APSIPA Transactions on Signal and Information Processing, 2023, 12, e202

Overview Paper

Plantar Space-Gait Cycle Transformer for Early Parkinson Disease Detection

Xiaoyue Wang[1,2], Teng Li[1], Haoqiang Hua[3], Lin Shu[3] and Xiaofen Xing[3*]

[1]*AHU-IAI AI Joint Laboratory, Anhui University, China*
[2]*Institute of Artificial Intelligence, Hefei Comprehensive National Science Center, China*
[3]*South China University of Technology, China*

ABSTRACT

Parkinson's disease (PD) is a chronic and long-term disease that seriously affects patients' quality of life. In underdeveloped areas, early detection of PD is primarily based on medical observation and patient self-description. Early diagnosis of PD can effectively reduce the disease's progression. Recent studies have suggested that the motor symptoms of PD can be reflected in plantar pressure. However, traditional machine learning models require manual feature selection, which can be time-consuming. Furthermore, although deep learning has seen rapid development, many clinical characteristics have not been taken into consideration. To address these limitations, a dual self-attention Transformer model is proposed to explore the spatial correlation of plantar space and the temporal correlation of the gait cycle. Considering the presence of symptoms such as foot tremors in PD patients, a masking mechanism is designed to focus locally on the unilateral foot during the support phase. An experimental paradigm is designed to evaluate the model's generalization capability across different subjects. The experimental results demonstrate that the proposed model achieves superior classification performance for the early detection of PD based on plantar pressure data.

*Corresponding author: Xiaofen Xing, xfxing@scut.edu.cn.

Received 28 April 2023; Revised 01 August 2023
ISSN 2048-7703; DOI 10.1561/116.00000227

Keywords: Parkinson, Transformer, Self-Attention, Plantar Pressure, VGRF.

1 Introduction

Parkinson's disease (PD) is the second most common neurodegenerative disease, after Alzheimer's disease (AD) [22]. It was first described by James Parkinson in his essay published in 1817 [28]. There are currently over 6 million PD patients worldwide, and the majority of them are middle-aged or older adults [5]. The disease is caused by a deficiency of dopamine neurons in the substantia nigra of the brain [18]. While a cure for PD is not currently known, treatments are available to alleviate the symptoms associated with it. Consequently, computer-assisted techniques have emerged to assist in the diagnosis and monitoring of PD.

Abnormal gait is one of the most apparent symptoms of PD progression [21, 30]. Micó-Amigo *et al.* [25] demonstrated that gait characteristics can serve as a predictor of an increased risk of conversion to PD. Since chronic abnormal gait is elusive, early self-detection and diagnosis of PD are challenging. As a result, regular gait assessments in specific populations become critical.

Presently, a clinical examination for PD patients relies heavily on patient self-reports and questionnaires conducted by clinicians using tools such as the Unified Parkinson's Disease Rating Scale (UPDRS) [32], the Freezing Of Gait Questionnaire (FOG-Q) [12], and the Hoehn&Yahr scale (H&Y scale) [13]. However, this method is highly subjective, time-consuming, and challenging for professionals. The variability of results between different physicians constrains the diagnostic and monitoring processes.

Brain imaging techniques, such as Positron Emission Tomography (PET) and Single-Photon Emission Computed Tomography (SPECT), are capable of detecting PD. However, these techniques have their inherent limitations. They rely on expensive and specialized medical equipment, which makes them difficult to use widely. It hinders their application in large-scale screening.

In recent years, wearable sensor technology has rapidly advanced, with miniaturized and portable devices being developed for daily use, especially in remote health monitoring [15, 29, 37]. Among these devices, wearable plantar pressure shoes have emerged as a promising tool that is portable, low-cost, and enables large-scale data collection compared to brain image techniques. They allow for the continuous, non-intrusive collection of Vertical Ground Reaction Force (VGRF) data, enabling tracking of daily activities and disease progression in PD patients [8]. VGRF is the force exerted by the ground on a person's foot during walking or any other weight-bearing activity. Due to possible motor impairments, postural balance issues, and other problems in PD patients, their walking patterns may exhibit greater individual variations. As

shown in Figure 2, it is evident that the walking pattern of the PD patient does not exhibit a distinct bimodal pattern typically observed in healthy individuals. This indicates an abnormal distribution of forces during walking in the PD patient. In addition, plantar pressure has been utilized to explore fall risk assessment in the elderly in several studies [33, 35].

Regarding the PD diagnosis task, there are many computer-assisted diagnosis systems based on plantar pressure data, most of which are hand-crafted features combined with traditional machine learning methods. Balaji *et al.* [7] extracted several temporal and spatial features from the data, which were utilized as inputs for PD diagnosis and severity rating through some supervised machine learning models. Abdulhay *et al.* [1] extracted various time features, such as stride time, stance phase, and swing phase, then employed a Support Vector Machine (SVM) to differentiate between PD patients and healthy controls. The machine learning approach of manual feature extraction requires more complex and time-consuming feature selection.

Deep learning methods for the task of PD diagnosis have gained significant momentum recently. One of the advantages that these methods offer over traditional machine learning methods is that they eliminate the need for manual analysis and selection of characteristic parameters. This advantage becomes more pronounced as the data volume and complexity of problems increase. For instance, Zhao *et al.* [39] proposed a two-channel model that combined Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) to extract temporal and spatial information from VGRF data. El Maachi *et al.* [9] utilized 1D CNNs to analyze 18 1D VGRF data points simultaneously from the soles for PD detection. Transformer has been successfully applied in the field of natural language processing and image recognition in recent years. Nguyen *et al.* [27] proposed a new method for PD detection based on gait analysis using automatic feature extraction through 1D Transformer. However, these methods do not adequately consider the clinical characteristics of plantar pressure data. Furthermore, CNN and LSTM architectures may struggle to capture the global context and manage long-term dependencies effectively, potentially impacting their performance in PD diagnosis. Therefore, this paper proposes a PD detection algorithm based on a dual self-attention mechanism. The mechanism is based on the idea of the Transformer model [34]. The parallel extraction of spatio-temporal features from plantar pressure data enhances analysis and preserves important information. Additionally, the proposed model incorporates a masking mechanism to consider clinical characteristics. The proposed approach leverages the advantages of the Transformer model, such as capturing long-range dependencies and providing interpretability, surpassing the limitations of LSTM. The cross-subject experimental results on a public dataset demonstrate the state-of-the-art performance of the proposed model, achieving an accuracy of 82.35% and a sensitivity of 86.77%.

In summary, the main contributions of this paper are as follows:

1. Considering the spatial-temporal characteristics of the plantar pressure data and the Transformer model, a novel Transformer model is designed to explore the variability characteristics of plantar pressure in PD patients. In particular, a dual self-attention encoder obtains the spatial correlation of plantar space and the temporal correlation of the gait cycle.

2. For the clinical characteristics of unilateral foot tremors that appear early in PD, the masking mechanism of channel self-attention is designed so that different heads use different masks in multi-head self-attention. Thus, the model gives more local attention to the spatial correlation of the unilateral foot in the support phase.

3. In terms of experimental design, most of the current work is not strictly cross-subject, which may lead to data leakage. Therefore, this paper adopts a strict cross-subject experimental paradigm and validates the cross-subject generalization of the model using a public dataset.

2 Relate Work

In detecting PD motor symptoms, many studies [9, 27, 36, 39] have been conducted on plantar pressure data. Plantar pressure analysis is a non-invasive and cost-effective method used to assess foot function and diagnose pathologies related to the foot. It provides valuable information about pressure distribution and timing during gait, which can aid in the evaluation of therapeutic interventions and improve patient outcomes [31].

The plantar pressure space is the distribution of forces between the human foot sole and the contact surface during walking or movement. By measuring and analyzing plantar pressure data, the force on the sole of the foot in different regions can be obtained, including pressure magnitude, time variation, and force distribution maps [23]. Additionally, studying the changes in pressure across the plantar aspect at different time points and regions enables the extraction of gait cycle information [26]. Figure 4 illustrates a standard gait cycle for a normal individual. Most studies have also extracted spatio-temporal feature information from this plantar space and gait cycle. We categorize the relevant methodologies into traditional machine learning methods and deep learning methods.

Traditional machine learning-based methods: The detection of PD motor symptoms currently rely on hand-crafted features combined with traditional machine learning techniques. Alam *et al.* [2] selected meaningful features based on sequential forward feature selection, oscillation time, step time variability, and center of pressure features and used different classifiers

to build models to classify PD patients and healthy controls. Farashi [10] used a comprehensive set of temporal, frequency, and time-frequency features extracted from VGRF data to distinguish. Zhao *et al.* [40] used the Ensemble K-Nearest Neighbor algorithm for the diagnosis of PD severity. Alkhatib *et al.* [3] performed a spatial and temporal signal analysis of VGRF data to classify gait as balanced and unbalanced. Simple features such as correlations were then used to further distinguish between balanced-health controls and balanced-PD patients. Finally, a linear decision boundary was used for classification. However, these methods have their limitations, as they require expert knowledge and are not robust to data distribution changes or domain shifts.

Deep learning-based methods: Deep learning has been rapidly developing and has shown great potential in PD detection. A significant advantage of deep learning models is their ability to automatically extract features. For instance, Xia *et al.* [36] believed that the left and right lower limbs have different kinematic characteristics during walking. Therefore, CNN and LSTM were used to model the left and right feet respectively, and finally the extracted features were combined for classification. In a similar way, Liu *et al.* [24] proposed a two-branch hybrid model for PD diagnosis. The model extracted spatial and temporal information using CNN and LSTM in the left and right feet, respectively. After the features were retrieved from the left and right feet, the LSTM model was used to merge them. Alle and Priyakumar [4] employed a 1D CNN with deeply separable convolutions to diagnose PD. They derived discriminative patterns from VGRF data by using linear prediction residuals (LPR). Aversano *et al.* [6] used deep neural networks for early PD detection using plantar pressure sensor data. Zhao *et al.* [39] proposed a two-channel model that combined CNN and LSTM to extract temporal and spatial information. El Maachi *et al.* [9] constructed a Deep Neural Network (DNN) classifier using a one-dimensional CNN and obtained the final classification results by parallelizing the 1D-CNN and the fully connected network. Nguyen *et al.* [27] used Transformer to get spatio-temporal features for PD detection. In addition, Jane *et al.* [20] proposed a Q backpropagation time delay neural network (QBTDNN) model for diagnosing the severity of gait disorders. These deep learning methods show great potential for early PD detection because they can capture subtle changes in the data and achieve higher accuracy than traditional machine learning methods.

It can be seen that most existing deep learning methods mostly explore spatio-temporal features with CNN and LSTM, and the model design does not take clinical characteristics well into account. The idea of this paper is to take advantage of the Transformer to capture the spatio-temporal features of data more effectively. And considering the clinical characteristics, improve the capture of important gait features.

3 Methodology

3.1 Method Overview

The overall architecture of the proposed Plantar Space-Gait Cycle Transformer (PSGCTR) model is depicted in Figure 1. The purpose of the proposed model is to distinguish whether a subject is a healthy control or a PD patient based on plantar pressure. In contrast to existing methods that often overlook clinical characteristics, our model incorporates relevant clinical characteristics to enhance performance. Through a dual self-attention mechanism with a mask, the model can pay more attention to important feature information. The proposed PSGCTR model comprises five main components, as follows: data preprocessing module, data input module, dual self-attention encoders module, classification module, and voting module.

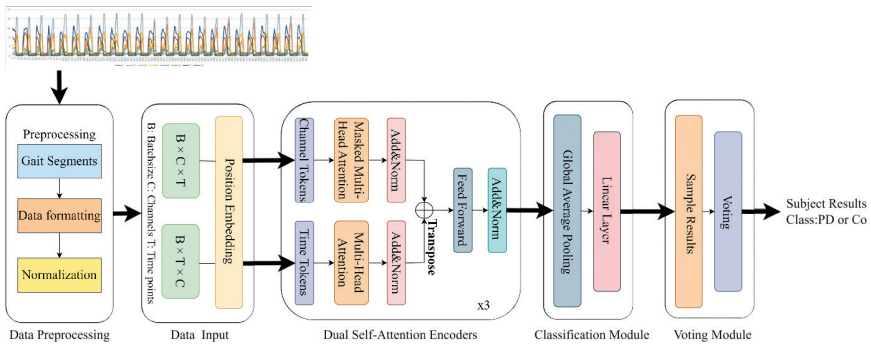

Figure 1: The Overall architecture of the proposed Plantar Space-Gait Cycle Transformer (PSGCTR) for PD detection. Each basic block of the PSGCTR mainly consists of five main components: data preprocessing, data input, dual self-attention encoders, classification, and voting.

The data preprocessing module slices the raw data into gait cycles and normalizes them. The data input module is divided into two forms: sensor channels and time points as tokens, respectively. Tokens are embedded into the network with corresponding positional information. The dual self-attention encoders module explores the correlation between plantar space and gait cycle time points. Lastly, the classification module obtains sample classification results and votes to obtain the subject results. Further details about each of these components are presented in the subsequent sections.

3.2 Data Preprocessing

The gait cycle refers to the time interval between consecutive heel strikes of the same foot during walking [19]. It comprises two primary phases, namely

the stance phase, which covers the interval when the foot is in contact with the ground, and the swing phase, which encompasses the interval when the foot is not in contact with the ground. Figure 4 presents an image of the entire gait cycle of an individual with a normal walking pattern.

In this study, each subject collected vertical ground reaction force (VGRF) data through eight plantar pressure sensors on both feet, with the total VGRF data from all eight sensors computed. The data is presented as a two-dimensional matrix of size $18 \times L$, where L denotes the duration of data collection. This matrix captures the pressure variations of the sensors located at different positions on the soles of both feet over time, encompassing the entire data collection process. In Figure 2, it can be seen that the VGRF data of PD patients may exhibit irregular fluctuations and lack a clear bimodal pattern. This may be due to impaired motor control in the PD patient, resulting in variability of the gait cycle and instability of walking. Based on this observation, we utilize prior knowledge to preprocess the data, enhancing its quality and making it ready for future analysis.

In the human body, there exist substantial differences in body weight between males and females, which can lead to notable variations in the observed vertical ground reaction forces (VGRF) among different subjects. To eliminate the potentially confounding effect of body weight on VGRF measurements, we perform a weight removal operation on the VGRF data. This process

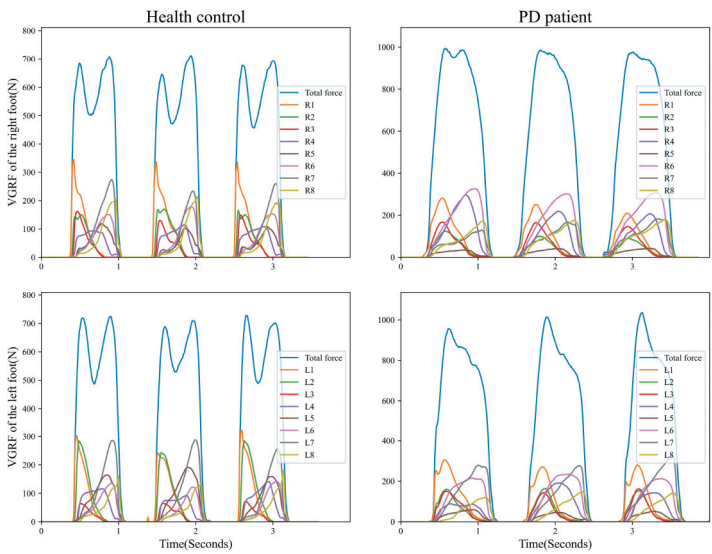

Figure 2: A graphical representation of the vertical ground reaction force (VGRF) data for the left and right feet of a typical PD patient and a healthy individual.

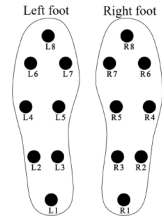

Figure 3: The force-sensitive sensors are positioned on the left and right feet.

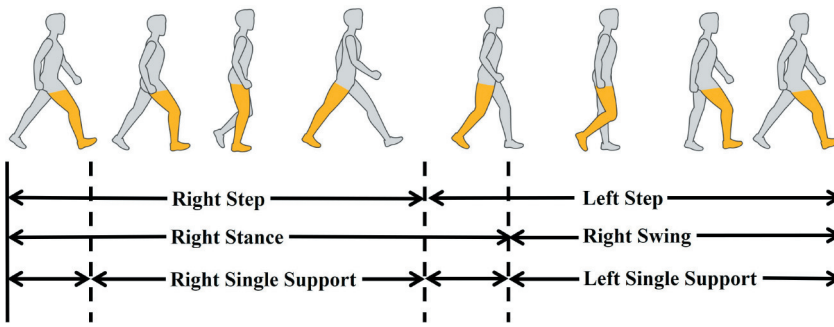

Figure 4: A diagram of a standard gait cycle.

involves subtracting the participant's body weight from the raw VGRF data to obtain net VGRF data that better reflects the actual forces exerted during gait.

In existing works [9, 27, 39], a fixed-length sliding window is commonly used to segment the data. However, since human walking is a periodic movement based on the gait cycle, it would be more reasonable to divide the subject's data into multiple samples according to the gait cycle. This segmentation method based on the gait cycle can make the data more aligned. Consequently, it enhances the reliability and accuracy of subsequent analyses and models.

Therefore, we proceed to split the data acquired from each participant into distinct samples based on the gait cycle to facilitate the analysis of the characteristics of gait. Additionally, we conduct an experimental investigation to examine the influence of sample length on the performance of our proposed model. Finally, to address the issue of imbalanced class samples after splitting the data, we utilize the Borderline-SMOTE algorithm [16] to oversample the healthy control group samples. This approach helps mitigate prediction bias problems that may arise in the model.

3.3 Position Embedding (PE)

The aim of the original Transformer model was to overcome the deficiency of inherent sequence information in Recurrent Neural Networks (RNNs). This model employs a different method to encode the positional information between sequence elements as a way of capturing long-range dependencies more effectively. In our proposed approach, the location information of each sensor time series is added before encoding to enhance the model's ability to comprehend the structure and context of the input sequence. Therefore, we follow the method in [34] to add relative position markers so that the model can make full use of the position information embedded in the sequences. The specific formula is as follows:

$$PE(pos, 2t) = sin\left(\frac{pos}{10000^{2t/d_{model}}}\right) \tag{1}$$

$$PE(pos, 2t+1) = cos\left(\frac{pos}{10000^{2t/d_{model}}}\right) \tag{2}$$

pos is the sensor position index, t is the time step, and d is the dimension of the embedding vector. At each time position of the vector, PE at even and odd time points is described by sine and cosine functions, respectively.

3.4 Plantar Space-Gait Cycle Dual Self-Attention

The Transformer model [34] was originally developed for natural language processing tasks, but it has been successfully applied in various domains, including computer vision and signal processing. The Transformer model utilizes a self-attention mechanism to capture global dependencies between different elements in the input sequence, which makes it particularly effective for modeling long-range dependencies. In the proposed Plantar Space-Gait Cycle Transformer (PSGCTR) model, dual self-attention is designed to capture the correlations between each sensor and each time point, enabling the model to effectively capture the complex interdependencies within the input signal. Additionally, a masking mechanism is incorporated based on clinical characteristics to improve local attention to the unilateral plantar space, which further enhances the model's ability to focus on important features.

The self-attention mechanism in the Transformer model calculates the importance of each element in a sequence with respect to all other elements, allowing the model to capture global dependencies between them. The PSGCTR model that extends the self-attention mechanism by introducing dual self-attention that considers both sensor-channel and time-point dependencies is proposed in this study. Specifically, the PSGCTR model consists of two modules: channel self-attention (CSA) and time self-attention (TSA). The CSA operates on the channel tokens, which represent the sensor data input vector with the shape $\epsilon R^{c\times l}$, where c and l represent the number of sensor

channels and length of the time series, respectively. The TSA operates on time tokens, which are the transposition of the channel tokens. Multi-head self-attention (MSA) is then applied to both the CSA and TSA, allowing the model to perform multiple self-attention operations in parallel.

Time Self-Attention (TSA): The TSA is a technique used to explore correlations between distinct time steps. Since a single sample after segmentation contains multiple steps, TSA can help to discern correlations between consecutive gait cycles and capture dependencies between different time steps. It is a useful technique to investigate the temporal relationships in VGRF data effectively. By incorporating the TSA module, the PSGCTR model enhances its ability to capture the temporal dependencies present in VGRF (Vertical Ground Reaction Force) data. In this module, time points are treated as tokens, where each token's dimension corresponds to the VGRF values of different sensors at the same time point. This enables us to perform self-attention calculations on the relationships between time points, effectively exploring the temporal dependencies within the VGRF data from all sensors. Similar to the self-attention process described in [34]. The self-attention operation in TSA uses a set of weight matrices to obtain the query vector (Q), key vector (K), and value vector (V) from the input vector. The attention weights are then calculated by computing the dot product between Q and K and dividing it by the square root of the feature dimension. The self-attention calculation formula can be expressed as follows:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{3}$$

where d_k is the key dimensionality and T is transposed. MSA is a powerful mechanism that allows the model to attend to multiple parts of the input sequence simultaneously by dividing the input into multiple smaller components. By breaking down the input into smaller components, each head can focus on distinct features, enabling the model to capture intricate dependencies and interactions between different input components. Additionally, MSA allows the model to attend to different "representation subspaces" of the input, which can be useful for tasks where different aspects of the input are relevant at different positions. In this method, the input is divided into smaller components, referred to as heads, which will perform attention simultaneously. The formula can be explained as follows:

$$MSA = Concat(head_1, \ldots, head_w)W^o \tag{4}$$

$$head_i = Attention(Q_i, K_i, V_i) \tag{5}$$

where i=head number.

Channel Self-Attention (CSA): The CSA is a mechanism that aims to explore the correlations between different channel tokens. This method

enables the model to capture the dependencies and relationships between different sensors, thereby extracting the most important information. In contrast to TSA, CSA employs a masking mechanism that integrates clinical prior knowledge.

The masked multi-head self-attention (MMSA) mechanism incorporates both preprocessed data characteristics and clinical prior knowledge. Each preprocessed sample contains three steps for the left foot and two steps for the right foot, as people typically alternate between their left and right feet while walking, as illustrated in Figure 4. However, PD patients may exhibit unilateral foot abnormalities during the support phase due to tremors, slow movement, and postural balance disorders. Therefore, the design is consistent with the characteristics of the sample. Different masks are used for different heads to enable the model to pay more local attention to the channel tokens of the support phase. Because only the plantar sensors in the support phase cycle have data generation, while the swing phase has no data generation because the foot is suspended. So we need to get the data of the support phase cycle more accurately. This method is highly relevant to the preprocessed sample information since normal individuals alternate between their left and right feet in the same pattern. By masking the channel tokens of the right foot plantar space when the left foot is used as the support phase, the model can capture different "representation subspaces" more effectively and better distinguish the difference between normal controls and PD patients.

Figure 5 illustrates the process of calculating the multi-mask QK^T matrix in MMSA. In contrast to the self-attention mechanism process described in [34]. This process, known as masked self-attention, enables the model to attend to relevant information while disregarding irrelevant information. The formula for masked self-attention is as follows:

$$Mask\ Attention(Q, K, V) = softmax(M \times \frac{QK^T}{\sqrt{d_k}})V \tag{6}$$

MMSA differs from MSA in that it uses masked self-attention for different heads, which are then concatenated. Next, the features obtained from MSA and MMSA are summed after layer normalization (LN) and residual concatenation to obtain the fused features. Finally, the features are fed into the feedforward neural network to get the final result of this encoder. The encoder process can be expressed as follows:

$$x' = (LN(MMSA(x)) + x) + (LN(MSA(x^T)) + x^T)^T \tag{7}$$

$$y = LN(MLP(x')) + x' \tag{8}$$

where x represents the input of the encoder and y represents the output of the encoder. After getting the encoder output result, it will pass into the next encoder.

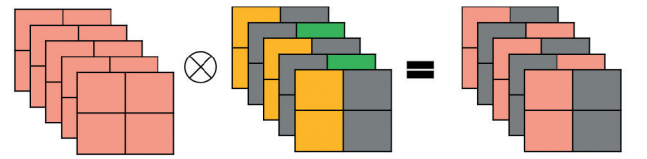

Figure 5: A multi-mask QK^T matrix.

3.5 Voting Module

This voting module is designed to decide whether the subject is a PD patient or not. Due to the small number of subjects in the dataset, the data collected from each subject was too long. Therefore, existing works [9, 24, 27, 36, 39] split the data of a subject into multiple segmented samples, and our preprocessing method does the same. The input of the network is a segmented sample, and the output of the network is the classification result of the segmented sample. Therefore, after getting the classification results of all the segmented samples of an object, we follow the majority vote to determine whether the object is PD or not. Specifically, if more than 50% of all the segmented samples of this object are predicted to be PD, then this object will be predicted to be PD.

4 Experiments

4.1 Dataset

In this study, the dataset used is obtained from PhysioNet[1] [14], which is a public repository of physiological and biomedical data. The dataset consisted of contributions from three different researchers, namely Ga [38], Ju [17], and Si [11]. The three researchers had different collection times and collection subjects, but the same collection equipment. The main difference was that the Ju and Si groups recorded normal walking at a self-selected speed. Ga repeated this part and added an additional task for each participant in which they performed a dual task while walking [38].

Participants were instructed to walk normally for approximately two minutes while wearing shoes equipped with pressure sensors. These shoes had eight pressure sensors on each foot, and their locations are depicted in Figure 3. The pressure data were sampled at a frequency of 100 Hz. Table 1 shows the demographic statistics of the subjects included in the dataset.

[1] https://physionet.org/content/gaitpdb/1.0.0/.

Table 1: Statistics about the dataset.

Groups	Total subjects	Male	Female
PD patients	93	58	35
Healthy controls	73	40	33

4.2 Experimental Setup

All experiments were performed on NVIDIA gtx1080ti GPU. We trained the baseline networks according to the implementation procedures described in the respective papers. In most existing works [9, 24, 36, 39], cross-subject training and testing were often not performed, which could have led to overfitting and inaccurate evaluation results. To address this issue, we adopt a strict cross-subject method for the experiments. Specifically, the dataset is randomly split into training, validation, and testing sets based on the subject ID, with a ratio of 8:1:2 for each class. This ensures that samples from the same subject do not appear in both the test and training sets. To reduce the impact of randomness, each experiment is repeated 10 times, and the average result is used as the final performance metric. The cross-entropy loss function is optimized using the Stochastic Gradient Descent (SGD) algorithm with a momentum of 0.9. Training uses 100 periods with early stops. Early stops are monitored using validation loss. The remaining model parameters are shown in Table 2. Subject-level results are obtained by averaging the results of individual samples belonging to the same subject, with a threshold of 50% used for classification. Additionally, experiments without cross-subject evaluation are also conducted to demonstrate the importance of cross-subject evaluation. The results are evaluated by specificity (Sp), sensitivity (Se), and accuracy (Acc).

Table 2: Parameters of the model.

Parameter	Value	Parameter	Value
Input shape	64×1×18×390	Channel heads	5
Batchsize	64	Time heads	2
Encoders	3	Mlp_dim	1024
Learning rate	0.001	Dropout	0.2

4.3 Results

4.3.1 Sample Length Analysis

One advantage of the Transformer model is its ability to capture long-range dependencies in input data. However, in the case of multi-step data such

as walking time, taking one step of gait splitting as the same sample may not account for the correlation between asynchronism, which refers to the time correlation between different gait cycles. To address this problem, we conducted experiments on asynchronism splitting as the sample and found that the model achieved the best performance when three steps of the left foot were taken as the sample. It should be noted that the masking mechanism was not used in this experiment.

When determining the length of the samples, previous studies have shown that the length of a gait cycle is typically between 100-160 steps [24]. Moreover, the stance phase is approximately 1.5 times longer than the swing phase [1]. Therefore, we set the length of a gait cycle to 150 steps, with 90 steps for the supporting phase and 60 steps for the swinging phase. To unify the lengths, zeros are used to fill those samples where the gait cycle length is less than the set length. The sequence lengths to be used in each step were calculated based on these parameters and are presented in Table 3.

Table 3: The influence of different sample lengths as input on the results.

Sample Size	*Sp(%)*	*Se(%)*	*Acc(%)*	*Length*
1	74.42	82.45	78.92	100
2	67.99	86.31	78.23	240
3	71.42	**86.46**	**79.83**	**390**
4	68.33	84.21	77.43	540

4.3.2 Comparison Analysis

The proposed method was compared with representative current methods. All the methods were classified into two groups based on the experimental design. The first group used a strict cross-subject experimental setup, while the other group used a non-cross-subject experimental setup to compare and emphasize the importance of cross-subject experiments. The cross-subject experimental design is to divide the training set and test set by subject ID. There were 132 subjects in the training set and 34 in the test set, with 55% PD patients. Due to conducting multiple experiments, the training set and test set subjects are not the same, and the number of collections per subject is also different. After dividing the number of samples, there are fluctuations in the training set and test set sample sizes, so the average training set sample is 22425 and test set sample is 5823.

In addition to our proposed method, we carefully select a set of baseline techniques that are widely recognized and representative of the existing methods used in the field of PD detection. These baseline techniques encompass

both deep learning models and traditional handcrafted feature models, all of which were evaluated on the same task and dataset as our proposed approach. We compared eight baseline techniques: four deep learning models (1D-Transformer [27], DUAL-CLSTM [24], 1D-Convnet [9], and CNN-LSTM [39]) and four traditional handcrafted feature models (Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Gradient Boosting Decision Tree (GBDT), and Random Forest (RF)). For the traditional methods, we employ feature selection as proposed by Alam *et al.* [2]. The results of the cross-subject experiments are shown in Table 4, and the non-cross-subject experiments are shown in Table 5. The results evaluation indices include Specificity (Sp), Sensitivity (Se), Positive Predictive Value (PPV), and Accuracy (Acc).

Table 4: Cross-subject results for different methods.

Methods	*Sp(%)*	*Se(%)*	*PPV(%)*	*Acc(%)*
1D-Convnet	78.18±9.9	79.91±11.5	82.68±7.4	79.14±7.7
1D-Transformer	70.02±13.7	81.57±7.8	77.55±8.2	76.47±5.6
CNN-LSTM	66.67±9.8	86.32±7.1	77.07±5.9	78.57±4.9
DUAL-CLSTM	71.43±9.7	84.21±8.1	78.89±7.2	78.78±4.2
SVM	73.98±8.3	82.89±7.3	80.29±5.5	78.94±6.3
KNN	82.85±6.1	71.42±3.8	83.11±4.9	76.46±2.2
GBDT	78.37±9.4	78.41±8.9	81.12±6.5	78.36±7.3
RF	75.55±13.1	79.83±8.5	80.31±8.8	77.94±6.1
PSGCTR(ours)	74.29±9.4	**86.77**±9.2	82.94±8.5	**82.35**±3.1

Table 5: Non-cross-subject results for different deep learning methods.

Methods	*Sp(%)*	*Se(%)*	*PPV(%)*	*Acc(%)*
1D-Convnet	92.92±2.1	99.12±0.3	96.82±1.5	97.23±0.5
1D-Transformer	94.95±0.8	98.87±0.4	97.61±0.6	97.67±0.4
CNN-LSTM	94.07±1.7	97.32±1.5	97.40±1.2	95.45±0.7
DUAL-CLSTM	94.06±1.1	97.34±0.8	97.38±0.9	95.25±0.5
PSGCTR(ours)	**95.45**±0.7	98.79±0.3	**97.99**±0.5	**97.77**±0.4

After comparing cross-subject experiments with those without cross-subject, we observed a significant performance difference of more than 10%. Therefore, cross-subject validation is essential for meaningful results and to avoid inflated outcomes. It ensures the model's ability to generalize effectively to unseen individuals and extends its performance beyond the training data.

Table 6: Ablation studies for various key components.

Methods	*Sp(%)*	*Se(%)*	*Acc(%)*
PSGCTR(ours)	74.29 ± 9.4	86.77 ± 9.2	82.35 ± 3.1
-M	71.42 ± 7.8	86.46 ± 4.1	79.83 ± 3.3
-TSA	65.99 ± 8.6	88.42 ± 7.7	78.52 ± 3.7
-CSA	76.19 ± 3.2	80.45 ± 6.6	78.57 ± 2.4

4.3.3 Ablation Study

In addition to the main experiment, we conducted ablation experiments to evaluate the effectiveness of key components in PSGCTR. Specifically, we removed the channel self-attention (-CSA) encoders, the temporal self-attention (-TSA) encoders, and the masking (-M) mechanism. The results of the ablation experiments are shown in Table 6.

As can be seen, removing either the -CSA or the -TSA encoders significantly decreases the performance of the model, indicating the importance of both types of self-attention. Removing the -M mechanism also results in decreased performance, suggesting that the masking is helpful in learning meaningful representations from the data. Overall, these ablation experiments demonstrate the effectiveness of the key components in PSGCTR.

4.3.4 Model Interpretability Analysis

The dual self-attention mechanism in PSGCTR can automatically learn the weights of different plantar spaces and gait cycle time points, which reflects their importance for PD detection. This allows professionals to focus on crucial plantar spaces and time points with the assistance of the weight information, which can improve the detection performance and model interpretability.

To further explicate this idea, we selected a PD patient sample and visualized the attention matrix in the CSA of the last encoder. The heat map of the attention matrix weights is shown in Figure 6 (a), which provides valuable insight into the decision-making process of the model and helps interpret its predictions. The heat map shows that the learned weights are concentrated in the lower right area, which is the right foot area. To confirm this, we presented the data of the right foot of the preprocessed sample, as depicted in Figure 6 (b). Based on the figure, it is evident that the subject exhibited a flat-footed landing and small broken step anomalies. Thus, we conclude that the sample of this subject is likely to exhibit abnormal performance on the right foot.

The distribution of samples in the dataset is illustrated in Figure 7, where the feature dimensions are reduced to 2 dimensions for visualization using the t-Distributed Stochastic Neighbor Embedding (t-SNE) technique. Figure 7 (a)

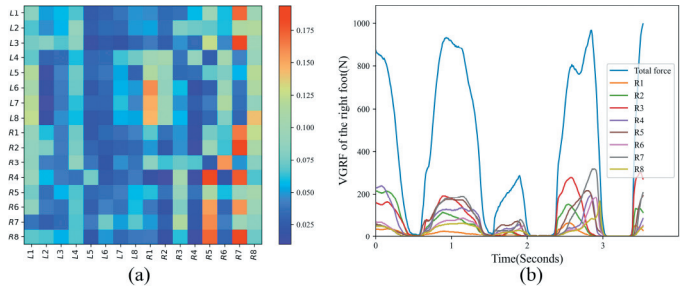

Figure 6: (a) The average weights of the heads in the CSA. (b) The right foot VGRF data.

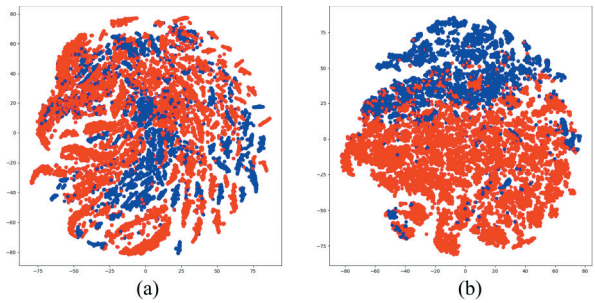

Figure 7: Visualization of the t-SNE results on the dataset, with samples of PD patients denoted by red dots and samples of healthy controls denoted by blue dots.

displays the feature representation of each sample after preprocessing. Due to the segmentation of each subject's data into multiple samples, the t-SNE visualization reveals that the samples from each subject tend to form distinct clusters in small localized regions. As a result, it becomes challenging to differentiate between PD patients and healthy controls using a linear classifier. In contrast, Figure 7 (b) shows that features extracted by the PSGCTR model can linearly separate PD patients and healthy control samples more effectively. The proposed PSGCTR model is able to better capture the PD-related features between samples, thus achieving better discriminatory ability.

4.3.5 Result Analysis

The sensitivity of the proposed PSGCTR model is 86.77%, indicating a high ability to correctly identify PD patients. However, false-negative results in medical diagnosis can be critical, potentially causing misdiagnosis and delaying treatment. To further evaluate the performance of the proposed model for early PD detection, we analyze the correctness percentage for different PD levels, using the Hoehn&Yahr scale (H&Y scale) recorded in the dataset.

However, the dataset only includes three PD grades. Therefore, we compare the performance of the proposed PSGCTR model with traditional machine learning methods.

Table 7 shows the correct percentage of different PD levels in the test set. The results indicate that the proposed PSGCTR model achieves the highest accuracy in all PD levels when compared to traditional machine learning methods, suggesting its superior performance in early PD detection. For example, for PD level 2, the proposed PSGCTR model achieves a correct percentage of 86.11%, while the traditional machine learning methods only achieve 66.35%–72.89%. This result demonstrates the potential of the proposed PSGCTR model for early PD detection, which can help clinicians diagnose PD patients earlier and provide better medical care.

Table 7: Severity Level statistics: PD results in the test set.

Methods	*2(%)*	*2.5(%)*	*3(%)*
1D-Convnet	71.58	87.74	96.67
1D-Transformer	64.53	97.14	100
CNN-LSTM	81.57	94.44	100
DUAL-CLSTM	80.05	93.12	100
SVM	72.27	96.65	95.83
KNN	66.35	79.82	80.95
GBDT	72.89	85.05	95.00
RF	69.75	95.29	88.85
PSGCTR(ours)	**86.11**	**97.14**	**100**

5 Discussion

In most of the existing works, rigorous cross-subject experiments were not done. Specifically, the data of all subjects were split into a number of samples according to certain rules and then directly disordered into training and testing sets. This results in the appearance of a training set and a test set with different samples of data from one subject at the same time. This practice will cause the result to be inflated. So we perform both cross-subject and non-cross-subject experiments, and we can find that the cross-subject experiment results will be more than 10% lower.

As shown in Table 4, the proposed PSGCTR model achieved the best performance. Compared to the deep learning models, the 1D-Convnet [9] and the 1D-Transformer [27] extracted time-domain features of different sensor channels through convolution and self-attention mechanisms, respectively,

and then fused all the information. However, neither of these approaches adequately explores the correlation between different sensor channels at the same moment, which is crucial for PD patients. For instance, PD patients may appear to land on their entire foot while walking. As in Figure 7 (b), all plantar sensor VGRF values reached maximum at the same moment. For traditional machine learning methods, gait statistical features are used [2]. For example, the Coefficient of Variation (CV) of swing time, standard deviation of Center of Pressure (CoP) of x-coordinate, and mean peak force at heel strike. The feature dimensions are relatively small. In addition, the hand-extracted features can only extract the global features, but cannot capture the spatio-temporal features with more subtle variations. Furthermore, the methods require manual feature extraction and professional knowledge. This approach is complex and time-consuming.

The proposed PSGCTR model is a Transformer model with a dual self-attention mechanism, which incorporates clinical characteristics. This model allows extracting of features from two spatial-temporal dimensions simultaneously. Compared to the 1D-Convnet [9] and the 1D-Transformer [27], by cascading. Our proposed model extracts the two-dimensional features in parallel and then fuses them. Our approach provides more comprehensive access to the spatio-temporal relationships in the data. And we use a different data segmentation method than in previous work, by using a cut-by-step approach rather than a sliding window cut. This preprocessing method is more reasonable and makes the different samples more aligned. And we discuss the effect of the number of sample steps on the model results, which are shown in Table 7.

However, this paper also has limitations. First, the number of subjects in the dataset is relatively small. More representative subjects are needed so that the generalization ability of the model can be better verified. Second, the model does not address the issue of individual variability among subjects well. In the gait data, there are differences among different subjects, which may be influenced by various individual factors such as individual height, weight, and personality. Finally, with only eight different position sensors provided per foot at the time of data acquisition, the spatial granularity may not be sufficient. And there may be other aspects of PD symptoms. Collecting only plantar pressure data may not be sufficient.

In real life, a false-positive result means misdiagnosing a healthy person as a PD patient, which may lead to a healthy person going to the hospital for additional medical tests. However, the task of our study was to perform early PD screening, and the final diagnosis needs to be confirmed by a specialized hospital. Accurate identification of individuals with PD is crucial in the early screening task, as early intervention and treatment can significantly improve the prognosis of patients. Although the sensitivity of the proposed PSGCTR model is higher than other methods, the specificity still needs to be improved.

6 Conclusion and Future Work

In this paper, we propose a PD detection model based on a transformer with a dual self-attention mechanism. The model takes into account the correlation between plantar space and gait cycle, while also considering clinical characteristics. We further introduce a masking mechanism to prioritize important information during the support phase. To evaluate the model's effectiveness, we conduct experiments using a cross-subject approach to assess its generalization ability.

For future work, we will be divided into two parts. Firstly, we aim to address the challenge of individual variability in plantar pressure data to enhance the robustness of our model. Second, we will collect and use a multimodal dataset that takes into account that PD symptoms are not limited to lower limb motor abnormalities, but may also include voice disorders, upper limb tremors, and memory loss. Multimodal data will be more favorable to distinguishing PD patients, thus further improving PD detection.

References

[1] E. Abdulhay, N. Arunkumar, K. Narasimhan, E. Vellaiappan, and V. Venkatraman, "Gait and tremor investigation using machine learning techniques for the diagnosis of Parkinson disease," *Future Generation Computer Systems*, 83, 2018, 366–73.

[2] M. N. Alam, A. Garg, T. T. K. Munia, R. Fazel-Rezai, and K. Tavakolian, "Vertical ground reaction force marker for Parkinson's disease," *PloS one*, 12(5), 2017, e0175951.

[3] R. Alkhatib, M. O. Diab, C. Corbier, and M. El Badaoui, "Machine learning algorithm for gait analysis and classification on early detection of Parkinson," *IEEE Sensors Letters*, 4(6), 2020, 1–4.

[4] S. Alle and U. D. Priyakumar, "Linear prediction residual for efficient diagnosis of Parkinson's disease from gait," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part V 24*, Springer, 2021, 614–23.

[5] M. J. Armstrong and M. S. Okun, "Diagnosis and treatment of Parkinson disease: a review," *Jama*, 323(6), 2020, 548–60.

[6] L. Aversano, M. L. Bernardi, M. Cimitile, and R. Pecori, "Early detection of parkinson disease using deep neural networks on gait dynamics," in *2020 International Joint Conference on Neural Networks (IJCNN)*, IEEE, 2020, 1–8.

[7] E. Balaji, D. Brindha, V. K. Elumalai, and K. Umesh, "Data-driven gait analysis for diagnosis and severity rating of Parkinson's disease," *Medical Engineering & Physics*, 91, 2021, 54–64.

[8] A. Channa, N. Popescu, and V. Ciobanu, "Wearable solutions for patients with Parkinson's disease and neurocognitive disorder: a systematic review," *Sensors*, 20(9), 2020, 2713.

[9] I. El Maachi, G.-A. Bilodeau, and W. Bouachir, "Deep 1D-Convnet for accurate Parkinson disease detection and severity prediction from gait," *Expert Systems with Applications*, 143, 2020, 113075.

[10] S. Farashi, "Distinguishing between Parkinson's disease patients and healthy individuals using a comprehensive set of time, frequency and time-frequency features extracted from vertical ground reaction force data," *Biomedical Signal Processing and Control*, 62, 2020, 102132.

[11] S. Frenkel-Toledo, N. Giladi, C. Peretz, T. Herman, L. Gruendlinger, and J. M. Hausdorff, "Treadmill walking as an external pacemaker to improve gait rhythm and stability in Parkinson's disease," *Movement disorders: official journal of the Movement Disorder Society*, 20(9), 2005, 1109–14.

[12] N. Giladi, J. Tal, T. Azulay, O. Rascol, D. J. Brooks, E. Melamed, W. Oertel, W. H. Poewe, F. Stocchi, and E. Tolosa, "Validation of the freezing of gait questionnaire in patients with Parkinson's disease," *Movement disorders: official journal of the Movement Disorder Society*, 24(5), 2009, 655–61.

[13] C. G. Goetz, W. Poewe, O. Rascol, C. Sampaio, G. T. Stebbins, C. Counsell, N. Giladi, R. G. Holloway, C. G. Moore, G. K. Wenning, *et al.*, "Movement Disorder Society Task Force report on the Hoehn and Yahr staging scale: status and recommendations the Movement Disorder Society Task Force on rating scales for Parkinson's disease," *Movement disorders*, 19(9), 2004, 1020–8.

[14] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals," *circulation*, 101(23), 2000, e215–e220.

[15] B. R. Greene, I. Premoli, K. McManus, D. McGrath, and B. Caulfield, "Predicting Fall Counts Using Wearable Sensors: A Novel Digital Biomarker for Parkinson's Disease," *Sensors*, 22(1), 2021, 54.

[16] H. Han, W.-Y. Wang, and B.-H. Mao, "Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning," in *Advances in Intelligent Computing: International Conference on Intelligent Computing, ICIC 2005, Hefei, China, August 23-26, 2005, Proceedings, Part I 1*, Springer, 2005, 878–87.

[17] J. M. Hausdorff, J. Lowenthal, T. Herman, L. Gruendlinger, C. Peretz, and N. Giladi, "Rhythmic auditory stimulation modulates gait variability in Parkinson's disease," *European Journal of Neuroscience*, 26(8), 2007, 2369–75.

[18] G. U. Höglinger, P. Rizk, M. P. Muriel, C. Duyckaerts, W. H. Oertel, I. Caille, and E. C. Hirsch, "Dopamine depletion impairs precursor cell proliferation in Parkinson disease," *Nature neuroscience*, 7(7), 2004, 726–35.

[19] M. Iosa, A. Fusco, F. Marchetti, G. Morone, C. Caltagirone, S. Paolucci, and A. Peppe, "The golden ratio of gait harmony: repetitive proportions of repetitive gait phases," *BioMed research international*, 2013, 2013.

[20] Y. N. Jane, H. K. Nehemiah, and K. Arputharaj, "A Q-backpropagated time delay neural network for diagnosing severity of gait disturbances in Parkinson's disease," *Journal of biomedical informatics*, 60, 2016, 169–76.

[21] J. Jankovic, "Parkinson's disease: clinical features and diagnosis," *Journal of neurology, neurosurgery & psychiatry*, 79(4), 2008, 368–76.

[22] A. I. Kaplin and M. Williams, "How common are the "common" neurologic disorders?" *Neurology*, 69(4), 2007, 410–1.

[23] F. Khan, M. F. Chevidikunnan, E. A. BinMulayh, and N. S. Al-Lehidan, "Plantar pressure distribution in the evaluation and differentiation of flatfeet," *Gait & Posture*, 101, 2023, 82–9.

[24] X. Liu, W. Li, Z. Liu, F. Du, and Q. Zou, "A dual-branch model for diagnosis of Parkinson's disease based on the independent and joint features of the left and right gait," *Applied Intelligence*, 2021, 1–12.

[25] M. E. Micó-Amigo, I. Kingma, S. Heinzel, S. M. Rispens, T. Heger, S. Nussbaum, R. C. van Lummel, D. Berg, W. Maetzler, and J. H. van Dieën, "Potential markers of progression in idiopathic Parkinson's disease derived from assessment of circular gait with a single body-fixed-sensor: a 5 year longitudinal study," *Frontiers in human neuroscience*, 13, 2019, 59.

[26] D. M. Mohan, A. H. Khandoker, S. A. Wasti, S. Ismail Ibrahim Ismail Alali, H. F. Jelinek, and K. Khalaf, "Assessment methods of post-stroke gait: A scoping review of technology-driven approaches to gait characterization and analysis," *Frontiers in Neurology*, 12, 2021, 650024.

[27] D. M. D. Nguyen, M. Miah, G.-A. Bilodeau, and W. Bouachir, "Transformers for 1D Signals in Parkinson's Disease Detection from Gait," *arXiv preprint arXiv:2204.00423*, 2022.

[28] J. Parkinson, "An essay on the shaking palsy," *The Journal of neuropsychiatry and clinical neurosciences*, 14(2), 2002, 223–36.

[29] A. B. Pinto, G. A. de Assis, L. C. Torres, T. Beltrame, and D. M. Domingues, "Wearables and detection of falls: a comparison of machine learning methods and sensors positioning," *Neural Processing Letters*, 54(3), 2022, 2165–79.

[30] M. Pistacchi, M. Gioulis, F. Sanson, M. Ennio De Giovannini, G. Filippi, F. Rossetto, and S. Z. Marsala, "Gait analysis and clinical correlations in early Parkinson's disease," *Functional neurology*, 32(1), 2017, 28.

[31] D. Rai and L. Aggarwal, "The study of plantar pressure distribution in normal and pathological foot," *Pol J Med Phys Eng*, 12(1), 2006, 25–34.

[32] M. D. S. T. F. on Rating Scales for Parkinson's Disease, "The unified Parkinson's disease rating scale (UPDRS): status and recommendations," *Movement Disorders*, 18(7), 2003, 738–50.

[33] Z. Song, J. Ou, L. Shu, G. Hu, S. Wu, X. Xu, and Z. Chen, "Fall risk assessment for the elderly based on weak foot features of wearable plantar pressure," *IEEE transactions on neural systems and rehabilitation engineering*, 30, 2022, 1060–70.

[34] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, 30, 2017.

[35] S. Wu, J. Ou, L. Shu, G. Hu, Z. Song, X. Xu, and Z. Chen, "MhNet: Multi-scale spatio-temporal hierarchical network for real-time wearable fall risk assessment of the elderly," *Computers in Biology and Medicine*, 144, 2022, 105355.

[36] Y. Xia, Z. Yao, Q. Ye, and N. Cheng, "A dual-modal attention-enhanced deep learning network for quantification of Parkinson's disease characteristics," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 28(1), 2019, 42–51.

[37] M. Yağanoğlu, "Real time wearable speech recognition system for deaf persons," *Computers & Electrical Engineering*, 91, 2021, 107026.

[38] G. Yogev, N. Giladi, C. Peretz, S. Springer, E. S. Simon, and J. M. Hausdorff, "Dual tasking, gait rhythmicity, and Parkinson's disease: which aspects of gait are attention demanding?" *European journal of neuroscience*, 22(5), 2005, 1248–56.

[39] A. Zhao, L. Qi, J. Li, J. Dong, and H. Yu, "A hybrid spatio-temporal model for detection and severity rating of Parkinson's disease from gait data," *Neurocomputing*, 315, 2018, 1–8.

[40] H. Zhao, R. Wang, Y. Lei, W.-H. Liao, H. Cao, and J. Cao, "Severity level diagnosis of Parkinson's disease by ensemble K-nearest neighbor under imbalanced data," *Expert Systems with Applications*, 189, 2022, 116113.