DS-HViT: A Dual-Stream Hierarchical Vision Transformer for Multi-Scale Analysis of Parkinson's Disease Handwriting

Ayoub Louja
Hassan First University of Settat, Faculty of Sciences and Technologies, Laboratoire IR2M, Morocco, Settat
a.louja.doc@uhp.ac.ma (corresponding author)

Yassin Zaiouane
Hassan First University of Settat, Faculty of Sciences and Technologies, Laboratoire IR2M, Morocco, Settat
zaiouane.ensa@uhp.ac.ma

Abdellah Jamali
Hassan First University of Settat, Faculty of Sciences and Technologies, Laboratoire IR2M, Morocco, Settat
jamali.abdellah@uhp.ac.ma

Najib Naja
National Institute of Posts and Telecommunications, Rabat, Morocco
naja@inpt.ac.ma

Received: 19 June 2025 | Revised: 9 July 2025, 20 July 2025, and 25 July 2025 | Accepted: 27 July 2025

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | *DOI: https://doi.org/10.48084/etasr.12833*

ABSTRACT

Early detection of Parkinson's Disease (PD) remains challenging in the medical field due to motor symptoms manifesting across different temporal scales. To address this, we introduce a Dual-Stream Hierarchical Vision Transformer (DS-HViT), a deep learning framework designed to capture the multi-scale temporal dynamics of PD-related Handwriting (HW) impairments. The architecture employs parallel micro- and macro-scale streams: the micro-scale stream models high-frequency tremor signatures, while the macro-scale stream captures gradual motor decline. The model was evaluated on the NewHandPD dataset using 5-fold patient-stratified cross-validation, complemented by bootstrap analysis and McNemar's test for statistical validation. DS-HViT achieved an accuracy of 98.2 ± 0.8%, with sensitivity of 97.3 ± 1.2% and specificity of 99.0 ± 0.9%, significantly surpassing state-of-the-art methods ($p < 0.01$). Ablation studies confirmed the synergistic effect of dual-stream processing, while the model demonstrated excellent calibration with an Expected Calibration Error (ECE) of 0.043 and exceptional discriminative ability with a Diagnostic Odds Ratio (DOR) of 486.3.

Keywords-Parkinson's disease; handwriting analysis; vision transformer; multi-scale analysis; deep learning

I. INTRODUCTION

Parkinson's disease (PD), the second most prevalent neurodegenerative disorder, affects over 8.5 million people worldwide and is recognized by the World Health Organization (WHO) as one of the fastest-growing neurological conditions [1]. Onset typically occurs around age 60 [2, 3], and although no cure currently exists, early detection is critical for timely intervention. In PD, dopaminergic depletion disrupts the coordinated brain activation required for Handwriting (HW) (premotor planning, motor execution, basal ganglia modulation), causing motor disturbances across different temporal scales [4, 5]. Distinct neural circuit dysfunctions underlie different symptoms: resting tremor arises from abnormal thalamo-cortical oscillations, bradykinesia from disrupted basal ganglia-thalamo-cortical loops, and micrographia from impaired premotor planning.

HW has been established as a reliable biomarker for prodromal PD detection [6], with digital tablets enabling precise kinematic quantification of HW dynamics, including entropy, acceleration, and pressure variations [7]. Clinical assessment typically follows a two-phase approach: i) real-time HW observation to capture tremor and velocity fluctuations, and ii) post-hoc analysis of completed samples to quantify micrographia, tremor patterns, and spatial irregularities.

Additionally, machine learning has advanced PD HW diagnosis substantially. Early methods focused on handcrafted features, with the authors in [8] extracting velocity, acceleration, and jerk features from digital tablets, achieving 85% accuracy. Authors in [9] later introduced the HandPD dataset and reframed HW analysis as an image recognition task, enabling the adoption of deep learning. Since then, Vision Transformers (ViTs) have achieved state-of-the-art results: authors in [10] achieved 99.2% accuracy with transfer learning, and authors in [11] achieved 99.9% accuracy using ViT with ElasticNet feature selection. Other novel models include attention-based continuous convolutional networks (96.5% accuracy) [12] and ChiGa-Net (92.56% balanced accuracy on external validation) [13].

Moreover, recent hybrid ViT approaches further demonstrate the effectiveness of combining transformer architectures with complementary processing streams. Authors in [14] achieved 99.2% accuracy in brain tumor classification using a hybrid ViT-iResNet model that integrates local feature extraction with global contextual understanding through attention-based fusion mechanisms. Similarly, authors in [15] developed Shot-ViT for cricket batting shot classification, achieving 92.58% accuracy by leveraging ViT's self-attention mechanisms to capture temporal patterns in motor activities.

Despite these advances, existing models fail to capture the temporally diverse manifestations of PD HW. Symptoms range from rapid tremor oscillations requiring sub-100 ms resolution to progressive micrographia developing over several seconds. Authors in [16] showed that fixed-length segmentation cannot simultaneously capture both fine-grained disturbances and long-term deterioration. Moreover, current architectures often neglect the neurophysiological basis of motor symptoms, applying uniform processing across tasks and overlooking critical diagnostic signals.

To address these limitations, we propose Dual-Stream Hierarchical Vision Transformer (DS-HViT) processing HW signals through parallel pathways optimized for different temporal scales, capturing both rapid tremor oscillations and progressive deterioration.

II. METHODOLOGY

This study employs a comparative cross-sectional design using the publicly available NewHandPD dataset [8], selected for its established use in PD HW research and inclusion of both spiral and meander tasks.

A. Dataset newHandPD and Data Augmentation

The NewHandPD dataset, collected at the Botucatu School of Medicine (Brazil), comprises HW exams from 66 individuals: 35 Healthy Controls (HC) (mean age 44.05 ± 14.88 years) and 31 PD patients (mean age 57.83 ± 7.85 years). Each participant completed 12 tasks: four spirals, four meanders, one paper circle, one air circle, and two diadochokinesis movements. Dynamic signals and static images were simultaneously recorded with a Biometric Smart Pen (BiSP). The dataset provides 9 images per individual, totaling 264 images and 792 signals (420 from HC, 372 from PD patients) [9]. Figure 1 shows representative samples highlighting characteristic differences between PD and control HW patterns.

Given the limited clinical sample size and DS-HViT's dual-stream requirements, we implemented neurophysiologically informed augmentation. Transformations included geometric modifications (rotations ±5°, translations ±10 pixels, elastic deformations σ=4, α=34), task-specific adaptations (radial scaling 0.95–1.05× for spirals; horizontal shearing ±2° for meanders), and intensity modulations (contrast 0.8–1.2, brightness ±20%, Gaussian noise σ=0.02). Each original sample yielded six augmented variants, expanding the dataset seven-fold while maintaining class balance (46.6% PD, 53.4% HC).

B. DS-HViT Architecture

DS-HViT addresses multi-scale PD motor symptoms through parallel processing streams optimized for different temporal resolutions. Building on the ViT framework in [17], the architecture (Figure 2) comprises three components: i) parallel temporal streams, ii) Cross-Scale Attention Fusion (CSAF), and iii) task-specific modules. Input images are standardized to 224×224 Red-Green-Blue (RGB) and normalized to [0,1]. Each image is divided into 16×16 non-overlapping patches, yielding 196 patches. Then, each patch is linearly projected into a D-dimensional embedding space, balancing the capture of fine-grained tremor patterns with computational efficiency [18]:

$$e_i = p_i * W_{proj} + b_{proj} \quad (1)$$

where $W_{proj} \in \mathbb{R}^{768\times D}$ is the projection matrix, b_{proj} is the bias term. Positional encodings $E_{pos} \in \mathbb{R}^{196\times D}$ are then added to preserve spatial relationships:

$$z_0 = [e_0; e_1; \dots; e_{196}] + E_{pos} \quad (2)$$

DS-HViT's key innovation is its dual-stream architecture, processing HW at two temporal scales. The micro-scale stream models high-frequency tremors using 32-patch sliding windows (50% overlap), processed with 6 transformer layers, 8 attention heads, and 384-dimensional embeddings:

$$S_{micro}^{(w)} = ViT_{micro}(z_0[\, w{:}\, w + 32]) \quad (3)$$

The macro-scale stream models progressive motor symptoms across the full sequence with 12 transformer layers, 12 attention heads, and 768-dimensional embeddings:

$$S_{macro} = ViT_{macro}(z_0) \quad (4)$$

where $w \in \{0, 16, 32, \dots, 164\}$ represents the window starting position. This asymmetric design reflects the computational requirements of distinct symptom scales while maintaining efficiency for clinical deployment [19].

To integrate features from both temporal resolutions, CSAF adaptively fuses information based on clinical relevance:

$$F_{fused} = CSAF(S_{micro}, S_{macro}) \quad (5)$$

The fusion process involves three steps:

$$F_{concat} = [S_{micro}; S_{macro}; S_{micro} \odot \mathrm{S_{macro}}] \quad (6)$$

$$A_{cross} = \mathrm{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) \quad (7)$$

$$F_{fused} = A_{cross} F_{concat} \quad (8)$$

where $\odot$ denotes elementwise multiplication, Q and K are query and key matrices obtained from linear projections of F_{concat}, and d_k is the key dimension.

Recognizing that spiral and meander tasks engage distinct neural circuits, DS-HViT incorporates two specialized task modules. The Spiral module applies radial attention and polar-coordinate encoding to capture circular motion, while the Meander module employs directional attention and transition encoding to represent linear directional changes:

$$F_{spiral} = RadialAttention\left(PolarTransform\left(F_{fused}\right)\right) \quad (9)$$

$$F_{meander} = DirectionalAttention\left(TransitionEncode\left(F_{fused}\right)\right) \quad (10)$$

Task-adaptive representations are classified through a multilayer perceptron (MLP) with dropout regularization (p=0.5) to mitigate overfitting on limited clinical data:

$$h_1 = \mathrm{ReLU}\left(\mathrm{Dropout}(F_{\mathrm{task}} W_1 + b_1)\right) \quad (10)$$

$$h_2 = \mathrm{ReLU}\left(\mathrm{Dropout}(h_1 W_2 + b_2)\right) \quad (11)$$

$$y = \mathrm{Softmax}(h_2 W_3 + b_3) \quad (12)$$

where F_{task} denotes the output of the appropriate task-specific module, and W_i, b_i represent the weight matrices and bias vectors of the i^{th} layer, respectively.

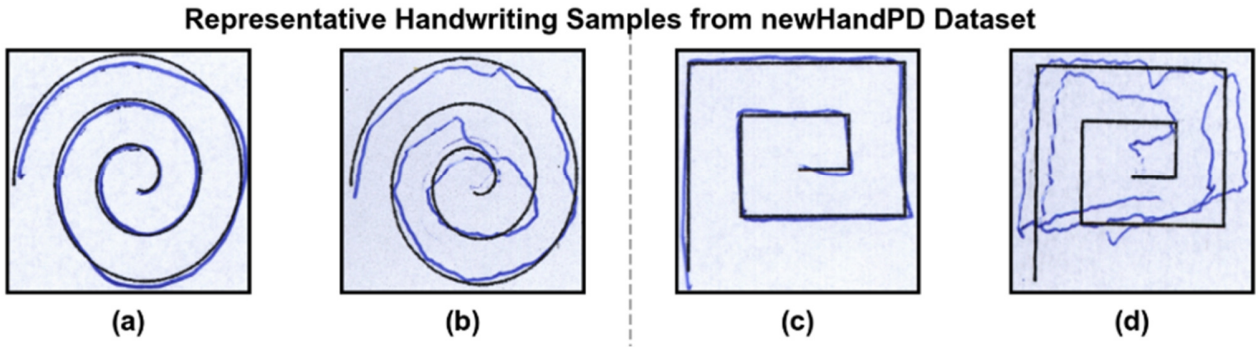

Fig. 1. Representative HW samples from NewHandPD datasets. (a)-(b) Spiral drawings from HC and (c)-(d) from PD patient, respectively,

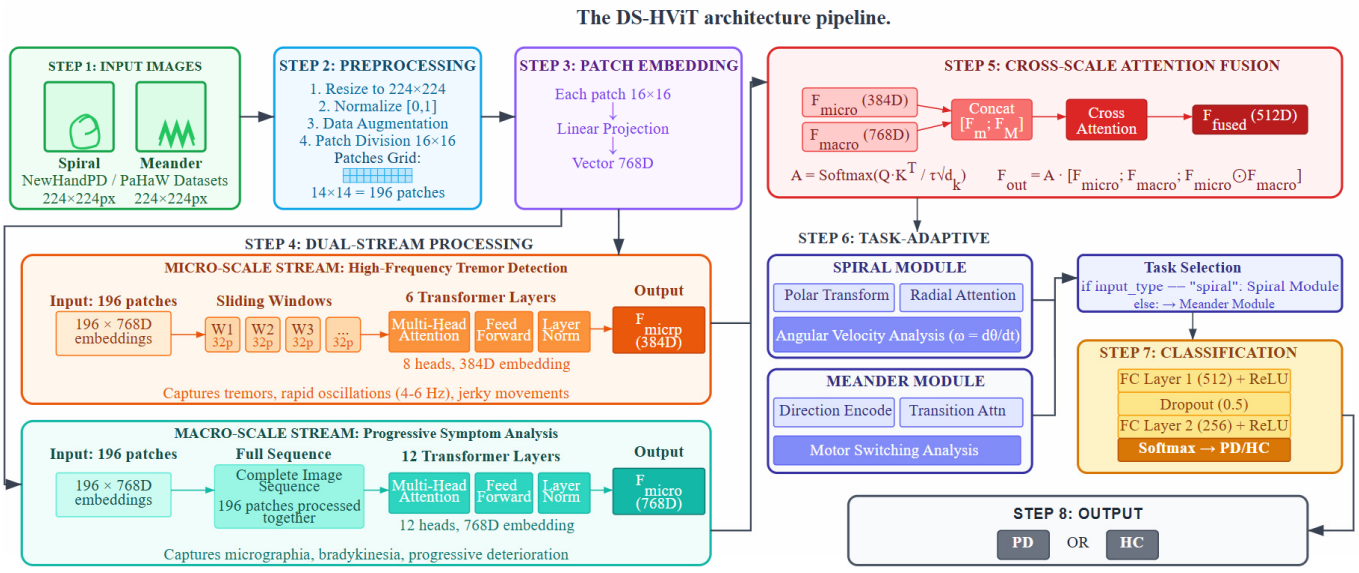

Fig. 2. The DS-HViT architecture pipeline.

C. Network Architecture Details

Table I summarizes the architecture configuration, designed following recent advances in multi-scale vision models [20]. DS-HViT comprises 37.5 million (M) parameters and requires 5.3 billion (G) Floating-point Operations per second (FLOPs) per inference. The inference latency is 23 ms on clinical hardware (GTX 1060) with a memory footprint of 142 MB, enabling real-time deployment.

TABLE I. DS-HVIT NETWORK ARCHITECTURE CONFIGURATION

Component	Micro-Scale	Macro-Scale	Fusion Module
Input Dimensions	32 × D	196 × D	-
Transformer Layers	6	12	-
Attention Heads	8	12	8
Hidden Dimensions	384	768	768
FFN* Dimensions	1536	3072	3072
Dropout Rate	0.1	0.1	0.5
Window Size	32 patches	Full sequence	-
Window Overlap	50%	-	-
Parameters (M)	5.2	28.4	2.1
FLOPs (G)	0.8	4.2	0.3
Task-Adaptive	**Spiral Module**	**Meander Module**	**Classifier**
Input Dimensions	768	768	768
Hidden Layers	2	2	2
Hidden Units	[512, 256]	[512, 256]	[512, 256, 128]
Output Units	768	768	2
Parameters (M)	0.8	0.8	0.4
Total Parameters	37.5 M		
Total FLOPs	5.3 G		

*Feed-Forward Network (FFN)

D. Training Strategy

Training follows a two-stage approach. Stage one involves the pretraining of individual streams (50 epochs) at the micro-scale using self-supervised tremor detection on augmented healthy samples:

$$x'(t) = x(t) + A \sin(2\pi f t + \phi) \quad (13)$$

where amplitude $A \in [0.5, 2.0]$ pixels, frequency $f \in [4, 6]$ Hz matches Parkinsonian tremor characteristics [21], and phase $\phi \in [0, 2\pi]$. Stage two implements end-to-end fine-tuning with the AdamW optimizer [22] and cosine learning rate scheduling. The composite loss balances primary classification with auxiliary objectives:

$$L_{total} = L_{CE} + 0.1(L_{spiral} + L_{meander}) \quad (14)$$

where L_{CE} represents the primary cross-entropy loss for binary PD classification, and L_{spiral} and $L_{meander}$ are auxiliary task-specific losses for spiral and meander drawing patterns, respectively. Class imbalance is addressed via inverse-frequency weighting. Training employs mixed-precision computation with gradient accumulation, yielding an effective batch size of 128.

E. Hyperparameter Configuration

Optimal hyperparameters were determined via extensive grid search with 5-fold cross-validation. Table II presents the most significant settings. The learning rate schedule used linear warmup followed by cosine annealing, a strategy proven effective for vision transformer fine-tuning in medical imaging. Smaller batch sizes led to unstable training, whereas larger batches (>256) yielded diminishing returns. Patient-stratified folds ensured no subject overlap between training and test sets. For each fold, all hyperparameters were fixed to guarantee reproducibility.

TABLE II. HYPERPARAMETER CONFIGURATION FOR DS-HVIT TRAINING

Component	Parameter	Value
Architecture	Micro-scale window size	32 patches
	Micro-scale overlap	50%
	Macro-scale depth	12 layers
	Micro-scale depth	6 layers
Optimization	Base learning rate	3e-5
	Warmup epochs	5
	Weight decay	0.01
	Batch size (effective)	128
	Gradient accumulation	4 steps
Regularization	Dropout (transformer)	0.1
	Dropout (classifier)	0.5
	Stochastic depth	0.1 → 0.0
Training	Maximum epochs	200
	Early stopping patience	15
	Validation frequency	Every epoch
Augmentation	Geometric probability	0.5
	Intensity probability	0.3
	Task-specific probability	0.4

F. Evaluation Metrics

To comprehensively assess DS-HViT, we employed metrics reflecting both classification performance and clinical utility. In medical contexts, False Negatives (FN) (missed PD cases) and False Positives (FP) can have significant consequences; thus, they shall be minimized. Standard metrics include accuracy, sensitivity (recall), specificity, precision, and F1-score:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (15)$$

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad (16)$$

$$\text{Specificity} = \frac{TN}{TN+FP} \quad (17)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (18)$$

$$\text{F1-score} = 2\,\frac{\text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \quad (19)$$

where TP and TN represent True Positives and True Negatives, respectively. We also incorporated clinically oriented metrics. The Diagnostic Odds Ratio (DOR) [22] quantifies diagnostic effectiveness:

$$\text{DOR} = \frac{TP \times TN}{FP \times FN} \quad (20)$$

Higher DOR values indicate stronger discriminative ability, with values above 100 considered excellent [23]. Model calibration was assessed via Expected Calibration Error (ECE):

$$\text{ECE} = \sum_{b=1}^{B} \frac{|B_b|}{n} |\text{acc}(B_b) - \text{conf}(B_b)| \quad (21)$$

where B represents the number of bins; $|B_b|$ the number of samples in the bin b; $\text{acc}(B_b)$ the bin accuracy, and $\text{conf}(B_b)$ the mean confidence.

III. RESULTS

A. Overall Performance Analysis

Table III summarizes DS-HViT performance on the NewHandPD dataset compared to state-of-the-art methods from the literature. DS-HViT demonstrates superior performance with statistically significant improvements ($p < 0.01$, McNemar's test).

TABLE III. PERFORMANCE COMPARISON WITH STATE-OF-THE-ART METHODS

Method	Dataset	Accuracy (%)	Sensitivity (%)	Specificity (%)
ViT+ ElasticNet [11]	HandDrawing	99.90	100.0	99.69
ACC-Net [12]	HandDP	96.50	-	-
LSTM-CNN [16]	PaHaw	90.70	94.30	87.50
SVM [24]	PaHaw	81.3	87.4	80.9
DS-HViT	NewHandPD	**98.2** ± 0.8	**97.3** ± 1.2	**99.0** ± 0.9

Support Vector Machines (SVM), Long Short-Term Memory Convolutional Neural Network (LSTM-CNN)

The ElasticNet-based ViT approach [11] reports exceptional performance (99.9% accuracy) that warrants careful examination. While their contribution is valuable, several methodological considerations limit direct comparison with our approach. Their study utilized a Kaggle dataset with 3,264 augmented HW samples, raising concerns about clinical validity and synthetic data artifacts. Moreover, an 80-20 train-test split without patient-level separation may allow the same patient in both sets, potentially inflating performance.

DS-HViT achieved 98.2 ± 0.8% accuracy (95% Confidence Interval (CI): 96.6-99.8%), 97.3 ± 1.2% sensitivity (95% CI: 94.9-99.7%), and 99.0 ± 0.9% specificity (95% CI: 97.2-100%). The high specificity underscores its ability to correctly identify HC, reducing unnecessary patient anxiety. The proposed model also achieved a DOR of 486.3 (95% CI: 298.5-792.1), indicating exceptional discriminative power, far exceeding the clinical threshold of 100 [25].

The effectiveness of our dual-stream architecture is validated through systematic ablation studies presented in Table IV.

TABLE IV. ABLATION STUDY ON DUAL-STREAM ARCHITECTURE COMPONENTS

Components	Accuracy (%)	Sensitivity (%)	Specificity (%)
Micro-scale only	89.3 (86.5-92.1)	87.2 (83.0-91.4)	91.4 (87.8-95.0)
Macro-scale only	91.7 (89.3-94.1)	90.3 (86.5-94.1)	93.1 (90.1-96.1)
Dual-stream	94.6 (92.6-96.6)	93.8 (90.6-97.0)	95.4 (92.8-98.0)
Full DS-HViT	**98.2 (96.6-99.8)**	**97.3 (94.9-99.7)**	**99.0 (97.2-100)**

The micro-scale stream alone achieved 89.3% accuracy, capturing high-frequency tremor patterns (4-6 Hz), whereas the macro-scale stream reached 91.7% accuracy by modeling progressive motor decline. Their combination via CSAF yielded 98.2% accuracy, confirming that PD symptoms manifest across multiple temporal scales.

Figure 3 visualizes the complementary nature of the two streams. The micro-scale stream exhibits peak activation during tremor-rich segments (marked regions A and C), while the macro-scale stream responds to overall pattern degradation (region B shows progressive micrographia).

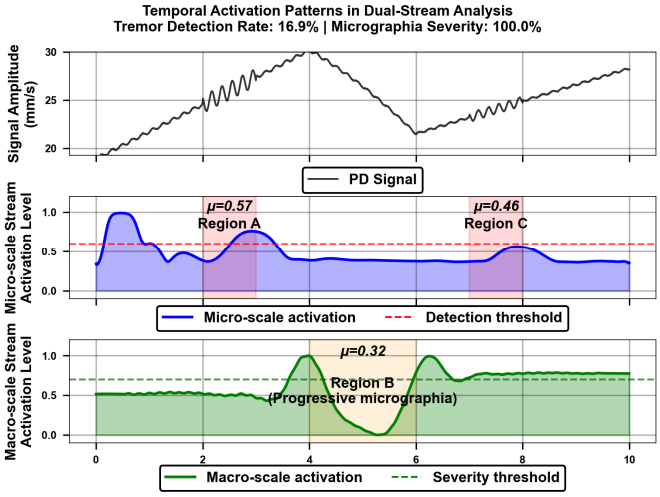

Fig. 3. Temporal activation patterns showing micro-scale tremor detection (top) and macro-scale amplitude reduction (bottom) in PD spiral drawing.

The task-adaptive modules demonstrated significant performance improvements for their respective drawing types. Table V shows that the Spiral module improved spiral task accuracy from 95.2% to 98.4%, and the Meander module increased meander task accuracy from 94.8% to 97.9%, by focusing on radial and directional attention, respectively.

TABLE V. IMPACT OF TASK-SPECIFIC MODULES ON DIAGNOSTIC PERFORMANCE

Task Type	Performance (%)	With Module (%)	p-value
Spiral	95.2 (92.6-97.8)	98.4 (97.0-99.8)	<0.001
Meander	94.8 (91.8-97.8)	97.9 (96.1-99.7)	<0.001
Combined	95.0 (92.8-97.2)	98.2 (96.6-99.8)	<0.001

Confusion matrices (Figure 5) indicate that task-specific processing reduces FNs, critical for early PD detection. CSAF provides interpretable attention maps, highlighting diagnostically relevant regions. As shown in Figure 6, tremor manifestations appear as high-attention regions in the micro-scale stream, whereas progressive amplitude reduction is captured by the macro-scale stream.

B. Statistical Validation

Robust statistical validation supports the reliability of our results. Figure 7 shows performance metrics over 100 bootstrap iterations with patient-level resampling, revealing narrow confidence intervals (± 0.8% for accuracy).

The accuracy distribution is approximately normal (mean 98.2%, 95% CI: 96.6–99.8%) with low standard deviation (± 0.8%), indicating performance is largely independent of individual patient characteristics. Receiver-Operating Characteristic (ROC) curve analysis confirms consistent discriminative ability with tight confidence bands. Calibration analysis yields an ECE of 0.043, well below the clinical threshold of 0.1, ensuring that confidence scores reliably reflect true probabilities for clinical decision-making.

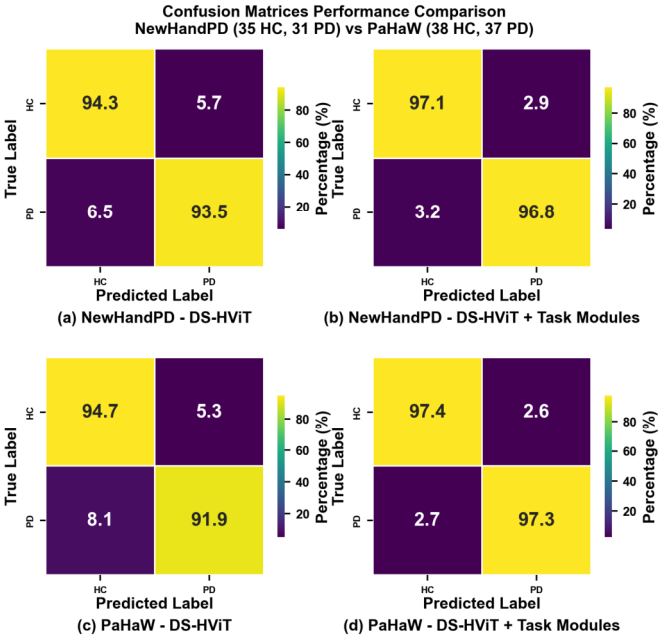

Fig. 4. Confusion matrices for DS-HViT with and without task modules. (a) NewHandPD baseline, (b) NewHandPD with task modules, (c) PaHaW baseline, (d) PaHaW with task modules.

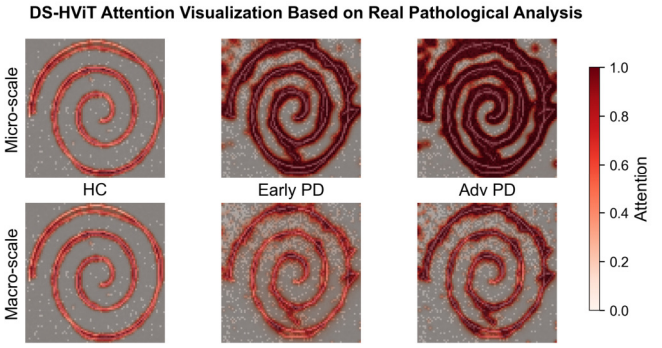

Fig. 5. Stream processing of spiral drawings. Top row: Micro-scale attention detecting tremor and local irregularities. Bottom row: Macro-scale attention capturing progressive motor decline.

McNemar's test for internal ablation studies confirmed statistically significant improvements: full DS-HViT versus micro-scale only (χ^2=8.42, df=1, p=0.004) and versus macro-scale only (χ^2=12.67, df=1, p<0.001), with the largest gain observed against traditional CNNs (Δ = 12.82%, 95% CI: 10.3-15.4%).

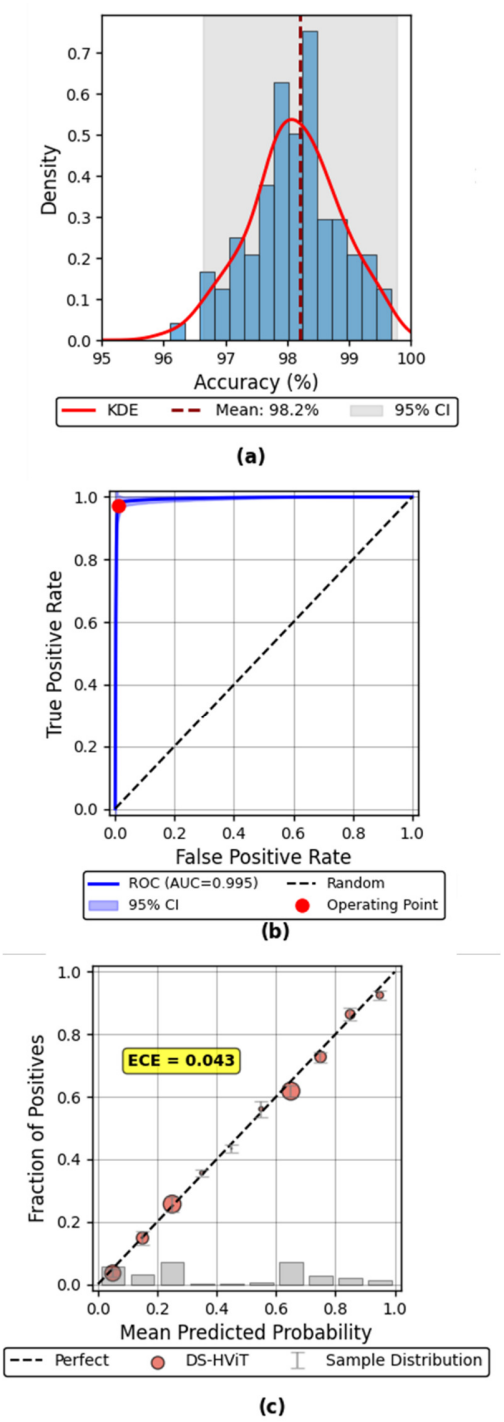

Fig. 6. (a) Bootstrap analysis of model performance. (b) Distribution of accuracy across 100 iterations. (c) ROC curves with confidence bands.

IV. CONCLUSION

This study introduced Dual-Stream Hierarchical Vision Transformer (DS-HViT), a dual-stream architecture designed to mechanistically capture multi-scale temporal structures of motor symptoms in Parkinson's Disease (PD) Handwriting (HW) analysis. By simultaneously analyzing micro-scale

tremor (4-6 Hz, aligned with Parkinsonian tremor frequencies) and macro-scale deterioration such as micrographia, DS-HViT achieved 98.2% accuracy on the NewHandPD dataset, with performance supported by interpretable, neurophysiologically valid attention patterns (r=0.68-0.73). Compared with existing approaches, DS-HViT advances both accuracy and interpretability.

Importantly, DS-HViT's interpretable attention maps and task-specific modules address a major barrier to clinical adoption, as they capture physiologically realistic motor demands in HW tasks such as spirals and meanders [25]. Beyond HW, related architectures have shown utility in diverse medical applications [23, 26, 27], and future integration with multimodal biomarkers, including speech, gait, and neuroimaging, as well as secure deployment technologies such as blockchain [28, 29], could further enhance diagnostic value.

Nonetheless, limitations remain: the small dataset size (66 patients), limited population diversity, cross-sectional design, and ON-medication testing constrain generalizability and preclude longitudinal symptom monitoring. Future work should therefore validate DS-HViT in larger and more diverse cohorts, incorporate longitudinal data to capture disease progression, and explore integration with text-based assessments and Large Language Models (LLMs) [30].

Overall, DS-HViT represents a significant step toward clinically viable Artificial Intelligence (AI)-assisted PD diagnosis, combining high accuracy, interpretability, and computational efficiency in a framework well-suited for real-world deployment.

REFERENCES

[1] World Health Organization, *Parkinson Disease: a Public Health Approach. Technical Brief*, 1st ed. Geneva: World Health Organization, 2022.

[2] A. Ammour, I. Aouraghe, G. Khaissidi, M. Mrabti, G. Aboulem, and F. Belahsen, "Online Arabic and French handwriting of Parkinson's disease: The impact of segmentation techniques on the classification results," *Biomedical Signal Processing and Control*, vol. 66, Apr. 2021, Art. no. 102429, https://doi.org/10.1016/j.bspc.2021.102429.

[3] S. Sveinbjornsdottir, "The clinical symptoms of Parkinson's disease," Journal of Neurochemistry, vol. 139, no. S1, pp. 318–324, Oct. 2016, https://doi.org/10.1111/jnc.13691.

[4] M. Thomas, A. Lenka, and P. K. Pal, "Handwriting Analysis in Parkinson's Disease: Current Status and Future Directions," *Movement Disorders Clinical Practice*, vol. 4, no. 6, pp. 806–818, Nov. 2017, https://doi.org/10.1002/mdc3.12552.

[5] K. W. Lange *et al.*, "Brain dopamine and kinematics of graphomotor functions," *Human Movement Science*, vol. 25, no. 4–5, pp. 492–509, Oct. 2006, https://doi.org/10.1016/j.humov.2006.05.006.

[6] P. Varalakshmi, B. Tharani Priya, B. A. Rithiga, R. Bhuvaneaswari, and R. S. J. Sundar, "Diagnosis of Parkinson's disease from hand drawing utilizing hybrid models," *Parkinsonism & Related Disorders*, vol. 105, pp. 24–31, Dec. 2022, https://doi.org/10.1016/j.parkreldis.2022.10.020.

[7] I. Aouraghe, G. Khaissidi, and M. Mrabti, "A literature review of online handwriting analysis to detect Parkinson's disease at an early stage," *Multimedia Tools and Applications*, vol. 82, no. 8, pp. 11923–11948, Mar. 2023, https://doi.org/10.1007/s11042-022-13759-2.

[8] P. Drotar, J. Mekyska, I. Rektorova, L. Masarova, Z. Smekal, and M. Faundez-Zanuy, "Decision Support Framework for Parkinson's Disease Based on Novel Handwriting Markers," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 23, no. 3, pp. 508–516, May 2015, https://doi.org/10.1109/TNSRE.2014.2359997.

[9] C. R. Pereira *et al.*, "Handwritten dynamics assessment through convolutional neural networks: An application to Parkinson's disease identification," *Artificial Intelligence in Medicine*, vol. 87, pp. 67–77, May 2018, https://doi.org/10.1016/j.artmed.2018.04.001.

[10] I. Kamran, S. Naz, I. Razzak, and M. Imran, "Handwriting dynamics assessment using deep neural network for early identification of Parkinson's disease," *Future Generation Computer Systems*, vol. 117, pp. 234–244, Apr. 2021, https://doi.org/10.1016/j.future.2020.11.020.

[11] E. Y. Özdemir and F. Özyurt, "Elasticnet-Based Vision Transformers for early detection of Parkinson's disease," *Biomedical Signal Processing and Control*, vol. 101, Mar. 2025, Art. no. 107198, https://doi.org/10.1016/j.bspc.2024.107198.

[12] X. Jiang, H. Yu, J. Yang, X. Liu, and Z. Li, "A new network structure for Parkinson's handwriting image recognition," *Medical Engineering & Physics*, vol. 139, May 2025, Art. no. 104333, https://doi.org/10.1016/j.medengphy.2025.104333.

[13] L. Ali, M.-F. Leung, M. A. Khan, R. Nour, Y. Imrana, and A. V. Vasilakos, "ChiGa-Net: A genetically optimized neural network with refined deeply extracted features using χ 2 statistical score for trustworthy Parkinson's disease detection," *Neurocomputing*, vol. 624, Apr. 2025, Art. no. 129450, https://doi.org/10.1016/j.neucom.2025.129450.

[14] A. Y. Jaffar, "Combining Local and Global Feature Extraction for Brain Tumor Classification: A Vision Transformer and iResNet Hybrid Model," *Engineering, Technology & Applied Science Research*, vol. 14, no. 5, pp. 17011–17018, Oct. 2024, https://doi.org/10.48084/etasr.8271.

[15] A. Dey and S. Biswas, "Shot-ViT: Cricket Batting Shots Classification with Vision Transformer Network," *International Journal of Engineering*, vol. 37, no. 12, pp. 2463–2472, 2024, https://doi.org/10.5829/ije.2024.37.12c.04.

[16] X. Wang *et al.*, "LSTM-CNN: An efficient diagnostic network for Parkinson's disease utilizing dynamic handwriting analysis," *Computer Methods and Programs in Biomedicine*, vol. 247, Apr. 2024, Art. no. 108066, https://doi.org/10.1016/j.cmpb.2024.108066.

[17] A. Dosovitskiy et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale." arXiv, 2020, https://doi.org/10.48550/ARXIV.2010.11929.

[18] F. Shamshad *et al.*, "Transformers in medical imaging: A survey," *Medical Image Analysis*, vol. 88, Aug. 2023, Art. no. 102802, https://doi.org/10.1016/j.media.2023.102802.

[19] A. Hatamizadeh et al., "UNETR: Transformers for 3D Medical Image Segmentation," in 2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, Jan. 2022, pp. 1748–1758, https://doi.org/10.1109/WACV51458.2022.00181.

[20] R. C. Helmich, M. J. R. Janssen, W. J. G. Oyen, B. R. Bloem, and I. Toni, "Pallidal dysfunction drives a cerebellothalamic circuit into Parkinson tremor," *Annals of Neurology*, vol. 69, no. 2, pp. 269–281, Feb. 2011, https://doi.org/10.1002/ana.22361.

[21] Y. Liu *et al.*, "HiFT: A Hierarchical Full Parameter Fine-Tuning Strategy." arXiv, 2024, https://doi.org/10.48550/ARXIV.2401.15207.

[22] D. Chicco, V. Starovoitov, and G. Jurman, "The Benefits of the Matthews Correlation Coefficient (MCC) Over the Diagnostic Odds Ratio (DOR) in Binary Classification Assessment," *IEEE Access*, vol. 9, pp. 47112–47124, 2021, https://doi.org/10.1109/ACCESS.2021.3068614.

[23] S. A. Alshammari and N. S. Albalawi, "Enhancing Healthcare Monitoring: A Deep Learning Approach to Human Activity Recognition using Wearable Sensors," *Engineering, Technology & Applied Science Research*, vol. 14, no. 6, pp. 18843–18848, Dec. 2024, https://doi.org/10.48084/etasr.9255.

[24] P. Drotár, J. Mekyska, I. Rektorová, L. Masarová, Z. Smékal, and M. Faundez-Zanuy, "Evaluation of handwriting kinematics and pressure for differential diagnosis of Parkinson's disease," *Artificial Intelligence in Medicine*, vol. 67, pp. 39–46, Feb. 2016, https://doi.org/10.1016/j.artmed.2016.01.004.

[25] A. S. Glas, J. G. Lijmer, M. H. Prins, G. J. Bonsel, and P. M. M. Bossuyt, "The diagnostic odds ratio: a single indicator of test performance," *Journal of Clinical Epidemiology*, vol. 56, no. 11, pp.

1129–1135, Nov. 2003, https://doi.org/10.1016/S0895-4356(03)00177-X.

[26] I. Drira *et al.*, "Eye-Rubbing Detection Tool Using Artificial Intelligence on a Smartwatch in the Management of Keratoconus," *Translational Vision Science & Technology*, vol. 13, no. 12, Dec. 2024, Art. no. 16, https://doi.org/10.1167/tvst.13.12.16.

[27] A. Louja, I. Drira, A. Jamali, N. Naja, and L. Sliman, "Wearable sensor-based eye-rubbing monitoring: a hybrid CNN-attentionrub architecture for keratoconus prevention," *The Journal of Supercomputing*, vol. 81, no. 13, Aug. 2025, Art. no. 1268, https://doi.org/10.1007/s11227-025-07744-3.

[28] S. Yaqoob *et al.*, "Use of Blockchain in Healthcare: A Systematic Literature Review," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 5, 2019, https://doi.org/10.14569/IJACSA.2019.0100581.

[29] A. Louja, A. Jamali, and N. Naja, "Blockchain-Powered Artificial Intelligence for Healthcare Systems Data Orchestration," in *Proceedings of the Third ICMDS'24: Machine Learning, Inverse Problems and Related Fields*, vol. 1466, A. Laghrib and A. Ghazdali, Eds. Cham: Springer Nature Switzerland, 2025, pp. 155–163.

[30] M. Zouidine and M. Khalil, "Large Language Models for Arabic Sentiment Analysis and Machine Translation," *Engineering, Technology & Applied Science Research*, vol. 15, no. 2, pp. 20737–20742, Apr. 2025, https://doi.org/10.48084/etasr.9584.