Cognitive Computation (2026) 18:46
https://doi.org/10.1007/s12559-026-10595-6

RESEARCH

Parkinson's Disease Detection From Electrical Stimulations WiFi Signals Using Information Fusion of Proposed Neural Networks

Zeeshan Habib[1] · Muhammad Attique Khan[2] · Zain Hussain[3] · Nathan Ng[4] · Ameer Hamza[5] · Ahmed Ibrahim Alzahrani[6] · Nasser Alalwan[6] · Zeshan Iqbal[7]

Received: 13 December 2025 / Accepted: 4 May 2026

Abstract

Parkinson's disease (PD) is a degenerative, chronic neurological condition that impairs a person's ability to move normally. People may experience difficulties with speaking, writing, walking, or performing basic tasks if dopamine-generating neurons in the brain are injured or die. Using traditional techniques for PD analysis is time-consuming and challenging, as the evaluation process is prone to high misclassification rates. Therefore, we proposed a deep learning-based architecture for classifying PD using WiFi signals in this work. The data is generated at the initial stage using WiFi signals. After that, we proposed two deep learning architectures from scratch. The first architecture, named E3-ST transformer, is based on a three-stage encoding scheme, and the second two residual-attention-block-based networks are named PD-RAN2. Both models are trained on generated WiFi signal data, and the hyperparameters are optimized using Bayesian Optimization (BO). In the next phase, trained models are used, and deep features are incorporated, employing a new method termed serial-based attention-weighted. The fused features are finally classified using neural network classifiers. The output is in label classes such as slow walking, fast walking, sitting on a chair, standing still, and FOG episodes. The Medium Neural Network (MN2) classifier achieved the best accuracy of 97.78%, whereas the individual models achieved 97.0% and 97.70%, respectively. A comparison with recent techniques shows that the proposed architectures achieve improved performance.

Keywords Parkinson disease · Wifi signals · Deep learning · Information fusion · Neural networks

✉ Muhammad Attique Khan
172649@korea.ac.kr

✉ Zeshan Iqbal
zeshan.iqbal@sivas.edu.tr

Zeeshan Habib
zeeshan.habib@hitecuni.edu.pk

Zain Hussain
zain.hussain@glasgow.ac.uk

Nathan Ng
nathannghuang@gmail.com

Ameer Hamza
ameer.hamza@ktu.edu

Ahmed Ibrahim Alzahrani
ahmed@ksu.edu.sa

Nasser Alalwan
nalalwan@ksu.edu.sa

[1] Department of Computer Science, HITEC University, Taxila, Pakistan

[2] Center of Artificial Intelligence, Prince Mohammad bin Fahd University, Al- Khobar, KSA, Saudi Arabia

[3] School of Infection and Immunity, College of Medicine, Veterinary and Life Sciences, University of Glasgow, Glasgow, UK

[4] Nothern Imaging Victoria, Melbourne, Australia

[5] Centre of Real Time Computer Systems, Kaunas University of Technology, Kaunas, Lithuania

[6] Department of Computer Science and Engineering, College of Applied Studies, King Saud University, P.O. Box 22459, Riyadh 11495, Saudi Arabia

[7] Department of Computer Engineering, Sivas University of Science and Technology, Sivas, Turkey

Published online: 13 May 2026

Introduction

Parkinson's disease (PD), a neurological condition that progresses over time, is the second most common neurodegenerative ailment and is increasing in global Prevalence [1, 2]. It is characterized by several progressive motor and non-motor symptoms [3], including bradykinesia, stiffness, resting tremor, and postural instability [4]. PD is the most common and is among the leading causes of disability globally, posing a significant burden to both the person and the community [5–7]. Among the many hitches that people with PD encounter, Freezing of Gait (FoG) [8] is one of the most concerning and incapacitating symptoms, which could be fatal [9]. FoG causes abrupt, sporadic stops or shuffling steps while walking, which can seriously impede mobility and increase the risk of falling [10]. More than half of the patients diagnosed with PD experience FoG, a debilitating motor sign of Parkinson's disease [11, 12].

Patients describe FOG as a sudden motor impairment that feels like their feet are glued to the ground and blocked from forward locomotion despite the intention to do so [13]. It can manifest in various ways, including shaky legs, shuffling gait, or complete akinesia. Fall incidents are the most common result of gait freezing in PD that can be fatal [14–16]. Falls and fear of falling can cause injuries, fractures, and ultimately, low quality and quantity of life in FoG [17–20]. In certain situations, gait freezing manifests [21]. For instance, when the effect of medication therapy wears off in PD patients, the occurrence, duration, and number of FOG episodes increase. Stress and performing cognitive and multiple-motor tasks can increase the likelihood of FoG [22, 23].

FOG typically occurs when there is an irregular stride, such as navigating a tight space or around corners. The frequency of FOG rises with the severity of Parkinson's disease [24]. Even if its milder stages are difficult to recognize early, more than half of FoG is what PD patients will eventually experience [25, 26]. FoG manifests in two different ways: ON-FoG or ON-state, which happens when the drug is fully effective, and OFF-FoG, which occurs when the drug's effectiveness is diminishing. Notably, ON-FoG poses unique challenges because it is resistant to numerous medicines. Although often responsive to dopaminergic treatment, mild to moderate FOG is frequently present in the ON state [27, 28]. It is activated when the person tries to turn what are known as motor blocks. The patient must change directions by taking multiple steps rather than twisting on one foot. FoG causes a sudden, temporary loss of mobility, resulting in a freeze. For a while, the patient experiences hesitancy when walking or moving their feet. The loss of motor abilities is exacerbated by fear of falling and loss of movement control [29].

One of the main reasons these patients fall is because of their postural instability and FOG, which increases their risk of hip fractures and injuries in elderly patients [30]. Thus, PD differs from other neurodegenerative diseases affecting the autonomic nervous system [31]. Due to the rise in the aging population worldwide, early PD diagnosis and prediction are essential goals in healthcare to offer timely preventive measures and enable patients to live everyday lives. The traditional techniques for assessing PD rely on physical clinical examination and, in most cases, may be misleading. Advancements in artificial intelligence (AI) and machine learning (ML) have significantly improved the accuracy of assessing and predicting FoG in PD patients [32]. Parkinson's disease currently has no known cure due to its complexity [33]. Experts may slow down the progression of PD by using treatments or deep brain stimulation to reawaken the brain's dopamine-producing neurons after an early diagnosis. On the other hand, early detection and appropriate treatment can help patients live everyday lives by reducing the symptoms of tremors and imbalance.

Recently, deep learning has achieved significant success in the medical domain for diagnosing and classifying conditions such as brain tumors [34], Parkinson's disease [35], Alzheimer's disease [36, 37], brain stroke, and others [38]. A deep learning architecture based on essential layers, including convolutional, relu, pooling, batch normalization, and fully connected. The convolutional layer extracts deep features, whereas the Softmax performs the final classification [39]. In recent years, several pre-trained models have been developed by computer vision researchers, such as GoogleNet, AlexNet, ResNet, MobileNet, and EfficientNet, to name a few more [40, 41]. These models are utilized for classification tasks in the medical domain and obtain notable performance.

Challenges

However, in Parkinson's disease, classifying from WiFi signals is difficult due to the nature of the input data and the design of the deep learning model. Therefore, designing a new model based on the challenges of PD using WiFi signals is essential. In this work, we proposed a fully automated deep-learning framework for classifying PD from WiFi signals. The proposed framework began with signal normalization and then proceeded to newly designed CNN architectures, which were later used for feature extraction, fusion, and classification.

Contributions

The significant contributions of this work are as follows.

- We proposed a new E^3-ST transformer, based on a deep learning architecture with a three-stage encoding scheme, for Parkinson's disease classification.
- We proposed a two-residual-attention-block network, PD-RAN^2, for PD classification. The PD-RAN^2 accepts the input size of dimension 200×1 from the input feature layer.
- Features are extracted from both designed models and are fused using a serial attention-weighted fusion technique rather than a simple serial fusion.
- A detailed comparison and ablation studies were performed to analyze the presented models for PD classification using WiFi Signals.

Related Work

The majority of people with PD exhibit a range of motor and non-motor symptoms as well as behaviors that progress with time [2, 12]. Specifically, FoG depresses their quality of life and makes them more likely to fall, which can be fatal [30]. By empowering the identification and prediction of FOG, researchers use smart technologies such as wearable sensors, mobile devices, and vision-based systems to better understand the intricate PD [32, 42]. Gait movement and stride variation in individuals, including PD patients, can be captured using sensors attached to various body positions [31]. Examples include wearable technology, cell phones, and sensors embedded in the ground. Several machine learning techniques were used to classify FoG for a chosen set of data or time frames, including ANNs, Random Forests, Support Vector Machines, Naïve Bayes, and K-nearest neighbors. The study reported [43] examined several machine learning algorithms to classify time- and frequency-domain features extracted from 2-, 3-, and 4-second time frames of accelerometer data for 10 patients into FoG, normal, and transition to FoG. These algorithms included Random Forest, Extreme Gradient Boosting, Gradient Boosting, Support Vector Machines using Radial Basis Functions, and two-layer ANNs. It was demonstrated that using 30 and 15 characteristics from 3-second time frames of accelerometer signals, ANNs could detect the transition to FoG intervals with a remarkably high sensitivity of 95.74%. This excellent performance was achieved by applying subject-dependent 10-fold cross-validation to the training set after splitting the dataset into 80% training and 20% testing.

Additionally, in [44], a subject-dependent 10-fold cross-validation sensitivity of up to 86%, 84%, and 81% was achieved at the waist, trouser pocket, and ankle positions, respectively, using a smartphone-based system. The signals from a smartphone accelerometer and gyroscope for 15 PD subjects were used to train and test various Machine Learning (ML) algorithms for Fog and no-Fog classification. Smartphone-based AdaBoost.M1 algorithm showed the best performance. In [45], the authors proposed a study using two distinct transfer learning models to detect FoG and No FoG in advance. The dataset was trained on most of the subjects' data and then retrained on the remaining data. Data were split into 50% training and 50% testing sets, and the model predicted FoG time stamps 1 and 3 s in advance, achieving subject-dependent accuracy of up to 87.5%. Natasa K. et al. [46] used various machine learning techniques to predict Freezing of Gait.

Accelerometer time series data from eight patients were analyzed, identifying 237 unique Freezing of Gait occurrences. These features were then used to train and test seven machine learning classifiers. Compared to benchmarked approaches, Support Vector Machines (SVM) with polynomial kernels outperformed the others. Using 18 features, the classification method was applied to 5-second windows. The Walk, FoG, and Transition classes yielded balanced accuracies (mean sensitivity and specificity) of 91%, 90%, and 82%, respectively. This was accomplished by dividing the dataset into training and test sets at 70% and 30%, respectively—the study in [47] used signal processing and deep learning methods to detect FoG.

The model was assessed using information from 15 PD subjects who also experienced FOG symptoms. An inertial measurement unit on the left side of the waist recorded signals from the tri-axial accelerometer, gyroscope, and magnetometer. The approach achieved validation performances with sensitivity and specificity of 88.6% and 78%, respectively. Camps et al. [48] presented a spectral stacking of two consecutive windows of 9-channel signals recorded from a waist-worn Inertial Measurement Unit (IMU) for 21 patients and classified into FoG and no FoG using an 8-layer one-dimensional CNN. The suggested model yielded a Leave-One-Subject-Out (LOSO) accuracy of 87.9%, a sensitivity of 92.6%, and a specificity of 86.7% with a window size of 2.5 s. Machine learning and deep learning are frequently used in pattern-based approaches to extract characteristics from CSI. While this methodology might not be as flexible as model-based approaches, it often yields good results for particular HAR tasks. It can identify a wide variety of human activity patterns. Most research on human activity recognition employs pattern-based techniques to automatically extract features from a denoised CSI signal and streamline the feature discovery process. Using the Intel 5300 tool, commercial WiFi equipment (IEEE 802.11n AP mode at 2.4 GHz), and machine learning, Wang et al. [49] were able to detect 91% of the various lip motions within six sentences. To convert gesture recognition difficulties into image classification problems.

Overall, most previous research on detecting signs of Parkinson's disease has focused on wearable sensors during day-to-day activities or on standard deep learning models that do not perform well with noisy signals and complex temporal data. In addition, prior work has focused primarily on feature fusion techniques to improve classification performance. The proposed framework addresses these issues by implementing two targeted architectures: the E^3-ST architecture captures temporal dependencies across multiple timescales via a multi-stage transformer, and the PD-RAN^2 architecture learns to enhance feature discrimination using residual attention mechanisms. This process will also lead to stronger and more accurately detected clinical FOG events.

Proposed Methodology

The proposed framework for classifying PD using WiFi Signals has been presented in this Section. Figure 1 presents the methodology of the proposed PD classification framework. In the initial step, the dataset is acquired and prepared in the CSI file (as discussed in the dataset section). After that, the proposed E3-ST and PD-RAN2 models are designed. The prepared datasets are fed to the developed model for training. After training, the models are used for feature extraction during testing, and the resulting information is later fused using a new serial attention-weighted method. After fusion, classification is performed using several methods, each applied in turn. A detailed description of each step is given in the subsections below.

Dataset Collection

This research used a dataset from an experimental protocol to examine variations in CSI resulting from multiple human activities, including freezing of gait (FOG) episodes [50]. Fifteen participants were recruited, with ten completing four activities considered normal, including slow walking, fast walking, voluntary stopping, and sit-to-stand transitions. The five remaining participants had clinically identifiable FOG episodes. Activities were completed in a randomized order among participants to ensure randomness and minimize potential bias. Data collected from the five participants who had FOG episodes were used to develop the ground truth, which was later used to train, validate, and benchmark models. The experimental protocol was designed to elicit perturbations in the wireless medium reflecting the intentional disturbances to the airways caused by various activities. These perturbations permitted the CSI signal to register minute variations in response to the activity being performed. These variations were further examined for potential capabilities to differentiate FOG episodes from activities that represent normal routines. The data collection section aimed to develop representative CSI signals that could reliably identify and classify FOG episodes in individuals living with Parkinson's disease.

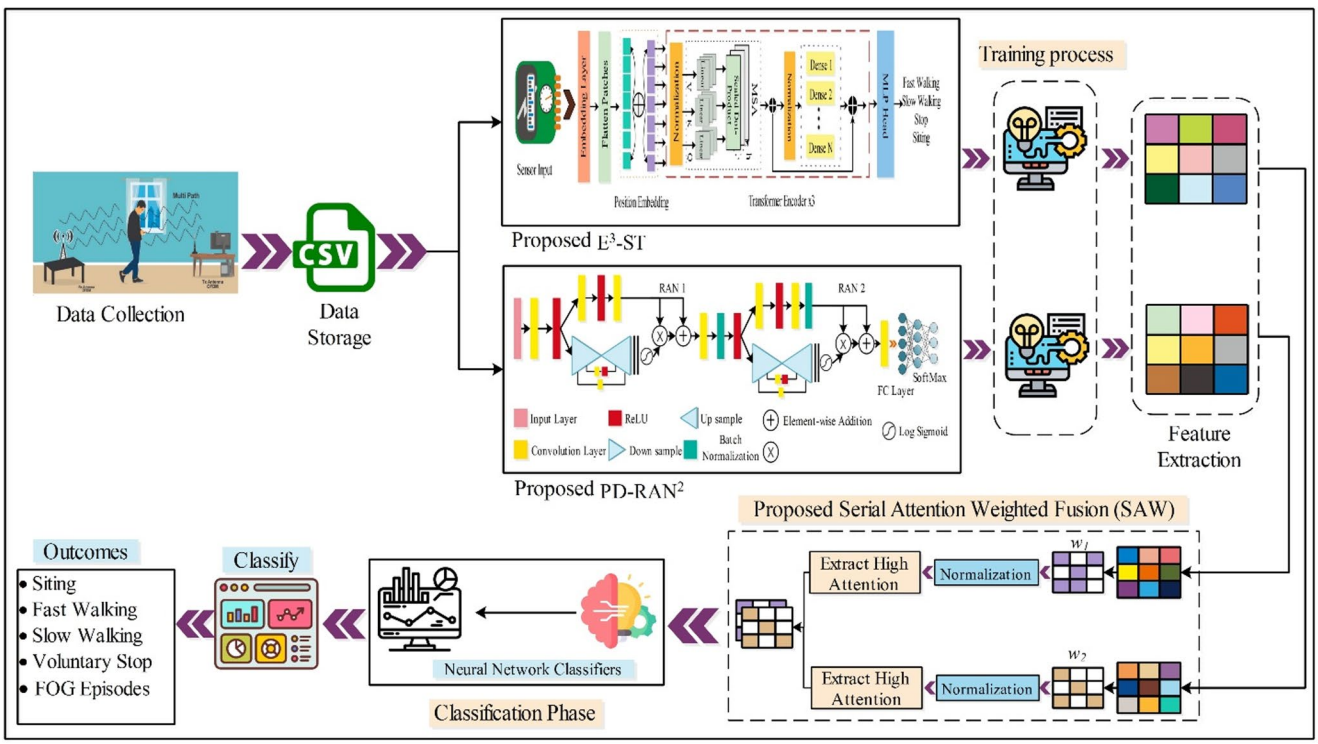

Fig. 1 Proposed framework based on deep learning for the classification of PD using WiFi signals

The wireless sensing system used a line-of-sight configuration comprising a WiFi transmitter (Tx) and a WiFi receiver (Rx). The Tx module transmitted a signal at 4.8 GHz, within the C-band, which includes frequencies typically used in 5G and even IoT-based applications. The Rx used an omnidirectional antenna connected to the Intel 5300 wireless network interface in a standard laptop running Ubuntu 14.10 OS, with 8 GB RAM. The CSI data obtained at the receiver was sent to a desktop workstation for post-processing and feature extraction. Because both Tx and Rx operate in the same frequency spectrum, this provided some channel consistency and improved signal fidelity. Mathematically, the acquired CSI dataset can be represented as:

$X = \{x_1, x_2, \ldots, x_n\}$, $x_i x_i \in \mathbb{R}^d$, where X is the set of CSI samples, *N* is the total number of collected observations, and *d* represents the number of features per sample. To reduce channel noise and ensure comparability across subjects, each CSI vector x_i is normalized as:

$$\begin{aligned} \widetilde{x}\,? &= \tfrac{(x?-\mu)}{\sigma}, \\ mu &= \left(\tfrac{1}{N}\right)\sum_{i=1}^{N} x?, \\ \sigma &= \sqrt{\left(\tfrac{1}{N}\right)\sum_{i=1}^{N}(x?-\mu)^2} \end{aligned} \tag{1}$$

Where μ and σ are the mean and standard deviation of the dataset, respectively, this preprocessing ensures that the input signals are standardized, highlighting activity-induced perturbations in CSI values while suppressing irrelevant fluctuations. Each CSI sample was labeled according to its corresponding activity, yielding five distinct labels. The scheme for labeled encoding and distribution of the sample rate is shown in Table 1.

Proposed E^3-ST

Sensor Transformers are specialized transformer models designed to analyze and process sensor data. Sensor data is treated as a sequence, and each data point and feature vector is considered a "token". These bindings are passed through the transformer encoder, which uses self-attention mechanisms to capture local and global dependencies over the sequence. The model can adapt to a range of tasks, including classification, regression, and forecasting. This approach is practical for PD classification, activity detection, and health monitoring applications. Sensor transformers focus on extracting meaningful patterns from sequential data and offer flexible, robust alternatives to traditional models, such as CNNs and RNNs, for analyzing sensor data. The motivation for the E^3-ST model is that it uses an encoding process comprising three progressive phases. The attention heads and dropout rates vary across phases to capture both local and global relationships within the CSI sequence. This gives the model the ability to create an environment to generalize from the data. Furthermore, the E^3-ST model uses convolutional operations throughout all encoder phases, which distinguishes it from standard transformer models. In addition, the E^3-ST provides stability for learning in the presence of signal noise via residual connections.

Table 1 Dataset classes and label encoding description

Label Encoding	Classes	Count	Samples
0	FOG episode	20	200
1	Sitting	20	200
2	Voluntary stop	20	200
3	Fast walking	20	200
4	Slow walking	20	200

In this work, we designed a new E^3-ST transformer based on a three-stage encoding scheme for Parkinson's disease classification. The proposed network starts with a feature input layer of 200 features. First, the raw feature vector x is mapped into an embedding space and enriched with positional encoding: $Z_0 = E(x) + P(x)$ where E(·) denotes the embedding function and P(·) represents positional encoding that preserves sequential information. After the mapping, the embedding layer converts the input into tokens. The first encoder starts with the position embedding layer; consider $P_D \in \mathbb{R}^{F\times 1}$. The embedding layer is defined as $\varnothing_{\epsilon} = \Phi(P_D)$, and the position embedding layer is formulated as $\varnothing_{p} = \varnothing_{\epsilon} + \kappa(P_D)$, where $\kappa(P_D)$ encodes the positional information. A dropout activation is employed with a dropout factor of 0.2, where $\varnothing_{d1} = \varphi_{D}(\varnothing_{p}, 0.2)$. Afterward, batch normalization is added to improve the convergence, which is defined as $\varnothing_{N} = \Pi_{N}(\varnothing_{d})$. A multi-head self-attention layer is employed to concurrently attend to multiple data chunks and to enable learning long-range dependencies and correlations. The operation of multi-head self-attention is $\varnothing_{SA}(Q,K,V) = \tau\left(\frac{QK^T}{\sqrt{S_e}}\right)V$, and concatenates the outcomes from heads, which are 3, and applies a linear transformation presented with $\varnothing_{MSA}(Q,K,V) = \bigcup(h_1,h_2,\ldots,h_3)\psi^{\circ}$ ..

After that, a dropout layer with a dropout factor of 0.2 is connected to the output of the multi-head self-attention layer. Later, a residual connection is employed using the given equation $\varnothing_{Res1} = \varnothing_{N} + \varnothing_{d}$. In the next step, the convolutional and GELU activation operations use $Conv1 = \lambda_{G}(\Psi(\varnothing_{Res1}))$. A dropout activation with a rate of 0.2 is applied to the GELU output. Furthermore, a convolutional operation is applied using $\varnothing_{Conv2} = \Psi(\varnothing_{d2})$ and a connected addition layer to create the second residual connection. The second residual block is created by adding a skip connection from the first

residual block to the last convolutional layer. This operation is presented as $\varnothing_{Res2} = \varnothing_{Conv2} \oplus \varnothing_{Res1}$.

The **second and third encoder stages** are constructed using the same mechanism, but with varying attention heads $(h = 4,6)$ and dropout factors $(0.1,0.3)$ to balance feature diversity and robustness. Finally, the output from the third encoder stage is passed through a multi-layer perceptron (MLP) head for classification. The final class probabilities are computed using the softmax function: $\widehat{y} = softmax\left(W_f \phi_{res}^{(3)} + b_f\right)$ where $\widehat{y}$ represents the predicted probability distribution across Parkinson's disease activity classes (slow walking, fast walking, sitting, standing, and FOG episodes). The model is trained by minimizing the categorical cross-entropy loss: $L = -\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c}\log\widehat{y}_{i,c}$, where $y_{i,c}$ is the ground-truth one-hot encoded label and $\widehat{y}_{i,c}$ is the predicted probability for class c. In total, the proposed E^3-T has 42 layers, three multi-head self-attention modules, and 1.6 M total trainable parameters. The proposed network is trained using a training data chunk on the selected dataset. The features are extracted from the trained model using the test data, and the extracted features have dimension $N \times 192$. The entire architecture is presented in Fig. 2. This figure shows that the sensor's input data is passed to the proposed network, which is trained with an MLP head and returns output classes for fast walking, slow walking, stop, and sitting.

Proposed PD-RAN^2

The motivation behind the PD-RAN^2 architecture is its residual attention, which employs both an adaptive method and a skip connection to identify the most salient regions within a signal while preserving the signal's raw feature representation. This method differs from conventional techniques for creating attention networks because multiple branches of attention are synthesized through progressive up- and down-sampling, affording it substantial advantages for identifying feature characteristics and recognizing patterns from CSI data. The residual attention mechanism combines the strength of residual connections and the attention mechanism to focus on the most relevant chunks of data while preserving original information. Unlike traditional networks that process data uniformly, residual Attention uses attention mechanisms, such as channel-wise attention, to dynamically identify and enhance critical features. By selectively improving key patterns while preserving the overall structure of signals, residual attention networks improve task performance, such as classification.

In this work, we designed two residual Attention block-based networks, PD-RAN^2, to classify Parkinson's disease. The PD-RAN^2 accepts the input size of dimension 200×1 from the input feature layer. After that, a convolutional layer with three filter size s, two strides, and 16 depths is added, followed by a ReLU activation to introduce non-linearity in the network. After that, the first residual attention module starts with the other two branches; the first branch has two convolutions with a filter size of $2,2$, a stride of $1,1$, a depth of $32,16$, and one ReLU activation. The other branch contains the convolutional configurations with three filter sizes, two strides, and 32 depths to downsample the feature map and transpose the convolutional configuration for the upsampled feature map. A sigmoid function is employed, and element-wise multiplication is performed with the first branch. The first branch is added with the outcome of the multiplication layer. The mathematical formulation of the residual attention module is defined as follows:

Consider an input feature map of $f \in \mathbb{R}^s$ that was obtained from the previous layers.

$$\psi_{conv1}(f)_{br1} = \Pi_{ReLU}(\omega_{c1} * f + b_{c1}) \tag{2}$$

$$\psi_{conv2}(\psi_{conv1}(f)_{b1})_{br1} = \omega_{c2} * \psi_{conv1}(f)_{br1} + b_{c2} \tag{3}$$

$$\psi_{ds}(f)_{br2} = \omega_{c3} * f + b_{c3} \tag{4}$$

$$\psi_{ups}(\Phi_{attbr2})_{br2} = \omega_{up}^{T} * \psi_{ds}(f)_{br2} + b_{up} \tag{5}$$

$$\Phi_{attbr2} = \sigma\left(\omega_{att}.ReLU\left(w.\psi_{ds}(f)_{br2} + b\right) + b_{att}\right) \tag{6}$$

$$\varphi_M = \psi_{conv2}(\psi_{conv1}(f)_{b1})_{br1} \odot \Phi_{attbr2} \tag{7}$$

$$\varphi_{skip} = \varphi_M ? \psi_{conv2}(\psi_{conv1}(f)_{b1})_{br1} \tag{8}$$

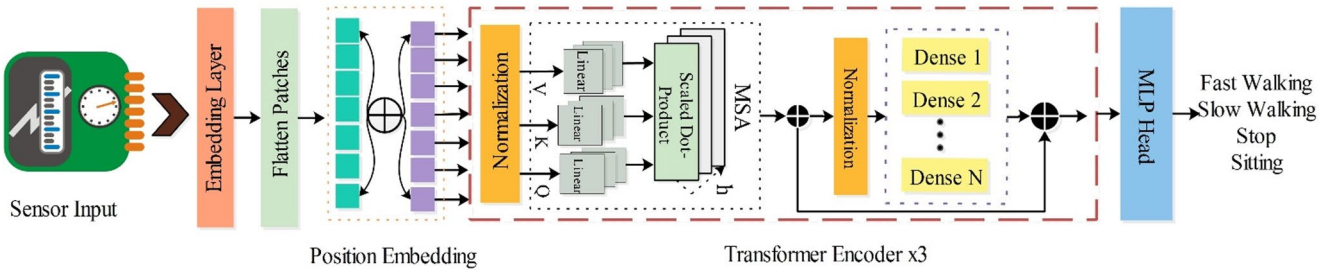

Fig. 2 Architecture of the proposed E^3-T transformer for the classification of Parkinson's disease

Where f is the feature map, $\omega_{c1}, \omega_{c2}, and\, \omega_{c3}$ are the convolutional weights, $b_{c1}, b_{c2}, and\, b_{c3}$ are the convolutional biases, ω_{up}^{T} is the weights of the upsampling convolutional operation, ω_{att} is the attention weights matrix, $\odot$ represents the element-wise multiplication, and ? is the addition operation, respectively. After the first module, a convolutional layer, batch normalization, and ReLU activation are used, followed by a second residual attention module, which is updated with one more batch normalization layer. Moreover, an additional convolutional layer is added. In the end, a fully connected layer and a softmax layer are used for classification. The mathematical definitions of these layers are:

$$\varphi_{Fc} = Softmax\left(\omega_{FC}.\psi_{conv11}\left(\varphi_{skip2}\right) + b_{FC}\right) \tag{9}$$

$$\varphi_{c} = \frac{e^{(\varphi_{Fc}, c)}}{\sum_{z=1}^{C} e^{(\varphi_{Fc}, z)}},\ z = 1,2,3,\ldots,C \tag{10}$$

Where the $\varphi_{c} \in \mathbb{R}^{C}$ represents the class probabilities for C classes, the proposed PD-RAN^2 has 32 layers with 1.2 M parameters. The designed model is trained on the training data and validated during training using the validation data. The features are extracted from the deep convolutional layer using the test data, and the extracted features have size $N \times 1024$. A systematic architecture is presented in Fig. 3. This figure shows that the input layer feeds into the designed network, which ends with a softmax layer for final classification.

Models Training and Features Extraction

After designing both models, training was performed on the selected database. A few hyperparameters are also initialized to train this model, such as initial learning rate, momentum, batch size, and dropout factor. Usually, these parameters are selected manually (brute-force); however, this is not a good idea because these hyperparameter values control the network training. Therefore, it is essential to initialize the hyperparameter using an optimization algorithm. The purpose of hyperparameter optimization is to minimize prediction error and maximize training accuracy. In this work, we utilized Bayesian Optimization [51] to select hyperparameter values. The hyperparameter range is given in Table 2. After selecting the best value, the model is trained and later used for feature engineering. In the feature engineering process, features are extracted, and the obtained vectors for these models are $N \times 192$ and $N \times 1024$, respectively. These features are fused using a serial attention-weighted fusion (SAW) technique, as described in Sect. 3.4.

Figures 4 and 5 provide the information on the training and validation performance of the proposed E^3-ST and PD-RAN^2. For E3-ST, there is a clear and consistent increase in both training & validation accuracy from approximately 35% to almost 99% over time, with almost no difference (very little gap) between the two different curves throughout this experiment. This close relationship provides strong evidence of excellent generalization during training. The loss curves for both training & validation were also very smooth, with a continual decline toward very low values, demonstrating stability during optimization and successful learning. For PD-RAN2, training accuracy increased rapidly during the initial epochs, reaching almost 99%; however, there was noticeable overfitting. Validation accuracy is lower than training accuracy at the midpoint of the

Table 2 Hyperparameter selection using Bayesian Optimization

Hyperparameter	Range	Resultant Value
Initial Learning Rate	10e-5 to 10e−1	0.004
Momentum	0–1	0.6
Dropout Rate	0.3 to 0.7	0.5
Batch Size	32, 64, 128	64
Optimizer	Adam, RMSprop, Adagrad	Adam

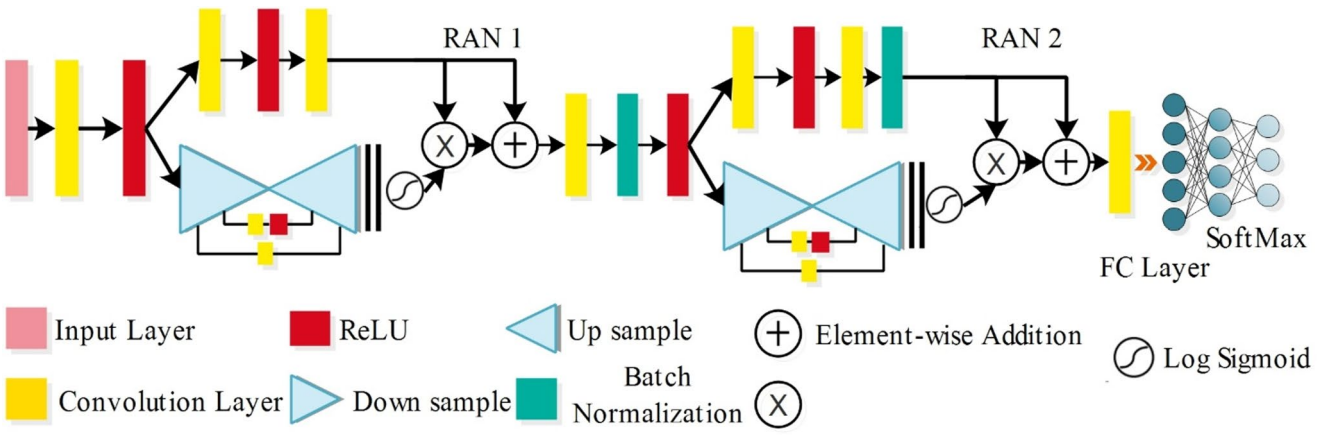

Fig. 3 Systematic architecture of proposed PD-RAN2 for Parkinson's disease classification

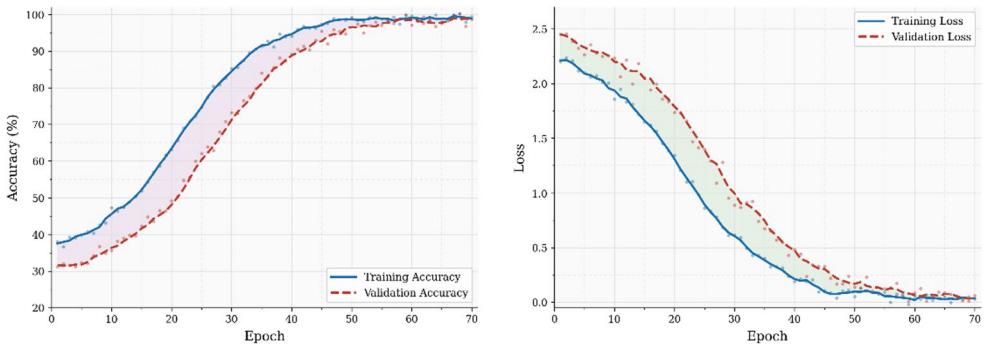

Fig. 4 Training validation curves of the proposed E^3-ST

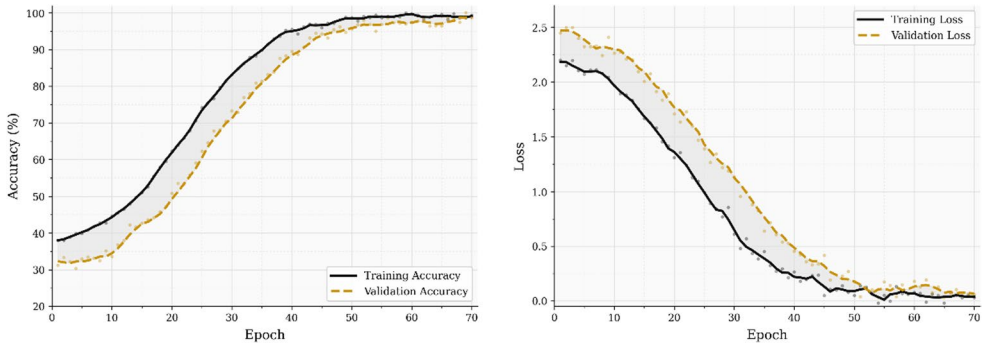

Fig. 5 Training validation curves of proposed PD-RAN^2

experiment. In addition to a continuous decline in both training & validation loss, the numerical gaps between them are much larger than in E^3-ST throughout most of the first and mid-training periods. They will lower toward the end of the experiment.

Proposed Serial Attention Weighted Fusion (SAW)

Combining features from different sources into a single source is called feature fusion [52]. In this study, we proposed a new fusion technique, Serial attention-weighted fusion (SAW), combining the most prominent features. The purpose of this fusion technique is to select the most important features and combine the most prominent ones for further classification. Mathematically, the formulation of this process is defined as follows:

Consider two feature vectors obtained from the proposed E^3-T and PD-RAN^2 with dimensions of $f_1 \in \mathbb{R}^{\alpha \times \beta_1}$ *and* $f_2 \in \mathbb{R}^{\alpha \times \beta_2}$, where α is the number of samples and β_1 and β_2 are extracted features—the values of $\beta_1 = 192$ *and* $\beta_2 = 1204$. The initial step calculates the attention value for each feature using Eq. (11).

$$\varphi_f(k) = |\omega(k)| \quad , \qquad k = 1,2,3,\ldots,\beta \tag{11}$$

After that, standard normalization is used to scale the attention values to the range $[]0,1$, and a threshold is applied to the normalized Attention to select the top features. This operation is performed by using Eqs. (12–14).

$$\mathbb{N}_f(k) = \frac{\varphi_f(k)}{\max(\varphi_f)} \tag{12}$$

$$\tau_{idx_1} = \{k \ | \mathbb{N}_{f1}(k) > \theta\} \tag{13}$$

$$\tau_{idx_2} = \{k \ | \mathbb{N}_{f2}(k) > \varphi\} \tag{14}$$

Where ω is the Attention weights matrix for f_1 and f_2, $\mathbb{N}_f(k)$ is the standard normalization outcome for f_1 and f_2, and τ_{idx_1} and τ_{idx_2} indicate the high attention values indices. After that, the feature values are computed from the close attention indices of the original vectors using Eqs. (14–15)—the dimensions of the gained features from f_1=114 and $f_2 = 976$, respectively. In the last step, the obtained high-attention features from both original vectors are combined using the serial method, as defined in Eq. (17).

$$f_{1\to\tau A} = f_1(:, \tau_{idx_1}) \tag{15}$$

$$f_{2\to\tau A} = f_2(:, \tau_{idx_2}) \tag{16}$$

$$F_{s_\tau} = \begin{bmatrix} f_{1\to\tau A} \\ f_{2\to\tau A} \end{bmatrix}_{\alpha \times 1090} \tag{17}$$

Where $f_{1\to\tau A}$ and $f_{2\to\tau A}$ are obtained, great attention feature vectors, and F_{s_τ} is the outcome of the serial method. The final output of this method is $\alpha \times 1090$. After the fusion, the resulting feature vector is passed to multiple neural network classifiers for final-phase classification.

Results and Analysis

Experimental Setup

The experimental setup of the proposed framework is described in this Section. To provide a fair and unbiased evaluation, a primary validation technique is performed using 10-fold cross-validation. This entails dividing the entire dataset into 10 different (non-overlapping) folds, which are used in 10 different iterations when performing testing and training. During each iteration, nine folds are used for training, and one fold is reserved for testing. In addition, part of the training set is set aside as a validation set to tune the model's parameters. To prevent data leakage during validation, an independent subject-based validation scheme is used. This means that all samples from a single participant are allocated to only one fold and cannot be used in any other fold. The use of this validation strategy ensures that the model's evaluation is conducted on unseen subjects, thereby providing a reliable means of assessing its generalization capabilities with reduced risk of overfitting.

Several hyperparameters are the learning rate, momentum, mini-batch size, number of epochs, dropout factor, and optimizer to train the models. The values for these hyperparameters are selected via BO, as presented in Table 2, and the cross-entropy loss function is employed in both proposed models. For the testing of the proposed architecture, five neural network classifiers are employed: Narrow Neural Network (N^3), Medium Neural Network (MN^2), Wide Neural Network (WN^2), Bi-layered Neural Network (BN^2), and Tri-layered Neural Network (TN^2). The experimental process was conducted in MATLAB 2024 on a desktop workstation with a 512 GB SSD, an NVIDIA A4500 20 GB graphics card, and 64 GB of RAM.

Proposed Results

This Section presents the proposed architecture results in numerical form and confusion matrices. The results are computed across three experiments: the proposed E^3-ST model features, PD-RAN^2, and serial-based attention-weighted fusion. For each model, results are described in the subsections below.

Results of Proposed E^3-ST

In this experiment, the proposed E^3-ST network is trained on the selected dataset, and prominent features are gained and fed to the neural network classifiers. The classification results of the proposed E^3-ST network on the chosen dataset have been presented in Table 3. As per this table, the WN^2 classifier achieved an outstanding accuracy of 97.1%. Furthermore, the WN^2 achieved 97.12% precision, 97.1% recall, and 97.11% F1 score. The confusion matrix in Fig. 6 further confirms the WN^2 classifier's results. The confusion matrix was used to verify the WN^2 classifier's accuracy and other computed metrics. The lowest accuracy of 90.2% is reported for TN^2 among all listed classifiers. The same computational measures are used for all the classifiers, and it is concluded that the WN^2 classifier outperformed all the other classifiers under study in terms of accuracy, precision, recall rate, and F1-score measure. It also reveals that the BN2 is the least computationally efficient, and N^3 takes the longest.

Results of Proposed PD-RAN^2

In this experiment, the proposed PD-RAN^2 model is trained on the selected dataset, and features are obtained from the deep convolutional activation. The obtained features are passed to the neural network classifiers. Table 4 presents the classification results. According to this table, the MN^2 classifier achieved the highest accuracy of 97.71%, precision of 97.72%, recall of 97.72%, and F1-score of 97.72% with a training time of 2.63 s. The other classifier also performed well, gaining accuracy in the range of (92.00 to 97.21). In Fig. 7, the confusion matrix is presented to verify the MN^2 classifier's results further. Moreover, the training time of each classifier is also observed, and the BN^2 classifier has the lowest training time of 2.042 s with a 97.22% precision rate. The N^3 classifier has the highest training time of 2.98 s with a 96.72% precision rate.

Results of the proposed Serial Attention Weighted Fusion (SAW)

In the last experiment, the extracted features from both proposed networks are fused using the proposed serial attention-weighted fusion technique. In this fusion,

Table 3 Classification results of proposed E^3-T CNN features using neural network classifiers

Classifiers	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)	Training Time (sec)	Prediction Speed (obs/sec)
N^3	95.91	95.92	95.9	95.91	2.200	148.97
MN^2	97.00	97.02	97.00	97.01	1.538	171.03
WN^2	**97.14**	**97.12**	**97.1**	**97.11**	**1.570**	**167.53**
BN^2	93.35	93.38	93.3	93.33	1.096	163.10
TN^2	90.23	90.22	90.2	90.21	1.389	171.52

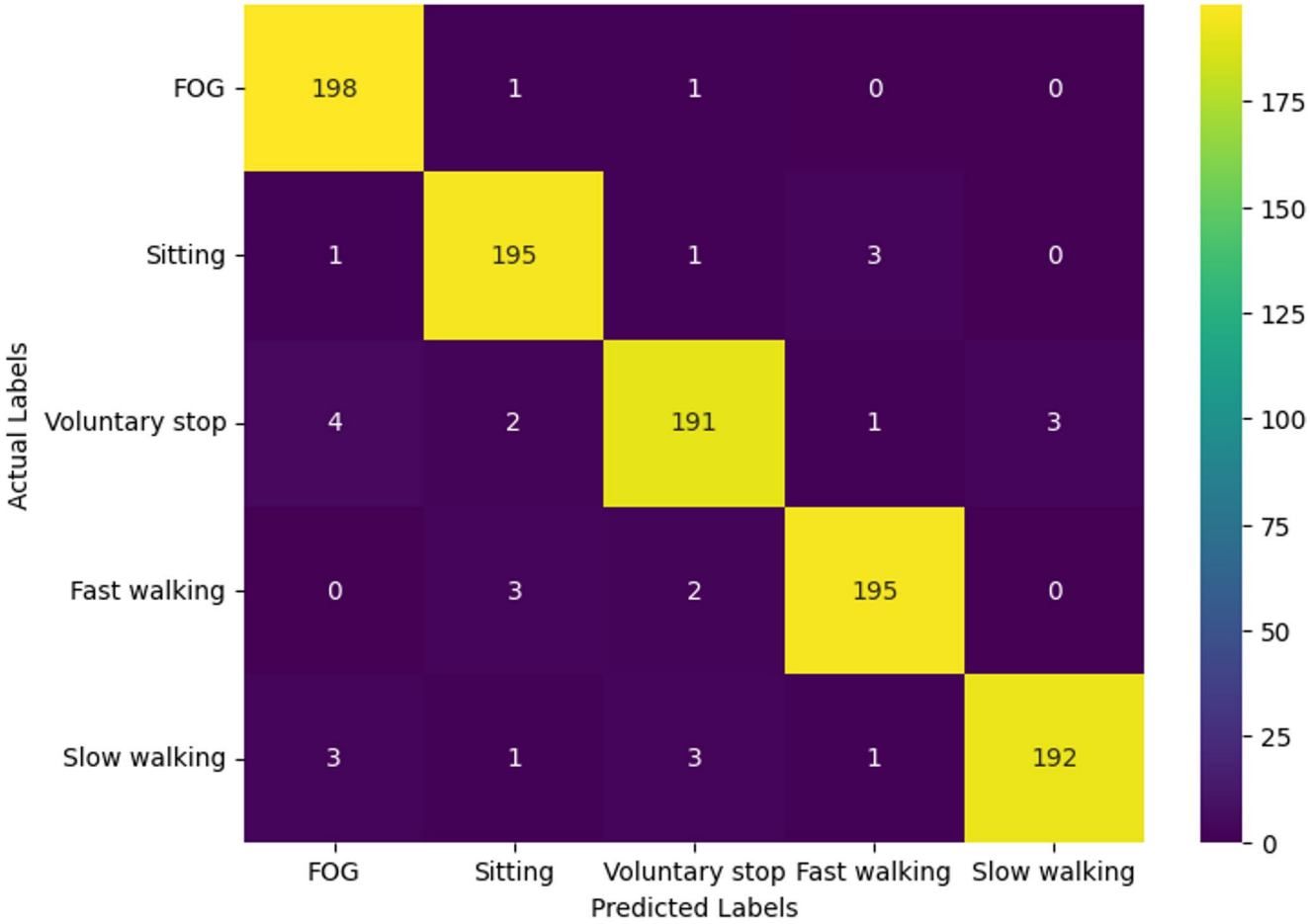

Fig. 6 Confusion matrix of proposed E^3-ST CNN features on WN^2 classifiers by utilizing the Parkinson's disease WiFi signals dataset

Table 4 Classification results of proposed PD-RAN^2 CNN model features using neural network classifiers

Classifiers	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC	Training Time (Sec)	Prediction Speed (obs/sec)
N^3	96.47	96.42	96.4	96.41	0.999	2.9891	118.79
MN^2	**97.71**	**97.72**	**97.72**	**97.72**	**0.995**	**2.6313**	**165.32**
WN^2	97.21	97.22	97.21	97.21	0.999	2.1882	**167.51**
BN^2	94.65	94.6	94.66	94.64	0.999	2.0244	136.71
TN^2	92.00	92.02	92.21	92.01	0.993	2.0429	174.14

attention weights are computed for both features, and only the high-attention features are fused into the outcome. The classification report of the obtained high-attention features is presented in Table 5. This table shows that the MN^2 classifier achieved the highest overall accuracy of 97.78%, precision of 97.79%, recall of 97.74%, and F1-score of 97.79%. Also, it is demonstrated that the rapid prediction speed (429.071 oz/sec) and the relatively low training time (3.19 s) are achieved for this experiment. This makes it the most efficient model for prediction performance and efficiency. The confusion matrix of this classifier is also shown in Fig. 8 for further investigation. On the other hand, the BN^2 and TN^2 classifiers have the lowest performance, with 94.26% and 93.15% accuracy, respectively, despite their high prediction speed (450 obs/sec). The WN2 classifier has a high accuracy of 97.21% and a high prediction speed (450.662 obs/sec), but is slightly behind the MN2 classifier in all measures.

Class-Wise Evaluation

Class-wise performance has been shown in Table 6. This table provides a clearer indication of how well the proposed architectures differentiate between different activity classes. The E3-ST and PD-RAN2 achieve high precision, recall, and F1 scores across all classes, indicating that their learned feature representations are strong and generalize well to unseen classes. Furthermore, when comparing the

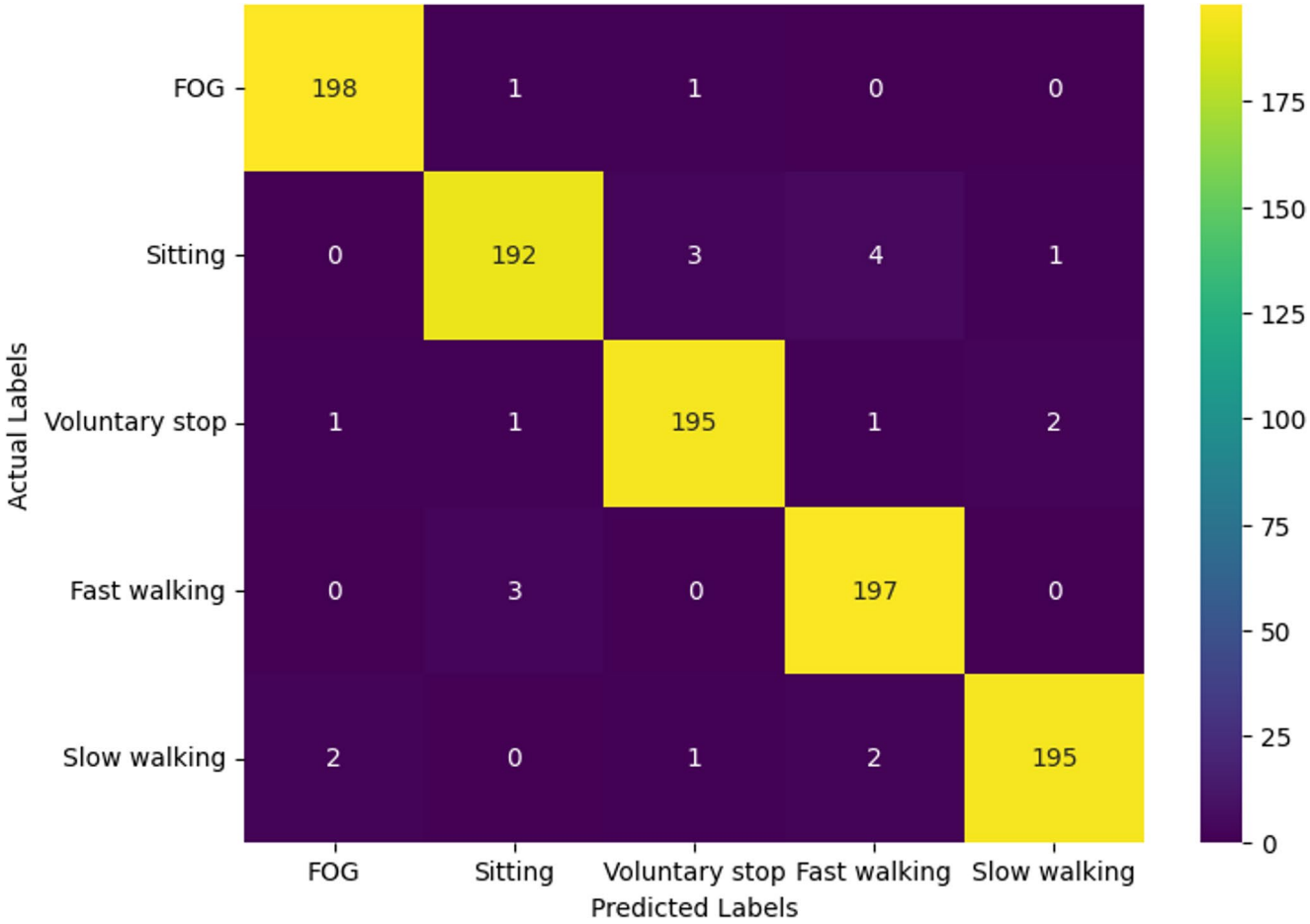

Fig. 7 Confusion matrix of proposed PD-RAN^2 features on MNN classifiers by utilizing the PD dataset

Table 5 Classification results of proposed serial attention weighted fusion features using neural network classifiers

Classifiers	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Training Time (sec)	Prediction speed(obs/sec)
N^3	96.31	96.32	96.3	96.37	5.13	363.281
MN^2	**97.78**	**97.79**	**97.74**	**97.79**	**3.19**	**429.071**
WN^2	97.21	97.21	97.24	97.28	2.59	450.662
BN^2	94.26	94.21	94.24	94.27	2.69	451.953
TN^N	93.15	93.16	93.14	93.12	4.43	454.312

two models, it's evident that the PD-RAN2 outperforms the E3-ST in the majority of classes, which is likely due to the PD-RAN2's use of a residual attention mechanism that allows the architecture to place increased emphasis on informative signal features and reduce the amount of noise in the CSI data.

From a clinical perspective, detecting Freezing of Gait (FOG) is the most important activity class for effective patient monitoring and intervention. When assessing both models, they performed well; however, the PD-RAN2 achieved the best precision and recall among the other classes in the FOG class. This demonstrates how effective the new architectures are at detecting subtle signal variations associated with freezing episodes. The improved performance in detecting FOG indicates that attention mechanisms are critical for highlighting temporal discriminative patterns that distinguish pathological events from typically occurring activities.

For the remaining activity classes, there is only slight variability in the performance metrics across the classes. Some minor differences in the performance metrics may be attributed to overlap between the slow walking and voluntary stopping classes, given their similar motion patterns. However, the differences in the performance metrics were negligible and would not adversely affect the overall classification of either model. The consistency in overall

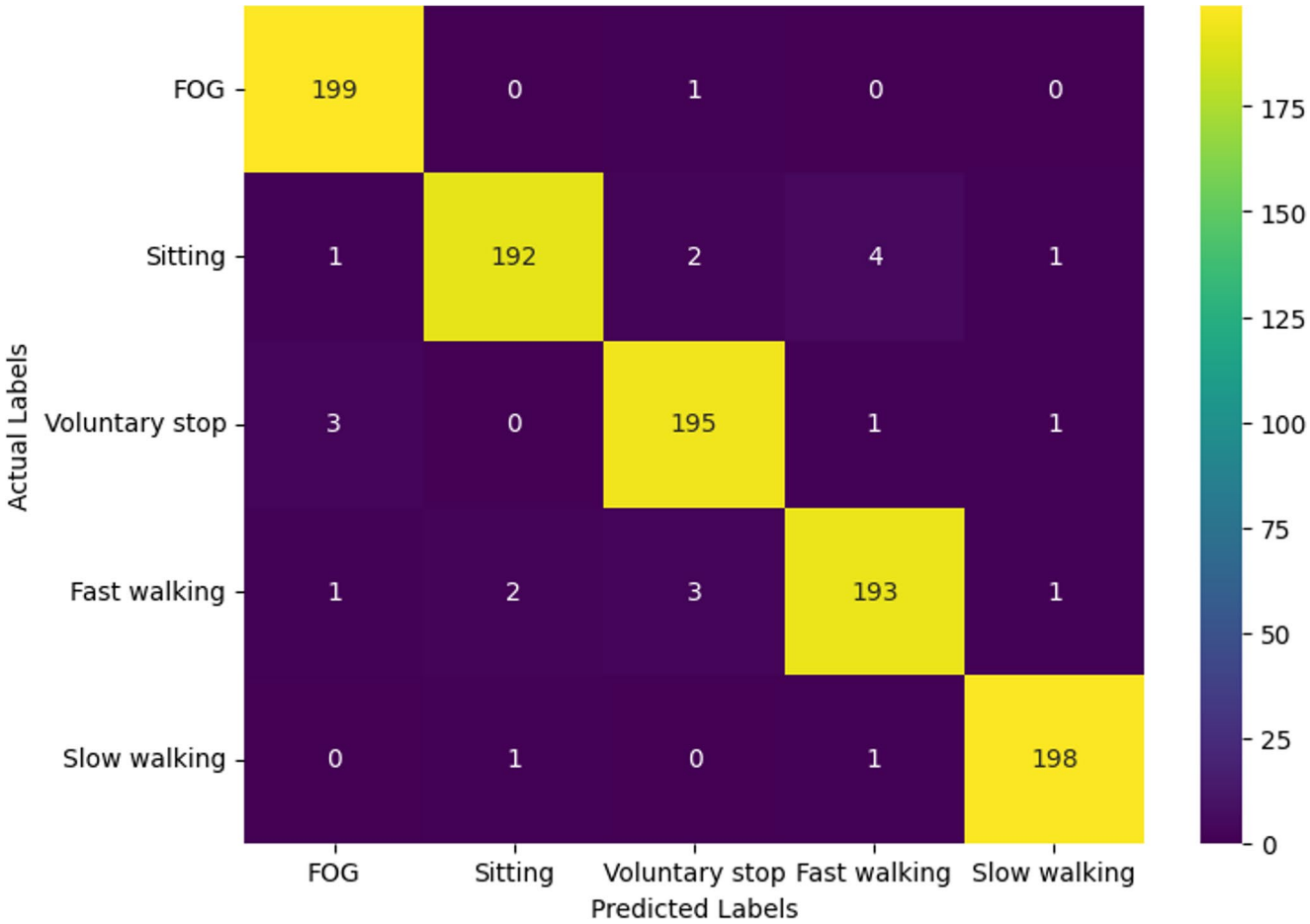

Fig. 8 Confusion matrix of proposed Serial attention-weighted Fusion features on MN^2 classifiers by utilizing the PD dataset

Table 6 Class-wise results obtained using the proposed E^3-ST and PD-RAN^2

Classes	E^3-ST Precision	E^3-ST Recall	E^3-ST F1	PD-RAN^2 Precision	PD-RAN^2 Recall	PD-RAN^2 F1
FOG	97.8	97.6	97.7	98.1	97.9	98.0
Sitting	96.9	97.0	96.9	97.5	97.6	97.5
Voluntary Stop	97.0	96.8	96.9	97.6	97.4	97.5
Fast Walking	97.2	97.1	97.1	97.8	97.7	97.7
Slow Walking	97.1	97.0	97.0	97.6	97.5	97.5

performance metrics across the remaining classes indicates that the architectures effectively learned inter-class boundaries, despite the high correlation among the different human motion signals within each class.

Discussion

This Section presents a detailed discussion of the proposed architecture in ablation studies, compared with pre-trained models and state-of-the-art (SOTA) techniques. The proposed architecture is based on two CNN models, E^3-ST and PD-RAN^2. Features extracted from both models are fused using a new serial-based weighted attention technique. For the E^3-ST model, WN^2 achieved the highest precision of 97.12%, as shown in Table 3; Fig. 6. The second proposed model, PD-RAN^2, further improved precision, achieving 97.72% (see results in Table 4 and the confusion matrix in Fig. 7). When fusing both models at the feature level using the proposed serial-based weighted attention approach, the resultant precision rate is 97.79%. A minor improvement was observed across all the classifiers listed in Table 5 compared to Tables 3 and 4. Also, comparing the confusion matrices across all models shows that the fused model achieves a higher prediction rate. Overall, the proposed model performs better at classifying PD using the WiFi Signals dataset. To further validate the performance of the proposed architecture, we conducted several ablation studies, as listed below.

Ablation Study

In this experiment, a comparative analysis is conducted to highlight the strengths of various models in terms of accuracy and efficiency, as shown in Table 7. The SVM provides robust accuracy (0.76), a fast inference time (0.0054 s), and is suitable for real-time predictions. Random Forest has improved slightly in SVM, achieving 0.78 accuracy and faster training time (0.64 s), which benefits from scalability. The decision tree achieves an impressive accuracy of 0.87, but training time is slightly longer (0.84 s) while still providing rapid conclusions (0.0021 s). DNN has the highest accuracy (0.92), but the training time (16.74 s) and inference time (1.0108 s) are much longer and better suited to less critical training. The proposed E^3-T (0.974 accuracy) and PD-RAN^2 (0.977 accuracy) models achieve the best performance with reasonable training and inference times, indicating improvements in predictive power without sacrificing efficiency.

In the following ablation study, the experiment uses the fusion technique to evaluate its impact, as described in Table 8. According to this table, the parallel fusion with the N×1216 dimension ensures a dense accuracy of 95.47% while maintaining a high estimation speed of 314.094 obs/sec. It provides a good balance between performance and efficiency. The serial fusion configuration achieves 97.21% accuracy but is slightly slower (309.742 obs/sec). It is more accurate, but this comes at the cost of a slight decrease in speed. The serial attention weight fusion further increases accuracy to 97.78% and time to 429.071 obs/sec, signifying that attention mechanisms improve performance but require more computational resources. Finally, the parallel attention-weighted fusion achieved 96.46% accuracy. Still, the inference time increased significantly to 613.146 obs/sec, illustrating the trade-off between accuracy and inference speed when parallel configurations use attention mechanisms.

Table 7 Comparison of proposed models with DNN and ML models

Models	Accuracy	Training Time (sec)	Inference Time (sec)
SVM	0.76	0.86	0.0054
Random Forest	0.78	0.64	0.0061
Decision Tree	0.87	0.84	0.0021
DNN	0.92	16.74	0.108
Proposed E^3-T	0.971	1.57	0.184
Proposed PD-RAN^2	0.977	2.63	0.169

Table 8 Comparison between proposed and other SOTA fusion techniques

Configurations	Dimensions	Accuracy	Inference Time (obs/sec)
Parallel Fusion	$N \times 1216$	95.47	314.094
Serial Fusion	$N \times 1216$	97.21	309.742
Serial Attention Weighted Fusion	$N \times 1090$	97.78	429.071
Parallel Attention-Weighted Fusion	$N \times 1130$	96.46	613.146

Comparison with SOTA

In this Section, a comprehensive comparison is conducted among recent and proposed frameworks, based on their methods and datasets. Table 9 presents the comparison of the proposed accuracy with SOTA techniques. The proposed framework for using DNNs on self-created datasets achieves an impressive 97.78% accuracy, exceeding that reported in several recent studies. In particular, Saha et al. [53] in 2024 proposed a GCN network for analyzing Parkinson's disease data, achieving 97.3% accuracy. Lee et al. [24] applied machine learning to the MiRNA dataset and achieved 97% accuracy. In 2025, Garehdaghi et al. [54] achieved 90.2% with handcrafted features, and Volkmann et al. [29] presented a framework based on machine learning and deep learning models, achieving 88.0% accuracy. In addition, Camacho et al. [55] presented a deep learning model trained on a self-created dataset, achieving an accuracy of 80.8%. The proposed model achieved an accuracy of 97.78%, surpassing the SOTA listed in this table.

Compared with Pre-Trained Models

A comparison of the proposed architecture's accuracy is also conducted with recent SOTA pre-trained models, using both training and test accuracy. Table 10 presents the results of this experiment. This table shows that the AlexNet model

Table 9 Comparison of proposed architecture accuracy with SOTA techniques

Ref	Year	Dataset	Methodology	Accuracy
Volkmann et al. [29]	2025	Private dataset	Methods of DL and ML for the identification of regions	88.0%
Garehdaghi et al. [54]	2025	Open Neuro recorded	Handcrafted Features	90.2%
Lee et al. [24]	2024	MiRNA dataset	Machine Learning	97.0%
Saha et al. [53]	2024	Parkinson's disease	GCN Network	97.4%
Camacho et al. [55]	2024	Self-Created dataset	Explainable Deep Learning	80.8%
Proposed Framework		Self-Created Dataset	Fusion of E^3-ST and PD-RAN^2 with Optimized Hyperparameters	97.78%

Table 10 Comparison of the proposed CNN architectures' accuracy with pre-trained CNN models using the WiFi signals dataset

CNN Model	Softmax Classifier	MN^2
	Training Accuracy (%)	Testing Accuracy (%)
Proposed E^3-T	**97.24**	**97.00**
Proposed PD-RAN^2	**98.10**	**97.71**
AlexNet	91.42	90.56
VGG16	90.50	90.26
ResNet50	92.60	91.84
InceptionV3	93.65	92.04
MobileNetV2	93.46	93.10
Densenet201	95.56	95.24

trained on the dataset generated in this work achieved a training accuracy of 91.42%. The VGG16 and ResNet50 achieved accuracy of 90.50% and 92.60%, respectively. Accuracy was further improved using InceptionV3, MobileNetV2, and DenseNet201 architectures, reaching 93.65%, 93.46%, and 95.56%, respectively. This indicates that the DenseNet architecture outperforms other pre-trained models by learning from complex data. The proposed models achieved training accuracy of 97.24% and 98.10%, respectively. In the testing phase, the MN2 classifier achieves accuracies of 97.0% and 97.71% for the Proposed E3-T and Proposed PD-RAN2 architectures, respectively. The test accuracies for the remaining models are 90.56%, 90.26%, 91.84%, 92.04%, 93.10%, and 95.24%, respectively.

Statistical and Analytical Evaluation

To ensure that the improvements in the proposed framework are not only empirical but also statistically significant, a detailed statistical analysis is conducted.

Hypothesis testing

To validate whether the observed improvements of the proposed method over the baseline are statistically significant, we perform paired hypothesis tests on the performance metric of interest (e.g., accuracy, F1, AUC) across cross-validation folds. We denote the metric of the proposed model on fold i by $m_i^{(prop)}$ and the metric of the baseline by $m_i^{(base)}$, for $i = 1,2,3 \ldots ..n$ (Typically *n*=10 for 10-fold CV). Our null and alternative hypotheses for the paired comparison are:

Paired t-test (parametric):

$H_0 : \mu_d$ versus $H_1 : \mu_d \neq 0$, where μ_d is the population mean of differences.

Wilcoxon signed-rank test (nonparametric):

$$H_0 : \text{median}\left(m^{(pop)} - m^{(base)}\right) = 0 \text{ versus } H_1 : \text{median} \neq 0$$

1. Compute per-fold differences

Compute the vector of paired differences: $d_i = m^{(pop)} - m^{(base)}$, i = 1,2,..., n.

1. Check normality of differences

Test normality of $\{d_i\}$ (recommended: Shapiro–Wilk). If the Shapiro–Wilk $(p_{sw} > 0.05)$ we do not reject normality and may use the paired t-test. If normality is violated $(p_{sw} \leq 0.05)$, prefer the Wilcoxon signed-rank test.

1. Paired t-test (if the normality assumption holds)

- **Mean and standard deviation of differences**:

$$\bar{d} = \frac{1}{n} \sum_{i=1}^{n} d_i \tag{18}$$

$$s_d = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (d_i - \bar{d})^2} \tag{19}$$

- **Test statistic**:

$$t = \frac{\bar{d}}{s_d/\sqrt{n}} \tag{20}$$

- **Degrees of freedom**:

$$d_{f=n-1} \tag{21}$$

- **Two-sided p-value**:

$$p = 2(1 - F_t(|t|; d_f)) \tag{22}$$

Where is the $F_t(.; d_f)$ cumulative distribution function.

- **Paired Cohen's d (effect size for paired samples)**:

$$d_{cohen} = \frac{\bar{d}}{s_d} \tag{23}$$

Interpretation thresholds: d≈0.2 (small), d≈0.5d (medium), d≥0.8 (large).

Decision rule: reject H_0 if $p < \alpha$ (commonly $\alpha = 0.05$). Report t(df) = value, p, 95% CI, and d_{cohen}.

1. Wilcoxon signed-rank test (if normality is violated)

Use the Wilcoxon signed-rank as a robust non-parametric alternative.

- Exclude pairs where di = 0d_i = 0. Let n′n' be the number of nonzero differences.
- For each remaining pair, compute |di||d_i| and assign ranks rir_i (average ranks for ties).
- Let R^+ be the sum of ranks for positive differences and R^- the sum for negative differences:
- **Test statistic**:

$$\boldsymbol{W = min(R^+, R^-)} \tag{24}$$

- For small n′ (typical cutoff n′≤20), compute the exact p-value from the Wilcoxon distribution. For larger n′, use normal approximation:

Mean and variance of W:

$$\mu_w = \frac{n'(n'+1)}{4} \tag{25}$$

$$\sigma_w \sqrt{\frac{n'(n'+1)(2n'+1)}{24}} \tag{26}$$

Standardized z (with continuity correction of 0.5 recommended):

$z = \frac{W - \mu_w - 0.5}{\sigma_w}$ (for two-sided test, use sign as appropriate).

Two-sided p-value:

$$p = 2(1 - \Phi(|z|)) \tag{27}$$

where Φ\Phi is the standard normal CDF.

- **Effect size for Wilcoxon**:

$$r = \frac{z}{\sqrt{n'}} \tag{28}$$

Interpret r as: 0.1 small, 0.3 medium, 0.5 large.

Decision rule: reject H_0 if p < α. Report W (or z), p, and effect size r.

1. Notes on robustness and alternatives

- The paired t-test is reasonably robust to mild departures from normality; for heavy skewness or outliers, use Wilcoxon or permutation tests (exact paired permutation test) for added robustness.
- When the sample size (number of folds) is small, prefer exact Wilcoxon or permutation tests rather than asymptotic approximations.

Conclusion

This paper proposed a deep learning framework for classifying Parkinson's disease from WiFi signals. We collected a database of WiFi signals for PD classification in the proposed framework. The proposed framework comprises two novel CNN architectures: E^3-T and PD-RAN2. Both models are trained on the prepared database, and their hyperparameters are initialized via Bayesian Optimization (BO). Subsequently, the trained models are used in the testing phase, where deep features are extracted from both models and subsequently fused using a novel serial weighted attention technique. The fused features are finally classified using neural network classifiers. The MN^2 classifier achieved 97.78% accuracy after fusing deep features. Detailed ablation studies and comparisons are conducted with recent techniques, and the following points:

- Sensor-transformer-based proposed models, such as E^3-T, performed well on the selected WiFi data and achieved improved accuracy and precision in PD classification.
- The proposed model, PD-RAN^2, combined residual and Attention mechanisms for PD classification. The combination process helps improve accuracy and reduces the model's computational time during training.
- Fusion of deep features using the proposed serial-based weighted attention technique improved the accuracy and precision rate compared to simple serial and parallel fusion techniques.

However, the proposed models have some limitations: they are evaluated on a specific dataset, limiting their generalizability to real-world clinical settings, and their computational cost is relatively high, which may limit their deployment in low-resource environments. In the future, the proposed model will be validated on additional datasets generated from WiFi signals. In addition, the model's parameters will be further optimized for low-resource settings.

Acknowledgements The Authors would like to thank the Ongoing Research Funding program (ORF-2025-157), King Saud University, Riyadh, Saudi Arabia.

Author Contributions Zeeshan Habib, Muhammad Attique Khan, Zain Hussain , Nathan Ng, Ameer Hamza, Ahmed Ibrahim Alalahrani, Nasser Alalwan, Zeshan Iqbal. All authors contributed equally in this work.

Funding Open access funding provided by the Scientific and Technological Research Council of Türkiye (TÜBİTAK). The Authors would like to thank the Ongoing Research Funding program (ORF-2025-157), King Saud University, Riyadh, Saudi Arabia.

Data Availability The dataset for this work is available upon request.

Declarations

Competing interests The authors declare no competing interests.

References

1. Payne T et al. Multimodal assessment of mitochondrial function in Parkinson's disease. Brain. 2024;147(1):267–280.
2. Shanthini S, Chandrasekar A. Multimodal Parkinson's disease classification using voice signals and hand-drawn images based on deep learning enabled by chronological tangent search optimization. Biomed Signal Process Control. 2025;102:107274.
3. Wu X, et al. Natural Neurobiological Active Compounds in Parkinson's Disease: Molecular Targets, Signaling Pathways, and Therapeutic Prospects. Int J Mol Sci. 2026;27(3):1301.
4. Tanveer M, Rashid AH, Kumar R, Balasubramanian R. Parkinson's disease diagnosis using neural networks: Survey and comprehensive evaluation. Inf Process Manag. 2022;59(3):102909.
5. Schiess N, et al. Six action steps to address global disparities in Parkinson disease: a World Health Organization priority. JAMA Neurol. 2022;79(9):929–36.
6. Collaborators GN. Global, regional, and national burden of neurological disorders, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Neurol. 2019;18(5):459.
7. Chin-Chan M, Navarro-Yepes J, Quintanilla-Vega B. Environmental pollutants as risk factors for neurodegenerative disorders: Alzheimer and Parkinson diseases. Front Cell Neurosci. 2015;9:124.
8. Choi M, Jo J, Jeong J. Digital gait biomarkers for Parkinson's disease: subject-wise validated explainable AI framework using vertical ground reaction force signals. Bioengineering. 2026;13(3):360.
9. Havemann J. A life shaken: my encounter with Parkinson's disease. JHU; 2002.
10. Borzì L, Sigcha L, Olmo G. Context recognition algorithms for energy-efficient freezing-of-gait detection in Parkinson's disease. Sensors. 2023;23(9):4426.
11. Weiss A, Herman T, Giladi N, Hausdorff JM. New evidence for gait abnormalities among Parkinson's disease patients who suffer from freezing of gait: insights using a body-fixed sensor worn for 3 days. J Neural Transm. 2015;122:403–10.
12. Navita, et al. Gait-based Parkinson's disease diagnosis and severity classification using force sensors and machine learning. Sci Rep. 2025;15(1):328.
13. Vercruysse S, Gilat M, Shine JM, Heremans E, Lewis S, Nieuwboer A. Freezing beyond gait in Parkinson's disease: a review of current neurobehavioral evidence. Neurosci Biobehavioral Reviews. 2014;43:213–27.
14. Bikias T, Iakovakis D, Hadjidimitriou S, Charisis V, Hadjileontiadis LJ. DeepFoG: an IMU-based detection of freezing of gait episodes in Parkinson's disease patients via deep learning. Front Rob AI. 2021;8:537384.
15. Okuma Y, de Lima ALS, Fukae J, Bloem BR, Snijders AH. A prospective study of falls in relation to freezing of gait and response fluctuations in Parkinson's disease. Parkinsonism Relat Disord. 2018;46:30–5.
16. Michałowska M, FIszer U, Krygowska-Wajs A, Owczarek K. Falls in Parkinson's disease. Causes and impact on patients' quality of life. Funct Neurol. 2005;20(4):163–8.
17. Bloem BR, Hausdorff JM, Visser JE, Giladi N. Falls and freezing of gait in Parkinson's disease: a review of two interconnected, episodic phenomena. Mov disorders: official J Mov Disorder Soc. 2004;19(8):871–84.
18. Nutt JG, Horak FB, Bloem BR. Milestones in gait, balance, and falling. Mov Disord. 2011;26(6):1166–74.
19. Borrione P, Tranchita E, Sansone P, Parisi A. Effects of physical activity in Parkinson's disease: A new tool for rehabilitation. World J Methodol. 2014;4(3):133.
20. Liu WY, Tung TH, Zhang C, Shi L. Systematic review for the prevention and management of falls and fear of falling in patients with Parkinson's disease. Brain Behav. 2022;12(8):e2690.
21. Kumar K, Ghosh R. Parkinson's disease diagnosis using deep learning model by analyzing the channels of electroencephalography signals from substansia niagra and ventral tegmental area regions of human brain. Int J Inform Technol. 2025;11:1–24.
22. Giladi N, Hausdorff JM. The role of mental function in the pathogenesis of freezing of gait in Parkinson's disease. J Neurol Sci. 2006;248(1–2):173–6.
23. Vandenbossche J, et al. Freezing of gait in Parkinson's disease: disturbances in automaticity and control. Front Hum Neurosci. 2013;6:356.
24. Lee J, Kouznetsova VL, Kesari S, Tsigelny I. Selective diagnostics of Amyotrophic Lateral Sclerosis, Alzheimer's and Parkinson's Diseases with machine learning and miRNA. Metab Brain Dis. 2025;40(1):79.
25. Miyahara Y, Phokaewvarangkul O, Kerr S, Anan C, Toriumi H, Bhidayasiri R. Comparing the efficacy of therapeutic Thai acupressure on plantar acupoints and laser cane therapy on freezing of gait in Parkinson's disease: a randomized non-inferiority trial. Front Neurol. 2024;15:1327448.
26. Gan J, et al. Prevalence and clinical features of FOG in Chinese PD patients, a multicenter and cross-sectional clinical study. Front Neurol. 2021;12:568841.
27. Zoetewei D et al. On-demand cueing for freezing of gait in Parkinson's disease: a randomized controlled trial. Mov Disord. 2024;35(5):876–886.
28. Gao C, Liu J, Tan Y, Chen S. Freezing of gait in Parkinson's disease: pathophysiology, risk factors and treatments. Translational neurodegeneration. 2020;9:1–22.
29. Volkmann H, et al. MRI classification of progressive supranuclear palsy, Parkinson disease and controls using deep learning and machine learning algorithms for the identification of regions and tracts of interest as potential biomarkers. Comput Biol Med. 2025;185:109518.
30. Li W et al. PIDGN: An explainable multimodal deep learning framework for early prediction of Parkinson's disease. J Neurosci Methods. 2025;6(1):110363.
31. Kumar K, Ghosh R. A multi-modal Parkinson's disease diagnosis system from EEG signals and online handwritten tasks using grey wolf optimization based deep learning model. Biomed Signal Process Control. 2025;100:106946.

32. Dattola S, Ielo A, Quartarone A, De Cola MC. Integrating wearable sensor signal processing with unsupervised learning methods for tremor classification in Parkinson's disease. Bioengineering. 2025;12(1):37.
33. Morris HR, Spillantini MG, Sue CM, Williams-Gray CH. The pathogenesis of Parkinson's disease. Lancet. 2024;403(10423):293–304.
34. Xu X, Yang J, Hu D, Por LY, Li C. Diffusion model with relation-aware attention and edge-aware constraint for multi-modal brain tumor segmentation. IEEE J Biomedical Health Inf. 2025.
35. Pilli R, Goel T, Murugan R, Tanveer M. Brain age estimation of Alzheimer's and Parkinson's affected individuals using self-attention based convolutional neural network. In: International Conference on Pattern Recognition. Springer; 2024. pp. 81–96.
36. Abbas MJ, Hussain A, Ayouni S, Maddeh M, Alhayan F. XRD-Net: a Novel Explainable Residual Dense Fusion Network for Alzheimer's Disease Recognition from MRI Images. Cogn Comput. 2025;17(6):174.
37. Mushtaq M, Hussain Z, Ayouni S, Maddeh M, Alhayan F. A Network-Level fused DenseInc226 lightweight architecture for alzheimer's disease prediction from magnetic resonance imaging. Cogn Comput. 2025;17(6):1–25.
38. Baruah B, Kumar A, Halder A. Deep Sequential Learning Architecture for Parkinson's Disease Prediction using Amino Acid Descriptors from Protein Sequences. Cogn Comput. 2025;17(6):1–30.
39. Rivera-Guzmán EF, Guerrero-Vásquez LF, Robles-Bykbaev VE. Quantization of deep neural networks for medical image analysis: a systematic review and meta-analysis. Technologies. 2026;14(1):76.
40. Abbas MJ, et al. C3BAM-XAI: Convolutional Block Attention Module Enhanced Explainable Artificial Intelligence-Based Parkinson's Disease Stage Classification. Cogn Comput. 2025;17(3):111.
41. Queiroz D, Carlos A, Anjos A, Berton L. Fair foundation models for medical image analysis: Challenges and perspectives. ACM Trans Comput Healthc. 2026;7(2):1–24.
42. Tuncer T, Tasci I, Tasci B, Hajiyeva R, Tuncer I, Dogan S. TPat: Transition pattern feature extraction based Parkinson's disorder detection using FNIRS signals. Appl Acoust. 2025;228:110307.
43. Kleanthous N, Hussain AJ, Khan W, Liatsis P. A new machine learning based approach to predict Freezing of Gait. Pattern Recognit Lett. 2020;140:119–26.
44. Kim H et al. Unconstrained detection of freezing of Gait in Parkinson's disease patients using smartphone. In: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE; 2015. pp. 3751–3754.
45. Torvi VG, Bhattacharya A, Chakraborty S. Deep domain adaptation to predict freezing of gait in patients with Parkinson's disease. In: 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA). IEEE; 2018. pp. 1001–1006.
46. Orphanidou NK, Hussain A, Keight R, Lishoa P, Hind J, Al-Askar H. Predicting freezing of gait in Parkinsons disease patients using machine learning. 2018 IEEE Congress on Evolutionary Computation (CEC). IEEE; 2018. pp. 1–8.
47. Camps J et al. Deep learning for detecting freezing of gait episodes in Parkinson's disease based on accelerometers. In: Advances in Computational Intelligence: 14th International Work-Conference on Artificial Neural Networks, IWANN 2017, June 14–16, 2017, Proceedings, Part II 14. Springer; 2017. pp. 344–355.
48. Camps J, et al. Deep learning for freezing of gait detection in Parkinson's disease patients in their homes using a waist-worn inertial measurement unit. Knowl Based Syst. 2018;139:119–31.
49. Wang G, Zou Y, Zhou Z, Wu K, Ni LM. We can hear you with WiFi! In: Proceedings of the 20th annual international conference on mobile computing and networking. 2014. pp. 593–604.
50. Habib Z, Mughal MA, Shabaz M. WiFOG: Integrating deep learning and hybrid feature selection for accurate freezing of gait detection. Alexandria Eng J. 2024;86:481–93.
51. Victoria AH, Maragatham G. Automatic tuning of hyperparameters using Bayesian optimization. Evol Syst. 2021;12(1):217–23.
52. Ross AA, Govindarajan R. Feature level fusion of hand and face biometrics. Biometric technology for human identification II. SPIE. 2005;5779:196–204.
53. Saha U, Ahamed IU, Ahamed IU, Hossain A-A. Graph convolutional network-based approach for parkinson's disease classification using euclidean distance graphs. In: 2024 7th International conference on informatics and computational sciences (ICICoS). IEEE; 2024. pp. 532–537.
54. Garehdaghi F, Sarbaz Y. A robust method for parkinson's disease diagnosis: Combining electroencephalography signal features with reconstructed phase space images. Med Eng Phys. 2025;135:104276.
55. Camacho M, Wilms M, Almgren H, Amador K, Camicioli R. Exploiting macro-and micro-structural brain changes for improved Parkinson's disease classification from MRI data. npj Parkinsons Dis; 2024.