare monitoring and diagnostic practices. This literature review explores various innovative approaches and their contributions to improving PD detection and management (Summary in Table I).

Tyler D. Alexander and his team [10] made notable strides in predicting changes in the quality of life (QoL) for Parkinson's patients through the use of a machine learning algorithm. Utilizing follow-up data from 630 patients, they developed an artificial neural network (ANN) model featuring sigmoidal activation functions. The model demonstrated a commendable 90% specificity in identifying patients with stable QoL, but its sensitivity was lower at 56%. This discrepancy suggests that while the model is effective in minimizing false positives, it may require further refinement to enhance its ability to detect significant QoL deterioration. The study highlights the promise of ANNs in monitoring QoL but also underscores the need for improved sensitivity to better capture those experiencing notable changes.

Alex John Sahaya Rani [11] explored the potential of gait analysis as an early detection method for Parkinson's disease. By creating a gait dataset from 81 patients, Rani applied various machine-learning algorithms to analyze motor patterns associated with PD. The study achieved the highest accuracy of 91.90% using the K-Nearest Neighbors (KNN) model, underscoring the value of wearable devices in capturing and analyzing gait data. This approach not only offers a non-invasive means of monitoring but also holds significant potential for early diagnosis, allowing for timely intervention and management of Parkinson's symptoms. The high accuracy achieved through KNN demonstrates the effectiveness of gait analysis in identifying motor changes that are indicative of early-stage PD.

C. Xiao et al. [9] investigated Parkinson's disease classification using a tree-based model derived from the PPMI database. Their approach involved the Gradient Boosting Decision Tree (GBDT) model to identify key features related to the disease. This was followed by a neural network with two hidden layers to refine classification. The study's results were remarkable, with a classification accuracy of 99.74% and an F1 score of 99.86%. This impressive performance highlights the importance of focusing on critical features, such as motor functions and cognitive assessments, to achieve high accuracy in PD classification. The use of GBDT for feature selection followed by a neural network for classification demonstrates a robust methodology for leveraging critical data to enhance diagnostic precision.

Michela Russo et al. [12] examined gait analysis to differentiate between Parkinson's patients with and without mild cognitive impairment (MCI). Their study, involving 80 participants, demonstrated that gait analysis could effectively identify PD patients with MCI with over 80% accuracy. This finding supports the use of gait analysis not only in the early diagnosis of Parkinson's but also in distinguishing between PD patients with varying cognitive impairments. The ability to accurately identify PD-MCI patients has significant implications for personalized treatment strategies and early intervention, highlighting the role of gait analysis in comprehensive PD management.

Haewon Byeon [13] developed a random forest-based model to predict Parkinson's disease with mild cognitive impairment using 96 records. This model integrated various health-related factors to identify early-stage Parkinson's dementia. The study emphasized the efficacy of random forests in managing complex health data due to their ability to handle numerous variables and interactions. By leveraging the strengths of random forests, Byeon's model provided a valuable methodology for predicting Parkinson's disease progression and cognitive decline, contributing to early diagnosis and targeted treatment approaches.

Maitane Martinez–Eguiluz et al. [14] focused on combining datasets from Biocruces and PPMI to enhance Parkinson's disease screening. By integrating these diverse data sources and applying feature selection techniques, the study achieved high accuracy rates. The use of Support

Vector Machines (SVM) and Multi-Layer Perceptron (MLP) models resulted in accuracy rates of 86.3% and 84.7%, respectively. This approach underscores the importance of data integration in improving screening accuracy and highlights the potential of combining multiple datasets to enhance the performance of machine learning models in PD diagnosis.

Aditi Govindu and her team [15] developed a telemedicine-based method for detecting Parkinson's disease using audio samples. Their evaluation of multiple machine learning models, including SVM and random forests, revealed that the SVM model achieved the highest accuracy of 91.84%. This study demonstrates the potential of leveraging audio-based diagnostics for remote and non-invasive PD detection. The use of telemedicine and audio samples offers a promising avenue for expanding diagnostic capabilities, especially in settings where traditional in-person evaluations may be challenging.

Kamal and Venkata [16] employed data from the PPMI study to develop a Boosted Logistic Regression model for early Parkinson's disease prediction. Their model achieved an accuracy of 97.16%, showcasing the effectiveness of advanced machine-learning techniques in early diagnosis. This high level of accuracy emphasizes the potential of boosted logistic regression in identifying early-stage Parkinson's disease, highlighting its role in predictive analytics and early intervention strategies.

Arti Rana et al. [17] investigated various supervised classification techniques for Parkinson's disease using voice features from 195 patients. Their approach, which employed neural networks, achieved the highest accuracy of 96.7%. This finding indicates the effectiveness of artificial neural networks in classifying Parkinson's disease based on vocal characteristics.

The study highlights the utility of voice features in PD detection and the capability of neural networks to accurately classify the disease based on subtle voice alterations.

Marimuthu et al. [18] utilized the UCI ML Parkinson's dataset and the XGB Classifier algorithm to achieve the highest accuracy of 94.87%. This study underscores the effectiveness of ensemble methods in Parkinson's disease prediction. The use of ensemble techniques like XGB Classifier demonstrates how combining multiple models can enhance predictive performance, contributing to more accurate and reliable PD diagnosis.

Pankaj Kumar Keserwani et al. [20] reviewed the application of Artificial Intelligence (AI) in Parkinson's disease detection, emphasizing AI's capability to process large datasets and make precise predictions. Their review highlights the potential of AI-driven techniques, including deep learning and meta-heuristic algorithms, in advancing PD diagnostics. The integration of AI into PD detection processes represents a significant advancement, offering improved accuracy and efficiency in managing and diagnosing Parkinson's disease.

Khushal Thakur and colleagues [19] developed machine learning models based on speech characteristics to classify Parkinson's disease. Their models, which included Extra Trees and Random Forest Classifiers, demonstrated high accuracy rates in classifying Parkinson's disease based on speech features. This approach underscores the potential of speech-based diagnostics in PD detection, offering an additional non-invasive method for monitoring and diagnosing the disease.

This literature review highlights the significant progress made in Parkinson's disease detection through the integration of diverse data sources and advanced machine-learning techniques. The studies reviewed demonstrate a range of methodologies, from gait analysis and audio-based diagnostics to ensemble methods and AI-driven approaches. Collectively, these advancements underscore ongoing opportunities for enhancing diagnostic accuracy and patient care in Parkinson's disease. The continuous evolution of these technologies promises further improvements in early detection, personalized treatment, and overall disease management.

Despite the promising advancements in multi-modal Parkinson's Disease (PD) detection, several research gaps require attention. Current studies often rely on controlled datasets like PPMI, which may not fully represent the diverse manifestations of PD across different populations, necessitating research on larger, more inclusive datasets incorporating various demographics, disease stages, and comorbidities. Furthermore, the lack of standardized protocols for data acquisition and preprocessing hinders the development of generalizable models and calls for establishing consistent guidelines across studies. While combining modalities improves accuracy, the interpretability of these complex models often remains a black box, emphasizing the need for explainable AI techniques to understand the contribution of each modality and identify underlying disease mechanisms. Additionally, translating these research findings into practical clinical tools and integrating them into routine healthcare workflows remains a challenge, requiring the development of user-friendly and accessible tools for clinicians and patients. Finally, addressing ethical considerations related to data privacy, algorithmic bias, and patient autonomy is crucial to ensure responsible and equitable implementation of these technologies. By addressing these research gaps, future studies can contribute to the development of more accurate, reliable, and interpretable multi-modal PD detection systems, ultimately leading to earlier diagnosis, personalized treatment, and improved quality of life for individuals with Parkinson's Disease.

III. Materials and Methods

The objective of this research paper is to develop and evaluate an advanced multimodal approach for the detection and classification of Parkinson's disease (PD) using the PPMI dataset, which encompasses both MRI images and tabular data containing UPDRS scores. The project aims to enhance diagnostic accuracy by integrating features extracted from these diverse data types. Specifically, MRI images were analyzed using Vision Transformer (ViT) technology for image-based feature extraction (Fig 2). The image features extracted using ViT are fused or concatenated with the tabular data. These extracted features will then be fused to create a unified machine-learning classification model.

TABLE I RELATED WORK SUMMARY

Data Source	Sample Size	Data type	Approach	Result	Advantage	Limitation
PPMI	751	Tabular data	Applied GBDT and ANN as feature ranking and classification techniques, respectively	99.74%	Detected PD with few key features	Deep Feature Selection and Bayesian Nonparametric Methods
PPMI	630	Tabular data	Applied UPDRS II	Specificity 90%	Excellent Prediction Sensitivity	Insufficient Data and Retrospective Design
PPMI	168	Tabular data	Applied SVM, RF, KNN, LR, and XGBoost	91.90%	An Affordable Wearable Gadget System for Gathering Motor Data	Limited Data
UCS, Italy	80	Tabular data	Applied DT, RF, NB, KNN, SVM	85.7%	Strong Correlation Between Walking Performance and Cognitive Dysfunction in PD	Only Focused on Gait Patterns
Parkinson's Dementia Clinical Epidemiology	526	Tabular data	Multiple LR, DT, and RF	67.70%	MCI Prediction Model Using National Survey Examination Data	Insufficient Examination of Obsessive-compulsive Symptoms and Inability to Evaluate PD Medication Use
Biocruce and PPMI	783	Tabular	Applied LR, ADAboost, DT, KNN, MLP, RF and SVM	86.30%	Applied RIPPER Algorithm to Detect Rates	Only focus on comparing controls without motor disorders Indicative of PD
PPMI and UCI	195	Audio data	Applied PCA, SVM, RF, LR and KNN	91.84%	Enhance the use of ML in Telemedicine	Only collected audio data
PPMI		Tabular data	SVM (RBF), SVM(Linear), SVM(Sigmoid), RF, ADA, GBDT, ET	100% (Single and Multimodality), 91.41% (NN)	Single and Multimodality analysis and identified key factors	Only Tabular Data

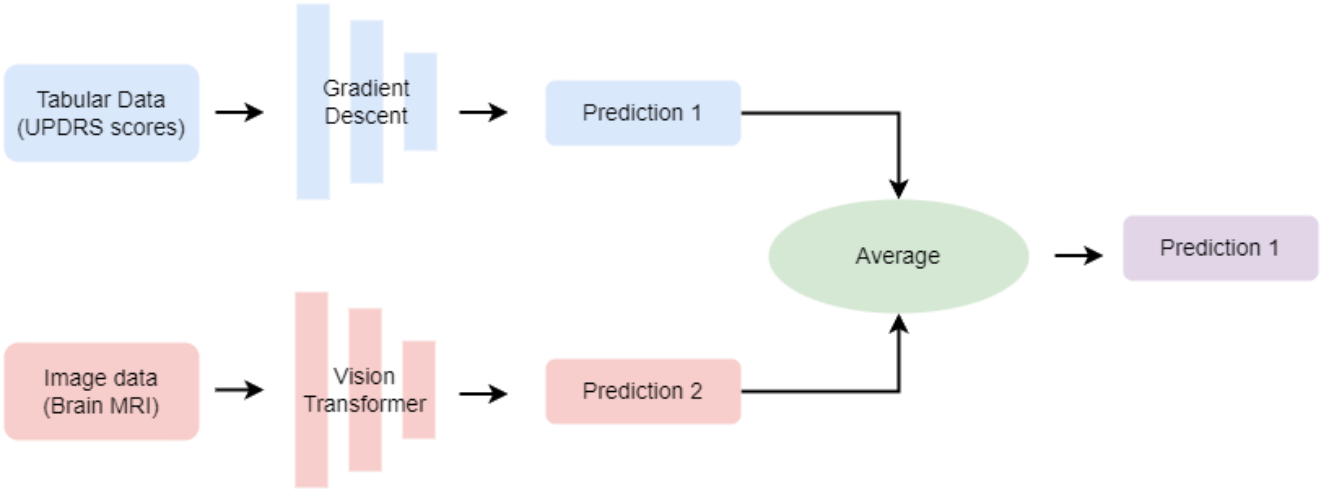


Fig. 2. Late Fusion Multimodal Classifier Architecture

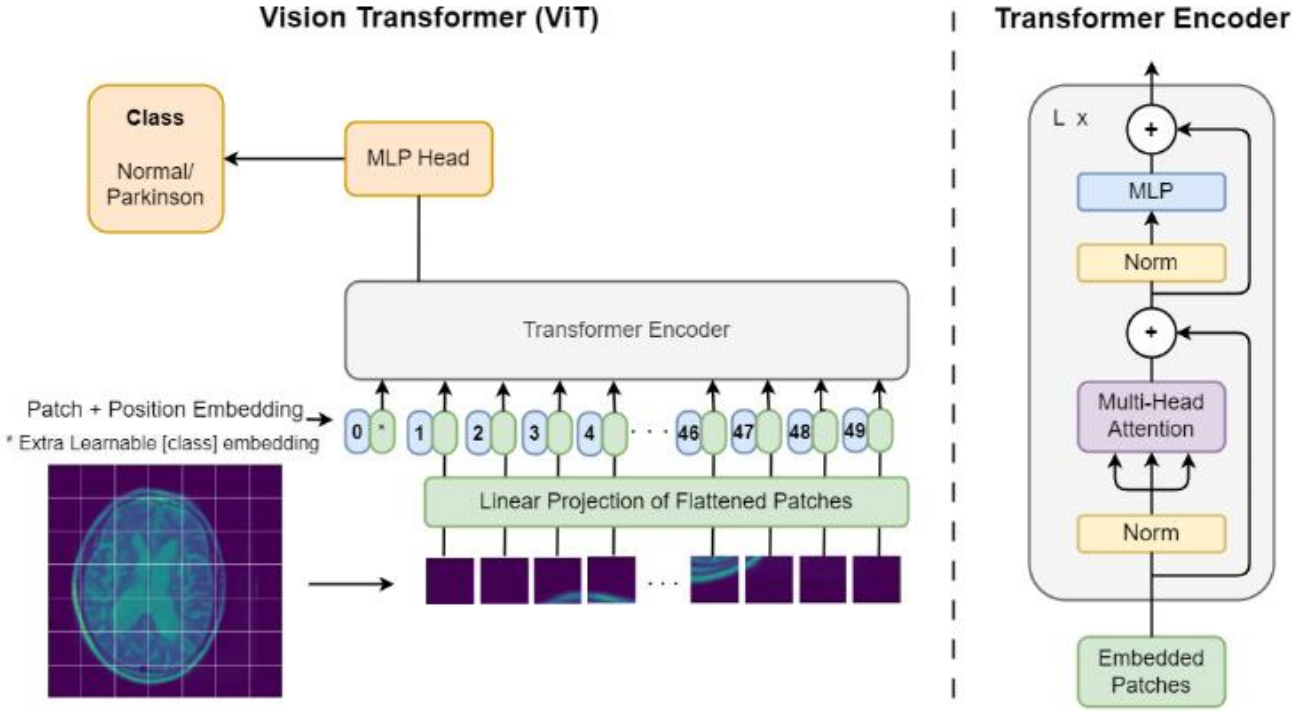


Fig. 3. Vision Transformer (ViT) architecture for learning MRI images

TensorFlow has been extensively used for implementing the ViT and the multimodal learning model. TensorFlow offers several key functions that facilitate the implementation of multimodal machine learning models, which integrate different data types such as images, text, and numerical data. The *tf.keras.Input* function is essential for defining multiple input tensors, allowing the model to accept different modalities. Various layers are provided by *tf.keras.layers*, such as Dense and Conv2D handle the specific data processing required for each input type. To combine the processed features from different modalities, the *tf.keras.layers.concatenate* function is used, enabling the fusion of multimodal data.

A Vision Transformer (ViT) model was implemented using a custom patch extraction layer using the *patchify* package instead of a Conv2D layer that is mostly used in other image processing models. Using *patchify*, an image was divided into smaller patches, which were then flattened and embedded using `Dense` layers. The embedded patches are passed through multiple *`tf.keras.layers.MultiHeadAttention`* and *`LayerNormalization`* blocks to capture global dependencies across the image. Positional encodings are added to retain spatial information. Finally, the transformer output is processed by `Dense` layers for classification or other tasks.

In the ViT model (Fig 3) implemented for reading the brain MRI images, MRI images were resized to 224×224 pixels and a patch size of 32×32 pixels was used with a stride size of 32 pixels, ensuring that there is no overlap between the patches. An image is thus divided into 49 patches. The patches are then flattened and embedded with patch, position, and learnable (class) information. The embedding is then passed through multiple Transform Encoder layers. After the transformer processes the image patches through attention mechanisms and the internal MLP layers, the resulting feature representations are passed to the MLP head, which typically consists of one or more dense layers. The MLP head maps the final transformer output to the desired number of output classes. It acts as the final step in the model, converting the learned features into predictions (e.g., class labels or probabilities).

A. Dataset

The tabular dataset (Table II) utilized in this study is sourced from the Parkinson's Progression Markers Initiative (PPMI) database. It consists of 3,420 records, each representing the Unified Parkinson's Disease Rating Scale (UPDRS) score totals across all four parts of the UPDRS evaluation (I: Mentation, Behavior, and Mood; II: Activities of Daily Living; III: Motor Examination; IV: Complications of Therapy). These records correspond to 3,420 participants, reflecting data from their most recent clinical visit.

In addition to the tabular data, the image dataset comprises 831 MRI scans. These images correspond to a subset of the participants in the UPDRS dataset, allowing for a multimodal analysis by combining clinical assessments with neuroimaging data. The inclusion of MRI data enables a deeper exploration of potential relationships between structural brain changes and disease progression, offering valuable insights into Parkinson's Disease characterization.

TABLE II. TABULAR DATA FILE LAYOUT

Column Name	Detabuscription
PATNO	Participant number
NP1RTOT	Total score of MDS-UPDRS Part I (Non-Motor Aspects of Experiences of Daily Living). Includes scoring on Anxious Mood, Apathy, Cognitive Impairment, Features-Dopamine Dysregulation Syndrome, Depressed Moods, Hallucinations And Psychosis
NP2PTOT	Total score of MDS-UPDRS Part II (Patient Questionnaire: Motor Aspects of Experiences of Daily Living) Includes scoring on dressing, eating tasks, freezing, Doing hobbies and other activities, handwriting, Hygiene
NP3TOT	Total score of MDS-UPDRS Part III (Motor Examination) Includes scoring on Dyskinesias, Global spontaneity of movement, Facial expression, Freezing of gait, Hand movements, Rigidity.
NP4TOT	Total score of MDS-UPDRS Part IV (Motor Complications)
CLASS	0 for ‘normal’ and 1 for ‘parkinson’ participant

The summary of the image dataset is presented in Table III below. Each MRI image file is named to reflect the corresponding participant’s identification number, ensuring clear traceability to the associated clinical data. The organization of the images into folders further facilitates classification, with each folder representing a diagnostic label—either 'normal' for healthy controls or 'parkinson' for individuals diagnosed with Parkinson’s Disease. This structured labeling scheme ensures the separation of the two classes, allowing for straightforward supervised learning and comparative analysis between the control and Parkinson’s groups.

TABLE III. MRI IMAGE INFORMATION

Number	831
Size (in pixels after resizing for standardization)	224×224
Number of channels	1
Format	PNG (Portable Network Graphic)
Patch size for ViT	32×32
PATNO identifier	File name
Label (class)	Folder name

B. Preprocessing

Before utilizing the dataset for machine learning, several crucial preprocessing steps were carried out to ensure data quality and relevance, with the ultimate goal of optimizing the dataset for effective model training. The following comprehensive preprocessing techniques were applied:

- Handling Missing Values and Duplicates

The first step involved identifying and removing duplicate records and entries containing missing or null values. These accounted for only a small fraction of the dataset, and their

removal did not result in any significant loss of critical information. As no substantial gaps in the data were detected, the imputation of missing values was unnecessary. This process was essential to maintain the integrity of the data and ensure that inconsistencies or redundancies would not bias the machine learning model.

- Exploratory Data Analysis (EDA)

A detailed exploratory data analysis (EDA) was conducted to gain deeper insights into the distribution, trends, and characteristics of the data. This analysis included visualizations to examine data patterns, identify potential outliers, and better understand the dataset's structure. One of the key observations was that the majority of Parkinson's Disease (PD) patients were male, predominantly within the age range of 60 to 70 years (see Fig. 4). This demographic insight provides a clearer context for the model and highlights potential biases in the dataset's composition.

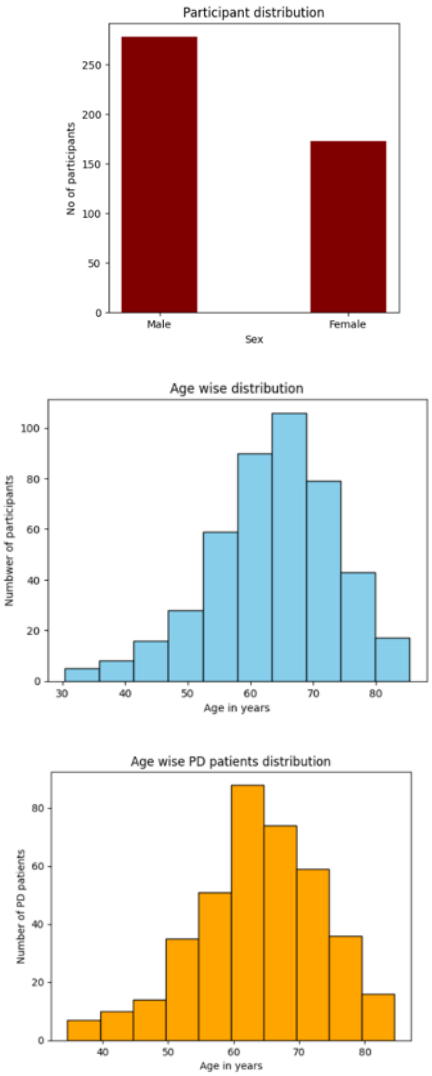
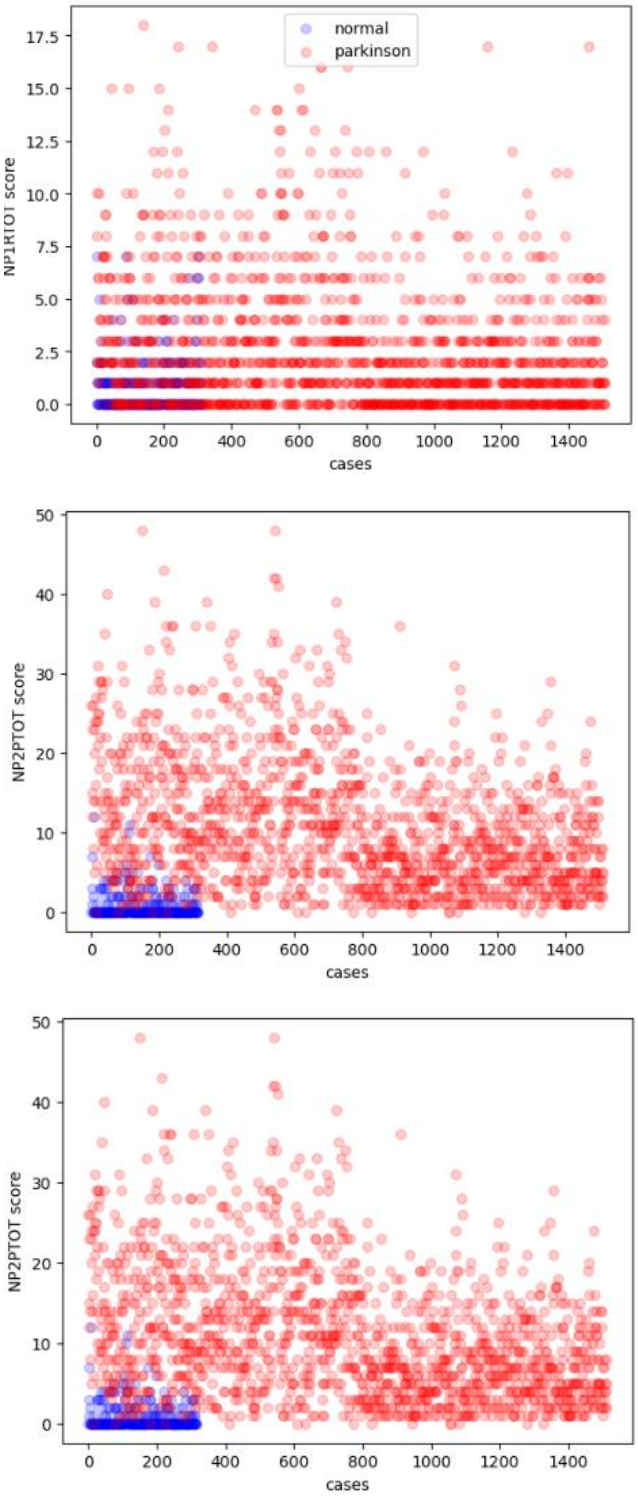


Fig. 4. Distribution of PD patients based on sex and age-wise distribution of participants and PD patients.

Moreover, the EDA revealed notable differences in UPDRS scores between PD patients and the control group. PD patients consistently showed higher scores across all four parts of the UPDRS evaluation compared to the normal participants. This was clearly reflected in the scatter plots (see Fig. 5), which visualize the distribution of UPDRS scores for each group. These score distributions provide valuable insights into the disease's progression and severity, which can inform the development of more accurate models.



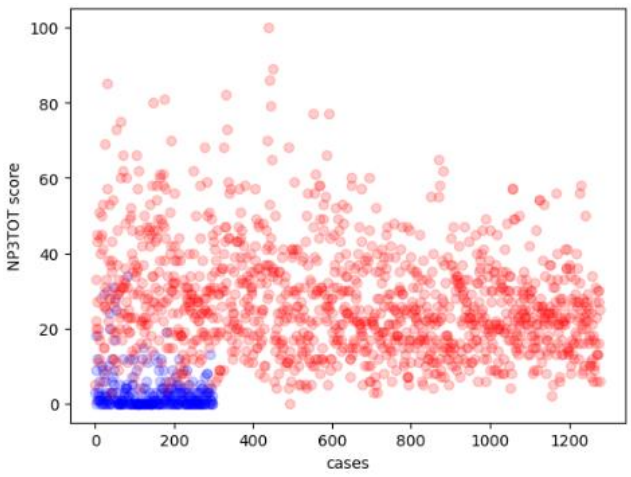


Fig. 5. Scatter plots showing the total scores from the four-part UPDRS evaluation for normal (blue) vs Parkinson patients. (red). The participants having Parkinson's disease have higher values for these scores than the normal (Healthy Control) participants.

C. Train and Test the Models

In this study, the tabular and image data were processed using different machine-learning models tailored to their respective data types. The tabular data was modeled using a Gradient Descent algorithm, while the image data was learned using a Vision Transformer (ViT). Both models were implemented using the TensorFlow framework, which provided the necessary tools for efficient training and integration.

A key advantage of the late fusion approach employed in this study is its flexibility in handling different types of data with potentially varying sample sizes. Since the tabular and image models are trained independently, the two datasets do not need to have the same number of records. This allows for the inclusion of additional data sources, even when the modalities are imbalanced in terms of size or availability. After training, the predictions from both models are combined by averaging their output probabilities, producing the final multimodal prediction. This method not only accommodates the inherent differences between the modalities but also preserves their individual strengths, leading to a more comprehensive and accurate prediction. The late fusion approach thus enhances the model's adaptability to real-world healthcare scenarios, where data from different modalities may not always be perfectly aligned in terms of quantity or structure. [29].

The tabular data had 3420 records with 2577 normal and 843 PD participants containing total scores from each of the four-part UPDRS evaluations. The model was trained and tested and its performance was evaluated by accuracy, precision, recall, and F1 scores. As late fusion is being used

The image data consisted of 834 T1 and T2 weighted brain MRI images, out of which 610 were normal MRI images and 221 were Parkinson's images. The images were organized into 2 folders named as per the class, and images were individually named with the participant numbers. The images were standardized into 224 x 224 pixels before being used in the model. The parameters of the model are summarized in Table IV.

TABLE IV. SUMMARY OF MACHINE LEARNING PARAMETERS FOR VISION TRANSFORMER MODEL

Parameter	Value
Image Size	224 × 224 pixels
Number of Channels	1
Number of Patches per image	49
Batch Size	32
Test Size	0.25
Learning Rate	0.0001
Number of Epochs	15
Number of classes	2
Number of Transformer Encoder Layers	12
Number of heads	4
Dropout Rate	0.1
Activation Function	'GELU' in the MLP layer, 'Softmax' in the classification layer
Loss Function	Categorical Crossentropy

In the late fusion model, multimodal predictions can be generated through various methods, such as averaging or taking a weighted average of the predictions from the individual modalities. For this study, a simple averaging approach was employed. Specifically, the predicted probabilities were first computed separately for the tabular data model using Gradient Descent and the image-based model using Vision Transformer (ViT). These predictions, obtained from a common test dataset, were then combined by averaging the output probabilities from both models. This straightforward averaging method ensures that both modalities contribute equally to the final prediction, without introducing bias toward one modality. Although more sophisticated techniques such as weighted averaging could potentially refine the prediction process by prioritizing the modality with stronger performance, the simple averaging approach was chosen to maintain simplicity and balance between the two data streams. This method provides an intuitive way to fuse the outputs, demonstrating the effectiveness of integrating tabular and image data in a unified multimodal framework.

IV. RESULTS AND DISCUSSION

The model performances were calculated for both the unimodal models, and the multimodal model are determined.

The train and test records for the Gradient descent model were split with a test ratio of 10% implying that 10% of records were used for validation and the rest used for training.

The ViT model training was run in batches of 32 over 15 epochs. The whole training process was run once without an attention layer, and then with a self-attention layer. The training process tracks and reports the accuracy and loss

function values while the training is in progress. The accuracy and loss functions from the training trace were plotted against the epochs for both test and train (Fig. 6).

The accuracy of the classifier represents the ratio of correctly classified instances to the total instances in the dataset. It provides an overall measure of the model's correctness and is calculated using the formula:

$$Accuracy = \frac{\text{True Positives + True Negatives}}{\text{Total Number of Predictions}} \quad (1)$$

Precision is a measure of how accurate are the positive predictions made by the model. It is the ratio of true positives to the sum of true positives and false positives:

$$Precision = \frac{\text{True Positives}}{\text{True Positives+False Positives}} \quad (2)$$

Recall (or Sensitivity) quantifies the ability of the model to capture all relevant instances. It is the ratio of true positives to the sum of true positives and false negatives:

$$Recall = \frac{\text{True Positives}}{\text{True Positives+False Negatives}} \quad (3)$$

The F1-Score is the harmonic mean of precision and recall, providing a balanced measure that considers both false positives and false negatives. It is calculated using the formula:

$$F1\ score = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision+Recall}} \quad (4)$$

The performance data of the two unimodal models and the fusion model are summarized in

TABLE V. PERFORMANCE SUMMARY

	Gradient Descent (Tabular data)	Vision Transformer (Image data)	Fusion Model (Multi-modal data)
Accuracy	0.9601	0.9613	0.9807
Precision	0.9062	0.9821	0.9833
Recall	0.9560	0.8871	0.9516
F1 Score	0.9305	0.9322	0.9672
AUC	0.9601	0.9462	0.9723

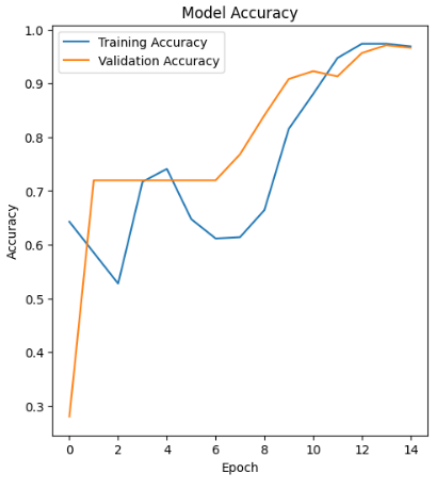


Fig. 6a. Accuracy Function Graph for Train and Validation of ViT model

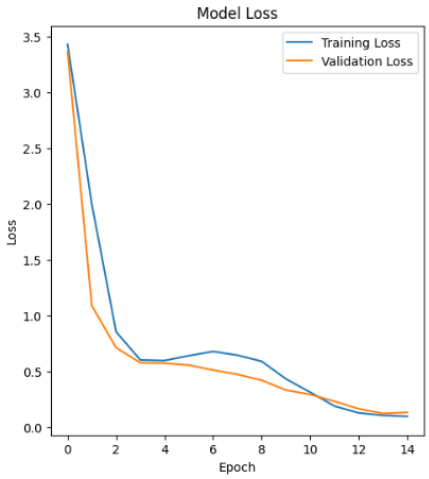


Fig. 6b. Loss Function Graph for Train and Validation of ViT model

The results demonstrate a clear advantage of the multimodal model over the unimodal models. Across most evaluation metrics, the fusion model outperforms both the tabular and image-based models, providing stronger predictive capabilities. Notably, all performance metrics—such as accuracy, precision, and F1-score—are consistently higher for the multimodal approach. The only exception is the recall score, which, while still competitive, is slightly lower than that achieved by the unimodal gradient descent model. This minor discrepancy may suggest that while the multimodal approach excels in overall predictive power, the gradient descent model may be better at identifying true positives in certain cases. Nonetheless, the fusion model's superior performance across most metrics reinforces the value of integrating multiple data modalities for a more comprehensive and balanced classification in healthcare applications like Parkinson's Disease diagnosis.

The Multiclass Receiver Operating Characteristics (ROC) were determined and plotted as ROC curves for each model and are shown in Fig.7.

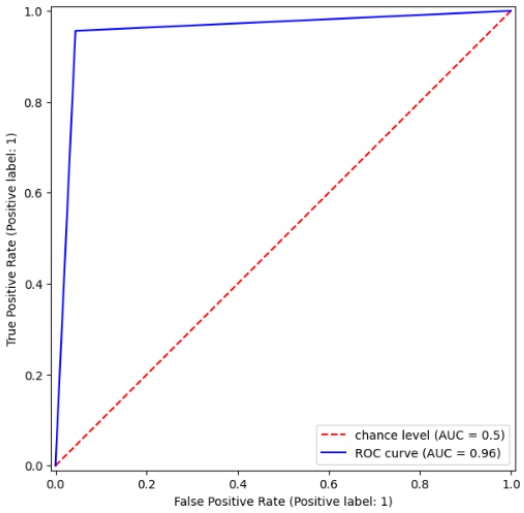


Fig. 7. ROC characteristics for the Gradient Descent model.

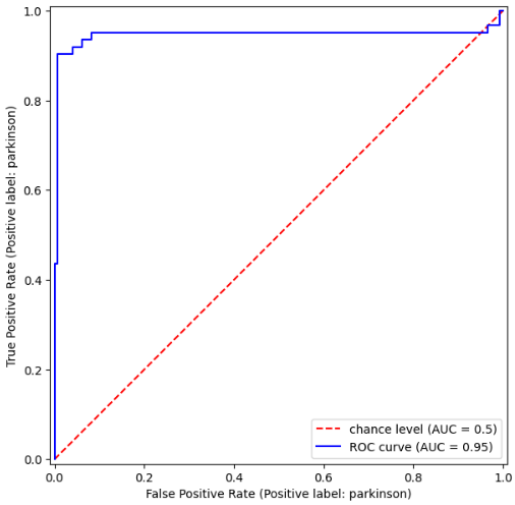


Fig. 7b. ROC characteristics for the ViT model.

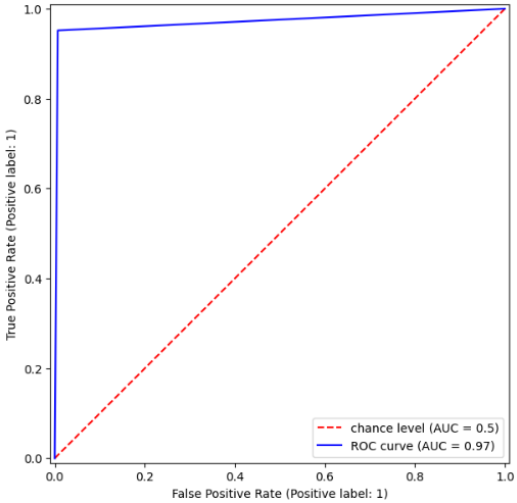


Fig. 7c. ROC characteristics for the Fusion model.

V. CONCLUSION

This study explored a multimodal approach for Parkinson's Disease classification, integrating tabular and image data to enhance predictive accuracy. By comparing unimodal models—one using only tabular data and the other using solely image data—with a fused model that combines both modalities, the study demonstrated that the fusion model consistently outperformed the unimodal approaches. The findings underscore the potential of multimodal machine learning in the healthcare domain, where different types of patient data are often available and can complement each other. By harnessing both structured and unstructured data, this method enables more robust and accurate decision-making, which is critical for complex conditions like Parkinson's Disease. The results highlight the promise of multimodal models in improving diagnostic precision and advancing personalized healthcare.

REFERENCES

[1] D. Bereczki, "The description of all four cardinal signs of Parkinson's disease in a Hungarian medical text published in 1690," Parkinsonism & Related Disorders, vol. 16, no. 4, pp. 290–293, 2010.

[2] J. Parkinson, "An essay on the shaking palsy," J. Neuropsychiatry Clin. Neurosci., vol. 14, no. 2, pp. 223–236, 2002.

[3] GBD 2016 Parkinson's Disease Collaborators, "Global, regional, and national burden of Parkinson's disease, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016," The Lancet Neurology, vol. 17, no. 11, pp. 939-953, 2018. [Online]. Available: https://doi.org/10.1016/S1474-4422(18)30295-3.

[4] T. Ternák et al., "Antibiotic consumption patterns in European countries are associated with the prevalence of Parkinson's disease; the possible augmenting role of the narrow-Spectrum Penicillin," Antibiotics, vol. 11, no. 9, p. 1145, 2022.

[5] R. Abdullah et al., "Parkinson's disease and age: the obvious but largely unexplored link," Exp. Gerontol., vol. 68, pp. 33–38, 2015.

[6] J. D. Gazewood, D. R. Richards, and K. Clebak, "Parkinson disease: an update," Am. Family Physician, vol. 87, no. 4, pp. 267–273, 2013.

[7] G. D. Santos et al., "Present and future of Parkinson's disease in Spain: Parkinson-2030 Delphi project," Brain Sci., vol. 11, no. 8, p. 1027, 2021.

[8] A. Toulouse and A. M. Sullivan, "Progress in Parkinson's disease—where do we stand?" Progress Neurobiol., vol. 85, no. 4, pp. 376–392, 2008.

[9] C. Xiao et al., "Key Marker Selection for the Detection of Early Parkinson's Disease using Importance-Driven Models," in 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2018.

[10] T. D. Alexander, C. Nataraj, and C. Wu, "A machine learning approach to predict quality of life changes in patients with Parkinson's Disease," Ann. Clin. Transl. Neurol., vol. 10, no. 3, pp. 312–320, 2023.

[11] J. S. R. Alex, "Early Detection of Parkinson's Disease using Motor Symptoms and Machine Learning," arXiv preprint arXiv:2304.09245, 2023.

[12] M. Russo et al., "Identification of a Gait pattern for detecting mild cognitive impairment in Parkinson's disease," Sensors, vol. 23, no. 4, p. 1985, 2023.

[13] H. Byeon, "Is the random forest algorithm suitable for predicting Parkinson's disease with mild cognitive impairment out of Parkinson's disease with normal cognition?" Int. J. Environ. Res. Publ. Health, vol. 17, no. 7, p. 2594, 2020.

[14] M. Martinez-Eguiluz et al., "Diagnostic classification of Parkinson's disease based on non-motor manifestations and machine learning strategies," Neural Comput. Appl., vol. 35, no. 8, pp. 5603–5617, 2023.

[15] A. Govindu and S. Palwe, "Early detection of Parkinson's disease using machine learning," Proc. Comput. Sci., vol. 218, pp. 249–261, 2023.

[16] K. N. R. Challa 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), IEEE, 2016.

[17] A. Rana et al., "An Efficient Machine Learning Approach for Diagnosing Parkinson's Disease by Utilizing Voice Features," Electronics, vol. 11, no. 22, p. 3782, 2022.

[18] M. Marimuthu et al., "Detection of Parkinson's disease using machine learning approach," Ann. Romanian Soc. Cell Biol., vol. 25, pp. 2544–2550, 2021.

[19] K. Thakur et al., "Diagnosis of Parkinson's Disease Using Machine Learning Algorithms," in Congress on Intelligent Systems, Singapore: Springer Nature Singapore, 2022.

[20] P. K. Keserwani, S. Das, and N. Sarkar, "A comparative study: prediction of Parkinson's disease using machine learning, deep learning and nature-inspired algorithms," Multimedia Tools Appl., vol. 1, pp. 1–49, 2024.

[21] C. G. Goetz et al., "Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results," Movement Disorders, vol. 23, no. 15, pp. 2129–2170, 2008.

[22] T. J. Hirschauer, H. Adeli, and J. A. Buford, "Computer-aided diagnosis of Parkinson's disease using enhanced probabilistic neural network," J. Med. Syst., vol. 39, p. 1–12, 2015.

[23] M. Visser et al., "Assessment of autonomic dysfunction in Parkinson's disease: the SCOPA-AUT," Movement Disorders, vol. 19, no. 11, pp. 1306–1312, 2004.

[24] M. W. Johns, "A new method for measuring daytime sleepiness: the Epworth Sleepiness Scale," Sleep, vol. 14, no. 6, pp. 540–545, 1991.

[25] F. V. Lopez et al., "Does the Geriatric Depression Scale measure depression in Parkinson's disease?" Int. J. Geriatric Psychiatry, vol. 33, no. 12, pp. 1662–1670, 2018.

[26] K. Stiasny-Kolster et al., "The REM Sleep Behavior Disorder Screening Questionnaire – a new diagnostic instrument," Movement Disorders, vol. 22, no. 16, pp. 2386–2393, 2007.

[27] I. Obeso et al., "Semantic and phonemic verbal fluency in Parkinson's disease: Influence of clinical and demographic variables," Behavioural Neurol., vol. 25, no. 2, pp. 111–118, 2012.

[28] H.-J. Yang et al., "Measuring anxiety in patients with early-stage Parkinson's disease: Rasch analysis of the State-Trait Anxiety Inventory," Front. Neurol., vol. 10, p. 49, 2019.

[29] C. Cui *et al.*, "Deep multimodal fusion of image and non-image data in disease diagnosis and prognosis: a review," *Progress in Biomedical Engineering*, vol. 5, no. 2, p. 022001, Mar. 2023, doi: 10.1088/2516-1091/acc2fe.