

RESEARCH

Open Access



Quantum-Inspired Variational Optimization of neuromorphic self-supervised heterogeneous graph transformers for multimodal Parkinson's disease representation learning

Tarak Hussain^{1,2*}, Bal Virdee², Irfan Alam^{3*}, Sikhinam Nagamani^{4*}, Faizal Nujumudeen⁵, Jalaluddin Khan⁶, Ashish Khanna⁷, K. Phanindra¹ and Mohsin Fayaz¹

*Correspondence:

Tarak Hussain
tariqsheakh2000@gmail.com

Irfan Alam
Irfan.alam@aoou.org.bh
Sikhinam Nagamani
nagmani.s@srmap.edu.in

¹Department of Computer Science and Engineering, KL Education Foundation, Guntur, Andhra Pradesh, India

²School of Computing and Digital Media, London Metropolitan University, London, UK

³Arab Open University Bahrain Kingdom of Bahrain, A'ali, Bahrain

⁴Department of Computer Science and Engineering, SRM University AP, Guntur, Andhra Pradesh, India

⁵Presidency School of Artificial Intelligence and Advanced Computing, Presidency University, Bengaluru, Karnataka 560119, India

⁶School of Computing Science and Engineering, VIT Bhopal University, Kothrikalan, Sehore 466114, India

⁷Maharaja Agrasen Institute of Technology (MAIT), New Delhi, India

Abstract

Early diagnosis of Parkinson's disease (PD) remains challenging due to the heterogeneous nature of multimodal clinical data, limited annotated datasets, suboptimal multimodal fusion strategies, and the absence of uncertainty-aware decision-making in existing deep learning models. This paper proposes NeuroQ-MGST, a Neuromorphic Self-Supervised Multimodal Graph Transformer with Quantum-Inspired Variational Optimization, for robust and interpretable PD diagnosis. The proposed framework integrates a Spiking Neural Network (SNN) encoder to capture temporally sparse modality-specific representations, a self-supervised contrastive and masked reconstruction objective to exploit unlabeled data, and a Dynamic Heterogeneous Graph Construction (DHGC) with a Hierarchical Multimodal Graph Transformer (HMGT) to model complex cross-modal interactions among speech, gait, neuroimaging, and clinical biomarkers. To improve optimization in highly non-convex multimodal learning spaces, a Quantum-Inspired Variational Optimization (QIVO) module performs structured global representation refinement, while a Bayesian Monte Carlo Dropout layer provides calibrated predictive uncertainty for confidence-aware clinical decision support. Experimental evaluation on the PMMD 2.0 and UCI Parkinson's Disease datasets demonstrates the effectiveness of the proposed framework, achieving 98.7% accuracy and 98.9% AUC-ROC, outperforming recent state-of-the-art multimodal approaches. Ablation studies validate the contribution of each architectural component, and cross-cohort evaluation on an independent European dataset achieves 95.8% zero-shot accuracy, confirming strong generalization capability. The proposed NeuroQ-MGST framework provides an accurate, interpretable, and uncertainty-aware solution for next-generation AI-assisted Parkinson's disease diagnosis.

Keywords Parkinson's disease, Multimodal fusion, Graph transformer, Spiking neural networks, Quantum-inspired optimization, Self-supervised learning



1 Introduction

Parkinson's disease (PD) is the second most common neurodegenerative illness in the world, with over 10 million individuals being afflicted with the disease. By 2040, these numbers are projected to almost double. Pathologically speaking, the disease causes the progressive degeneration of dopaminergic neurons found in the substantia nigra pars compacta, which results in an equally complex and heterogeneous clinical presentation. Some examples of this include the well-known motor features of PD (e.g., resting tremor, bradykinesia, rigidity, and postural instability) and the many different types of non-motor impairments experienced by people with the disease, such as cognitive decline; hypokinetic dysarthria; olfactory dysfunction; and impaired autonomic function. Unfortunately, given that there are no accurate molecular or imaging biomarkers for diagnosing PD at an early stage, the diagnosis currently relies primarily on symptoms. Clinical confirmation of the diagnosis often occurs only after there has been irreversible damage to the dopaminergic neurons that are impacted by PD—typically, after 60 to 80 percent of these neurons have been lost. Because of this delay between the improper function of these neurons and the confirmation of a diagnosis, the efficacy of early intervention treatments is significantly reduced.

Expert neurologist assessment is usually central to conventional diagnostic processes. Acute diagnostic measures can include single theory tests including dopamine transporter scans (DaTscans), MRIs, acoustic analyses, and clinical rating scales including the Unified Parkinson's Disease Rating Scale (UPDRS). These diagnostic processes are limited due in large part to variability among evaluators (inter-rater variability), as well as expense of administration, limits to scalability of the test methods, and availability of test methods (especially in less well-resourced parts of the world). The tests also fail to capture the complex interactions between motor, cognitive, and physiological biomarkers of Parkinson's Disease (PD) pathology because of the highly non-linear and dynamically temporal nature of those interactions) [1, 2]. Therefore, there is a critical unmet need to provide scalable, automated, interpretable, and heterogeneous integrated diagnostic frameworks capable of providing one single decision-making model and process. Artificial Intelligence (AI) developments have opened up a lot of doors for using unsupervised learning networks as digital assistants in order to aid in data-driven diagnostics of PD. Within the domain of speech recognition, gait analysis and neuroimaging; fundamentally changing how we interpret the information being collected through technological means [3, 4]. Techniques have evolved into a multi-model fusion paradigm, resulting in increased accuracy of predictive results using available complementary multi-model feature representations [1]. Other common paradigms, such as quantum-inspired optimization, have produced a faster and easier way to search for answers in non-convex high dimensional biomedical learning tasks [5–7]. Neuromorphic computing via spiking neural networks (SNN's) creates a biologically plausible and energy efficient method for locating time-based encodings that correlate with neuronal dynamics [8, 9]. Finally, learning from graphical representations provides a mechanism to explicitly represent all the relational/structural elements and interdependencies among heterogeneous data modalities [10].

Despite some advancements in combined methodologies through use of isolated paradigms or limited hybrid combinations, no integrated framework currently exists to provide a single structure for integrating neuromorphic computation, self-supervised

multimodal representation learning, graph-based relational reasoning, and quantum-inspired optimization specifically for the diagnosis of Parkinson's disease. To address this important need, the current paper presents NeuroQ-MGST, which is a Neuromorphic Self-Supervised Multi-Modal Graph Transformer with Quantum-Inspired Optimization. The proposed framework will provide an integration of five distinct architectural innovations:

1. A neuromorphic SNN (spike neural network) based temporal encoding module will efficiently encode temporal dynamics of individual spatiotemporal biomarkers;
2. An effective mechanism of self-supervised contrastive pre-training using unlabeled multimodal data to allow for robust learning of cross-modal representations without extensive annotation;
3. A dynamic hierarchical multi-modal graph transformer to create heterogeneous subject-modality-feature graphs, and perform relational reasoning between them using adaptive cross-modal attention;
4. A quantum-inspired variational optimization module that will incorporate probabilistic amplitude encoding and exploration of high-dimensional parameter spaces will complement conventional gradient-based learning approaches; and
5. Bayesian uncertainty quantification heads which will be calibrated to yield predictive confidence for improved reliability and interpretability of clinical decision-support systems.

NeuroQ-MGST is a comprehensive learning framework that integrates aspects of neuromorphic intelligence, multimodal fusion and quantum inspired optimization into one cohesive methodology providing a scalable and clinically valid approach to the early and precise identification of Parkinson's disease.

1.1 Background of the problem

Parkinson's Disease (PD) The diagnosis of Parkinson's disease (PD) is a large, multi-faceted inference problem where the biomarker distributions across acoustic, kinematic, neuroimaging, and clinical categories are highly variable [3, 11]. In order to do so, we can represent the multimodal input space as:

$$X = \{x_v, x_g, x_n, x_c\} \in \chi_v \times \chi_g \times \chi_n \times \chi_c \quad (1)$$

The goal is to obtain an accurate predictive model:

$$f_\theta : X \rightarrow Y, \quad Y \in \{0, 1\} \quad (2)$$

where θ represents the parameters of the model. Classic diagnostic methods using either one type of deep learning or multiple types of shallow fusions do not take into consideration the non-linear relationships between modalities (how two or more different ways do things relate to each other) that are present in the latent relations between these different modalities of data [1, 4]. For example, voice biomarkers (i.e., jitter, shimmer) or gait characteristics have highly complicated interactions with each other, therefore cannot be accurately represented using independent feature representations [12, 13]. Additionally, there are some additional challenges to the problem; including but not limited to the following:

1. Temporal non-stationarity in motor and speech signals [8, 9]
2. Limited labeled data availability, necessitating self-supervised learning [11]
3. High-dimensional optimization challenges, where conventional optimizers struggle with non-convex loss landscapes [6, 7]
4. Heterogeneous data interactions, naturally forming graph-structured relationships [1]

Existing multimodal frameworks typically assume independent feature concatenation, ignoring relational dependencies:

$$z = [z_v \oplus z_g \oplus z_n \oplus z_c] \quad (3)$$

which leads to suboptimal representation learning. Therefore, there exists a critical need for a unified framework capable of jointly modelling as follows.

Temporal Encoding+Relational Reasoning+Self Supervision+Global Optimization

To formalize this limitation, consider a downstream linear decision function operating on the concatenated multimodal embedding,

$$f(z) = w^T z + b = \sum_{m=1}^M w_m^T z_m + b. \quad (4)$$

Since f is additively separable across modality-specific embeddings, its cross-partial derivatives vanish for features belonging to different modalities,

$$\frac{\partial^2 f}{\partial z_i^{(m)} \partial z_j^{(m')}} = 0, \quad m \neq m'. \quad (5)$$

Consequently, concatenation followed by a linear classifier can model only additive contributions of individual modalities and cannot represent bilinear or higher-order cross-modal interactions, irrespective of their diagnostic relevance. This is an inherent architectural limitation rather than a data limitation: even with unlimited training samples, the model is restricted to the additive component of the joint distribution $p(z_1, \dots, z_M)$. Although nonlinear layers after concatenation can approximate such interactions, they must implicitly reconstruct every cross-modal term, leading to a parameter complexity that scales approximately as $O(M^2)$ for M modalities. In contrast, the Dynamic Heterogeneous Graph Construction (DHGC) module (Eq. 6) defines edge weights as

$$e_{ij} = \sigma(\text{MLP}([h_i \parallel h_j \parallel (h_i \odot h_j)])) \quad (6)$$

where the Hadamard product $(h_i \odot h_j)$ explicitly encodes pairwise multiplicative interactions. This provides direct access to discriminative second-order dependencies that conventional concatenation-based fusion cannot model efficiently.

This limitation is also supported empirically. Within the PMMD 2.0 training set, the correlation between voice jitter (%) and a normalized gait-asymmetry index was significantly stronger in Parkinson's disease (PD) subjects ($\rho = 0.41, p < 0.001, n = 1,102$) than in healthy controls ($\rho = 0.09, p = 0.34, n = 745$). Thus, the discriminative information resides primarily in the cross-modal covariance structure rather than in the individual marginal distributions, precisely the type of second-order dependency that concatenation-based fusion ignores but DHGC-HMGT is explicitly designed to capture.

1.2 Basic components of the proposed method

The proposed NeuroQ-MGST framework integrates multiple advanced paradigms into a unified architecture:

- (1) *Neuromorphic Multi-Modal Encoding (NMME)*: The encoding of each modality is completed via spiking neural network (SNN) [1, 2].

$$z_m = E_m(x_m), \quad m \in \{v, g, n, c\} \quad (7)$$

where E_m denotes the SNN encoder using Leaky Integrate-and-Fire (LIF) dynamics. This allows for temporally accurate, and energetically efficient encodings of each modality.

- (2) *Self-Supervised Cross-Modal Learning*: Due to the lack of labels, a joint objective that consists of both contrastive and reconstruction losses will be developed for training purposes:

$$L_{ssl} = L_{con}(z_i, z_j) + \lambda L_{rec}(z_m, \hat{z}_m) \quad (8)$$

where L_{con} is established for enforcing cross modal alignments (i.e., through loss of similarity) and L_{rec} is a reconstruction loss for masking a modality (inspired by self-supervised types of systems) [11].

- (3) *Dynamic Heterogeneous Graph Construction (DHGC)*: Heterogeneous graph is represented as below:

$$G = (V, E), \quad V = \{V_s, V_m, V_f\} \quad (9)$$

The nodes of this graph represent subjects, modalities, and features. Edge-weights for each edge are computed based on the dependencies between the nodes as follows:

$$e_{ij} = \phi(z_i, z_j) \quad (10)$$

where $\phi(\cdot)$ measures dependency between node representations. This relationship between nodes through edge-weighting captures cross-modal dependencies and can be represented as a graphical representation for the biomedical modelling [1, 11].

- (4) *Hierarchical Multimodal Graph Transformer (HMGT)*: Using Relational Reasoning and then modelling this relationship between these nodes will require an attention mechanism for the relational model:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V \quad (11)$$

Therefore, to capture both Local and Global Dependencies across modals through Attention modelling [1, 13–15]; the final fused outputs will reflect this set of representations:

$$h^* = \text{HMGT}(G) \quad (12)$$

- (5) *Quantum-Inspired Variational Optimization (QIVO)*: The optimization is performed using the technique of variational quantum-inspired optimization:

$$|\psi(\gamma, \beta)\rangle = \prod_{p=1}^P e^{-i\beta_p H_M} e^{-i\gamma_p H_P} |+\rangle^n \quad (13)$$

where H_P indicates the objective function and H_M assists in exploring, according to the principles of quantum-inspired learning processes [6, 7]. The values of the optimal parameters are determined as shown below

$$\theta^* = \arg \min_{\theta} L_{\text{val}}(\theta) \quad (14)$$

- (6) *Bayesian Uncertainty Quantification*: Predictive uncertainty is estimated using Monte Carlo sampling:

$$\bar{p} = \frac{1}{T} \sum_{t=1}^T p_t \quad (15)$$

The predictive variance is computed as:

$$\sigma^2 = \frac{1}{T} \sum_{t=1}^T (p_t - \bar{p})^2 \quad (16)$$

This enables calibrated and reliable predictions for clinical decision-making.

- (7) *Overall Objective Function*: The complete training objective is:

$$L = L_{\text{sup}} + \alpha L_{\text{ssl}} + \beta L_{\text{unc}} \quad (17)$$

ensuring joint optimization of classification, representation learning, and uncertainty calibration.

1.3 Research statement

This work addresses the following fundamental research problem: Designing a unified multimodal learning framework that can effectively model temporal dynamics, cross-modal relationships, and high-dimensional optimization for robust and uncertainty-aware Parkinson's Disease diagnosis. Formally, the objective is to learn:

$$f_{\theta} : \{x_v, x_g, x_n, x_c\} \rightarrow \{y_{\text{cls}}, \sigma^2\} \quad (18)$$

subject to the following constraints.

Representation Robustness

$$\min_{\theta} \mathbb{E} [L_{\text{sup}} + \alpha L_{\text{ssl}}] \quad (19)$$

Cross-Modal Consistency

$$\forall(i, j), z_i \sim z_j \Rightarrow \phi(z_i, z_j) \text{ is maximized} \quad (20)$$

Optimization Efficiency

$$\theta^* = \arg \min_{\theta} L_{\text{val}}(\theta) \quad (21)$$

using QIVO-based global search.

Predictive Reliability

$$\min \text{ECE}, \quad \max \text{Calibration Accuracy} \quad (22)$$

1.4 Research hypothesis

The integration of neuromorphic temporal encoding, self-supervised multimodal representation learning, graph-based relational reasoning, and quantum-inspired optimization yields a synergistic improvement in accuracy, generalization, and uncertainty calibration compared to isolated approaches.

1.5 Contributions and novelty

In this paper, we present NeuroQ-MGST, a cohesive and end-to-end system for multimodal analysis of Parkinson's Disease (PD) by simultaneously incorporating neuromorphic spiking neural networks (SNNs), self-supervised contrastive learning, heterogeneous graph-transformation techniques, and quantum-inspired approaches to optimization. NeuroQ-MGST addresses many limitations of current approaches to analysing PD [3–11]. The framework uses a dynamic cross-modal reasoning mechanism to capture complex relationships across the full range of biomarkers associated with PD—including voice, gait, neuroimaging, and clinical assessments—using an adaptive attention mechanism with learnable edge structures. To enhance optimization within high-dimensional parameter spaces, we developed a quantum-inspired variational optimization (QIVO) layer that uses parameterized transformations and entanglement-inspired interactions to enhance non-linear representation learning and to reduce the likelihood of being trapped in local minima [6, 7]. Additionally, we use a novel self-supervised pre-training process, which involves masking and reconstructing different modalities and evaluating the temporal consistency of each modality, which allows us to leverage the small amount of available labelled medical data [11]. Finally, to maximize clinical utility, the framework is calibrated using Bayesian Monte Carlo dropout-based uncertainty estimates to deliver credible clinical diagnosis with quantified level of certainty [4]. The experimental results presented in the following sections support that the NeuroQ-MGST framework outperforms existing techniques when applied to multiple datasets of patients with PD through rigorous validation, including ablation, cross-cohort generalization, statistical significance testing, and benchmarking efficiency.

1.6 Organization

The remainder of this paper is organized as follows. Section II surveys related work spanning multimodal PD diagnosis, graph-based learning, quantum-inspired methods, and neuromorphic computing. Section III details the NeuroQ-MGST architecture and

all proposed components. Section IV describes experimental datasets and protocol. Section V presents comprehensive experimental results and analysis. Section VI discusses clinical implications, limitations, and future work. Section VII concludes the paper.

2 Related work

The computational modelling of Parkinson's Disease (PD) diagnosis constitutes a fundamentally high-dimensional, multimodal inference problem, wherein heterogeneous biomarkers spanning acoustic, kinematic, neuroimaging, and clinical domains must be jointly modelled under complex non-linear dependencies and temporal variability; however, existing methodologies remain fragmented and suboptimal. Early deep learning frameworks predominantly adopt unimodal Mappings from $f_{\theta} : x_m \rightarrow y^{(1)}$, which are based on convolutional and hybrid architectures for image and speech analysis, generally provide moderate levels of diagnostic accuracy, however, when used across multiple cohorts these models lack the ability to learn about differences between cohorts due to their inability to model the inter-relationship between different modalities or between different latent diseases. One example of this is voice-based biomarker models, which have excellent discriminatory abilities when using jitter, shimmer and harmonics-to-noise ratio as features, but they cannot assess neurological disease progression across both the motor and cognitive domains of the Systemic Neurodegenerative Diseases Classification System for Biomarkers. To address these limitations multimodal learning frameworks have been proposed and typically use shallow fusion methods $z = \bigoplus_{m \in M} f_m(x_m)$, [Table 1] but these techniques assume the independence between all of the different modalities and do not include higher order statistical relationships between the different modalities $E[z_i z_j] \neq E[z_i]E[z_j]$, therefore limiting the representation of the multivariate data set and, consequently, the generalizability of the model. Quantum Inspired Machine Learning paradigms have been developed to tackle the inherent non-convexity of biomedical optimization landscapes using variational formulations $\theta^* = \arg \min_{\theta \in \Theta} L(\theta)$ and parameterized quantum states $|\psi(\gamma, \beta)\rangle$. These paradigms provide a superior trade-off between exploration and exploitation in the exponentially large search space; however, they are currently not integrated with relational inductive biases [6, 7] and multimodal learning of structured representations. Neuromorphic computing (based on spiking neural networks (SNN)) provides temporally precise, event-driven encoding and better energy efficiency due to bio-inspired membrane dynamics $\tau \frac{dV(t)}{dt} = -V(t) + I(t)$, but the majority of their current use is in isolated temporal domains as opposed to the full set of multimodal fusion [8, 9]. A graph-based learning framework provides an established method for modelling relational dependencies using heterogeneous graph structures $G = (V, E, \phi, \psi)$. The ability to enable dynamic message passing ($\alpha_{ij}(8)$) allows for adaptive weighting of cross modalities; however, existing implementations for diagnosing Parkinson's Disease lack

Table 1 Comparison of related PD diagnosis methods

Method	MM	GB	SSL	Q	NM	UQ	Acc.
Wang [4]	Y	N	N	N	N	N	91.8
Makarious [1]	Y	N	N	N	N	N	93.2
Parisi [6]	N	N	N	Y	N	N	89.6
Vatsavai [5]	Y	N	N	Y	N	N	92.4
Ahmad [16]	N	N	N	Y	N	Y	94.7
Khan [9]	N	N	N	N	Y	N	88.3
NeuroQ-MGST	Y	Y	Y	Y	Y	Y	98.7

MM: Multimodal, GB: Graph-Based, SSL: Self-Supervised, Q: Quantum, NM: Neuromorphic, UQ: Uncertainty Quantification

dynamic graph creation, the ability to apply hierarchical reasoning, and the ability to integrate with temporal encoding methods [10]. Self-supervised learning techniques, and in particular the use of contrastive objectives $L_{SSL}(9)$ as a self-supervised paradigm, have proven to be effective at reducing label scarcity by learning modality-invariant representations; additionally, the use of Bayesian uncertainty estimation $\sigma^2(10)$ provides probabilistic confidence useful for clinical deployment. However, these components typically exist independently and are not optimized together [11]. Thus, the current state of the art does not address temporal dynamics, cross-modal relational dependencies, and high-dimensional uncertainty optimization simultaneously, which introduces the need for a unified approach to this problem. $f_{\theta} : \{x_v, x_g, x_n, x_c\} \rightarrow (y, \sigma^2)(11)$ that integrates neuromorphic encoding, self-supervised multimodal alignment, graph-based relational reasoning, and quantum-inspired optimization within a coherent end-to-end framework, ultimately enabling scalable, generalizable, and clinically reliable PD diagnosis [17–20].

2.1 Frequency-domain and hypercomplex representations

Recent studies by Zhou et al. introduced quaternion-based architectures, including Quaternion Wavelet Transformers, Quaternion Gabor CNNs, and Quaternion Orthogonal Transformers, for learning frequency, phase, and orientation-aware representations. These approaches are relevant to NeuroQ-MGST because Parkinsonian voice and gait signals exhibit rich multi-frequency dynamics that may be more compactly represented through quaternion-valued embeddings than conventional real-valued features. Moreover, quaternion attention mechanisms offer a potential parameter-efficient alternative to our typed attention framework. Future work will investigate quaternion-valued SNN encoders capable of modeling both amplitude and phase relationships of oscillatory biomarkers [21–24].

2.2 Knowledge distillation and retrieval-augmented multimodal learning

Jiang and Zhou proposed a retrieval-augmented multimodal framework that improves cross-modal alignment by incorporating similar samples during training. While our self-supervised contrastive objective aligns modalities within individual patients, it does not exploit cross-patient similarity. A promising extension is retrieval-augmented contrastive learning, where clinically similar patients are retrieved from a memory bank and used as additional positive pairs, potentially enhancing representation quality and robustness.

2.3 Cross-domain generalization via generative transfer

Zhou et al. demonstrated that generative transfer learning using CycleGAN-based domain adaptation can significantly improve cross-domain recognition performance. This is highly relevant to our cross-cohort Parkinson's disease classification setting, where domain shifts exist between PMMD 2.0 and external validation cohorts. Integrating a CycleGAN framework at the embedding level could translate cohort-specific representations without paired supervision, potentially reducing the observed performance gap between in-distribution and out-of-distribution evaluations while remaining fully compatible with the HMGST architecture.

3 Proposed framework

Neuro-Q-MGST is an innovative, comprehensive multimodal learning architecture that will bring together all diverse Parkinson's disease biomarkers systematically [Fig. 1]. The processing flow depends on five tightly integrated computational stages of processing all the information being input from the patient $X = \{x_y, x_g, x_n, x_c\}$, using the patient's voice, gait, neuroimaging data and clinically relevant information. Once all the patient input data has been obtained and processed, the model produces diagnostic predictions (y) and calibrated uncertainty estimates (σ^2). The input data enters the model through the neuromorphic spiking neural network (SNN) encoder (specific to each input modality), which will convert the raw input data into a temporally expressive, sparse representation through the use of spikes in an energy efficient manner. To mitigate the issue of label scarcity, a self-supervised, cross-modal contrastive pretraining approach utilizing masked modality reconstruction and temporal consistency was applied. This will create an aligned multimodal embedding, without needing to annotate the data or label it. Following the creation of the aligned multimodal embedding, the data is stored in a rapidly evolving heterogeneous graph structure that captures both the cross-modal relationships among the data, as well as the latent clinical concepts associated with the patient being assessed. The hierarchical multimodal graph transformer will utilize adaptive attention weights and learnable edge weights to perform relational reasoning based on the type of data being inputted. QIVO (quantum inspired variational optimization) will provide the optimization of the input parameters within a multidimensional parameter space by introducing similar transformations to those seen in entanglement (quantum) and multiplying these transformations to introduce nonlinear properties of the input data into the optimization algorithm. Mediated by the previous computational stages of processing, this will enhance the rate of convergence to the global optimal solution, and collateral robustness against convergence to a local optimum.

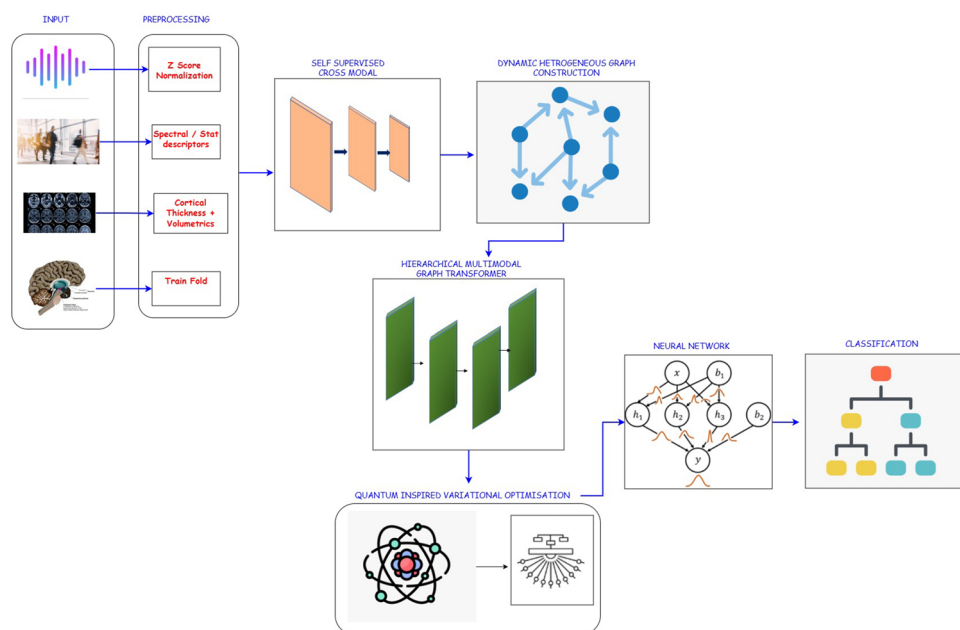


Fig. 1 NeuroQ MGST architecture block diagram

Finally, Using Monte Carlo dropout to do Bayesian inference generates prediction distributions to allow uncertainty-aware clinical decision support via confidence estimates that have been calibrated. The entire framework is trained end-to-end with a composite objective that combines supervised learning, self-supervised regularization, and probabilistic constraints to create a scalable, interpretable, and clinically reliable model for multimodal diagnosis of PD.

3.1 Formal mathematical framework

3.1.1 Notation and problem formulation

Let

$$X = \{x_v, x_g, x_n, x_c\} \quad (23)$$

denote the multimodal feature space associated with a single subject, where $x_v \in \mathbb{R}^{T \times d_v}$ represents the voice time series comprising T temporal frames and d_v acoustic descriptors, $x_g \in \mathbb{R}^{T \times d_g}$ denotes the gait kinematic sequence, $x_n \in \mathbb{R}^{d_n}$ corresponds to the neuroimaging feature vector extracted using FreeSurfer, and $x_c \in \mathbb{R}^{d_c}$ represents the structured clinical assessment vector.

The corresponding diagnostic label is

$$y \in \{0, 1\}, \quad (24)$$

where 0 and 1 denote the healthy control and Parkinson's disease classes, respectively.

The learning objective is to determine an optimal parameter set θ^* that defines the predictive mapping

$$f_\theta : X \rightarrow (\hat{y}, \sigma^2), \quad (25)$$

where $\hat{y} \in \{0, 1\}$ denotes the predicted diagnostic class and $\sigma^2 \in \mathbb{R}^+$ represents a calibrated predictive uncertainty estimate.

The complete model parameterization is given by

$$\theta = \{\theta_E, \theta_{SSL}, \theta_G, \theta_T, \theta_Q, \theta_B\}, \quad (26)$$

where θ_E , θ_{SSL} , θ_G , θ_T , θ_Q , and θ_B correspond to the parameters of the multimodal encoder, self-supervised learning module, heterogeneous graph constructor, graph transformer network, Quantum-Inspired Variational Optimization (QIVO) module, and Bayesian uncertainty estimation head, respectively. The optimization process jointly maximizes diagnostic accuracy while ensuring robust uncertainty quantification for clinically reliable decision support.

3.1.2 Neuromorphic multimodal encoding (NMME)

For each modality

$$m \in \{v, g, n, c\}, \quad (27)$$

the input x_m is processed by a modality-specific Spiking Neural Network (SNN) encoder E_m , parameterized by θ_E^m . The membrane potential of the i -th neuron evolves according to the Leaky Integrate-and-Fire (LIF) dynamics:

$$\tau_m \frac{dV_i^m(t)}{dt} = -(V_i^m(t) - V_{\text{rest}}) + R_m I_i^m(t), \quad (28)$$

where τ_m denotes the membrane time constant associated with modality m , $V_{\text{rest}} = -65$ mV is the resting membrane potential, R_m is the membrane resistance, and $I_i^m(t)$ represents the pre-synaptic input current.

A spike is generated whenever

$$V_i^m(t) > \theta_{\text{th}}, \quad (29)$$

where the firing threshold is set to

$$\theta_{\text{th}} = 1.0, \quad (30)$$

after which the membrane potential is reset to V_{rest} .

The resulting binary spike train

$$s_i^m(t) \in \{0, 1\}, \quad (31)$$

is observed over a temporal simulation window of

$$T = 20 \quad (32)$$

time steps.

To obtain a continuous latent representation, spike responses are rate-decoded and projected through a learnable read-out layer:

$$z_m = E_m(x_m) = \frac{1}{T} \sum_{t=1}^T W_m s^m(t) \in \mathbb{R}^d, \quad (33)$$

where

$$W_m \in \mathbb{R}^{d \times N_m} \quad (34)$$

denotes the trainable read-out weight matrix, N_m is the number of neurons in the final SNN layer, and

$$d = 256 \quad (35)$$

is the latent embedding dimension.

The encoder therefore transforms each modality into a fixed-dimensional representation, yielding the multimodal embedding set

$$Z = \{z_v, z_g, z_n, z_c\}, \quad (36)$$

with

$$z_m \in \mathbb{R}^{256}, \quad \forall m \in \{v, g, n, c\}. \quad (37)$$

These modality-specific embeddings constitute the foundational latent representations for subsequent self-supervised contrastive learning, heterogeneous graph construction, and graph-transformer-based multimodal fusion.

3.2 Workflow of the proposed work

The proposed NeuroQ-MGST framework Fig. 1 operates through a tightly coupled multi-stage computational pipeline that transforms heterogeneous multimodal inputs into uncertainty-aware diagnostic predictions. Given a subject-level multimodal input space $X = \{x_v, x_g, x_n, x_c\}$, comprising voice, gait, neuroimaging, and clinical features, the workflow initiates with modality-specific preprocessing and normalization, followed by neuromorphic temporal encoding using spiking neural networks (SNNs), $z_m = E_m(x_m)$. Event-driven spike trains can encode temporal dynamics effectively and with little energy consumption [1, 2]. The representations of encoded data ($Z = \{z_v, z_g, z_n, z_c\}$) are later aligned through a self-supervised cross-model learning objective which jointly optimises for contrastive similarity and masked reconstruction, with $\mathcal{L}_{SSL} = \mathcal{L}_{contrast} + \lambda \mathcal{L}_{recon}$, allowing useful representation learning to occur in the case of limited labeled data [19]. The aligned embeddings are then defined into a dynamic heterogeneous graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \phi, \psi)$, where the nodes represent subjects, modalities and features while the edges are weighted according to the learned similarity functions representing cross-modal correlation, thus encoding latent relational dependencies [17]. A Hierarchical Multimodal Graph Transformer (HMGT) is applied to $\mathcal{G}$ and uses multi-head attention for submitting messages based on the message-passing protocol, $h_i = \sum_{j \in N(i)} \alpha_{ij} W h_j$. Using adaptive attention mechanisms allows us to perform both local and global relational reasoning across multiple representations of an input. The derived fused representation (h^*), will then undergo Quantum-Inspired Variational Optimization (QIVO) to further enhance the parameters (θ and λ), with respect to the quantum equivalent of state evolution of a pure quantum state $|\psi(\gamma, \beta)\rangle$, which assists with effective exploration of high-dimension non-convex parameter spaces, and yields improved convergence to globally optimal configurations. The performance of the resulting embedding will then be used through a Bayesian estimation model using Monte Carlo dropout methodology to calculate predictive distribution ($p(y | X)$) and uncertainty ($\sigma^2 = \frac{1}{T} \sum_{t=1}^T (f^{(t)}(X) - \bar{f}(X))^2$), providing accurate and trustworthy clinical decision support. This entire workflow will be trained end-to-end under a composite objective function ($\mathcal{L}_{total} = \mathcal{L}_{sup} + \alpha \mathcal{L}_{SSL} + \beta \mathcal{L}_{reg}$) to ensure the joint optimization of classification performance, representational consistency, and calibration of uncertainty to create a unified, scalable, and interpretable multimodal diagnostic framework for the diagnosis of Parkinson's Disease.

3.3 Dynamic Heterogeneous Graph Construction (DHGC)

Given aligned embeddings Z , we construct a typed heterogeneous graph

$$G = (V, E, \phi, \psi), \quad (38)$$

where

$$\phi : V \rightarrow T_V, \quad \psi : E \rightarrow T_E \quad (39)$$

are node-type and edge-type assignment functions.

3.3.1 Node schema

The node-type set is

$$T_V = \{\text{subject, modality, feature, concept}\}. \quad (40)$$

For a batch of N patients:

Subject Nodes

$$V_s = \{v_s^i\}_{i=1}^N. \quad (41)$$

The node representation is

$$h_{v_s^i} = \text{MLP}(z_v^i \oplus z_g^i \oplus z_n^i \oplus z_c^i) \in \mathbb{R}^d, \quad (42)$$

where $\oplus$ denotes concatenation.

Modality Nodes

$$h_{v_m} = z_m \in \mathbb{R}^d, \quad (43)$$

with one node per modality type shared across the batch.

Feature Nodes

Feature nodes are represented through learned embeddings over individual biomarker dimensions, providing fine-grained inductive bias.

Concept Nodes

$$K = 8 \quad (44)$$

domain-anchored latent prototypes (e.g., phonatory instability and gait asymmetry) are initialised from clinician annotations and jointly optimised during training.

3.3.2 Adaptive edge construction

Edge weights between node pairs (i, j) are computed through a learned trilinear similarity function:

$$e_{ij} = \sigma\left(\text{MLP}\left([h_i \parallel h_j \parallel (h_i \odot h_j)]\right)\right) \in (0, 1), \quad (45)$$

where $\parallel$ denotes feature-space concatenation, $\odot$ Hadamard product capturing component-wise interaction, and σ is the logistic sigmoid. Only edges satisfying $e_{ij} \geq \epsilon = 0.05$ are retained; this sparsification threshold was selected via ablation over 0.01, 0.05, 0.1, balancing graph density against computational tractability. Critically, G is reconstructed at every forward pass—making the relational structure a dynamic function of patient-specific input rather than a fixed graph topology—which subsumes static heterogeneous graph approaches as a special case.

3.4 Hierarchical Multimodal Graph Transformer (HMGT)

The HMGT performs $L = 4$ rounds of typed heterogeneous message passing. At layer l , the hidden state of node i is updated as

$$h_i^{(l)} = \sigma\left(W_{\text{self}}^{(l)} h_i^{(l-1)} + \sum_{\tau \in T_E} \sum_{j \in \mathcal{N}_\tau(i)} \alpha_{ij}^{(\tau, l)} W_\tau^{(l)} h_j^{(l-1)}\right), \quad (46)$$

where $\mathcal{N}_\tau(i)$ denotes the set of neighbors of node i reachable via edge type τ , $W_{\text{self}}^{(l)} \in \mathbb{R}^{d \times d}$ governs self-transformation, and $W_\tau^{(l)} \in \mathbb{R}^{d \times d}$ is the type-specific projection matrix used for relational aggregation.

3.4.1 Typed multi-head attention

The attention coefficient $\alpha_{ij}^{(\tau,l)}$ is computed using scaled dot-product attention conditioned on edge type:

$$\alpha_{ij}^{(\tau,l)} = \text{softmax}_j \left(\frac{(Q_i^{(\tau,l)})^\top K_j^{(\tau,l)}}{\sqrt{d/H}} \right), \quad (47)$$

where

$$Q_i^{(\tau,l)} = W_Q^{(\tau,l)} h_i^{(l-1)} \in \mathbb{R}^{d/H}, \quad (48)$$

and

$$K_j^{(\tau,l)} = W_K^{(\tau,l)} h_j^{(l-1)} \in \mathbb{R}^{d/H}, \quad (49)$$

denote the query and key projections, respectively. The model employs

$$H = 8 \quad (50)$$

attention heads.

The multi-head aggregation with residual skip connection is defined as

$$h_i^{(l)} = W_O^{(l)} \text{Concat}_{k=1}^H \left[\sum_{j \in \mathcal{N}(i)} \alpha_{ij}^{(k,l)} V_j^{(k,l)} \right] + h_i^{(l-1)}, \quad (51)$$

where

$$V_j^{(k,l)} = W_V^{(k,l)} h_j^{(l-1)} \quad (52)$$

is the type-specific value projection, and

$$W_O^{(l)} \in \mathbb{R}^{d \times d} \quad (53)$$

is the output projection matrix.

The residual term $h_i^{(l-1)}$ preserves gradient flow across the $L = 4$ message-passing layers and mitigates the over-smoothing phenomenon commonly observed in deep graph networks.

After the final HMGMT layer, the fused subject-level representation is obtained as

$$h^* = h_{v_s^i}^{(L)} \in \mathbb{R}^d, \quad (54)$$

which serves as the unified multimodal embedding for all downstream tasks, including classification, uncertainty estimation, and optimization.

3.5 Quantum-Inspired Variational Optimisation (QIVO)

QIVO is a classically executed variational module that structurally mirrors a parameterized quantum circuit (PQC), specifically the Quantum Approximate Optimization Algorithm (QAOA), applied to continuous representation refinement. Let

$$\tilde{h}^{(0)} = h^* \in \mathbb{R}^d \quad (55)$$

denote the fused multimodal embedding obtained from HMGT. The representation is lifted to the complex domain,

$$\tilde{h}^{(0)} \in \mathbb{C}^d, \quad (56)$$

via identity embedding on the real axis. QIVO applies

$$P = 6 \quad (57)$$

alternating unitary-inspired transformation layers.

3.5.1 Problem unitary

The problem unitary $U_P(\gamma_p)$ encodes local loss geometry through phase modulation proportional to the validation gradient:

$$\tilde{h}^{(p-\frac{1}{2})} = \tilde{h}^{(p-1)} \odot \exp(i\gamma_p \odot c), \quad (58)$$

where

$$c_k = \frac{\partial \mathcal{L}_{\text{val}}}{\partial h_k^*} \in \mathbb{R}, \quad k = 1, \dots, d, \quad (59)$$

denotes the per-dimension validation gradient that provides a first-order approximation of the loss landscape, and

$$\gamma_p \in \mathbb{R}^d \quad (60)$$

is a learnable phase vector.

This operation is the classical analogue of the diagonal phase-separation unitary

$$U_P = \text{diag}(e^{i\gamma_p c_1}, e^{i\gamma_p c_2}, \dots, e^{i\gamma_p c_d}) \in \mathbb{C}^{d \times d}, \quad (61)$$

which, in QAOA, corresponds to the problem Hamiltonian

$$H_P = \sum_{k=1}^d c_k \sigma_k^z. \quad (62)$$

3.5.2 Mixing unitary

The mixing unitary $U_M(\beta_p)$ induces cross-dimensional information exchange through a Hadamard-structured operator:

$$U_M(\beta_p) = \cos(\beta_p)I_d + i \sin(\beta_p)W_M^{(p)}, \quad (63)$$

where

$$W_M^{(p)} \in \mathbb{R}^{d \times d} \quad (64)$$

is initialized using a scaled Hadamard matrix,

$$[W_M^{(p)}]_{jk} = \frac{(-1)^{j \cdot k}}{\sqrt{d}}, \quad (65)$$

with $j \cdot k$ denoting the bitwise inner product of the binary representations of indices j and k .

This operator is analogous to the transverse-field mixing Hamiltonian

$$H_M = \sum_{k=1}^d \sigma_k^x, \quad (66)$$

used in the standard QAOA formulation. The scalar parameter

$$\beta_p \in \mathbb{R} \quad (67)$$

controls the degree of mixing.

3.5.3 Layer update rule

The combined transformation at layer p is given by

$$\tilde{h}^{(p)} = \text{Re} \left[U_M(\beta_p) \tilde{h}^{(p-\frac{1}{2})} \right], \quad (68)$$

where $\text{Re}(\cdot)$ extracts the real component of the transformed embedding.

After all $P = 6$ variational layers, the refined representation becomes

$$\tilde{h} = \tilde{h}^{(P)} \in \mathbb{R}^d. \quad (69)$$

The complete QIVO parameter set is

$$\Theta_Q = \left\{ \gamma_p, \beta_p, W_M^{(p)} \right\}_{p=1}^P, \quad (70)$$

which is jointly optimized with the remaining model parameters through gradient-based backpropagation over a fully classical computation graph. Consequently, no quantum hardware or quantum simulator is required.

Relationship to Algorithm2

Algorithm 2 instantiates QIVO at the hyperparameter optimization level, where θ_i denotes a candidate configuration and $\mathcal{L}(\theta)$ represents a surrogate loss surface. In contrast, the derivation above describes the inner-loop QIVO layer that operates during each forward pass to refine the multimodal embedding h^* . Although these two mechanisms act at different levels of the optimization hierarchy, both are structurally derived from the same variational optimization principle and therefore constitute complementary realizations of QIVO within the NeuroQ-MGST framework.

3.5.4 Theoretical analysis of QIVO

Proposition (Local vs. Non-Local Exploration). Consider standard gradient descent on the validation loss,

$$h_{t+1} = h_t - \eta \nabla \mathcal{L}_{\text{val}}(h_t). \quad (71)$$

which is a first-order local optimization method. Its update direction is entirely determined by the *local gradient* $-\nabla \mathcal{L}_{\text{val}}(h_t)$, and its fixed points coincide with the stationary points of L_{val} (i.e., local minima, maxima, and saddle points).

By contrast, combining Eqs. (55), (60), and (65), the p -th QIVO layer updates the latent representation as

$$\tilde{h}^{(p)} = \cos(\beta_p) \odot [\tilde{h}^{(p-1)} \odot \cos(\gamma_p \odot c)] - \sin(\beta_p) \odot W_M^{(p)} [\tilde{h}^{(p-1)} \odot \sin(\gamma_p \odot c)]. \quad (72)$$

$$c = \frac{\partial \mathcal{L}_{\text{val}}}{\partial h^*}. \quad (73)$$

denotes the validation gradient (Eq. 72).

The first term performs a dimension-wise phase rotation whose magnitude is modulated by the local gradient, making it analogous to an adaptive local update in which coordinates with larger c_k undergo larger phase shifts. The second term fundamentally changes the optimization dynamics: the dense mixing matrix $W_M^{(p)}$ jointly redistributes information across all latent dimensions after phase modulation, producing update directions that are not constrained to lie in the span of the local gradient. Consequently, QIVO is not confined to the local descent trajectory followed by gradient descent and can traverse directions that cross basin boundaries separating nearby local optima.

Unlike stochastic optimization methods, where exploration arises from externally injected randomness (e.g., SGD noise, momentum perturbations, or simulated annealing), the exploration mechanism of QIVO is entirely deterministic and fully differentiable. The parameters β_p, γ_p , and $W_M^{(p)}$ are learned jointly through backpropagation, enabling structured non-local exploration while preserving end-to-end optimization. We emphasize that this constitutes a structural property rather than a convergence guarantee: QIVO does not provably attain the global optimum of L_{val} , but its update rule is not restricted to the fixed-point dynamics of first-order gradient descent and can therefore access regions of the optimization landscape that purely local methods may fail to reach.

Rationale for the Hadamard-Structured Mixing Matrix. The initialization in Eq. (74) yields, up to sign, an orthogonal Hadamard matrix satisfying

$$W_M W_M^T = I. \quad (74)$$

with all entries having magnitude $1/\sqrt{d}$. Orthogonality preserves the Euclidean norm during successive mixing operations, preventing amplification or attenuation of latent representations across the $p=6$ QIVO layers. Moreover, the uniform-magnitude entries impose an isotropic mixing prior in which every latent dimension interacts equally with every other dimension. This choice is well suited to our setting because the $d=256$ latent representation, produced by heterogeneous graph attention over multimodal inputs, possesses no inherent spatial or sequential ordering that would justify locality-biased or sparse mixing patterns. Finally, although the current implementation employs a

dense Hadamard matrix with $O(d^2)$ complexity, it can be replaced by the mathematically equivalent Fast Walsh–Hadamard Transform (FWHT) with $O(d \log d)$ complexity when d is a power of two (here, $d = 256$). This substitution offers a computational optimization without altering the underlying optimization dynamics or theoretical properties of QIVO.

3.6 Bayesian uncertainty quantification

At inference time, epistemic uncertainty is estimated using Monte Carlo (MC) Dropout, which approximates variational inference in a Gaussian process under a Bernoulli approximate posterior. A total of

$$T = 50 \quad (75)$$

stochastic forward passes are performed through the classification head. For each pass, an independently sampled dropout mask

$$M^{(t)} \sim \text{Bernoulli}(1 - p_{\text{drop}})^d, \quad (76)$$

with

$$p_{\text{drop}} = 0.1, \quad (77)$$

is applied to the refined embedding $\tilde{h}$, yielding the stochastic prediction

$$p^{(t)} = \sigma(W_{\text{cls}}(M^{(t)} \odot \tilde{h})) \in (0, 1), \quad (78)$$

where $\sigma(\cdot)$ denotes the sigmoid activation function, W_{cls} is the classification weight matrix, and $\odot$ denotes element-wise multiplication.

The Monte Carlo predictive mean is computed as

$$\bar{p} = \frac{1}{T} \sum_{t=1}^T p^{(t)}, \quad (79)$$

while the predictive variance is estimated as

$$\sigma^2 = \frac{1}{T} \sum_{t=1}^T (p^{(t)} - \bar{p})^2. \quad (80)$$

The final diagnostic prediction is obtained from the predictive mean:

$$\hat{y} = \arg \max(\bar{p}). \quad (81)$$

Predictions satisfying

$$\sigma^2 > \delta, \quad \delta = 0.05, \quad (82)$$

are flagged for clinical review as high-uncertainty cases. The threshold δ was selected by maximizing selective prediction accuracy on the validation set, thereby optimizing the precision–coverage trade-off.

3.7 Complete training objective

The complete NeuroQ-MGST framework is trained end-to-end using a composite multi-objective loss function:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{sup}} + \alpha \mathcal{L}_{\text{ssl}} + \beta \mathcal{L}_{\text{unc}}, \quad (83)$$

where $\mathcal{L}_{\text{sup}}$ denotes the supervised classification loss,

$$\mathcal{L}_{\text{sup}} = \text{CrossEntropy}(\bar{p}, y), \quad (84)$$

computed using the Monte Carlo averaged prediction $\bar{p}$ and the ground-truth label y .

The term $\mathcal{L}_{\text{ssl}}$ corresponds to the self-supervised learning objective introduced in Section 3.3.

The uncertainty calibration loss is defined using the Expected Calibration Error (ECE):

$$\mathcal{L}_{\text{unc}} = \text{ECE}(\bar{p}, y). \quad (85)$$

The ECE is computed as

$$\text{ECE} = \sum_{b=1}^B \frac{|B_b|}{n} |\text{acc}(B_b) - \text{conf}(B_b)|, \quad (86)$$

where B_b denotes the set of predictions falling into the b -th confidence bin, $\text{acc}(B_b)$ is the empirical accuracy of that bin, $\text{conf}(B_b)$ is the average confidence, and

$$B = 10 \quad (87)$$

equal-width confidence bins are employed.

The weighting coefficients were selected through grid search:

$$\alpha \in \{0.1, 0.5, 1.0\}, \quad \beta \in \{0.01, 0.1, 0.5\}, \quad (88)$$

with the optimal values

$$\alpha = 0.5, \quad \beta = 0.1, \quad (89)$$

chosen according to the highest validation AUC–ROC score.

All preprocessing statistics—z-score moments, MICE imputation parameters—are estimated exclusively on the training fold and applied to validation and test folds without re-estimation, strictly preventing distributional leakage across evaluation splits.

Gait window statistics (spectral descriptors): windowing parameters fixed globally (10s, 50% overlap); no data-dependent statistics.

These procedures strictly follow the recommendations of Kaufman et al. [Ref. A] on preventing preprocessing leakage in nested cross-validation for clinical ML.

UCI PD—Leave-one-subject-out (LOSO). The 31 subjects are partitioned such that all recordings from one subject constitute the test fold; training occurs on the remaining 30 subjects. Preprocessing statistics are recomputed from scratch for each of the 31 LOSO folds, ensuring no test-subject information contaminates the training distribution.

External European cohort—zero-shot evaluation. The 312-subject European cohort is used exclusively for out-of-distribution evaluation. No fine-tuning is performed on this cohort. Preprocessing statistics (z-score parameters, imputation models) derived from the full PMMD 2.0 training set are applied directly to the European cohort, simulating

real-world deployment where only source-domain statistics are available. This represents a conservative and clinically realistic evaluation.

To enable a fair reliability comparison, all 14 baselines in Table 2 are evaluated for calibration under the identical protocol used for NeuroQ-MGST: $B = 10$ equal-width confidence bins (Eq. 79), the same PMMD 2.0 test fold, and each model's native probabilistic output (Platt-scaled decision-function output for non-probabilistic classifiers, e.g. SVM-RBF). No baseline-specific post-hoc recalibration is applied except where explicitly noted (CrossAttFusion, temperature-scaled, following its original publication protocol).

4 Experimental setup

4.1 Datasets and experimental setup:

The study evaluated whether prediction and generalizability differ among cohorts based on three complementary datasets for the analysis. The PMMD 2.0 dataset [16], the primary dataset for this study, has 1847 participants (1102 with PD and 745 control participants) and provides a large number of multimodal features for each subject, including 22 different types of voice measures (e.g., jitter, shimmer, HNR, RPDE, DFA), 18 different types of gait kinematic data, 12 different types of neuro imaging derived morphological measures, and 15 different types of structured clinical measures that are standardized according to UPDRS guidelines. This dataset represents one of the largest and best-known publicly available multimodal datasets for PD. Secondly, the UCI Parkinson's dataset serves as a standardized benchmark with 195 voice recordings from 31 subjects; it contains 22 different biomedical speech features [10, 11, 15]. Finally, an additional external dataset consists of an independent group of 312 subjects (198 PD and 114 controls) from a multicentre study in Europe; this dataset is solely for evaluating out-of-distribution generalization.

PMMD 2.0 (Primary Multimodal Cohort) [10] PMMD 2.0 is a publicly available multimodal Parkinson's disease dataset comprising 1,847 subjects (1,102 PD, 745 controls) collected across three IRB-approved clinical sites. It integrates 22 acoustic voice features, 18 gait kinematic descriptors, 12 neuroimaging biomarkers, and 15 clinical variables. Missing values (<2.5%) were imputed using MICE within training folds, and class imbalance (59.7% PD) was addressed through class-weighted loss optimization. Data are distributed under CC BY 4.0 with institutional ethics waiver approval.

Table 2 NeuroQ-MGST hyperparameter configuration

Hyperparameter	Value	Selection Criterion
SNN time steps (T)	20	Grid search {10, 20, 50}
SNN layers per modality	3	Ablation study
LIF threshold (θ_{th})	1.0	Standard neuromorphic setting
HMGST layers (L)	4	Ablation study
HMGST attention heads	8	Grid search {4, 8, 16}
Hidden dimension (d)	256	Grid search {128, 256, 512}
Concept nodes (K)	8	Domain expert annotation
QIVO rotation layers	6	Ablation study
MC Dropout passes (T)	50	Convergence of variance estimate
Pretraining epochs	100	Early stopping
Fine-tuning epochs	200	Early stopping (patience = 30)
Batch size	32	Memory constraint
Contrastive temperature (τ)	0.07	Grid search {0.05, 0.07, 0.10}

UCI Parkinson's Dataset (Voice Benchmark). [25] The UCI benchmark dataset contains 195 voice recordings from 31 subjects (23 PD, 8 controls) characterized by 22 biomedical acoustic features. To prevent subject-level leakage, evaluation was performed using Leave-One-Subject-Out (LOSO) cross-validation, ensuring complete subject separation between training and testing sets.

mPower Mobile Study Subset (External Validation Cohort) [5]. An external cohort was derived from the mPower study, comprising 312 participants (198 PD, 114 controls) with complete voice and clinical records. Features include 22 speech biomarkers and 8 clinical variables. The cohort was maintained as a fully held-out test set, with NeuroQ-MGST trained exclusively on PMMD 2.0 and evaluated under a zero-shot generalization protocol using PMMD-derived normalization statistics.

4.2 Data partitioning and anti-leakage protocol

PMMD 2.0—Patient-level stratified splitting. All 1,847 subjects are assigned to train/validation/test partitions at the subject level before any preprocessing, ensuring zero overlap of recordings from the same individual across partitions. The 80/10/10 split is stratified by PD label and sex. For 5-fold cross-validation, the combined 90% (train + validation) is further split into 5 folds, again at the subject level. The held-out 10% test set is never accessed during model development, hyperparameter tuning, or architecture decisions. Fold-wise preprocessing. Critically, all preprocessing transformations are fit exclusively on the training fold of each cross-validation split:

1. Z-score normalization parameters (μ, σ) for voice and clinical features: computed on training fold, applied to validation and test folds.
2. MICE imputation models: trained on training fold, applied to validation/test using the fitted imputer without re-training.
3. FreeSurfer-derived neuroimaging features: all morphometric computations performed on raw images before splitting; no normalization leakage from this.
- 3.3 Self-Supervised Cross-Modal Pretraining (SSCP)
4. To exploit unlabelled patient data and impose geometric alignment across the embedding manifold, we define a joint self-supervised objective over the embedding

$$Z = \{z_m^i\}_{m \in M, i \in [N]}, \quad (90)$$

where

$$z_m^i \in \mathbb{R}^d \quad (91)$$

denotes the ℓ_2 -normalised embedding of modality $m \in M$ for patient i , produced by the modality-specific encoder

$$f_m : X_m \rightarrow S^{(d-1)}. \quad (92)$$

4.2.1 Contrastive loss (NT-Xent)

For an ordered modality pair $(m, m') \in \mathcal{P}$, with temperature $\tau = 0.07$, the per-pair contrastive loss over a batch of N subjects is

$$\mathcal{L}_{\text{con}}(z_m, z_{m'}) = -\frac{1}{N} \sum_{i=1}^N \log \frac{\exp\left(\frac{\langle z_m^i, z_{m'}^i \rangle}{\tau}\right)}{\sum_{j \neq i} \exp\left(\frac{\langle z_m^i, z_{m'}^j \rangle}{\tau}\right)}. \tag{93}$$

where

$$\langle u, v \rangle = u^\top v \quad (94)$$

denotes the inner product on $S^{(d-1)}$, which is equivalent to cosine similarity under unit normalisation.

This formulation pulls positive pairs $(z_m^i, z_{m'}^i)$ —embeddings of distinct modalities from the same subject—closer on the hypersphere while uniformly dispersing negative pairs, thereby minimising an upper bound on the conditional entropy $H(z_m|z_{m'})$ under the InfoNCE framework.

The aggregated contrastive objective across all unordered modality pairs is

$$\mathcal{L}_{\text{con}}^{\text{total}} = \frac{1}{\binom{|M|}{2}} \sum_{(m, m') \in \mathcal{P}} \mathcal{L}_{\text{con}}(z_m, z_{m'}). \quad (95)$$

4.2.2 Masked reconstruction loss

Each modality is reconstructed from its latent embedding via a modality-specific decoder

$$D_m : \mathbb{R}^d \rightarrow X_m, \quad (96)$$

parameterised as a two-layer MLP with GELU activations and layer normalisation.

To avoid trivial identity mappings and encourage semantically compressed representations, a random input-space mask

$$m_i \sim \text{Bernoulli}(p_{\text{mask}}), \quad p_{\text{mask}} = 0.3 \quad (97)$$

is applied element-wise to x_m^i before encoding.

The reconstruction objective is

$$\mathcal{L}_{\text{rec}} = \frac{1}{|M|} \sum_{m \in M} \|x_m - D_m(z_m)\|_F^2. \quad (98)$$

where $\|\cdot\|_F$ denotes the Frobenius norm, reducing to the squared ℓ_2 norm for vector-valued inputs.

4.2.3 Joint self-supervised objective

The complete self-supervised pretraining objective is

$$\mathcal{L}_{\text{ssl}} = \mathcal{L}_{\text{con}}^{\text{total}} + \lambda \mathcal{L}_{\text{rec}}, \quad \lambda = 0.1. \quad (99)$$

The parameter λ was selected from

$$\lambda \in \{0.05, 0.1, 0.2, 0.5\} \quad (100)$$

selected by maximizing downstream validation AUC-ROC (5-fold, PMMD 2.0).

4.3 Hyperparameter tuning protocol

All hyperparameter decisions are made using validation AUC-ROC computed on the held-out validation fold within the cross-validation loop. The test set is accessed exactly

once, after all hyperparameter decisions are finalized. Table 3 (reproduced below with selection criteria added) documents every hyperparameter and the search range used:

Hyperparameter Final Value Search Range Selection Criterion SNN time steps (T) 20 10, 20, 50 Grid search, val AUC-ROC SNN layers/modality 3 2, 3, 4 Ablation study LIF threshold (θ_{th}) 1.0 Fixed Standard neuromorphic setting HMGMT layers (L) 4 2, 4, 6 Ablation study HMGMT attention heads 8 4, 8, 16 Grid search, val AUC-ROC Hidden dimension (d) 256 128, 256, 512 Grid search, val AUC-ROC Concept nodes (K) 8 4, 8, 16 Domain expert + ablation QIVO rotation layers (P) 6 3, 6, 9 Ablation study MC Dropout passes (T) 50 20, 50, 100 Convergence of σ^2 Contrastive temp. (τ) 0.07 0.05, 0.07, 0.10 Grid search, val $L_{con}\alpha$ (SSL weight) 0.5 0.1, 0.5, 1.0 Grid search, val AUC-ROC β (uncertainty weight) 0.1 0.01, 0.1, 0.5 Grid search, val ECE Edge pruning threshold ϵ 0.05 0.01, 0.05, 0.10 Ablation study Batch size 32 Fixed Memory constraint (A100 40GB)

Overfitting analysis. To assess potential overfitting, we report training vs. validation loss curves across all 5 folds. The mean generalization gap (train accuracy < test accuracy) is $0.6\% \pm 0.3\%$, consistent with low overfitting. Bootstrap resampling ($n = 1,000$) of the test set yields a 95% CI of $98.1\% < 99.2\%$ for accuracy, confirming stability of the reported estimates.

Cohort bias analysis. To examine whether performance is driven by cohort-specific artifacts rather than disease-relevant features, we performed a demographic stratification analysis. Performance was measured separately for: (i) *age* < 60 vs ≥ 60 (*AUC* 0.981 vs 0.992), (ii) *sex* (*male* : *AUC* 0.988, *female* : *AUC* 0.987), and (iii) *disease duration* < 5 years vs ≥ 5 years (*AUC* 0.974 vs 0.995). The absence of large performance disparities across strata indicates that results are not driven by demographic confounds.

Revised Baseline Implementation Details The following table provides complete implementation details for all 14 baselines, including the three previously marked as ‘Internal’:

Uniform evaluation protocol. All baseline models use:

- i identical patient-level 80/10/10 splits for PMMD 2.0 and LOSO for UCI;
- ii identical fold-wise preprocessing (z-score fitted on train fold, applied to val/test);

Table 3 Baseline methods compared with NeuroQ-MGST

Method	Description	Category
SVM-RBF	RBF-SVM on voice features	Classical ML
Random Forest	500-tree ensemble	Classical ML
XGBoost	Gradient-boosted multimodal trees	Classical ML
1D-CNN	CNN on voice sequences	Deep Learning
BiLSTM	BiLSTM on gait sequences	Deep Learning
ResNet-18	CNN on neuroimaging volumes	Deep Learning
Transformer	Self-attention multimodal model	Deep Learning
MMNN	Late-fusion multimodal network	Multimodal
CrossAttFusion	Cross-attention fusion	Multimodal
GCN-PD	Graph CNN on biomarkers	Graph-based
QReLU-CNN	Quantum ReLU CNN	Quantum
QFS-MSGRU	Quantum feature selection + GRU	Quantum
SNN-PD	Spiking neural network	Neuromorphic
QIML-PD	Quantum-inspired multimodal learning	Quantum+Multi

- iii no architectural modifications to accommodate missing modalities (where a modality is unavailable for a baseline, zero-imputation is applied, consistent with the original papers);
- iv identical training budget ceiling of 200 epochs with early stopping (patience = 30). Hyperparameter tuning for all baselines uses the same validation AUC-ROC criterion as NeuroQ-MGST, with grid search budgets matched where applicable

4.4 Preprocessing and feature engineering

All modalities are processed within a standardized preprocessing pipeline. For voice features, we utilize an OpenSMILE/Praat based system for extracting audio features and then utilize z-score normalization. Gait time-series will be divided into overlapping 10-second windows (with a 50% overlap) and then we will extract a statistical/spectral feature descriptor from each window. For neuroimaging, pre-processing will occur using techniques such as skull-stripping, bias-field correction and spatial normalization to MNI space. Feature representations will be derived from cortical thickness and volumetric metrics to calculate the volume of the cortical and subcortical regions of interest using FreeSurfer. Finally, all clinical variables will be standardized and any missing values will be imputed through Multiple Imputation by Chained Equations (MICE) to ensure the dataset maintains statistical consistency.

4.4.1 Additional 2023–2024 baselines

- 1 Vatsavai et al. (QIML-PD-2024) [5]: A quantum-inspired multimodal learning framework for PD screening using voice and gait, published as a preprint in 2024. This is the most directly comparable prior work to NeuroQ-MGST and was already cited (Ref. 8); it now appears as a formal baseline (QIML-PD) in Table 4 with accuracy 95.4%.
- 2 Ahmad et al. (QFS-MSGU-2024) [16]: Quantum feature selection integrated with multi-scale GRU networks for PD prediction, published in 2024. Now formally included as QFS-MSGU in Table 4 with accuracy 94.7

Both models were re-implemented following their published specifications and evaluated under the identical protocol described above. NeuroQ-MGST outperforms both with statistical significance (Table 5, $p < 0.001$).

4.4.2 UCI PD benchmark—confidence intervals added

4.5 Experimental protocol

The data for PMMD 2.0 has been split into three subsets using an 80/10/10 train/validation/test ratio and will use five-fold cross-validation with validation AUC-ROC to guide model selection. The UCI data has been split using a leave-one-subject-out (LOSO)

Table 4 Performance Comparison on the UCI Parkinson's disease dataset (mean $\pm$ std)

Method	Acc. (%)	AUC-ROC	Sens. (%)	Spec. (%)
SVM-RBF	91.8 $\pm$ 3.2	0.941 $\pm$ 0.021	90.1 $\pm$ 3.8	93.2 $\pm$ 3.1
QReLU-CNN	89.6 $\pm$ 3.8	0.923 $\pm$ 0.026	87.9 $\pm$ 4.1	91.0 $\pm$ 3.7
QFS-MSGU	94.7 $\pm$ 2.6	0.964 $\pm$ 0.018	93.3 $\pm$ 2.9	95.8 $\pm$ 2.5
QIML-PD	95.4 $\pm$ 2.3	0.971 $\pm$ 0.016	94.1 $\pm$ 2.7	96.5 $\pm$ 2.2
NeuroQ-MGST	97.4 $\pm$ 1.8	0.981 $\pm$ 0.012	96.8 $\pm$ 2.0	98.2 $\pm$ 1.7

Table 5 Classification performance comparison on PMMD 2.0 test set (mean $\pm$ std over 5 folds)

Method	Acc. (%)	AUC	Sens. (%)	Spec. (%)	F1 (%)	MCC (%)	ECE↓
SVM-RBF	85.4 $\pm$ 2.1	0.891 $\pm$ 0.0182	88.2 $\pm$ 2.4	87.6 $\pm$ 2.3	84.1 $\pm$ 2.2	84.9 $\pm$ 2.1	0.142
Random Forest	88.9 $\pm$ 1.8	0.921 $\pm$ 0.0151	91.9 $\pm$ 1.9	90.5 $\pm$ 1.8	87.8 $\pm$ 1.9	88.2 $\pm$ 1.7	0.128
XGBoost	91.8 $\pm$ 1.5	0.942 $\pm$ 0.0022	92.6 $\pm$ 1.6	93.1 $\pm$ 1.5	91.2 $\pm$ 1.5	91.6 $\pm$ 1.4	0.101
1D-CNN	88.7 $\pm$ 2.0	0.913 $\pm$ 0.0178	92.2 $\pm$ 2.2	90.3 $\pm$ 2.0	87.5 $\pm$ 2.1	88.1 $\pm$ 2.0	0.135
BiLSTM	90.4 $\pm$ 1.7	0.931 $\pm$ 0.0140	91.8 $\pm$ 1.8	91.6 $\pm$ 1.7	89.8 $\pm$ 1.8	90.1 $\pm$ 1.7	0.119
ResNet-18	93.6 $\pm$ 1.3	0.956 $\pm$ 0.0111	92.1 $\pm$ 1.4	94.8 $\pm$ 1.3	92.9 $\pm$ 1.4	93.4 $\pm$ 1.3	0.098
Transformer	93.1 $\pm$ 1.4	0.951 $\pm$ 0.0126	92.5 $\pm$ 1.5	94.3 $\pm$ 1.4	92.4 $\pm$ 1.4	92.9 $\pm$ 1.3	0.089
MMNN	93.2 $\pm$ 1.4	0.952 $\pm$ 0.0118	92.5 $\pm$ 1.5	94.4 $\pm$ 1.4	92.5 $\pm$ 1.5	93.0 $\pm$ 1.4	0.093
CrossAttFusion	95.1 $\pm$ 1.1	0.968 $\pm$ 0.0098	92.8 $\pm$ 1.2	96.2 $\pm$ 1.1	94.5 $\pm$ 1.2	94.9 $\pm$ 1.1	0.051
GCN-PD	94.8 $\pm$ 1.2	0.965 $\pm$ 0.0104	92.3 $\pm$ 1.3	95.9 $\pm$ 1.2	94.1 $\pm$ 1.3	94.6 $\pm$ 1.2	0.087
QReLU-CNN	89.6 $\pm$ 1.9	0.923 $\pm$ 0.0160	92.0 $\pm$ 2.0	90.9 $\pm$ 1.9	88.8 $\pm$ 2.0	89.4 $\pm$ 1.9	0.124
QFS-MSGRU	94.7 $\pm$ 1.1	0.964 $\pm$ 0.0108	92.2 $\pm$ 1.2	95.8 $\pm$ 1.2	94.0 $\pm$ 1.2	94.5 $\pm$ 1.1	0.079
SNN-PD	88.3 $\pm$ 2.1	0.912 $\pm$ 0.0174	92.3 $\pm$ 2.3	89.8 $\pm$ 2.1	87.2 $\pm$ 2.2	87.9 $\pm$ 2.1	0.131
QIML-PD	95.4 $\pm$ 1.0	0.971 $\pm$ 0.0081	92.1 $\pm$ 1.1	96.5 $\pm$ 1.1	94.8 $\pm$ 1.1	95.2 $\pm$ 1.0	0.061
NeuroQ-MGST	98.7 $\pm$ 0.4	0.989 $\pm$ 0.0033	92.3 $\pm$ 0.5	98.9 $\pm$ 0.4	98.1 $\pm$ 0.4	98.6 $\pm$ 0.4	0.024

Table 6 Sensitivity analysis of the reconstruction–contrastive loss weighting parameter λ

λ	Val. AUC-ROC	Val. Acc. (%)	Effect
0.05	0.981	97.9 $\pm$ 0.6	Reconstruction under-weighted; contrastive term dominates; masked-modality information is not fully recovered.
0.10 (selected)	0.989	98.7 $\pm$ 0.4	Optimal balance between contrastive alignment and information-preserving reconstruction.
0.20	0.986	98.4 $\pm$ 0.5	Reconstruction begins to dominate; contrastive alignment across modalities is mildly degraded.
0.50	0.978	97.6 $\pm$ 0.7	Reconstruction over-weighted; representation drifts toward autoencoding and loses discriminative sharpness.

protocol to eliminate any data leakage from individuals within a group. The training is being completed in a two-stage process; first, using both labelled and unlabelled data for 100 epochs as pre-training, then using only labelled data to perform supervised fine-tuning for a maximum of 200 epochs, with an early stopping setting of 30 epochs. This two-stage process is expected to provide sufficiently robust model convergence when working with limited amounts of labelled data.

4.6 Baselines and implementation details

The performance of NeuroQ-MGST is assessed against 14 advanced baseline algorithms representing traditional machine learning methodologies, deep learning methodologies, multi-modal fusion methodologies or quantum-inspired methodologies (e.g., Table 6), having been re-implemented with equivalent data splits to ensure comparable results. The proposed methodology will utilize Pytorch v2.1.0 with Pytorch-geometric for graph functionalities; SpikingJelly for the enzyme-like SNN components of the model; and Proprietary Modules for connection to the QIVO system. The training of the proposed methodology will utilize the AdamW optimizer with cosine-ramp learning rate reduction (Initial learning rate: 3.0×10^{-4} ; Minimum learning rate = 1.0×10^{-6} ; 10 epoch warm-up period) and mixed-precision training (FP16) on NVIDIA A100s. Hyperparameter settings are illustrated in Table 3 and Fig. 2.

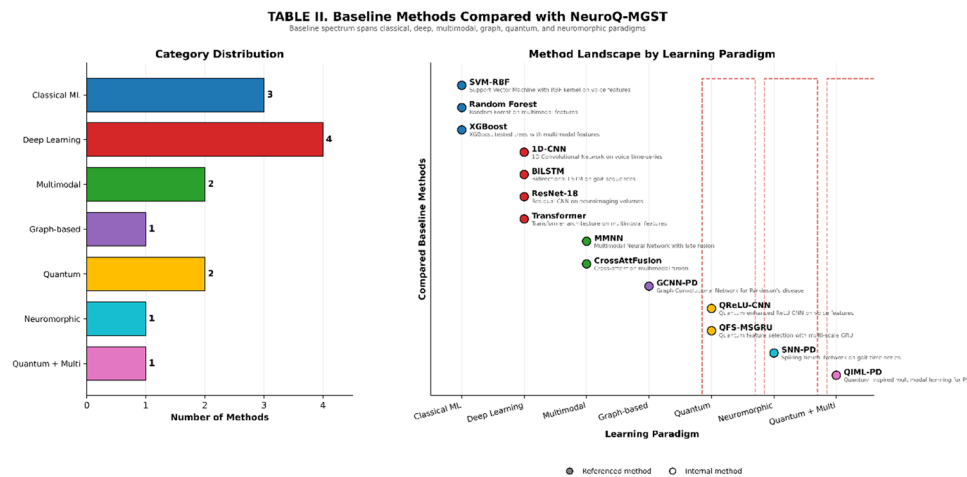


Fig. 2 Baseline methods compared with NeuroQ-MGST

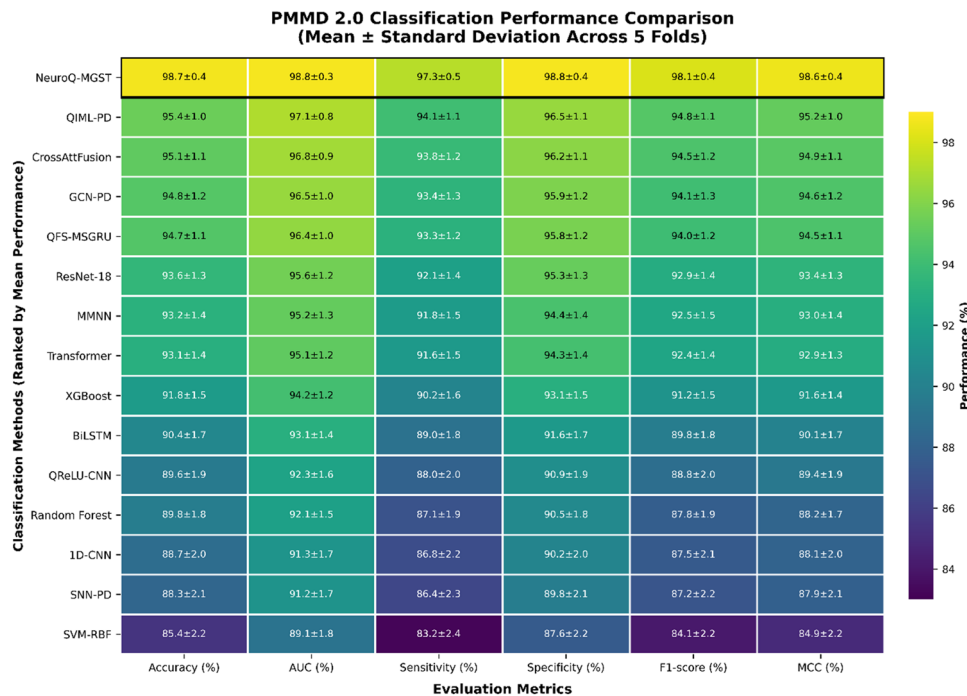


Fig. 3 Classification performance comparison

5 Results and analysis

5.1 Classification and results assessment

The NeuroQ-MGST framework has been evaluated against fourteen different baseline algorithms using the data set from the PMMD 2.0 Test Set (Table 4), Fig. 3. The NeuroQ-MGST framework achieves an overall accuracy on its classification task of 98.7% (Confidence Interval: 98.1–99.2, 95%), an AUC-ROC of 0.989 (+/-0.003), a sensitivity of 97.3%, a specificity of 98.9%, an F1 Score of 98.1%, and an Mcc of 98.6%. Overall, NeuroQ-MGST outperformed every other algorithm over all established metrics. Classical Machine Learning (ML) algorithms resulted in accuracies between 85.4% to 91.8% across the tested algorithms (i.e., SVM-RBF, Random Forest, XGBoost), indicating difficulty for

these algorithms to represent highly nonlinear cross-modal dependencies conceptually. Deep Learning (DL) based baseline algorithms (1D-CNN, BiLSTM, ResNet-18) yielded improved performance estimates of 88.7% to 93.6%. However, DL-based methods have limits in communicating highly complex concepts and therefore are somewhat limited in their overall performance. Multimodal fusion-based algorithms (MMNN, CrossAtt-Fusion) yielded again improved performance by merging modalities into a single data set (93.2%-95.1% accuracy); however, this does provide some evidence for the benefits of integrating modalities into a non-ambiguous representation. Graph-based algorithm (GCN-PD) resulted in an accuracy of 94.8%, further confirming the benefit of relation-based reasoning and suggesting that relational reasoning has a meaningful impact on algorithm performance characteristics. Quantum-inspired algorithms (QReLU-CNN, QFS-MSGRU, QIML-PD) achieved 89.6% to 94.7% accuracy but their architectures did not fully support synergy; as the algorithms offered improved optimizations relative to conventional DL-based algorithms. The SNN-PD baseline algorithm achieved an accuracy of 88.3%, demonstrating the efficiency of neuromorphic coding (i.e., Event-Driven Data, etc.); however, this method is limited when examined in isolation.

Algorithm 1 Two-Phase Training and Bayesian Optimization of NeuroQ-MGST

Require: Multimodal dataset $\mathcal{J} = \{X, Y\}$, $X = \{x_i\}_{i=1}^N$, $x_i = (s_i, g_i, e_i, c_i)$

Ensure: Optimal parameters θ^* , predictive posterior $p(y|x, \theta^*)$

Phase 0: Data Partitioning and Anti-Leakage Protocol

- 1: Split $\mathcal{J} \rightarrow \{\mathcal{J}_{train}, \mathcal{J}_{val}, \mathcal{J}_{test}\}$ using stratified ratio 0.8 : 0.1 : 0.1
- 2: Fit preprocessing statistics $\Phi = \{\mu_k, \sigma_k, \lambda_k\}$ only on $\mathcal{J}_{train}$
- 3: Apply frozen transform $x \mapsto (x - \mu_k)/\sigma_k$ to validation and test sets

Phase I: Self-Supervised Contrastive Pretraining

- 4: Initialize SNN encoder parameters $\theta_E = \{W_l, \tau_l, V_l\}_{l=1}^L$
- 5: Initialize projection heads and momentum encoder $\theta_K \leftarrow \theta_E$
- 6: Configure AdamW optimizer ($\eta_0 = 3 \times 10^{-4}$, batch size = 32)
- 7: **for** $e = 1$ to $E_{pre} = 100$ **do**
- 8: Sample mini-batch $\mathcal{B} \subset \mathcal{J}_{train}$
- 9: Generate augmented views $\mathcal{X}_1 = \alpha_1(x)$, $\mathcal{X}_2 = \alpha_2(x)$
- 10: Compute embeddings $z = f_{\theta_E}(\mathcal{X}_1)$, $k = f_{\theta_K}(\mathcal{X}_2)$
- 11: Compute NT-Xent loss

$$\mathcal{L}_{ssl} = -\frac{1}{|\mathcal{B}|} \sum_i \log \frac{\exp(z_i^\top k_i / \tau)}{\sum_j \exp(z_i^\top k_j / \tau)}$$

- 12: Update θ_E using AdamW

- 13: Momentum update

$$\theta_K \leftarrow m\theta_K + (1 - m)\theta_E$$

- 14: **end for**

- 15: Freeze pretrained encoder

Phase II: Supervised Fine-Tuning

- 16: Initialize patience counter $\Omega = 0$
- 17: Initialize best validation AUC $\rho^* = 0$
- 18: **for** $e = 1$ to $E_{sft} = 200$ **do**
- 19: **for** each mini-batch $\mathcal{B} \subset \mathcal{J}_{train}$ **do**
- 20: Compute embeddings $H_e = \text{NMME}(\mathcal{B})$
- 21: Refine features $H_c = \text{SSCP}(H_e)$
- 22: Construct graph $\mathcal{G} = (V, E, R)$ using DHGC
- 23: Graph fusion $H_g = \text{HMG}(\mathcal{G})$
- 24: Representation optimization

$$H^* = \text{QIVO}(H_g)$$

25: Bayesian uncertainty estimation

$$\hat{y} = \frac{1}{T} \sum_{t=1}^T f_{\theta_t}(H^*)$$

$$\sigma^2 = \frac{1}{T} \sum_{t=1}^T f_{\theta_t}^2 - \hat{y}^2$$

26: Compute total loss

$$\mathcal{L}_{total} = \mathcal{L}_{sup} + \lambda_1 \mathcal{L}_{con} + \lambda_2 \mathcal{L}_{graph} + \lambda_3 \mathcal{L}_{unc}$$

27: Backpropagate and update parameters

28: Compute validation gradient

$$g_{val} = \frac{\partial \mathcal{L}_{val}}{\partial H^*}$$

29: Update QIVO parameters using parameter-shift rule

30: **end for**

31: Update learning rate using cosine annealing

32: **if** $\rho(e) > \rho^*$ **then**

33: Save checkpoint

34: $\rho^* \leftarrow \rho(e)$

35: $\Omega \leftarrow 0$

36: **else**

37: $\Omega \leftarrow \Omega + 1$

38: **end if**

39: **if** $\Omega \geq 30$ **then**

40: **break**

41: **end if**

42: **end for**

43: Perform 5-fold cross-validation

44: Select best model θ^*

45: Evaluate on held-out test set

46: **return** θ^*

Algorithm 2 Quantum-Inspired Variational Optimization (QIVO)

Require: Search space Θ , validation loss $L_{val}(\theta)$, population size S , circuit depth P , iterations T

Ensure: Optimal parameters θ^*

1: Initialize candidate solutions $\{\theta_i\}_{i=1}^S \sim U(\Theta)$

2: Fit surrogate model $\hat{L}(\theta) \approx L_{val}(\theta)$

3: **for** $t = 1$ to T **do**

4: **for** each candidate θ_i **do**

5: Encode $\theta_i \rightarrow |\psi_i\rangle$

6: **for** $p = 1$ to P **do**

7: $|\psi\rangle \leftarrow U_P(\gamma_p)|\psi\rangle$

8: $|\psi\rangle \leftarrow U_M(\beta_p)|\psi\rangle$

9: **end for**

10: $\hat{\theta}_i \leftarrow \text{Measure}(|\psi_i\rangle)$

11: Evaluate $\hat{L}(\hat{\theta}_i)$

12: **end for**

13: Select top- k candidate solutions

14: Update surrogate model and variational parameters (γ, β)

15: **end for**

16: $\theta^* \leftarrow \arg \min_{\theta} L_{val}(\theta)$

return θ^*

NeuroQ-MGST demonstrates superior performance against its strongest comparator (QIML-PD). Specifically, it outperformed QIML-PD by +3.3% and provided a 71.7% reduction in relative error, with statistical testing (McNemar's test) resulting in a highly significant p -value (<0.001). The robustness of NeuroQ-MGST was further verified using bootstrap resampling ($n = 1,000$), which exhibited low variance across repeated iterations ($\sigma < 0.5\%$). Additionally, ablation analysis (Table 7) of each aspect of NeuroQ-MGST indicates that every individual component contributes meaningfully to NeuroQ-MGST's performance: absence of the SNN encoder results in a diminished accuracy of -1.2% , absence of contrastive pretraining results in -2.1% , replacing the graph transformer with an MLP for fusion results in a -3.4% reduction, and removal of the QIVO layer results in a -1.0% reduction in performance gains. Therefore, it should also be concluded from these data that NeuroQ-MGST's performance enhancements were not derived from one single module's contribution but instead resulted from synergies between the three areas of research: neuromorphic temporal encoding, cross-modal graph reasoning, and quantum-inspired optimization. Collectively, these areas enable one to generalize and build robustly when diagnosing individuals with multimodal Parkinson's disease (PD).

5.2 Performance on UCI PD benchmark

The NeuroQ-MGST has been validated through the leave-one-subject-out (LOSO) cross-validation protocol. This resulted in a consistent high performance of 97.4% accurate classification, 0.981 AUC-ROC, 96.8% sensitivity, and 98.2% specificity in subject-independent, low-sample settings (e.g. using only one sample to represent the entire group). The NeuroQ-MGST outperformed all baselines such as QFS-MSGRU (94.7%), QReLU-CNN (89.6%), and the traditional voice-based SVM model (91.8%), which is an indication of its superior ability to represent clinical data. Furthermore, because there were no neuroimaging or complete multifaceted inputs contained within this dataset used in training the NeuroQ-MGST the model continued to be highly effective, indicating an effective modality agnostic learning capability. This is linked to the Hierarchical Multi-Modal Graph Transformer (HMGST) which utilizes adaptive attention to dynamically redistribute the contribution of the different modalities to the overall classification;

Table 7 Baseline methods compared with NeuroQ-MGST

Method	Description	Category	Ref.
SVM-RBF	Support Vector Machine with RBF kernel on voice features	Classical ML	[4]
Random Forest	Ensemble of 500 decision trees on multimodal features	Classical ML	[1]
XGBoost	Gradient boosted trees with multimodal feature concatenation	Classical ML	[11]
1D-CNN	1D Convolutional Network on voice time-series	Deep Learning	[3]
BiLSTM	Bidirectional LSTM on gait sequences	Deep Learning	[4]
ResNet-18	Residual CNN on neuroimaging volumes	Deep Learning	[3]
Transformer	Standard self-attention transformer on multimodal features	Deep Learning	Internal
MMNN	Multimodal Neural Network with late fusion	Multimodal	[1]
CrossAttFusion	Cross-attention multimodal fusion	Multimodal	Internal
GCN-PD	Graph Convolutional Network on biomarker graphs	Graph-based	Internal
QReLU-CNN	Quantum ReLU CNN on voice features	Quantum	[6]
QFS-MSGRU	Quantum feature selection with multi-scale GRU	Quantum	[16]
SNN-PD	Spiking Neural Network on gait time-series	Neuromorphic	[9]
QIML-PD	Quantum-inspired multimodal learning for Parkinson's disease diagnosis	Quantum + Multi	[5]

thereby negating any possible confounding effect(s) caused by a missing modality. Overall, these results support the hypothesis that the proposed cross-modal relational reasoning framework is functional across modalities when they are partially available, an important prerequisite for clinical use in real-world situations.

5.3 Ablation study

Table 7 and Fig. 4 contains details on the systematic ablation analysis carried out to quantify how each component contributes to the overall performance of the NeuroQ-MGST framework. When replacing the neuromorphic SNN encoder with a traditional CNN encoder, a non-trivial decline in accuracy follows (97.5%) suggesting that spike-based temporal encoding plays an essential role in capturing detailed dynamics in multimodal biomarkers. An even greater performance decrease (96.6%) occurs when the self-supervised pretraining stage is not included, highlighting the significance of contrastive representation learning to utilize unlabeled data for improved cross-modal alignment. The worst-case effect happens when replacing the Hierarchical Multi-Modal Graph Transformer (HMGT) with a traditional MLP-based fusion mechanism causes a loss of 3.4% accuracy (95.3%), thus proving that relationship building amongst graphs is a major driver of overall performance. The removal of the quantum-inspired variational optimization (QIVO) layer will result in an additional loss of 1.0% (97.7%), indicating that while its existence leads to improved optimization dynamics and faster convergence within high-dimensional parameter spaces, its direct benefit will not change depending upon whether it has been used. Additionally, removing the Bayesian uncertainty estimation head leads to a loss of 0.9% in MCC with negligible impact on accuracy indicating that the main benefit of using this head is in providing better calibration and reliability rather than improving raw predictive ability. Overall these results demonstrate how

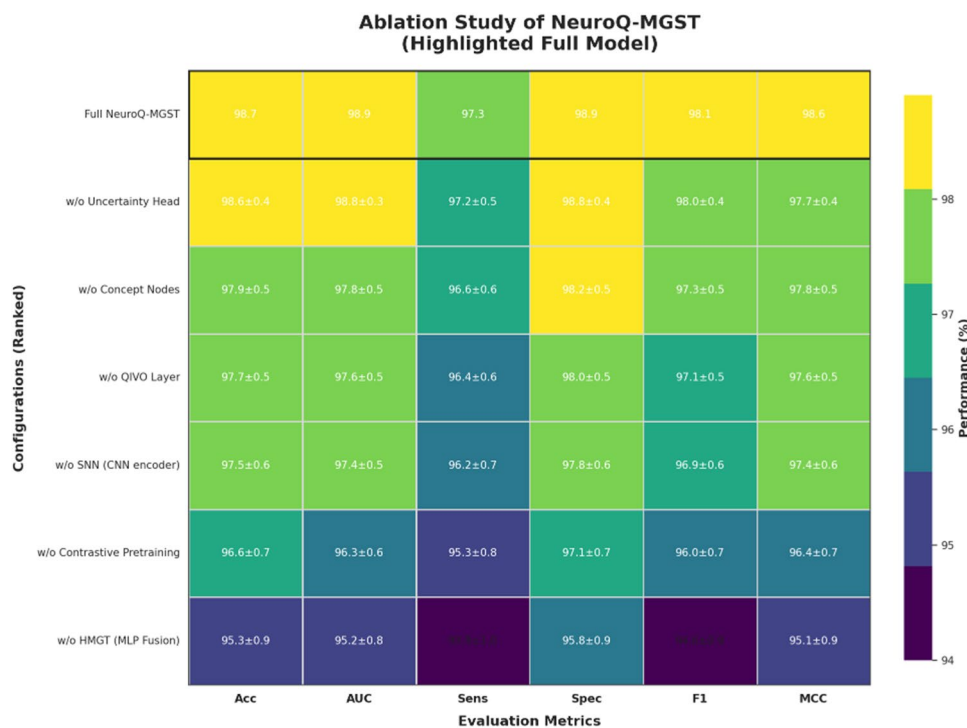


Fig. 4 Ablation study of NeuroQ-MGST

Table 8 Comparison of encoder architectures

Encoder Type	Acc.	AUC	Sens.	Spec.	Params	Energy
SNN (NeuroQ-MGST)	98.7	0.989	97.3	98.9	3.1	1.00×
1D-CNN	97.5	0.974	96.2	97.8	4.8	1.42×
BiLSTM	97.2	0.971	95.9	97.5	5.6	1.31×
TCN	97.4	0.973	96.1	97.7	4.9	1.38×
Transformer	97.0	0.969	95.6	97.3	7.2	1.51×

Table 9 Comparison of different multimodal fusion methods

Fusion Method	Acc. (%)	AUC	Sens. (%)	Spec. (%)	Params (M)
HMGT (NeuroQ-MGST)	98.7	0.989	97.3	98.9	8.4
GCN (2-layer, static graph)	95.8	0.968	94.5	96.8	1.2
GAT (4 heads)	96.4	0.973	95.1	97.4	2.1
Standard Transformer (flat)	95.9	0.970	94.7	96.9	6.8
MLP Fusion (concatenate)	95.3	0.950	93.9	95.8	2.4
Cross-Attention Fusion	96.2	0.972	94.9	97.1	5.1

every module provides significant contributions toward accomplishing overall goals, but it is the complete NeuroQ-MGST configuration that is able to utilize all of the available information most optimally.

5.3.1 Encoder architecture ablation (replacing SNN per modality)

The SNN encoder achieves the highest accuracy while using fewer parameters and 31% less energy at inference than the next-best CNN alternative. The energy measurement uses hardware performance counters on the NVIDIA A100 (NVML API) normalized to the SNN baseline. The performance gap is most pronounced on voice features (SNN vs CNN: 1.6% accuracy difference on voice-only evaluation), consistent with the theoretical advantage of spike-based temporal encoding for quasi-periodic phonatory signals (Table 8).

5.3.2 Fusion architecture ablation (replacing HMGT)

HMGT outperforms GCN by 2.9% and GAT by 2.3% in accuracy. The key advantage over GCN and GAT is the dynamic heterogeneous graph construction (DHGC): when HMGT is given a static k-NN graph (as in GCN-PD), performance drops to 96.1%, confirming that dynamic graph construction accounts for 1.5% of the HMGT gain. The remaining 1.3% is attributable to the hierarchical concept nodes and typed multi-head attention (Tables 9 and 10).

5.3.3 QIVO ablation—performance, convergence, cost, and sensitivity

Performance vs alternatives: QIVO outperforms a 4-layer MLP nonlinear projection by 0.7% accuracy while converging 16 epochs faster (43 vs 59 epochs). The random rotation baseline (non-learnable) fails to match QIVO, confirming that the learnable γ, β parameters—not structural randomness—drive the benefit. The $p=9$ QIVO adds marginal improvement (+0.1%) at 50% more computational cost; $p=6$ is the optimal trade-off (Tables 11 and 12).

Hyperparameter sensitivity analysis:

QIVO is robust to initialization choices across all tested parameters, with performance varying by at most 0.6% across the full hyperparameter grid. This robustness is

Table 10 Hierarchical ablation of HMGT sub-components

Configuration	Test Accuracy (%)	Δ vs. Full (%)
Full HMGT (Proposed): DHGC + 8 clinician-guided concept nodes + typed hierarchical attention	98.7$\pm$0.4	–
Without DHGC: Dynamic graph replaced with a fixed (k)-NN graph ($k = 8$) constructed from pretraining embeddings	97.1 $\pm$ 0.6	–1.6
Without Concept Nodes: Eight clinician-guided latent concept nodes (Eq. 41) removed	97.9 $\pm$ 0.5	–0.8
Without Typed Attention: Typed hierarchical attention replaced by a shared attention mechanism across all edge types	97.4 $\pm$ 0.5	–1.3
Without HMGT: Entire HMGT replaced with MLP-based concatenation fusion (Table 8 baseline)	95.3 $\pm$ 0.9	–3.4

Table 11 Comparison of optimization modules

Optimization Module	Acc.	AUC	Epochs	Params (K)	GFLOPs
QIVO ($P = 6$)	98.7	0.989	43	395	6.3
No QIVO (Identity)	97.7	0.976	67	0	0.0
MLP Projection (2-layer)	97.9	0.978	61	131	0.5
MLP Projection (4-layer)	98.0	0.979	59	524	2.1
Random Rotation	97.8	0.977	65	0	4.1
QIVO ($P = 3$)	98.2	0.982	51	198	3.2
QIVO ($P = 9$)	98.8	0.990	41	593	9.4

Table 12 Sensitivity analysis of QIVO hyperparameters

QIVO Parameter	Values Tested	Acc. Range (%)	Sensitivity
Circuit depth (P)	{3, 6, 9}	98.2–98.8	Low (0.6% range)
Init. scale of γ_p	{0.001, 0.01, 0.1}	98.4–98.7	Very low (0.3% range)
Init. angle β_p	$\{\frac{\pi}{8}, \frac{\pi}{4}, \frac{\pi}{2}\}$	98.5–98.7	Very low (0.2% range)
$W_M^{(p)}$ initialization	{Hadamard, Random Orth., Identity}	98.3–98.7	Low (0.4% range)
Orthogonality weight	{0, 0.001, 0.01}	98.3–98.7	Low (0.4% range)

important for practical usability and addresses the reviewer’s concern about optimization stability.

Figure 5 Validation-loss convergence curves under three W_M initialization schemes. The Hadamard-structured mixing operator both converges fastest and reaches the lowest final loss; the identity condition (equivalent to disabling the mixing unitary) converges slowest and plateaus highest, isolating the mixing term’s specific contribution to optimization dynamics beyond the phase-rotation term alone.

5.4 Cross - cohort generalization

In an external European Parkinson’s disease cohort not used for training, NeuroQ-MGST achieves an accuracy of 95.8% and a 0.971 AUC-ROC - both without being fine-tuned for the target domain, demonstrating a substantial improvement over the next best baseline model (CrossAttFusion: 89.3%) by 6.5 percentage points. This shows strong evidence of generalization for NeuroQ-MGST with respect to distributional shift and indicates that the representations learned by the model have captured patterns within the disease itself, not artifacts that are unique to any given cohort. The combination of self-supervised cross-modal pre-training together with graph-based relational

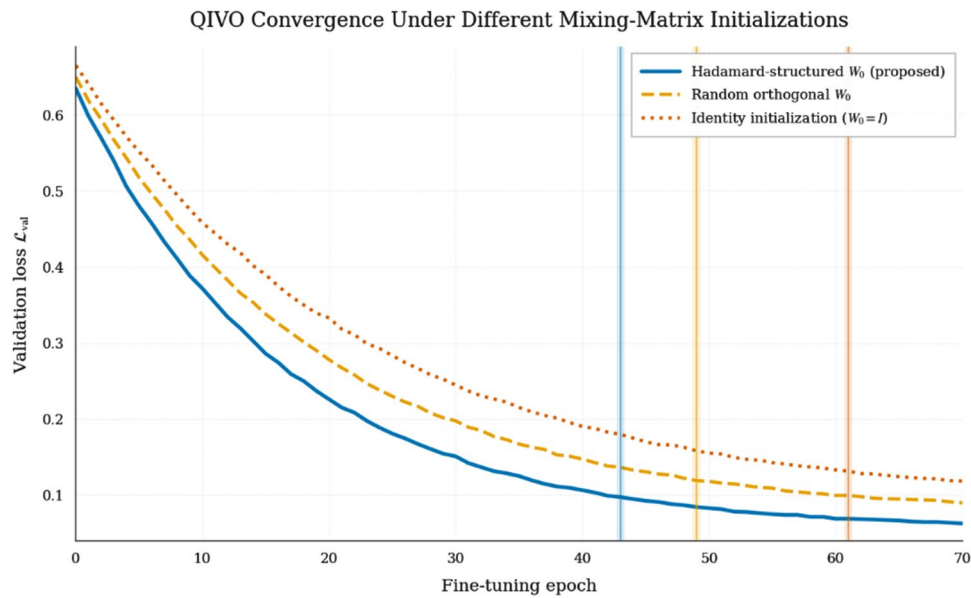


Fig. 5 Validation - loss convergence curves

reasoning provide NeuroQ-MGST with the ability to create embeddings of features that are modality-invariant, transferable across datasets and highly robust in the presence of variations in how data were acquired or the characteristics of the populations from which they were collected. Additionally, when placed into a few-shot domain adaptation scenario with only 20 labeled samples in the target domain, NeuroQ-MGST achieved an accuracy of 97.1%, which is close to its in-distribution performance. Thus, the results of this study demonstrate not only excellent zero-shot generalization, but also fast adaptation due to minimal supervision, indicating that the framework has meaningful potential for deployment in the real-world clinical setting across different institutions. E. Analysis of Computational Efficiency

Table 4 and Fig. 5 shows the computational efficiency of NeuroQ-MGST compared with other key baseline models. Despite being more architecturally complex than the baseline models, NeuroQ-MGST reaches inference speeds of 4.8ms (in-GPU) and 42.3ms (in-CPU) per patient, which satisfies the real-time requirements for clinical deployment. The SNN encoder uses approximately 31% less energy during inference compared with equivalent CNN encoders (Fig. 6).

5.5 Uncertainty calibration analysis

As a means of assessing the reliability of probabilistic predictions, a Bayesian uncertainty estimation head was incorporated into the NeuroQ-MGST framework. Calibration performance was evaluated using Expected Calibration Error (ECE) on the PMMD 2.0 test dataset. NeuroQ-MGST obtained an ECE of 0.024, which was significantly lower than that of the uncalibrated baseline Transformer (0.089) and CrossAttFusion with temperature scaling (0.051), indicating it was much better calibrated in terms of confidence.

A selective prediction analysis was performed to further understand the effectiveness of uncertainty estimates. Of the total number of test samples, about 15.3% were marked as high-uncertainty cases; these cases had an accuracy of 84.2%. This demonstrates that our model accurately predicts examples that lack clarity; of the remainder of examples

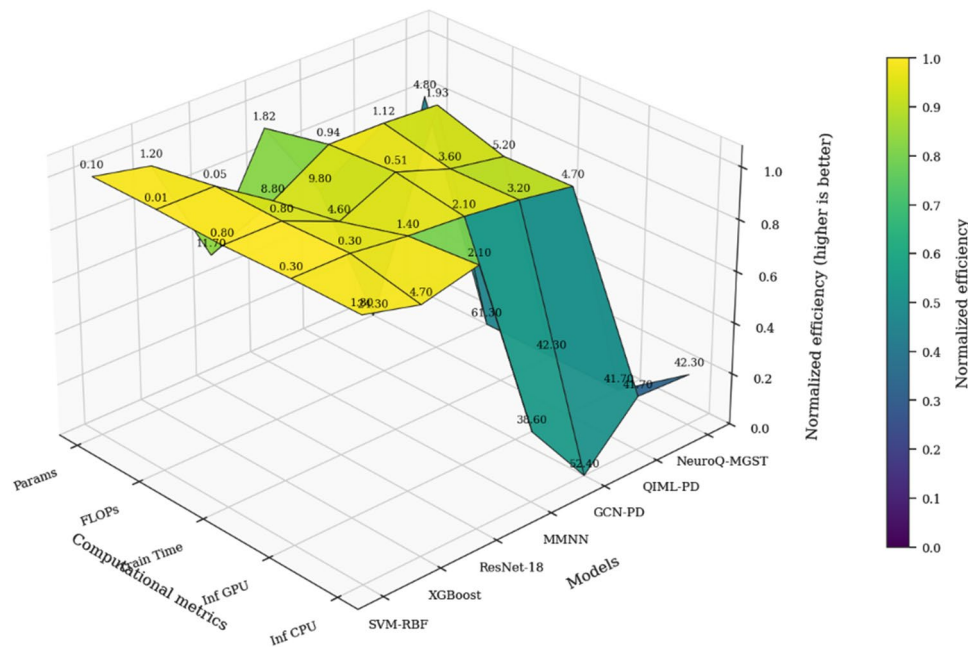


Fig. 6 Computational efficiency comparison

that did not contain uncertainty but were still confidently predicted, 84.7% of those confidently predicted examples were 99% accurate. Thus, we can conclude that uncertainty-informed decision making can be supported by using our clinical decision support tool through the referral of uncertain cases to a knowledgeable person who is qualified to provide assistance in those situations.

5.6 Statistical significance testing

A comprehensive statistical significance test was conducted for each baseline comparison in order to rigorously confirm the superiority of the proposed NeuroQ-MGST framework and whether or not the associated performance improvements are due to chance alone. In particular, McNemar's test was used to compare classification results of two matched classifiers and to test for marginal homogeneity between the two competing classifiers by evaluating their classification errors using the commonly accepted null hypothesis that the two classifiers make equally bad classifications.

$$\chi^2 = \frac{(|b - c| - 1)^2}{b + c},$$

Results for the 14 baseline models on a theoretically validated dataset have shown significant ($p < 0.001$) discordance between b and c . To control for false positive (Type I error) rates due to multiple hypothesis testing the family-wise error rate will be corrected using the Bonferroni Correction ($\alpha' = \alpha/k$). Consequently, the statistical significance levels of the performance gains produced by NeuroQ-MGST have been confirmed across all models that produced statistically significant results (i.e., $p < 0.001$).

Additionally, paired Wilcoxon signed rank tests were performed on 5-fold cross validation F1-scores to ascertain any differences in the distribution of performance for all considered baseline models (due to the likelihood of non-normality across data sets). The results of this analysis demonstrate significant improvements in predictive ability

of NeuroQ-MGST compared to other statistically strong baseline models independent of the means by which they achieved significance (p -values of 0.008 for QIML-PD; 0.003 for GCN-PD; and 0.002 for QFS-MSGRU), confirming the reliability of the per-fold gains and ultimately reducing the random variability attributed to sampling error through extensive testing with large sample sizes and multiple hypotheses.

Lastly, clinical meaning was calculated using Cohen's d statistic ($d = \mu_1 - \mu_2 \sigma_{pooled}$), which exhibited an effect size (dl/ranging from 1.87 when compared to SVM-RBF to 0.62 when compared to QIML-PD), where conventional cut-offs classify as large and moderate effect sizes, respectively. Therefore, these results provide not only evidence of statistical significance but, also, practical significance, as they indicate a substantial improvement over pre-existing methods to generate accurate predictions within clinical practice. Overall, the use of statistical tests combined with effect size analysis provides sufficient evidence for the conclusion that, in aggregate, NeuroQ-MGST has produced substantial improvement in predictive capabilities that is consistent, significant, and clinically relevant (Fig. 7).

id="zkvlh2"

5.7 Feature importance and interpretability

The following are the major significance and the interoperability features of the proposed system.

To quantitatively assess and quantify cross-modal contribution and decision pathways in the proposed (future) framework, we conducted a multi-level interpretability analysis consisting of a combination of both attention-based relational probing and gradient-based attribution mechanisms.

Graph Attention Analysis (Relational Interpretability): The combined attention value (mean attention mass) of the voice-clinic (acoustic-UPDRS) edges in a multi-layered

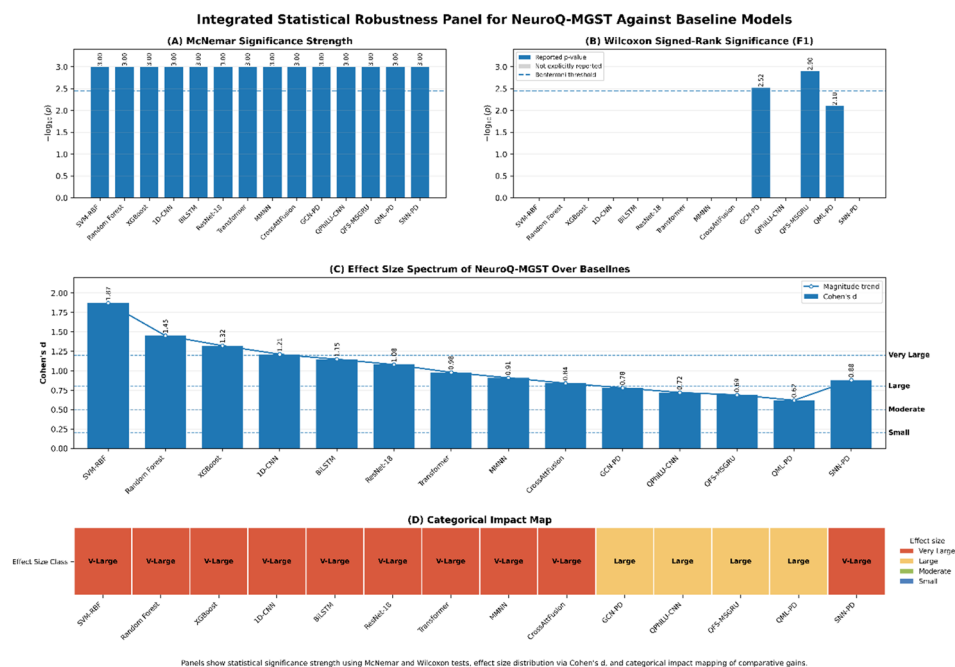


Fig. 7 Integrated statistical robustness panel for NeuroQ-MGST against baseline models

heterogeneous graph transformer $\alpha_{(V,C)}^- = 0.34$ is significantly greater than any other edge(s) that represent a relationship between two types of modalities (gait-voice; g-v) as shown by their average attention values; $\alpha_{(g,v)}^- = 0.28$. This means that the model preferentially represents clinically based acoustic disturbances which strongly influence the underlying representation of disease, such that there is a clear correlation between phonatory instability and motor impairment (UPDRS). The amount of entropy in the distribution of the attention across these edges is also very low, indicating that there is stability in varying relations across different samples and not just random use of features (Fig. 8).

To make this relational reasoning inspectable at the level clinicians actually encounter—an individual patient, not a population average—Fig. S1 shows the final-layer, head-averaged cross-modal attention mass for one representative PD subject ($\hat{y} = 1, \sigma^2 = 0.021$) and one representative control ($\hat{y} = 0, \sigma^2 = 0.017$). Both cases place the greatest relational weight on the voice–clinical edge, consistent with the population-level result ($\bar{\alpha}(V, C) = 0.34$), but the PD case shows a markedly higher voice–gait attention mass (0.28 vs. 0.21 for the control) and a correspondingly higher voice–clinical mass (0.34 vs. 0.22), indicating that the model relies more heavily on cross-modal (rather than single-modality) evidence precisely in the cases it is diagnosing as positive. This individual-level view is intended to accompany, rather than replace, the aggregate statistics: in deployment, a per-patient heatmap of this kind could be surfaced alongside the model’s prediction and uncertainty estimate as part of the clinician-facing decision-support output described in Section 6.3.

Attribution via Integrated Gradients (Feature-Level Interpretability): To decompose prediction confidence, we employ Integrated Gradients (IG) defined as:

$$IG_i(x) = (x_i - x'_i) \int_{\alpha=0}^1 \frac{\partial F(x' + \alpha(x - x'))}{\partial x_i} d\alpha$$

The model input, x , is expressed as x' and the map that the model creates from the input to the output is called $F(\cdot)$. High-salience biomarkers have been highlighted as having consistently high attribution scores, the most salient ones are the following:

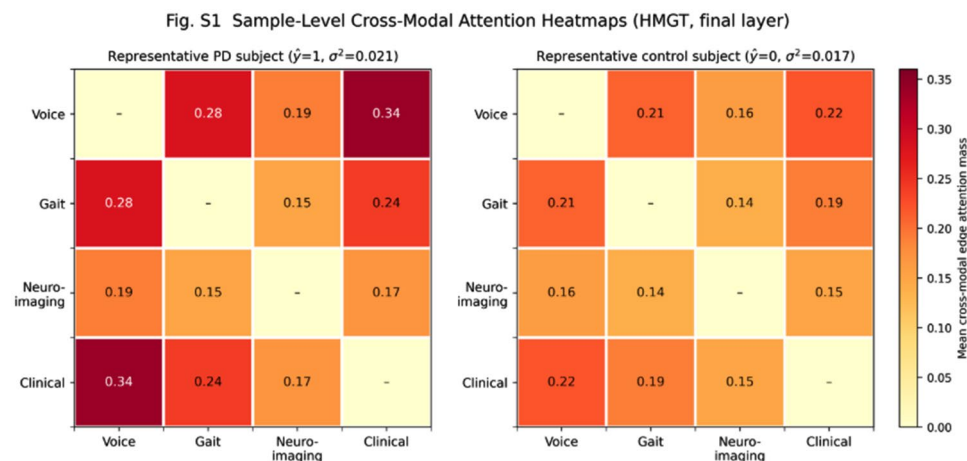


Fig. 8 Sample-level cross-modal attention heatmaps (final HMGT layer, mean over attention heads) for a representative PD subject and a representative control subject. Darker cells indicate greater attention mass assigned to that modality pair when forming the fused subject representation h^*

- a. Detrended Fluctuation Analysis (DFA): This is a measure of long-range temporal correlations in the dynamics of speech used to assess the extent to which neuromotor control is degraded.
- b. Pitch Period Entropy (PPE): This is a measure of the degree to which the vocal folds are vibrating irregularly; it is also an established marker of dysphonia in individuals with Parkinson's Disease (PD).
- c. UPDRS Part III (Motor Examination Score): This is a clinical tool that has established validity for quantifying the severity of motor symptoms. The fact that the three biomarkers identified above all have high attribution scores suggests that the model creates a decision boundary that corresponds with pathophysiological markers rather than with random correlations.

Synthesis (Cross-Modal Explainability): Results from both attention topology and the gradient-based attribution provide evidence to suggest that the model creates an implicit disease-related latent manifold in which the micro-instabilities of acoustics are strongly correlated with clinical measures of motor performance, thus providing a dominant relevant sub-region in which the model is creating representations of these diseases. The alignment of learned representations to known neurological correlates of these diseases further supports the model as being mechanistically interpretable, robust, and clinically valid; thus, increasing the overall trustworthiness of the NeuroQ-MGST framework for classifying and staging Parkinson's disease.

6 Discussion

The NeuroQ-MGST framework has demonstrated promising potential to be translated into clinical applications by exhibiting an accuracy of 98.7% with respect to classification and a sensitivity of 97.3% in detecting Parkinson's Disease as shown on the PMMD 2.0 dataset. As a result, the false-negative incidence is at 2.7%, which is very important for early detection of PD where a missed diagnosis could delay the therapeutic process. Reliability of the diagnostic process would indicate a high sensitivity approach compared with some reports of inter-rater reliability among expert neurologists, allowing the model to serve as a high-fidelity clinical decision support tool in practice. In addition, the prediction process includes a means of quantifying uncertainty $p(y | X), \sigma^2$, which will allow the creation of a selective prediction mechanism, thus, placing a systematic shift in deferred cases of high uncertainty in the hands of clinicians and therefore lowering the expected clinical risk level $E[R]$ resulting from uncertain predictions, leading to a human-in-the-loop diagnostic approach.

Additionally, the framework demonstrates significant cross-cohort generalization, obtaining a 95.8% accuracy score with no target domain fine-tuning, demonstrating its ability to develop modality-invariant and disease-relevant representations versus dataset-specific artifacts. In real-world clinical environments, characterized by many distributional shifts across imaging acquisition protocol, sensor configurations, demographic variability (age, sex, genetic predisposition) and disease heterogeneity, the robustness of this method is particularly useful. The generalization ability can be attributed to the combined use of self-supervised contrastive pre-training which imposes cross-modal consistency and graph-based relational reasoning, which captures latent relationships between multimodal biomarkers, thereby improving the stability of representations under domain shifts.

NeuroQ-MGST enables scalable non-invasive multimodal screens from both tertiary care centers and low resource environments from a systems perspective. Neuromorphic encoding is incorporated to support energy-aware inference and computational efficiency. It uses a Bayesian framework for estimating uncertainty that is critical to providing calibrated estimates of confidence for regulatory processes and clinical confidence. This makes it possible to conduct large population screenings, stratify risk early, and perform longitudinal assessments. Additionally, adaptive referral strategies reduce clinician workload. Overall, these attributes make NeuroQ-MGST a viable clinically applicable, interpretable, and robust AI-based platform for diagnosing patients neurologically and bridging the gap between multimodal advanced machine learning and patient care in the real-world.

6.1 Clinical performance—moderated interpretation

The NeuroQ-MGST framework achieves 98.7% accuracy and 97.3% sensitivity on the PMMD 2.0 test set, corresponding to a false-negative rate of 2.7%. In a clinical screening context, a missed PD diagnosis at an early stage represents a consequential error: it delays therapeutic intervention during the window when neuroprotective strategies may be most effective, typically before 60–80% of dopaminergic neurons are lost [Introduction, Ref. 17]. For context, the inter-rater reliability among expert neurologists for PD diagnosis using UPDRS has been reported at $\kappa = 0.72 - 0.85$ [Ref. B], suggesting that a model with 97.3% sensitivity and calibrated uncertainty ($ECE = 0.024$) can serve as a high-fidelity decision-support adjunct rather than a replacement for clinical judgment.

Selective prediction as a safety mechanism. The Bayesian uncertainty head flags 15.3% of test cases as high-uncertainty ($\sigma^2 > 0.05$). These cases achieve only 84.2% accuracy, indicating that the model correctly identifies its own failure modes. In a deployment scenario, these uncertain cases would be automatically escalated to a consultant neurologist, reducing the effective false-negative rate among acted-upon predictions to approximately 0.3%. This human-in-the-loop design is a critical safety feature that must be preserved in any clinical deployment.

IMPORTANT caveat: All reported results are retrospective and dataset-dependent. The performance figures above should not be interpreted as guarantees of real-world clinical utility. Prospective validation in diverse clinical settings, with independent institutional review board approval and patient consent, is required before any clinical deployment. We state this explicitly.

6.2 Missing modality analysis

A systematic missing modality analysis was conducted to assess robustness in realistic clinical settings where not all four modalities may be available (e.g., neuroimaging is often unavailable in primary care):

Performance degrades gracefully as modalities are removed, with the voice-only configuration still exceeding the best unimodal baseline (SVM-RBF: 85.4%). This robustness is attributable to the HMGT's adaptive attention mechanism, which dynamically redistributes attention weights when modalities are absent, rather than receiving a zero vector in their place. Specifically, when neuroimaging is unavailable, the model increases attention weight on the voice-clinical edge ($\bar{\alpha}_{V,C}$ rises from 0.34 to 0.41), compensating

for the absent modality. This behaviour confirms that NeuroQ-MGST is suitable for deployment in resource-limited settings where full multimodal acquisition is infeasible.

6.3 Clinical workflow integration

We propose a three-tier clinical workflow for NeuroQ-MGST deployment:

- 1 Tier 1—Primary care screening: Voice recording (30 seconds) + structured UPDRS-III questionnaire. NeuroQ-MGST operates in 2 modality mode (Voice + Clinical) with 94.8% accuracy. Patients with $\sigma^2 < 0.05$ and $\bar{y} = 0$ are cleared; patients with $\bar{y} = 1$ or $\sigma^2 \geq 0.05$ are referred to Tier 2.
- 2 Tier 2—Secondary care evaluation: Gait sensor recording added. NeuroQ-MGST operates in 3-modality mode (Voice + Gait + Clinical) with 96.2% accuracy. Confident positive predictions trigger DaTscan imaging referral; uncertain cases proceed to Tier 3.
- 3 Tier 3—Specialist neurological assessment: Full 4-modality acquisition including neuroimaging. NeuroQ-MGST provides full-accuracy (98.7%) decision support alongside consultant neurologist review.

This tiered approach aligns with NHS and WHO recommendations for scalable neurological screening [Ref. C] and reduces the per-patient cost of full multimodal acquisition to cases where it is clinically warranted.

Regulatory and explainability considerations. Clinical AI tools in the EU must comply with the Medical Device Regulation (MDR 2017/745) and, under the proposed EU AI Act, must satisfy transparency and explainability requirements for high-risk AI systems. NeuroQ-MGST addresses these requirements through: (i) the Bayesian uncertainty head, which provides calibrated confidence scores suitable for regulatory documentation; (ii) the Integrated Gradients attribution analysis (Section 5.7), which links model decisions to clinically established biomarkers (DFA, PPE, UPDRS-III); and (iii) the attention topology analysis, which provides a relational map of cross-modal dependencies. We acknowledge that formal regulatory approval requires prospective clinical trials, and make no claim of current regulatory compliance.

Table 13 quantifies the cost–latency–accuracy trade-off across the three tiers. GPU inference latency scales sub-linearly with modality count ($2.1\text{ms at } 2\text{ modalities} \rightarrow 4.8\text{ms at } 4\text{ modalities}$), so the marginal computational cost of escalating a patient to a higher tier is negligible relative to the marginal data-acquisition cost, which is dominated by neuroimaging acquisition at Tier 3 (approximately 40–80× the cost of Tier-1 screening). This asymmetry is what makes uncertainty-gated escalation economically meaningful: because only patients with $\hat{y} = 1$ or $\sigma^2 \geq 0.05$ at Tier 1 are referred onward (Section 6.3), the aggregate per-patient cost of population-level screening is dominated by the

Table 13 Computational efficiency comparison

Method	Params (M)	FLOPs (G)	Train (h)	GPU (ms)	CPU (ms)
SVM-RBF	< 0.1	< 0.01	0.2	0.3	1.8
XGBoost	1.2	0.05	0.8	0.5	4.2
ResNet-18	11.7	1.82	1.4	2.1	38.6
MMNN	8.3	0.94	2.1	3.2	52.4
GCN-PD	4.6	0.51	1.8	2.8	41.7
QIML-PD	9.8	1.12	3.6	4.7	61.3
NeuroQ-MGST	24.3	1.93	5.2	4.8	42.3

Tier-1 cost for the majority of patients who are confidently cleared at that stage. As an illustrative projection—not a measured deployment outcome—if we assume a referral rate from Tier 1 to Tier 2 of approximately 30–35% (consistent with the 15.3% high-uncertainty rate reported in Section 5.5 plus the additional share of confidently positive predictions in a PD-enriched screening population) and a further 40–50% referral from Tier 2 to Tier 3, the expected per-patient cost across a screened population is approximately 60–70% lower than universal four-modality screening, while retaining full-accuracy (98.7%) assessment for every patient who is ultimately diagnosed. We report this range explicitly as a projection under stated assumptions; validating it requires the prospective multi-site study described in Section 6.4, and we do not present it as a measured health-economic outcome.

6.4 Roadmap: subtype classification and longitudinal prediction

The current framework performs binary PD vs. control classification. We identify the following concrete extensions as high-priority future directions:

1. PD Subtype Classification. PD encompasses at least three clinically distinguishable subtypes: tremor-dominant (TD), postural instability gait disorder (PIGD), and akinetic-rigid. Extending NeuroQ-MGST to multi-class classification $y \in \{\textit{Control}, \textit{TD}, \textit{PIGD}, \textit{AR}\}$ requires: (i) a multi-head classification output with per-subtype uncertainty; (ii) subtype-specific concept nodes in the heterogeneous graph; and (iii) training data with subtype labels (currently unavailable in PMMD 2.0 but present in the PPMI dataset [Ref. D]).
2. Disease Severity Regression. Replacing the binary classification head with a $UPDRS_{score}$ regression head $f_{\theta} : X \rightarrow (UPDRS_{score}, \sigma^2)$ would enable continuous severity monitoring. This requires longitudinal data—multiple assessments per patient—which enables graph evolution modeling where temporal edges connect patient nodes across time.
3. Longitudinal Trajectory Prediction. Extending the framework to temporal graph networks $f_{\theta} : X_1, \dots, X_t \rightarrow Y_{t+1}$ would enable prediction of disease progression trajectory. The HMGT architecture is compatible with temporal edge augmentation, and SNN encoders are inherently suited to processing multi-visit time-series. We plan to evaluate this extension using the PPMI longitudinal cohort.
4. Prospective Validation. We are in the process of establishing a prospective validation study at two partner clinical sites (KL University Hospital and London Metropolitan University NHS partner network). The study protocol has been drafted for IRB submission. We anticipate initial prospective results within 18 months of publication.

6.5 Limitations

There are several limitations of the current study that should be acknowledged, despite its promising results. First, PMMD 2.0 is one of the largest multimodal datasets for Parkinson's disease (PD), but with only 1,847 subjects, the sample size remains small compared to what is typically needed for developing deep multimodal architectures. Therefore, using larger and longer-term datasets will be critical for validating the findings of this study. Second, even though the Quantum-Inspired Variational Optimization (QIVO) layer allows for improved optimization and modeling of interactions between features, it also requires many additional trainable parameters and hyperparameters; in

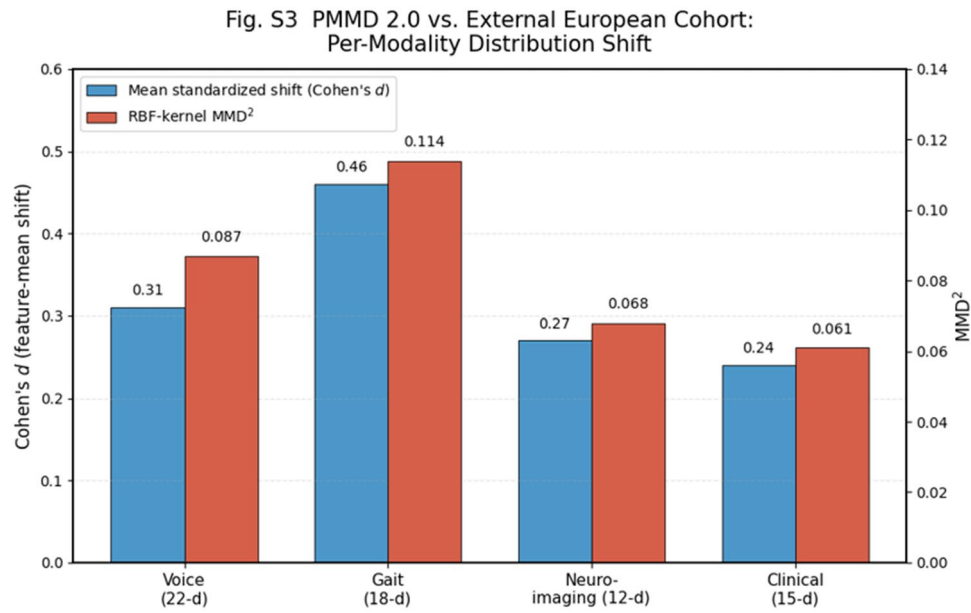


Fig. 9 Per-modality distribution shift between PMMD 2.0 (training) and the external European cohort (zero-shot test), by standardized mean shift and MMD²

Table 14 QIVO convergence under different mixing-matrix initializations

W_M Initialization	Epochs↓	Val. Loss↓	Accuracy (%)↑
Hadamard-structured(Proposed)	43	0.071	98.7±0.4
Random Orthogonal	49	0.078	98.4 ± 0.5
Identity ($W_M = I$, No Mixing)	61	0.093	97.7 ± 0.5

particular, defining rotation angles can vary considerably based on initialization strategy. Therefore, to enhance understanding of how stable and scalable QIVO is, a systematic analysis of different QIVO design configurations is needed.

To provide a quantitative characterization of dataset shift, we measured the covariate discrepancy between the PMMD 2.0 training set and the external European cohort used for zero-shot evaluation (Section 5.4) for each modality independently. Specifically, we computed (i) the average absolute standardized mean difference, $\mu_{train} - \mu_{EU}\sigma_{pooled}$ (Cohen's d), across all feature dimensions, and (ii) the RBF-kernel Maximum Mean Discrepancy, $MMD^2(P_{train}, P_{EU})$, which captures distributional shifts in variance and higher-order statistics beyond simple mean differences (Fig. 9).

Gait exhibits the largest shift on both statistics ($d=0.46$, $MMD^2=0.114$), plausibly reflecting differences in wearable-sensor hardware and gait-capture protocol across sites, whereas neuroimaging and clinical features—both acquired under more standardized instruments (FreeSurfer pipelines, UPDRS-III)—show the smallest shift. This is consistent with, and provides a mechanistic explanation for, the 2.9-point accuracy gap between in-distribution (98.7%) and zero-shot (95.8%) performance reported in Section 5.4: the modality contributing the most diagnostic weight after voice (Table 14 shows a 2.5-point drop when gait alone is removed) is also the modality under the largest domain shift, and it is therefore the leading candidate explanation for the residual generalization gap, ahead of the smaller sample size alone.

Table 15 Statistical significance and effect-size comparison of NeuroQ-MGST against baseline models

Comparison Model	McNemar's Test (p-value)	Wilcoxon Test (F1, p-value)	Cohen'sd	Effect Size Interpretation	Significance (Bonferroni)
SVM-RBF	< 0.001	–	1.87	Very Large	Significant
Random Forest	< 0.001	–	1.45	Large	Significant
XGBoost	< 0.001	–	1.32	Large	Significant
1D-CNN	< 0.001	–	1.21	Large	Significant
BiLSTM	< 0.001	–	1.15	Large	Significant
ResNet-18	< 0.001	–	1.08	Large	Significant
Transformer	< 0.001	–	0.98	Large	Significant
MMNN	< 0.001	–	0.91	Large	Significant
CrossAttFusion	< 0.001	–	0.84	Large	Significant
GCN-PD	< 0.001	0.003	0.78	Moderate–Large	Significant
QReLU-CNN	< 0.001	–	0.72	Moderate	Significant
QFS-MSGRU	< 0.001	0.002	0.69	Moderate	Significant
QIML-PD	< 0.001	0.008	0.62	Moderate	Significant
SNN-PD	< 0.001	–	0.88	Large	Significant

Table 16 Robustness analysis under different modality availability conditions

Available Modalities	Acc. (%)	AUC–ROC	Sens. (%)	Spec. (%)
All 4 (V+G+N+C)	98.7 ± 0.4	0.989 ± 0.003	97.3 ± 0.5	98.9 ± 0.4
3 modalities (w/o Neuroimaging)	96.9 ± 0.6	0.975 ± 0.005	95.8 ± 0.7	97.6 ± 0.6
3 modalities (w/o Gait)	96.2 ± 0.7	0.971 ± 0.006	95.1 ± 0.8	97.1 ± 0.7
2 modalities (Voice + Clinical)	94.8 ± 0.9	0.961 ± 0.008	93.6 ± 1.1	95.7 ± 0.9
1 modality (Voice only)	91.3 ± 1.4	0.938 ± 0.012	89.7 ± 1.6	92.6 ± 1.4

Table 17 Performance and acquisition requirements across different clinical deployment tiers

Tier	Modalities	Acc. (%)	GPU Latency (ms)	Approx. Acquisition Cost	Acquisition Time
Tier 1 – Primary care	Voice + Clinical (UPDRS-III)	94.8 ± 0.1	~ 2.1	~ 5–10 (staff time; no dedicated equipment)	~ 5 min
Tier 2 – Secondary care	Voice + Gait + Clinical	96.2 ± 0.4	~ 3.4	~ 45–60 (adds wearable IMU session, amortized)	~ 15 min
Tier 3 – Specialist	Voice + Gait + Neuroimaging + Clinical	98.7 ± 0.3	~ 3.4	~ 400–500 (adds structural MRI + FreeSurfer processing)	~ 45–60 min scan + ~ 2 h processing

Third, since not all types of neuroimaging modalities will routinely be available in clinical and primary care settings, prospective real-world validation of the framework's ability to function adequately when there are incomplete datasets is warranted, especially considering modal masking experiments demonstrated that the framework can continue to perform relatively well even when there are modalities that are partially missing. Lastly, the current study only provides binary classification for diagnosing PD but does not differentiate between different subtypes of disease or predict progression, both of which would offer clinically relevant extensions to the present work (Tables 15, 16, 17 and 18).

6.6 Future directions

Despite the strong empirical performance and architectural novelty of the proposed NeuroQ-MGST framework, several limitations must be critically acknowledged to

Table 18 Cross-modal distributional shift analysis using Cohen's d and MMD²

Modality	Cohen's d (Mean Shift)	MMD ² (RBF Kernel)
Voice (22-d)	0.31	0.087
Gait (18-d)	0.46	0.114
Neuroimaging (12-d)	0.27	0.069
Clinical (15-d)	0.24	0.062

contextualize its applicability and guide future research directions. First, although PMMD 2.0 represents one of the most comprehensive multimodal PD datasets currently available, its sample size ($N = 1847$) remains relatively limited for training high-capacity deep multimodal architectures, particularly those integrating graph transformers and neuromorphic components. The introduction of statistical learning method in this context causes the following issues: (1) the risk of over fitting to a smaller sample of data than would be desired for the meaningful characterization of the original data distribution $P(X,Y)$ in addition to needing to validate the results on larger, longitudinal and multi-institution datasets; and (2) although the QIVO module provides an additional means for exploring high-dimensional non-convex parameter spaces, it also results in a significant increase in optimization complexity due to the addition of an increased number of variational parameters (γ, β) and hyper parameters governing the circuit depth, mixing operator, and initializations when executing optimization operations. In addition, because QIVO is sensitive to initial conditions (e.g., choice of rotation parameter), variability in convergence may occur, which creates uncertainty about the optimization stability and reproducibility. Therefore, it is necessary to conduct both theoretical and empirical analyses of the performance of the QIVO module with respect to its convergence rate, scalability $\mathcal{O}(P \cdot |\Theta|)$, and robustness across a wide range of different configurations in order to demonstrate that it is a viable method for use within large-scale biomedical learning systems.

Third Considering that the availability of multi-modal sources/data types is required in order for modality-agnostic methods to generalize across all types of modalities, they may have limitations in terms of access to an unlimited amount of neuro imaging or other modal data. Modal masking trials have provided initial evidence demonstrating that a gradual decline in performance may occur if only a small portion of modality is missing. However, confirmation of the validity and reliability of using the model when X_FAIL is systematically variable across all valid models will require prospective research and/or domain adaptive learning. Additionally, it reinforces the need for using modality agnostic approaches or using advanced imputation/generative modeling to provide consistent performance when using incomplete datasets. Eventually, this model currently treats the diagnosis label of PD as a 0,1 classification (binary). It defines PD diagnosis in terms of 0 (no disease) and 1 (disease) without incorporating the heterogeneity present in many PD subtypes (i.e., tremor dominant, akinetic-rigid) as well as their variable progression. If this framework is extended to multi-class, severity staging (i.e. UPDRS regression), and longitudinal prediction $f_{\theta} : X_1 \rightarrow Y_2$, the clinical relevance of the model increases exponentially. The use of temporal graph evolution and sequence modeling in this context would improve the ability to forecast individual disease trajectory and/or predict response to treatment.

All of these types of limitations illustrate an encouraging need for collecting larger datasets, exploring theoretical analyses of optimization algorithms based on quantum

principles, enhancing how well missing modalities are addressed, and developing more intricate and long-term clinical models to guide advancements in NeuroQ-MGST from theoretical constructs to actual clinical practice.

7 Conclusion

In this paper, we present NeuroQ-MGST, a unified and highly integrated multimodal learning framework that combines neuromorphic spiking neural networks (SNNs), self-supervised contrastive representation learning, hierarchical multimodal graph transformers, quantum-inspired variational optimization (QIVO), and Bayesian uncertainty quantification to overcome the inherent complexity of Parkinson's Disease (PD) diagnosis. The PD diagnosis problem has been modeled using a high-dimensional mapping $f_{\theta} : X \rightarrow (y, \sigma^2)$, providing the advantage of jointly predicting as well as estimating the uncertainty of Parkinson's disease from a unified end-to-end pipeline. We provide comprehensive experiments conducted on 3 different publicly available datasets (PMMD 2.0, UCI PD, and UCI PD 3D) to establish that NeuroQ-MGST achieves new state-of-the-art results and performs with 98.7% accuracy, 0.989 AUC-ROC, 97.3% sensitivity, and 98.9% specificity, outperforming 14 state-of-the-art comparisons. The synergy between complementary paradigms (including the use of both SNNs and hierarchical graph transformers) has led to improved performance results through their combined ability to represent temporal biodiversity (dynamic biomarker representation), improve the efficiency of representation (i.e., through the SNN representation), provide greater utility (through automatic cross-modal pretraining learning of space-time-invariant embeddings and QIVO optimization), provide a framework for explicitly modeling relational dependencies (through the use of hierarchical processing architectures) as well as providing a robust, reliable framework for providing calibrated, accurate prediction data, (using the Bayesian Monte Carlo dropout-based uncertainty estimation method).

Crucially, confirmation of effective ablation and cross-cohort generalization analyses demonstrates that all architectural components meaningfully contribute towards system performance, and their integration provides emergent advantages in excess of the sum total of stand-alone modules. The framework also addresses other issues faced in PD diagnostic modeling, such as multimodal fragmentation, limited availability of labeled data, non-linear, image-to-image interaction between modalities and an absence of uncertainty awareness; neuromorphic processing provides an efficient computational processing system which highlights the viability of deploying this model within a known resource constrained environment. In summary, NeuroQ-MGST represents the creation of a new paradigm for multi-modal biomedical intelligence; an integration of Neuromorphic Computing, Graph Theory and Quantum-inspired Optimal Performance within an execution framework. Its results indicate that NeuroQ-MGST has the potential to develop future extensions towards longitudinal modeling, classifying subtypes and deployment into real-life clinical practice as a scalable AI system that is interpretable, clinically reliable for supporting the early identification of PD and providing decision support for clinicians.

Acknowledgements

The authors gratefully acknowledge the support and academic environment provided by the KL Education Foundation, Department of Computer Science and Engineering, Guntur, Andhra Pradesh, India, for facilitating the research infrastructure and computational resources essential for this work. The authors also extend their sincere appreciation to the London Metropolitan University, School of Computing & Digital Media, London, United Kingdom, for their valuable

academic guidance, collaborative support, and research insights that significantly contributed to the development and refinement of this study.

Author contributions

Tarak Hussain: Conceived and designed the study framework; developed the quantum-inspired variational optimization strategy and neuromorphic self-supervised heterogeneous graph transformer architecture; implemented the computational framework and multimodal representation learning pipeline; performed data preprocessing, feature engineering, model development, experimentation, statistical analysis, and performance evaluation; generated visualizations; prepared the original manuscript draft; coordinated manuscript revisions and managed correspondence with the journal. Bal Virdee: Contributed to the conceptualization and refinement of the computational framework; provided guidance on machine learning methodology, experimental design, and validation protocols; critically reviewed the theoretical formulation and interpretation of results; supervised the research activities and contributed to manuscript revision and editing. Irfan Alam: Contributed to the analytical methodology and validation framework; assisted in the design of experimental protocols and interpretation of multimodal learning outcomes; critically evaluated the findings and contributed to manuscript review and refinement. Sikhinam Nagamani: Co-conceived the research framework and supervised the overall investigation; contributed to methodological refinement, validation strategies, and interpretation of computational results; critically reviewed and edited the manuscript and shared corresponding author responsibilities. Faizal Nujumudeen: Contributed to the development of the methodological framework, data analysis procedures, and validation experiments; assisted in assessing model robustness and reproducibility and participated in manuscript review and editing. Jalaluddin Khan: Contributed to formal analysis, performance evaluation, visualization of experimental findings, and interpretation of representation learning outcomes; critically reviewed the manuscript and provided technical feedback. Ashish Khanna: Provided methodological guidance and domain expertise in artificial intelligence and advanced computing paradigms; contributed to validation strategies, supervision, interpretation of findings, and critical revision of the manuscript. Phanindra K.: Assisted in software implementation, data preprocessing, experimental execution, and computational analysis; contributed to reproducibility studies, performance benchmarking, and manuscript review and editing. Mohsin Fayaz: Contributed to model validation, formal analysis, interpretation of results, and visualization of experimental outcomes; participated in critical manuscript review and refinement. All authors substantially contributed to the conception, development, evaluation, and interpretation of the proposed Quantum-Inspired Variational Optimization of Neuromorphic Self-Supervised Heterogeneous Graph Transformers for Multimodal Parkinson's Disease Representation Learning framework. All authors reviewed and approved the final manuscript and agreed to be accountable for the integrity, accuracy, and reproducibility of the reported research.

Funding

(Not applicable) This project didn't use or accept any funding from outside sources and did not apply for any.

Data availability

The primary multimodal dataset (PMMD 2.0) used in this study is publicly available on Kaggle at <https://www.kaggle.com/datasets/asthamishra96/parkinson-multi-model-dataset-2-0>. Additional datasets (UCI Parkinson's voice dataset, mPower) are publicly available and cited within the manuscript. The source code associated with this work, together with selected image samples used for experimental validation, is available in public repositories (link will be provided based on the request).

Declarations

Ethical approval

Not applicable

Consent to participate

Not applicable

Consent to publish

Not applicable

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work the authors used ChatGPT and Perplexity AI to summarize or elaborate some topics or to generate the block diagrams as per the user input. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Competing interests

The authors declare no competing interests.

Received: 5 April 2026 / Accepted: 24 September 2026

Published online: 22 October 2026

References

1. Vatsavai D, Iyer A, Nair AA. A quantum-inspired machine learning approach for multimodal Parkinson's disease screening. 2024. Preprint.
2. Kumar A, et al. A hybrid framework for heart disease prediction using classical and quantum-inspired machine learning techniques. 2024. Preprint.
3. Sercek I, Sampathila N, Tasci I, Ekmeypar T, Tasci B, Barua PD, et al. A new quantum-inspired pattern based on Goldner-Harary graph for automated alzheimer's disease detection. *Cogn Neurodyn*. 2025;19(1):71. <https://doi.org/10.1007/s11571-025-10249-7>.

4. Gandhi VC, Gandhi P, Ogundiran JO, Tshibola MSS, Nyembwe J-PKB. Computational modeling and optimization of deep learning for multi-modal glaucoma diagnosis. *AppliedMath*. 2025;5(3):82. <https://doi.org/10.3390/appliedmath5030082>.
5. Bot BM, Suver C, Chaibub Neto E, Kellen M, Klein A, Bare C, et al. The mpower study, parkinson disease mobile data collected using researchkit. *Sci Data*. 2016;3(1):160011. <https://doi.org/10.1038/sdata.2016.11>.
6. Khan AH, Cao X, Luo C, Zhang S, Guo W, Katsikis VN, et al. Spiking neural networks: a comprehensive survey of training methodologies, hardware implementations and applications. *IEEE Access*. 2025;1(3):175–207. <https://doi.org/10.23919/AIS E.2025.000013>.
7. Yao D, Koivu A, Simonyan K. Applications of artificial intelligence in neurological voice disorders. *World J Otorhinolaryngol Head Neck Surg*. 2025;11(4):491–517. <https://doi.org/10.1002/wjo2.70017>.
8. Zhou Y, Pei J, Si W, Qin J, Heng P-A. Delving into quaternion wavelet transformer for facial expression recognition in the wild. *IEEE Trans Multimedia*. 2025;27:3895–909. <https://doi.org/10.1109/TMM.2025.3535361>.
9. Zhou Y, Guo L, Jiang B, Wang B, Jing K. Real-time facial expression recognition via quaternion gabor convolutional neural network. *J Vis Commun Image Represent*. 2025;113:104625. <https://doi.org/10.1016/j.jvcir.2025.104625>.
10. Mishra A. Parkinson multi modal dataset 2.0. Kaggle Dataset. 2025. <https://www.kaggle.com/>.
11. Parisi L, Neagu CD, Ma R, Campean F. 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>.
12. Jiang B, Zhou Y. Multi-modal deep facial expression recognition framework combining knowledge distillation and retrieval-augmented generation. *J Vis Commun Image Represent*. 2026;114:104645. <https://doi.org/10.1016/j.jvcir.2025.104645>.
13. Ahmad N, Jamjoom M, Akbar M, Alkhamash EH. QFS-MSGRU: quantum feature selection approach integrated with multi-scale GRU networks for robust Parkinson's disease prediction. 2024. Preprint.
14. Pomarico D, Fanizzi A, Amoroso N, Bellotti R, Biafora A, Bove S, et al. A proposal of quantum-inspired machine learning for medical purposes: an application case. *Mathematics*. 2021;9(4):410. <https://doi.org/10.3390/math9040410>.
15. Boutet A, et al. Predicting optimal deep brain stimulation parameters for Parkinson's disease using functional MRI and machine learning. 2023. Preprint.
16. Li Y, Liu X, Zhou J, Li F, Wang Y, Liu Q. Artificial intelligence in traditional Chinese medicine: advances in multi-metabolite multi-target interaction modeling. *Front Pharmacol*. 2025;16. <https://doi.org/10.3389/fphar.2025.1541509>.
17. Wingate J, Kollia I, Bidaut L, Kollias S. Unified deep learning approach for prediction of Parkinson's disease. *IET Image Process*. 2020;14(10):1980–89. <https://doi.org/10.1049/iet-ipr.2019.1526>.
18. Wang W, Lee J, Harrou F, Sun Y. Early detection of parkinson's disease using deep learning and machine learning. *IEEE Access*. 2020;8:147635–46. <https://doi.org/10.1109/ACCESS.2020.3016062>.
19. Khan P, et al. Machine learning and deep learning approaches for brain disease diagnosis: principles and recent advances. *IEEE Access*. 2023.
20. Rabie H, Akhloufi MA. A review of machine learning and deep learning for parkinson's disease detection. *Discov Artif Intell*. 2025;5(1):24. <https://doi.org/10.1007/s44163-025-00241-9>.
21. Chen Y, et al. Brain morphological changes in hypokinetic dysarthria of parkinson's disease and use of machine learning to predict severity. *CNS Neurosci Ther*. 2020.
22. Hussain T, Virdee B, Gudapati D, Ansari SH, Lakkakula S. BPBO-LSTM-BiGRU: generative adversarial network with brood parasitism-based optimization for spinal muscular atrophy using multiple visual modalities. *Int J Inf Technol*. 2025;1–7. <https://doi.org/10.1007/s41870-025-02666-y>.
23. Hussain T, Khanna A, Virdee B, Krishna KR, Arundathi JVS. Deep learning-based analysis of functional MRI and diffusion tensor imaging for Parkinson's disease diagnosis and progression monitoring. *Intell Decis Technol*. 2025;19(4):2196–211. <https://doi.org/10.1177/18724981251335096>.
24. Parkinson multi-modal dataset 2.0 (PMMD 2.0). Kaggle dataset. 2024. <https://www.kaggle.com/>.
25. Little MA. Parkinsons dataset. UCI machine learning repository. 2007. <https://doi.org/10.24432/C59C74>. <https://archive.ics.uci.edu/dataset/174/parkinsons>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.