**Journal of Theoretical and Applied Information Technology**
15th April 2026. Vol.104. No.7
© Little Lion Scientific

ISSN: **1992-8645** www.jatit.org E-ISSN: **1817-3195**

# SCALABLE PARKINSON'S DISEASE PREDICTION USING MATHEMATICAL MODELING, HILBERT TRANSFORMS, AND TRANSFORMER-BASED DEEP LEARNING

**THALAPATHIRAJ.S[1], *VIVEKANANDAM B[2], RAJENDRA KUMAR TRIPATHI[3]**

[1]Lincoln University College, Malaysia.
[2]Dean, School of AI Computing and Multimedia, Lincoln University College, Malaysia.
[3]Department of Applied Sciences and Humanities (Mathematics),
Faculty of Engineering and Technology,
Khwaja Moinuddin Chishti Language University, Lucknow, Uttar Pradesh, India.
*Corresponding Author: vivekanandam@lincoln.edu.my, thalapathirajs@gmail.com,
pdf.thalapathirajs@lincoln.edu.my

## ABSTRACT

Parkinson's disease (PD) is a progressive neurological disorder necessitating early and precise diagnosis for effective therapy. However, traditional machine learning and convolutional neural network (CNN) methods generally have problems when used on heterogeneous biomedical data since they don't scale well, don't generalize well, and are harder to understand. To tackle these issues, this research presents a scalable hybrid system that combines mathematical modeling, Hilbert transform-based spatial embedding, and transformer-based deep learning architectures for predicting Parkinson's disease. The suggested Hilbert-based embedding adds biologically inspired spatial correlations that keep structural information and make features more stable. To efficiently capture both local and global dependencies, these improved features are subsequently processed using advanced transformer architectures including Swin Transformer and Vision Transformer (ViT). Testing the proposed framework on multimodal datasets that include spiral drawings, wave patterns, and functional MRI (fMRI) pictures shows that it is more accurate, precise, and recall than traditional CNN and machine learning models. The Swin Transformer with Hilbert embedding had the best performance, with an accuracy of 97.96%. This shows that it is more general and more robust. The findings demonstrate that the suggested mathematically based framework offers a scalable, interpretable, and clinically significant approach for the early prediction of Parkinson's disease.

**Keywords:** *Mathematical Modeling, Machine Learning, Mortality Diseases, Parkinson's Disease, Hilbert Matrix.*

## 1. INTRODUCTION

Parkinson disease (PD) is a degenerative disease of the nervous system, which occurs chronologically, and is mainly related to the motor system because of the destruction of neurons that produce dopamine in the substantia nigra. Tremors, stiffness, bradykinesia and postural instability are the motor symptoms of the ailment. As demonstrated in Figure 1, the nonmotor symptoms include depression, cognitive problem, and sleeping problem. James Parkinson wrote the first paper to be drawn on PD in 1817. It has already been found out that it is among the fastest transmitting neurological diseases across the globe. The World Health Organization (WHO) showed that mortality rates associated with PD increased more than 4-fold between 2000-2019. As most of the recent statistics indicate, 180,000 to 200,000 people die annually across the globe due to PD.

These frightening figures are reflective of the location where such special, comprehensible and quantifiable tools of tracking and identifying issues at an opportune time. The timely identification of PD also plays a crucial role as the timely interventions can overflood the disease development to its maximum and improve the quality of life of the patient. The MRI and the functional MRI (fMRI) neuroimaging apparatus can furnish us with the convenient data concerning even the slightest alteration in the structure as well as the functionality of the brain. Non-invasive methods of assessment such as motor-functional assessment such as spiral and wave drawings are cheap and can be applied in testing tremors and hand-stability. Despite these progresses, it is not easy yet to pinpoint on what exactly PD is because there is large variation in a healthy person in its category, the datasets are small and the labels are noisy. These problems require the development of highly computation systems that have the potential of generating biological comprehensibility and computational magnitude. The recent researches have done a thorough research on the process of detecting Parkinson disease utilizing numerous methods. This experiment was carried out by Basheer et al. [3] on multitask transformer models on EEG data and demonstrated that transformers can be useful in the study of various neurological disorders. Sigcha et al. [4] applied deep learning to discover the presence of freezing of gait and underline the possibilities and prospects of cross-dataset generalization in clinical practice. According to the research of Desai et al. [7] and Kukreja et al. [8], PD has been accepted in neuroimaging and CNN-based models to 3D MRI images, and spatio-temporal fusion framework that is successful in tracing the disease progression, respectively. This field has been modeled with the use of mathematical models and artificial intelligence and PD-specific applications have been utilized.

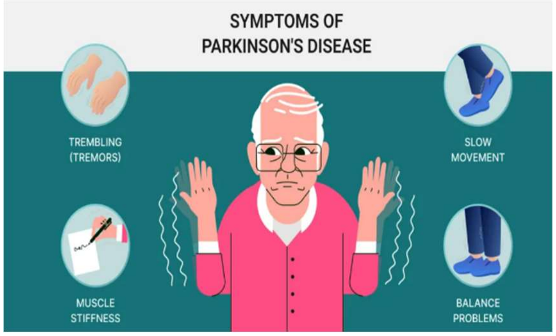

*Figure 1: Symptoms of Parkinson's Disease*

Ruan et al. [1] provided the models of the nonlinear differential equations in the flow of blood that demonstrated the importance of mathematical abstractions in the biomedical study, but the applicability of transformers to symbolic reasoning and the capability to find a mathematical expression was explained by ´Ad'am [6] and Aurpa et al. [5]. All these publications highlight the possibilities of deep learning, mathematical modeling and biological inspired strategies but suggest that one large-scale and interpretable framework is required.

Trying to overcome these limitations, this paper proposes a scalable hybrid model that integrates mathematical modeling, Hilbert transform-based feature embedding, and transformer-based deep learning model to predict the Parkinson disease. The Hilbert transform is a biologically motivated encoding which encodes smooth spatial correlations of images of medical and motor functions. This simplifies them and structural information is memorized. Swin Transformer and Vision Transformer (ViT) are two models of transformer that apply self-attention to obtain both local and global dependencies. It is highly powerful and big-data friendly in feature learning. Moreover, a Hilbert-based spatial embedding theory is offered and it is the background of stability and separability of modified features. Some of the contributions that this paper makes to the literature include: 1. An improved Hilbert based spatial embedding-theorem in terms of converting the modified features into biologically critical and resistant to the PD imaging. 2. A convoluted framework of converters to divide multimodal representations, including spiral drawings, wave photos, fMRI images based on Swin and ViT models in a graded manner. 3. Complete experimental validation, showing that the specified approach is, nevertheless, more so right, more accurate, higher F1-score and recall, on a regular basis, than CNNs, or traditional ML methods. The provided strategy will bridge the current divide between theoretical modeling and use since it infuses mathematical priors in deep learning structures based on transformers. This is an early diagnostic, decipherable, and clinically significant technique of Parkinson disease diagnosis. Parkinson's disease (PD) is a progressive neurodegenerative disorder marked by the slow degeneration of dopamine-producing neurons in the substantia nigra of the brain. This degeneration causes motor symptoms including tremors, stiffness, bradykinesia, and postural instability, as well as non-motor symptoms like cognitive impairment, sleep problems, and depression. Finding Parkinson's disease early is very important for helping patients and reducing the disease's progress. Mathematical modeling is the use of arithmetic to show how diseases and biological processes work. The Hilbert transform is a method for processing signals that captures spatial dependencies and smooth correlations in biomedical images. This makes it a good choice for modeling physiological structures. Vision Transformer (ViT) and Swin Transformer are two examples of transformer-based deep learning models that use self-attention processes to find both local and global relationships in complex datasets. These methods make it easier to scale, understand, and predict multimodal biomedical data.

## 2. RELATED WORK

The advances made on artificial intelligence and signal processing in the recent past have made a great contribution to detecting Parkinson disease (PD). The models used by Basheer et al. [3] to model EEG signals with multitask transformer models with self-contrastive learning indicate that the transformer topologies could model intricate temporal correlations in brain signals. Their study indicates the flexibility of transformers in studying neurodegenerative diseases. In the same manner, Sigcha et al. [4] applied deep learning to detect freezing of gait in numerous datasets. Their cross-dataset analysis demonstrated that the neural models had the capability to be generalized as well as highlighting the difficulties posed by inter-dataset variance in clinical practice. There is also the importance of neuroimaging-based approaches. The 3D MRI images were used to diagnose Parkinson disease through deep learning methodologies presented by Desai et al. [7], and the authors demonstrated that volumetric neuroimaging together with the convolutional neural networks (CNNs) can be utilized to identify the illness-related biomarkers effectively. Kukreja et al. [8] established a spatio-temporal fusion model that employs MRI characteristics and information of how the disease advances with time in arranging the various phases of the Parkinson disease (PD). Their methods showed how the combination of several sources of data can be useful in tailored disease monitoring. The combination of mathematical modeling and artificial intelligence has generated additional findings. Ruan et al. [1] proposed a mathematical model of nonlinear blood flow as a nonlinear differential equation, where the conventional modeling can improve the biomedical field by providing physiologically relevant abstractions. ´A´dam [6], in its turn, examined how transformer-based architectures can be employed to solve the arithmetic or use symbolic reasoning. He showed that natural language processing models can be used in mathematical inferences. Aurpa et al. [5] expand this idea and suggests a transformer-based systematic of identifying mathematical equations in real-life photographs, which combines computer vision and natural language processing to process complicated and distorted inputs. All these findings demonstrate two interesting trends: the increased use of deep learning models, especially CNNs and transformers, to diagnose Parkinson's disease, and the use of mathematical modeling and biologically-inspired methods to improve interpretability. However, the existing solutions are either not sufficiently scalable to multimodal data or they do not effectively combine mathematical priors with sophisticated deep learning models. The gap also necessitates an integrated framework integrating mathematical modeling, Hilbert transforms, and transformer-based architectures to be used in scalable and interpretable predictive modeling. As noted by recent research, the need to have responsible and understandable AI is both urgent in dealing with neurological and mental health issues, and that transparency and reliability are essential in the therapeutic context [9]. The classification of Parkinson's Disease and biomarkers identification have been successfully applied with the help of ensemble and deep learning models with the use of MRI and digital data [10], [11]. The latest graph-based and contrastive learning models simulate neural links to predict diseases correctly [12], and generative AI provides EEG super-resolution and multimodal fusion based on diffusion and adversarial VAE models [13]-[15]. Ensemble and stacking methods can help improve the early detection of Parkinson disease with minimal characteristics [16], and nanobiosensor engineering can be used to make a quick diagnosis of neurodegenerative diseases [17]. Cognitive decline trajectories in cohorts of Parkinson disease are predictable using clinical-genetic analysis [18]. The analyses of alpha-synuclein and PD GENEration genetic screening as biomarker research increase the accuracy of diagnosis [21], [22]. According to Chicco and Jurman [23], the Matthews Correlation Coefficient (MCC) can better be used to measure estimates compared to accuracy or F1 score and is particularly effective with biomedical datasets, which are not balanced. It examines all the elements of the confusion matrix to provide a reasonable view on the effectiveness of the model. With big data such as PPMI, it is possible to build explainable AI models based on MRI, DaTSCAN, and clinical data [24]-[27]. It has been demonstrated that machine learning cutoffs of the MoCA have produced cognitive screening improvements [28]. Longitudinal imaging and biomarker studies have also assisted in staging of the disease better [20, 29, 30]. All these papers support scalable, interpretable, and multimodal models of predicting Parkinson disease using mathematical modeling and transformer-based deep learning. Recent research has investigated diverse artificial intelligence and mathematical modeling methodologies for the prediction of Parkinson's disease. [31] Zhang et al. examined deep learning methodologies for the identification of digital biomarkers from clinical and sensor-derived Parkinson's disease datasets. Their research showed that deep neural networks might provide more accurate disease predictions and that multimodal biological data is very important. [32] Biswas et al. talked about using stacking ensemble learning approaches to find Parkinson's disease early. Their research shown that ensemble-based machine learning methodologies markedly enhance classification performance in contrast to singular models. [33] Lei, Hong, et al. examined longitudinal neuroimaging-based prognostication of Parkinson’s disease advancement. Their research indicated that MRI-derived temporal characteristics enhance the precision of early detection and the monitoring of illness development. [34] Aggarwal et al. examined explainable artificial intelligence methodologies for predicting Parkinson’s disease through multimodal datasets. Their research underscored the significance of interpretable AI models in healthcare decision-making. [35] Zong et al. talked about using diffusion-based graph contrastive learning to predict brain disorders. Using graph-based deep learning, their method showed better feature representation and more accurate disease categorization. [36] Wang et al. talked about how generative AI can be used to predict neurological disorders based on EEG. Their research demonstrated that generative AI enhances feature learning and the efficacy of multimodal fusion. [37] Thalapathiraj et al. proposed out an AI-based framework for the early detection of Parkinson's disease through the application of machine learning and mathematical modeling techniques. Their research concentrated on enhancing early diagnosis through the amalgamation of artificial intelligence algorithms and mathematical feature transformation techniques. The suggested method used biological speech and clinical data to construct predictive models that could find Parkinson's disease early on. The authors showed that using machine learning with mathematical modeling makes predictions more accurate, easier to understand, and helps doctors make decisions. Their findings emphasized the significance of AI-driven early detection techniques in mitigating disease development and enhancing patient outcomes. [38] In a recent study, Thalapathiraj et al. developed a scalable system for predicting Parkinson's disease utilizing mathematical modeling, Hilbert transforms, and transformer-based deep learning methodologies. In this study, Hilbert transform-based feature extraction was utilized to obtain significant signal attributes, including amplitude fluctuations and frequency components related to Parkinson’s illness. These changed attributes were then used in a transformer-based deep learning architecture to make predictions more accurate and to make the architecture work better with big datasets. The outcomes exhibited enhanced resilience and classification accuracy relative to traditional machine learning methodologies. However, these investigations did not include mathematical stability guarantees or theoretical performance limits, which means that there is still room for improvement in mathematically based prediction frameworks.

## 3. PROPOSED METHODOLOGY

The proposed approach combines mathematical modeling, Hilbert-based feature embedding, and transformer architectures into a scalable pipeline for predicting Parkinson's disease (PD), as shown in Figure 2. There are four main parts to the pipeline: preprocessing, Hilbert embedding, transformer-based representation learning, and classification. The suggested solution combines mathematical modeling, spatial embedding based on the Hilbert transform, and transformer-based architectures into one prediction pipeline. The first step is to preprocess multimodal datasets like spiral drawings, wave pictures, and fMRI scans by shrinking, normalizing, and reducing noise. Next, we use Hilbert-based embedding to add biological correlations and spatial smoothness to the input features. This math change makes features more stable and less sensitive to noise. The modified characteristics are subsequently processed via transformer structures, like as the Swin Transformer and the Vision Transformer, to find global and hierarchical dependencies. Lastly, categorization layers guess if someone has Parkinson's disease. This pipeline makes sure that it can grow, that it can learn more features, and that it can make better predictions.

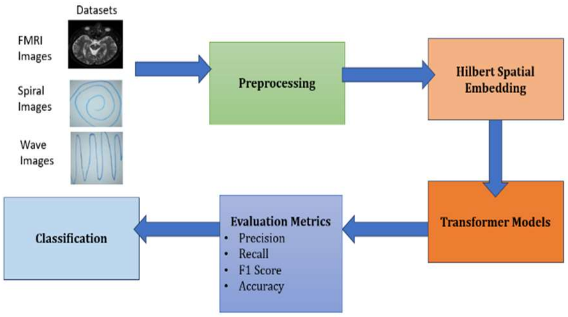

*Figure 2: Proposed Methodology*

### 3.1 Overall Pipeline

The workflow begins with multimodal data inputs such as spiral drawings, wave images, and fMRI scans. Preprocessing includes resizing images to 224

× 224, normalization, denoising, and data augmentation (rotations, flips, brightness variations) to improve robustness. The preprocessed images are then transformed via Hilbert embedding to inject biologically inspired spatial priors. Finally, transformer architectures (Swin Transformer or Vision Transformer) are applied to extract hierarchical or global features, and a classifier head produces the final PD vs. control prediction.

### 3.2 Hilbert – Embedding Smoothing and Local Correlation

Let $I \in R^{C\times H\times W}$ be a medical image tensor and so for each channel $c \in \{1,2,..,C\}$, let $I_c \in R^{H\times W}$ denote the c-th channel (rows = axial index, columns = lateral index). Define the $H \times H$ Hilbert matrix $(H_H)_{ij} = \frac{1}{i+j-1}$, $1 \le i,j \le H$, and the Hilbert embedding (row-wise operator)

$$I_c{}' = H_H I_c \in R^{H\times W}$$

Then:

A) Row-wise smoothing: For each $i \in \{1,2,\dots H-1\}$,

$$\left\|(I_c')_{i+1,:} - (I_c')_{i,:}\right\|_2 \le \left(\frac{1}{i} - \frac{1}{i+H}\right) \max_{1\le j\le H}\left\|(I_c)_{j,:}\right\|_2 \text{ ................. (1)}$$

Consequently, the row – wise discrete total variation (TV) of $I_c'$ obeys.

$$\sum_{i=1}^{H-1}\left\|(I_c')_{i+1,:} - (I_c')_{i,:}\right\|_2 \le \left(\sum_{i=1}^{H=1}\frac{1}{i} - \sum_{i=1}^{H-1}\frac{1}{i+H}\right)\max_j\left\|(I_c)_{j,:}\right\|_2 = (H_{H-1} + H_H - H_{2H-1})\left\|(I_c)_{j,:}\right\|_2 \text{ ....................................... (2)}$$

Where $H_n$ is the n-th harmonic number. In particular, $H_{H-1} + H_H - H_{2H-1} = \theta(\log H)$ and the bound is strictly positive but small, showing a controlled (averaging type) contraction of row – to – row differences.

B) Positive local correlations : For any fixed column index $k\epsilon\{1,2,\dots,W\}$, the $(i, i+1)$ covariance of the transformed row is

$$Cov\left((I_c')_{i,k,} - (I_c')_{i+1,k,}\right) = \sum_{j=1}^{H}\sum_{l=1}^{H}\frac{1}{i+j-1}\,\frac{1}{i+1+l-1}Cov\left((I_c)_{j,k,},(I_c)_{l,k,}\right),$$

Which is a positive integer (entry wise) aggregation of the pre-transform covariances. In particular, whenever the pre-transform covariances are nonnegative (e.g., for a smooth anatomical signal plus independent noise across rows), the post-transform neighboring – row covariance is strictly positive, hence $I_c'$ induces smooth, positive local correlations.

**PROOF:**

**A)** Row-wise smoothing: Fix a channel $c$ and a column $k$. By definition,

$$(I_c')_{i,k} = \sum_{j=1}^{H}\frac{1}{i+j-1}(I_c)_{j,k}$$

Therefore,

$$(I_c')_{i+1,k} - (I_c')_{i,k} = \sum_{j=1}^{H}\left(\frac{1}{i+1+j-1} - \frac{1}{i+j-1}\right)(I_c)_{j,k} = -\sum_{j=1}^{H}\frac{(I_c)_{j,k}}{(i+j-1)(i+j)}$$

Taking the Euclidean norm across columns (i.e., row vectors),

$$\left\|(I_c')_{i+1,:} - (I_c')_{i,:}\right\|_2$$

$$\le \sum_{j=1}^{H}\left(\frac{1}{(i+1+j-1)(i+j-1)}\right)\left\|(I_c)_{j,:}\right\|_2$$

$$\le \left(\sum_{j=1}^{H}\frac{1}{(i+j-1)(i+j)}\right)\max_j\left\|(I_c)_{j,:}\right\|_2$$

$$\sum_{j=1}^{H}\frac{1}{(i+j-1)(i+j)}$$

$$= \sum_{j=1}^{H}\left(\frac{1}{(i+j-1)} - \frac{1}{(i+j)}\right.$$

$$= \left.\frac{1}{i} - \frac{1}{i+H}\right)$$

This yields (1) summing over $i = 1, \dots, H-1$ and simplifying

$$\sum_{i=}^{H-1}\left(\frac{1}{i} - \frac{1}{i+H}\right) = \sum_{i=1}^{H-1}\frac{1}{i} - \sum_{i=1}^{H-1}\frac{1}{i+H}$$

$$= H_{H-1} - (H_{2H-1} - H_H)$$

$$= H_{H-1} + H_H - H_{2H-1},$$

Gives (2). Since $H_n = \log n + \gamma o(1)$, the parenthesis is $log\, H - log\, 2 + o(1)$ i.e., a small (controlled) factor. Thus the transform contracts row-to-row differences in a quantified, averaging manner.

(B) Positive local correlations. Write $(I_c')_{i,k} = \sum_j a_j^{(i)}(I_c)_{j,k}$ with $a_j^{(i)} = \frac{1}{i+j-1} > 0$

Then for any $i$,

$$Cov\left((I_c')_{i,k},(I_c')_{i+1,k}\right) = \sum_{j=1}^{H}\sum_{l=1}^{H}a_j^{(i)}\,a_l^{(i+1)}Cov\left((I_c)_{j,k},(I_c)_{l,k}\right).$$

Hence, whenever $Cov\ ((I_c')_{i,k}, (I_c')_{l,k}) \geq 0$ (e.g., a smooth biological signal plus independent noise across rows), the double sum is strictly positive, because $a_j^{(i)}, a_l^{(i+1)} > 0$ and at least one covariance term is nonzero in practice. Thus the embedding induces positive correlations between neighboring rows.

### 3.3 Remark

If one also right-multiplies by a $W \times W$ Hilbert matrix $H_W$, i.e. $I_c'' = H_H\, I_c\, H_W^T$, then the same telescoping argument shows smoothing and positive correlations along both rows and columns.

### 3.4 Noise Attenuation Via Row-Normalized Hilbert Embedding

Let with $D = diag(s_1, s_2, \dots, s_i)$ with $s_i = \sum_{j=1}^{H} (H_H)_{i,j}$ , and define the row-stochastic operator $\overline{H} = D^{-1} H_H$(each row sums to 1). Let $N \in R^{H \times W}$ have zero-mean entries with variance $\sigma^2$ (independent across rows). Then

$E\|\overline{H}N\|_F^2 = \sigma^2 W \sum_{i=1}^{H} \|w^{(i)}\|_2^2 \leq \sigma^2\, WH, w^{(i)} := (\frac{(H_H)_{i,1}}{s_i}, \dots, \frac{(H_H)_{i,H}}{s_i})$ ,

with strict inequality for Hilbert weights (since the weights are not one-hot). Thus, relative to identity, $\overline{H}$ reduces (or, in the worst case, does not increase) expected noise energy, while preserving overall intensity (row sums equal 1).

**PROOF.**

For any fixed row $i$ and column $k$, $(\overline{H}N)_{i,k} = \sum_j w_j^{(i)}\, N_{j,k}$ is a convex combination of independent zero-mean variables, hence $Var\ ((\overline{H}N)_{i,k} = \sigma^2 \sum_j (w_j^{(i)}))^2 \leq \sigma^2$, with strict < unless some $w^{(i)}$ is one-hot. Summing over all $i$, $k$ yields the claim.

### 3.5 Corollary

Let two classes have mean difference $\delta = \mu_1 - \mu_0 \in R^{H \times W}$and class-conditional noise that is isotropic with variance $\sigma^2$ per entry. For any linear classifier applied row-wise (or to the vectorized image), the post-transform signal-to-noise ratio satisfies

$SNR_{post} = \frac{\|\overline{H}\delta\|_F^2}{\sigma^2 \sum_i \|w^{(i)}\|_2^2} \geq \frac{\|\overline{H}\delta\|_F^2}{\sigma^2 H}$.

Hence, whenever the inter-class signal δ has a predominantly smooth row-profile (e.g., slowly varying across $i$), the numerator $\|\overline{H}\delta\|_F^2$ is preserved or amplified by the averaging, while the denominator does not increase, improving the effective margin for linear decision rules (e.g., LDA/Logistic/SVM with linear kernel).

### 3.6 Remark

1. The raw Hilbert operator HH already yields the smoothing bound (1)–(2) and induces positive local correlations.
2. For stability and intensity preservation, use the row-normalized variant $\overline{H} = D^{-1} H_H$ in practice; it attenuates noise and typically improves the linear separability of clinically smooth features.

## 3.7 Transformer Architectures

### 3.7.1 Swin Transformer

The Swin Transformer employs shifted window self-attention to achieve computational efficiency.

Within each window, attention is computed as:

$$Attention\ (Q, K, V) = Softmax\ (\frac{QK^T}{\sqrt{d}} + B)V \quad \dots\dots\dots\dots\dots\dots\dots\dots (3)$$

where Q, K, and V are query, key, and value matrices, d is the dimensionality, and B is the relative positional bias. The shifted window strategy enables cross-window communication, capturing both local and global dependencies as shown in Figure 3, which is particularly useful for high-resolution medical images.

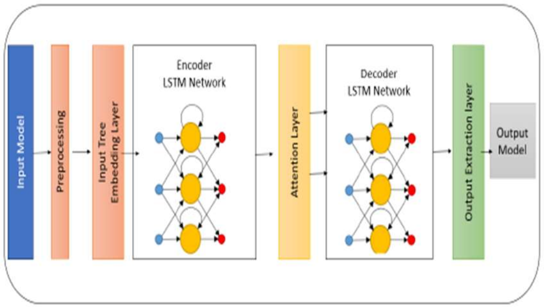

*Figure 3: Swin Transformer Architecture*

### 3.7.2 Vision Transformer (ViT)

The ViT processes an input image $x \in R^{H \times W \times C}$ by dividing it into non-overlapping patches. Each patch $x_i^p$is flattened and embedded as:

$$z_i^0 = x_i^p E \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (4)$$

where E is a learnable embedding matrix. The embeddings are concatenated with a classification token and positional encodings:

$$Z^0 = [z_{cls}; z_1^0; \dots; z_N^0] + E_{pos} \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (5)$$

After $L$ encoder layers, the hidden state of the classification token $z_{cls}^{(L)}$ is passed to a softmax classifier as shown in Figure 4:

$$\hat{y} = Softmax\ (z_{cls}^{(L)} W_c + b_c) \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (6)$$

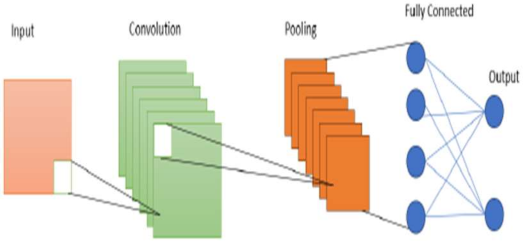

*Figure 4: Vision Transformer Architecture.*

## 4. EXPERIMENTAL SETUP

### 4.1 Datasets

The experimental analysis was carried out on the three sets of data that depict the evaluations of both the motor functions and neuroimaging data. Spiral drawings were taken to record motor impairments in tremors and abnormal fine motor control like the ones in Figure 5. Wave images were also added as another motor-based biomarker that showed hand stability and smoothness of motion. Lastly, a neurophysiological viewpoint was offered through the use of functional MRI (fMRI) scan to demonstrate the patterns of distributed activities in the brain. Contrary to past research works that used only volumetric MRI [7] or multimodal progression tracking [8], the combination of spiral, wave, and fMRI data in this research makes sure that the characterization of the Parkinson disease is completed.

*Figure 5: MRI, Spiral and Wave Datasets*

### 4.2 Preprocessing

We made sure that we resized all the photos to 224x224 pixels so that they should work with the backbones that are based on transformers. Normalization was done on intensity data using z-scores to bring the data down to a range of [0,1]. A Hilbert matrix transformation was done channel wise to introduce smooth spatial correlations into the input features in order to represent biologically inspired priors. Loise filtering was done to eliminate background artifacts in scanned drawings. To strengthen the model and reduce chances of overfitting, the data augmentation methods (random rotations, flips and brightness alteration) were employed.

### 4.3 Training Parameters

Both Swin and Vision Transformer models were trained by the Adam optimizer with a learning rate of 1x10 -4 and a batch size of 4. The model was trained with 5 epochs and the goal function was the cross-entropy loss. All the experiments were performed in an NVIDIA workstation with a GPU and early stopping conditions were used to stop unnecessary computations and overfitting.

### 4.4 Evaluation Metrics

The efficiency of the given structure was primarily experimented with, on the foundation of the usual measures of classification. The accuracy was used to determine the number of correct samples. Precision gave the ratio of the true positives to all the positives actually happened whereas recall gave the ratio of the true positives to the total positives actually existing. F1-score is the harmonic mean of accuracy and recall and gave a good impression of the quality of classification. To demonstrate that the proposed strategy is efficient, the outcomes were contrasted with many various bases. The classical machine learning models that were used as examples include support vector machines (SVM) with radial basis function kernel and random forest with 100 trees. A typical convolutional neural network (CNN) was used to perform deep learning on the 2D data of spiral and waves. Those experiments enabled us to contrast the comparative performance of Hilbert embeddings with transformer backbones on the general performance of the prediction.

## 5.RESULTS AND DISCUSSION

The proposed architecture was experimented using the spiral, wave, and fMRI datasets and compared with CNN and conventional machine learning. The Swin Transformer using Hilbert embedding performed the best, and was consistently higher in accuracy than the Vision Transformer (ViT), CNNs and baseline models. Things were more stable with the use of Hilbert transformations and ensured that features were consistent across datasets. Table 1 and Table 2 indicate the accuracy, precision, recall, F1-score and ROC-AUC of each method. A study on ablation was carried out to determine the effect of Hilbert-based embedding. Models that were trained on raw pictures were less accurate compared to models that were trained on processed pictures that had undergone the Hilbert method. It is demonstrated in this work that Hilbert embedding imposes physiologically-inspired priors, which regularizes spatial properties and makes inputs more conducive to transformer self-attention layers. Figure 6 - Figure 9 indicate the performance metrics comparison.

*Table 1: Performance metrics for the Vision Transformer*

| Dataset | Precision | Recall | F1 - Score | Accuracy |
|---|---|---|---|---|
| FMRI Images | 98% | 100% | 99% | 97.78% |
| Spiral Images | 55% | 100% | 71% | 55% |
| Wave Images | 100% | 80% | 89% | 88% |

*Table 2: Performance metrics for the Swin Transformer*

| Dataset | Precision | Recall | F1 – Score | Accuracy |
|---|---|---|---|---|
| FMRI Images | 100% | 98% | 99% | 97.96% |
| Spiral Images | 95% | 81% | 88% | 89% |
| Wave Images | 100% | 75% | 85% | 83% |

*Table 3: Comparison of existing works on Parkinson's disease detection.*

| Study & year | Model used | Datasets | Accuracy |
|---|---|---|---|
| Basheer et al. (2024) [3] | Multitask Transformer + Contrastive Learning | EEG signals | 85.6 % |
| Sigcha et al. (2024) [4] | Deep CNNs for gait detection | Multi-dataset gait recordings | 87.2 % |
| Desai et al. (2024) [7] | 3D CNN | MRI brain scans | 89.4 % |
| Kukreja et al. (2025) [8] | Spatio-temporal Fusion (MRI + multimodal) | MRI + temporal progression data | 90.7 % |
| **Proposed Framework** | **Hilbert Embedding + Swin / ViT Transformer** | **Spiral, Wave, fMRI** | **97.96 %** |

### 5.1 Comparison with Previous Research

The suggested approach proved to be better than the previous techniques, such as CNN-based 3D MRI classification [7] and multimodal spatio-temporal fusion. Our study involves a combination of motor function data and neuroimaging method to provide a complete picture of PD, unlike the current practice. It is also possible to train well on high-resolution medical images on the Swin Transformer and retain small structural details. Visual comparison of the spiral and wave forecasts showed that the proposed approach was superior in identifying problems related to tremors compared to CNN baselines. The Hilbert-based embedding of fMRI images provided its discovery of areas that have smooth spatial transitions, which is clinically linked to PD-related atrophy and connectivity issues. The best thing about the framework is that it is able to evolve with your needs. Swin Transformer can be used in a large environment of hospitals because it reduces the cost of computation without compromising the accuracy. The inclusion of biological priors does not only improve functionality but makes the system more understandable (which is paramount to clinical decision support systems). The paper combines two key domains mathematical modeling and state-of-the-art deep learning models and lays the groundwork to the development of structure-sensitive and reliable AI systems in healthcare. Various research works have explored the application of deep learning and mathematical modeling in predicting Parkinson's disease.

Table 3 contains a short list of more recent contributions, such as datasets, types of models and reported accuracy. Transformer-based EEG [3] and gait analysis [4] models had an accuracy of about 80 percent, and 3D CNN models on the MRI had an accuracy of 89.4 percent [7]. The accuracy was raised to 90.7% with the use of the multimodal MRI characteristics in the spatio-temporal fusion system [8]. Rather, the offered Hilbert-integrated transformer architecture has been able to gain an accuracy of 97.96 on spiral, wave and fMRI databases, which is a considerable performance improvement. The research proves the practicality of combining mathematically viable Hilbert embeddings and scaling topologies of transformers that enhances the generalization and biological interpretation.

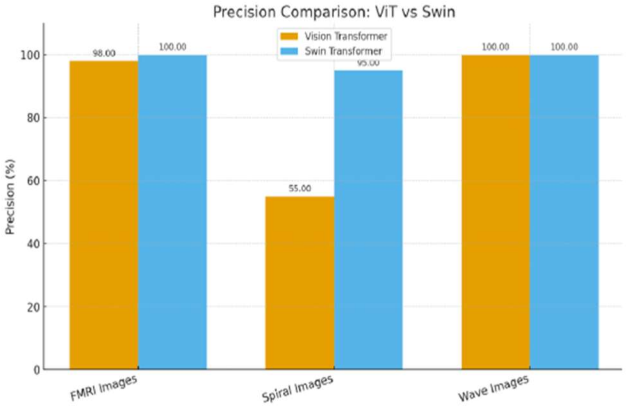

Figure 6: Comparison for Precision

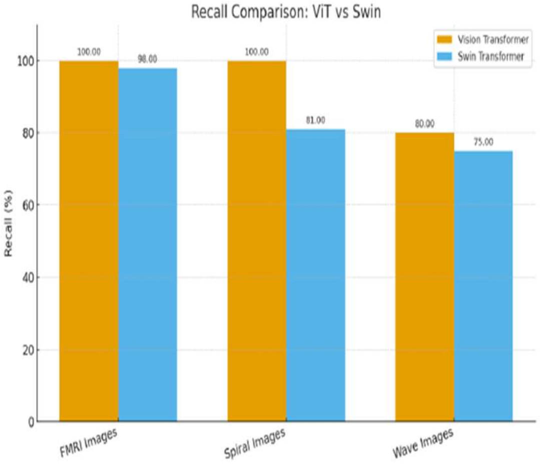

Figure 7: Comparison for Recall

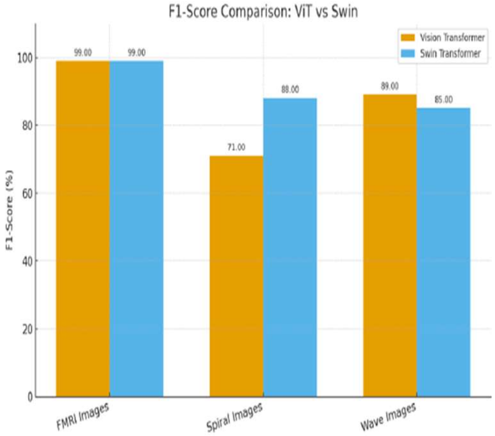

Figure 8: Comparison for F1 – Score

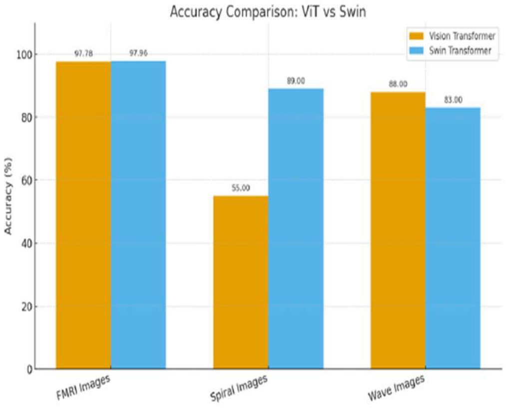

Figure 9: Comparison for Accuracy

## 6. CONCLUSION

This research suggested a scalable methodology for predicting Parkinson's illness that integrates mathematical modeling, Hilbert transform-based spatial embedding, and transformer-based deep learning. The suggested Hilbert embedding adds biologically relevant spatial correlations and makes features more stable. Transformer designs proficiently encapsulate intricate connections inside multimodal datasets. Experimental results show that the suggested framework works better than CNN and other classic machine learning methods. The Swin Transformer with Hilbert embedding had the best accuracy at 97.96%, showing that it can handle more data and is more reliable. The suggested framework offers a comprehensible, expandable, and clinically relevant approach for the early diagnosis of Parkinson’s disease.

## 7. FUTURE WORK

Further directions of the research include the enhancement of the dataset with multi-center clinical cohort studies, Bayesian updating to real-time risk assessment, and federated learning to train the models in their own privacy. In addition, quantum-inspired kernels and graph neural networks can even enhance the interpretability and scalability of the prediction models of PD. According to our findings, our hypothesis that mathematically based, biologically inspired embeddings, combined with state-of-the-art transformer structures, can be an effective method to early, consistent and scalable prediction of Parkinson disease can be compared to the achievement of our objective.

## REFERENCES:

[1]. Ruan, W., Clark, M. E., Zhao, M., Anthony Curcio, Department of Mathematics, Computer Science and Statistics, Purdue University Calumet, VasSol, Inc. (2002). A model of blood flow in a circulation network. https://doi.org/10.48550/arXiv.math/0210410.

[2]. Sanabria, D. E., Aman, J. E., Amaya, V. Z., Johnson, L. A., Farooqi, H., Wang, J., Hill, M., Patriat, R., Sovell-Brown, K., Molnar, G. F., Darrow, D., McGovern, R., Cooper, S. E., Harel, N., MacKinnon, C. D., Park, M. C., Vitek, J. L. (2022). Controlling pallidal oscillations in real-time in Parkinson's disease using evoked interference deep brain stimulation (eiDBS): Proof of concept in the human. Brain Stimulation, 15(5), 1111–1119. https: //doi.org/10.1016/j.brs.2022.07.047.

[3]. Basheer, S., Aldehim, G., Alluhaidan, A. S., & Sakri, S. (2024). Improving mental dysfunction detection from EEG signals: Self-contrastive learning and multitask learning with transformers. Alexandria Engineering Journal, 106, 52–59. https://doi.org/10.1016/j.aej.2024.06.058

[4]. Sigcha, L., Borz`ı, L., & Olmo, G. (2024). Deep learning algorithms for detecting freezing of gait in Parkinson's disease: A cross-dataset study. Expert Systems With Applications, 255, 124522. https://doi.org/10.1016/j.eswa.2024.124522

[5]. Aurpa, T. T., Fariha, K. N., Hossain, K., Jeba, S. M., Ahmed, M. S., Adib, M. R. S., Islam, F., Akter, F. (2024). Deep transformer-based architecture for the recognition of mathematical equations from real-world math problems. Heliyon, 10(20), e39089. https://doi.org/10.1016/j.heliyon.2024.e39089

[6]. ´Ad´am, F. (2024). Solving Mathematical Problems with Transformers. In ELTE Faculty of Science.

[7]. Desai, S., Chhinkaniwala, H., Shah, S., Gajjar, P. (2024). Enhancing Parkinson's Disease Diagnosis through Deep Learning-Based Classification of 3D MRI Images. Procedia Computer Science, 235, 201–213. https://doi.org/10.1016/j.procs.2024.04.023

[8]. Kukreja, V., Ahuja, V., Ati, M., Shanmugasundaram, H., Krishnamoorthy, M., Sharma, R., Bhattacherjee, A. (2025). A Spatio-Temporal Fusion-Based approach for Multi-Dimensional classification of Parkinson's disease progression using Multi-Modal dataset. Results in Engineering, 105317. https://doi.org/10.1016/j.rineng.2025.105317.

[9]. S.H. El-Sappagh, W. Nazih, M. Alharbi, T. AbuHmed, Responsible artificial intelligence for mental health disorders: current applications and future challenges, J. Disabil. Res. 4 (1) (2025) 20240101. https://doi.org/10.57197/JDR-2024-0101

[10]. A.R. Palakayala, Parkinson's disease classification based on enhanced ensemble learning and brain MRI, Majlesi J. Electr. Eng. 18 (1) (2024) 283–291. https://doi.org/10.30486/mjee.2024.2006443.1360

[11]. H. Zhang, K. Deng, H. Li, R.L. Albin, Y. Guan, Deep learning identifies digital biomarkers for self-reported Parkinson's disease, Patterns 1 (3) (2020) 100042. https://doi.org/10.1016/j.patter.2020.100042

[12]. Y. Zong, Q. Zuo, M.K.P. Ng, B. Lei, S. Wang, A new brain network construction paradigm for brain disorder via diffusion-based graph contrastive learning, IEEE Trans. Pattern Anal. Mach. Intell. 46 (12) (2024) 10389–10403. https://doi.org/10.1109/TPAMI.2024.3442811

[13]. S. Wang, T. Zhou, Y. Shen, Y. Li, G. Huang, Y. Hu, Generative AI enables EEG super-resolution via spatio-temporal adaptive diffusion Learning, IEEE Trans. Consum. Electron. 41 (10) (2025) 11762–11776. https://doi.org/10.1109/TCE.2025.3528438

[14]. Q. Zuo, N. Zhong, Y. Pan, H. Wu, B. Lei, S. Wang, Brain structure-function fusing representation learning using adversarial decomposed-VAE for analyzing MCI, IEEE Trans. Neural Syst. Rehabil. Eng. 31 (2023) 4017–4028. https://doi.org/ 10.1109/TNSRE.2023.3323432

[15]. Q. Zuo, Y. Shen, N. Zhong, C.P. Chen, B. Lei, S. Wang, Alzheimer's disease prediction via brain structural-functional deep fusing network, IEEE Trans. Neural Syst. Rehabil. Eng. 31 (2023) 4601–4612. https://doi.org/10.1109/TNSRE.2023.3333952 .

[16]. S.K. Biswas, A.N. Boruah, R. Saha, R.S. Raj, M. Chakraborty, M. Bordoloi, Early detection of Parkinson disease using stacking ensemble method, Comput. Methods Biomech. Biomed. Engin. 26 (5) (2023) 527–539. https://doi.org/10.1080/10255842.2022.2072683

[17]. T.S. Dhahi, A.K.Y. Dafhalla, A.W. Al-Mufti, M.E. Elobaid, T. Adam, S.C. Gopinath, Application of nanobiosensor engineering in

the diagnosis of neurodegenerative disorders, Results Eng. 24 (1) (2024) 102790. https://doi.org/10.1016/j.rineng.2024.102790.

[18]. G. Liu, et al., Prediction of cognition in Parkinson's disease with a clinical–genetic score: a longitudinal analysis of nine cohorts, Lancet Neurol. 16 (8) (2017) 620–629. https://doi.org/10.1016/S1474-4422(17)30122-9

[19]. U. Vignesh, R. Parvathi, K.G. Ram, Ensemble deep learning model for protein secondary structure prediction using NLP metrics and explainable AI, Results Eng. 24 (2024) 103435. https://doi.org/10.1016/j.rineng.2024.103435

[20]. H. Lei, et al., Early diagnosis and clinical score prediction of Parkinson's disease based on longitudinal neuroimaging data, Neural Comput. Appl. 35 (22) (2023) 16429–16455. https://doi.org/10.1007/s00521-023-08508-x

[21]. A. Atik, T. Stewart, J. Zhang, Alpha-synuclein as a biomarker for Parkinson's disease, Brain Pathol. 26 (3) (2016) 410–418. https://doi.org/10.1111/bpa.12370

[22]. L. Cook, et al., Parkinson's disease variant detection and disclosure: PD GENEration, a North American study, Brain 147 (8) (2024) 2668–2679. https://doi.org/10.1093/brain/awae142

[23]. D. Chicco, G. Jurman, The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation, BMC Genom. 21 (1) (2020) 37–63. https://doi.org/10.1186/s12864-019-6413-7

[24]. N. Aggarwal, G.J. Saxena, S. Singh, A. Pundir, Explainable AI approach for early detection of Parkinson's disease using PPMI Online data, Neural Comput. Appl. 36 (30) (2024) 19209–19230. https://doi.org/10.1007/s00521-024-10127-z

[25]. Martínez-Murcia, et al., Automatic detection of Parkinsonism using significance measures and component analysis in DaTSCAN imaging, Neurocomputing 126 (30) (2014) 58–70. https://doi.org/10.1016/j.neucom.2013.01.054

[26]. G. Solana-Lavalle, R. Rosas-Romero, Classification of PPMI MRI scans with voxelbased morphometry and machine learning to assist in the diagnosis of Parkinson's disease, Comput. Methods Programs Biomed. 198 (11) (2021) 105793. https://doi.org/10.1016/j.cmpb.2020.105793

[27]. S. Kanagaraj, M.S. Hema, M.N. Guptha, Optimized supervised learning approach to predict Parkinson's disease with minimal attributes using PPMI datasets, Multimed. Tools Appl. 83 (16) (2024) 48499–48520. https://doi.org/10.1007/s11042-023-17582-1

[28]. K. Baek, et al., Comparing MoCA performance in Parkinson's disease: age and education-adjusted cutoffs vs. machine learning, J. Mov. Disord. 17 (2) (2024) 171–180. https://doi.org/10.14802/jmd.23271

[29]. T. Simuni, et al., Longitudinal clinical and biomarker characteristics of nonmanifesting LRRK2 G2019S carriers in the PPMI cohort, npj Parkinson's Dis. 8 (1) (2022) 140–152. https://doi.org/10.1038/s41531-022-00404-w

[30]. F. Nabizadeh, K. Pirahesh, P. Valizadeh, REM sleep behavior disorder and cerebrospinal fluid alpha-synuclein, amyloid beta, total tau and phosphorylated tau in Parkinson's disease: a cross-sectional and longitudinal study, J. Neurol. 269 (9) (2022) 4836–4845. https://doi.org/10.1007/s00415-022-11120-z.

[31]. Zhang, Hui, Deng, K., Li, H., Albin, R. L., Guan, Y. (2020). Deep learning identifies digital biomarkers for self-reported Parkinson's disease. Patterns, 1(3), 100042. https://doi.org/10.1016/j.patter.2020.100042

[32]. Biswas, Sanjay Kumar, Boruah, A. N., Saha, R., Raj, R. S., Chakraborty, M., Bordoloi, M. (2023). Early detection of Parkinson disease using stacking ensemble method. Computational Methods in Biomechanics and Biomedical Engineering, 26(5), 527–539. https://doi.org/10.1080/10255842.2022.2072683

[33]. Lei, Hong, et al. (2023). Early diagnosis and clinical score prediction of Parkinson's disease based on longitudinal neuroimaging data. Neural Computing and Applications, 35(22), 16429–16455. https://doi.org/10.1007/s00521-023-08508-x

[34]. Aggarwal, N., Saxena, G. J., Singh, S., Pundir, A. (2024). Explainable AI approach for early detection of Parkinson's disease using PPMI data. Neural Computing and Applications. https://doi.org/10.1007/s00521-024-10127-z

[35]. Zong, Yong, Zuo, Q., Ng, M. K. P., Lei, B., Wang, S. (2024). Diffusion-based graph contrastive learning for brain disorder prediction. IEEE Transactions on Pattern Analysis and Machine Intelligence. https://doi.org/10.1109/TPAMI.2024.3442811

[36]. Wang, Shuo, Zhou, T., Shen, Y., Li, Y., Huang, G., Hu, Y. (2025). Generative AI enables EEG super-resolution via spatio-temporal adaptive diffusion learning. IEEE Transactions. https://doi.org/10.1109/TCE.2025.3528438

[37]. Thalapathiraj, S., Vivekanandam, Prof. (Dr)., & Tripathi, Dr. R. K. (2025). AI-Driven early identification of Parkinson's disease using machine learning and mathematical modelling. In SGS Engineering & Sciences, SGS Engineering & Sciences (Vol. 1, Issue 2). https://spast.org/index.php/techrep/index

[38]. Thalapathiraj, S., Vivekanandam, Prof. (Dr)., & Tripathi, Dr. R. K. (2025b). Scalable Parkinson's disease prediction using mathematical modeling, Hilbert Transforms, and Transformer-Based Deep Learning. SGS Engineering & Sciences, 1(4). https://spast.org/index.php/techrep/index