

# An Attention-based Temporal Network for Parkinson's Disease Severity Rating using Gait Signals

Huimin Wu<sup>1</sup>, Yongcan Liu<sup>1</sup>, Haozhe Yang<sup>1,2</sup>, Zhongxiang Xie<sup>1,2</sup>, Xianchao Chen<sup>1,2</sup>,  
Mingzhi Wen<sup>1,2</sup>, and Aite Zhao<sup>1\*</sup>

<sup>1</sup> College of Computer Science and Technology, Qingdao University  
Qingdao, 266071 China

[e-mail: 2021020693@qdu.edu.cn, liuyc15550884378@163.com, zhaoaite@qdu.edu.cn]

<sup>2</sup> Turing Innovation Team, Qingdao University  
Qingdao, 266071 China

[e-mail: 2019206624@qdu.edu.cn, 2019206688@qdu.edu.cn, 2019206773@qdu.edu.cn,  
2019206766@qdu.edu.cn]

\*Corresponding author: Aite Zhao

*Received March 1 2023; revised March 23, 2023; revised August 13, 2023; accepted September 11, 2023;  
published October 31, 2023*

---

## Abstract

Parkinson's disease (PD) is a typical, chronic neurodegenerative disease involving the concentration of dopamine, which can disrupt motor activity and cause different degrees of gait disturbance relevant to PD severity in patients. As current clinical PD diagnosis is a complex, time-consuming, and challenging task that relays on physicians' subjective evaluation of visual observations, gait disturbance has been extensively explored to make automatic detection of PD diagnosis and severity rating and provides auxiliary information for physicians' decisions using gait data from various acquisition devices. Among them, wearable sensors have the advantage of flexibility since they do not limit the wearers' activity sphere in this application scenario. In this paper, an attention-based temporal network (ATN) is designed for the time series structure of gait data (vertical ground reaction force signals) from foot sensor systems, to learn the discriminative differences related to PD severity levels hidden in sequential data. The structure of the proposed method is illuminated by Transformer Network for its success in excavating temporal information, containing three modules: a preprocessing module to map intra-moment features, a feature extractor computing complicated gait characteristic of the whole signal sequence in the temporal dimension, and a classifier for the final decision-making about PD severity assessment. The experiment is conducted on the public dataset PDgait of VGRF signals to verify the proposed model's validity and show promising classification performance compared with several existing methods.

---

**Keywords:** Parkinson's disease, Gait signals, Severity rating, Vertical ground reaction force, Transformer network.

## 1. Introduction

**P**arkinson's disease (PD) is a typical, chronic neurodegenerative disease involving the concentration of dopamine, a neurotransmitter in the brain with the ability to regulate and control emotion and muscles [1]. As the second most prevalent neurodegenerative disease after Alzheimer's, Parkinson's disease affects a large number of people severely and wreaks havoc on the quality of life for patients. With the persistent apoptosis of dopamine neurons in the midbrain areas which work with the mechanisms that release dopamine and project it to other cortical areas, these PD patients inevitably begin to experience deterioration in their manipulating capability of the muscle and manifest levels of dysfunctions including speech disabilities, gait abnormalities, attention-deficit disorders, and impulsivity [2]. Among these symptoms, gait abnormality is one of the major clinical characteristics of Parkinson's disease and its specific manifestation constantly evolves as the course of the illness progresses, which provides a significant foundation for synthesis diagnosis or state analysis. The current diagnosis of Parkinson's disease is commonly based on the specialist's observation and criteria of the patient's behavioral traits. Inevitably there are problems of subjectivity and time consuming with this mode of diagnosis, which demonstrates the significance of developing auxiliary diagnostic technology to monitor patient status and aid physician judgment.

Several kinds of subsidiary sensors [3], [4] are utilized in human gait data collection under various modalities with different working principles which helps with studies of PD auxiliary diagnosis. Among them, wearable sensors [5] worn on the human body that works by sensing inertia or pressure are more convenient and appropriate for practical application than those having requirements to be deployed in the environment which limits the subjects' scope of activity, due to their needs for daily living and long-term monitoring. Pressure sensors [6], [7] are efficiently applied to research by deploying them on the soles of the subject's feet measuring the vertical ground reaction force (VGRF) of the subject's natural walking.

Machine learning methods used in literature, such as random forest (RF) [8], support vector machine (SVM) [9], and naive Bayesian (NB) [10], have been explored fully to handle and analyze sensor signals for PD computer-assisted diagnosis and obtained promising achievement. Then, other studies developed sophisticated deep neural networks for PD severity recognition on account of gait signal data, and they exhibited considerable classification performance [11], [12]. Nevertheless, there are certain limitations in these approaches that most of them are not specifically designed for the data with temporal structures, whereas the gait signals recorded by wearable sensors hold important temporal information which is closely bound up with the severity of Parkinson's disease. In addition, some methods [6], [13] that are designed to be a recurrent structure tailored to sequential data inevitably incur high computational overhead when processing a long sequence.

Along with deep learning techniques progressing rapidly, the Transformer [14] was proposed with the goal of coping with long sequence input and reducing sequential computation in natural language processing, and has reached many other fields and performed well with its appropriate structural design whose self-attention mechanism breaks the rule of serial computation by calculating mutual correlation among feature vectors from intra-sequence time steps. The method of this structure and mechanism is eminently suitable for modeling gait sequential signals.

Hence, the Transformer is applied to the PD abnormal gait recognition task by creating an attention-based temporal network (ATN) with data input of VGRF signals in our work. Designed ATN contains three main submodules: a preprocessing module reducing the interference from outliers in source data and performing feature mapping within each time

step, a temporal feature extraction module to explore complex gait information from data in the direction of time series, and a decision classifier which maps gait-related discriminative characteristics into the category dimension to calculate the probability of belonging to each class and make the final decision about PD severity diagnosis.

The remainder of this paper is organized based on the following sections. Section 2 looks at the related works. Section 3 describes the proposed method. Section 4 gives a detailed presentation about the experiment setting and discusses the results. Finally, we conclude the whole paper and outline possible future works in Section 5.

## 2. Related Work

Advances in modern computer-aided diagnosis have been made thanks to the proposal and utilization of various machine learning algorithms, some of which have been explored and applied to PD diagnosis coupled with stage rating of Parkinson's disease under the supervision of several accepted integrated scales, such as Hoehn & Yahr (H&Y) [15] and unified Parkinson's disease rating scale (UPDRS) [16]. These studies mostly understand and explore gait-related sensor signal data by means of handcraft or automatic feature extraction.

Human gait can be regarded as a series of alternate movements of the lower limbs in a rhythmic pattern, which leads to the recorded sensor signal sequences necessarily exhibiting periodic characteristics consistent with the gait cycle [17]. Thus, some studies have taken this perspective to calculate several statistical parameters (such as median, mean, period length, standard deviation, and kurtosis) of sequential data as the feature vector for the distinction between PD patients and healthy subjects.

For example, Abdulhay *et al.* [18] manually calculated different statistics, including swing phases, stance time, and stride time, as temporal features from VGRF signals, and then use a medium Gaussian SVM to classify the extracted features and implement PD detection. In [19], statistical analysis was performed on gait time series data to identify salient features, and four different machine learning classifiers, including ensemble classifier (EC) [20], support vector machine (SVM) [21], Bayes classifier (BC) [22], and decision tree (DT) [23], were utilized to obtain the optimal classification performance.

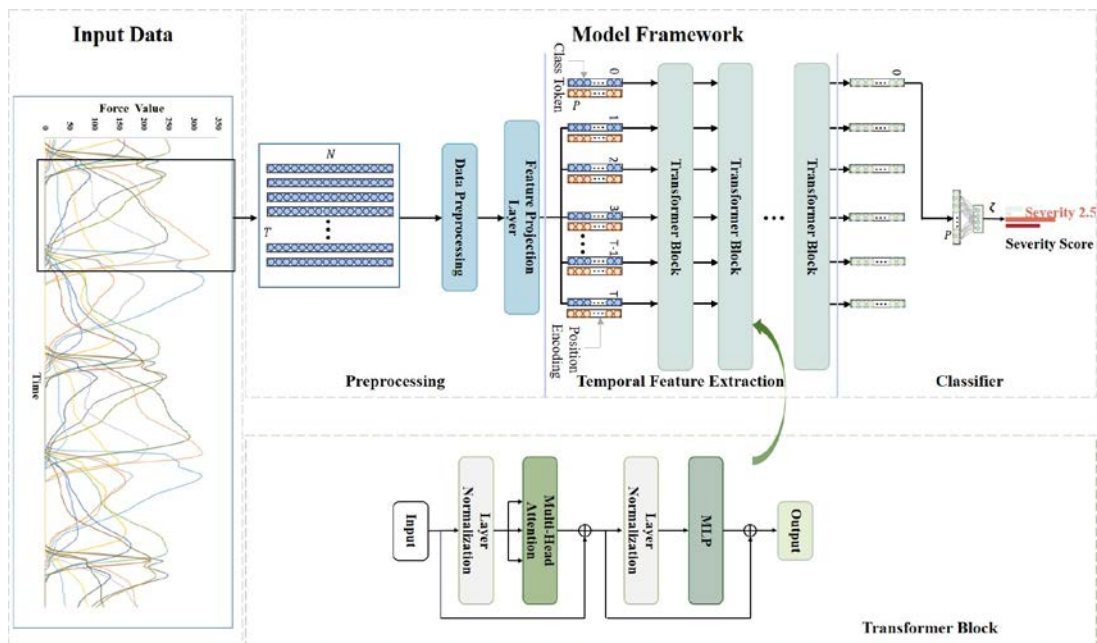
Driven by the shortfalls of handcraft methods, deep learning algorithms emerge with the ability to automatically select feature vectors in a way that minimizes the loss function. Recurrent and convolutional neural networks (RNNs and CNNs) are two kinds of typical network architectures for classification tasks [24], [25]. In [13], Flagg *et al.* utilized a bi-direction recurrent neural network (BiRNN) to extract gait features for the purpose of foot pressure data streaming evaluation. Nevertheless, the recursive computation order of RNNs retraced the past memory information using the output of the previous time step which makes its calculation clumsy and time-consuming. In order to improve the computational efficiency, Maachi *et al.* [26] utilized a 1D-CNN classifier and convolve along the temporal direction of the feature matrix A one-dimensional convolution kernel to integrate temporal information. This design can make models compute features for all time steps simultaneously, but it lacks flexibility for varying sequence lengths.

These researches have conducted exhaustive analysis of gait pressure data from wearable sensors and obtained good discrimination results with a view of the advantages of wearable equipment free from space constraints, which provides a great development for Parkinson's disease auxiliary diagnosis. However, these methods did not flexibly model the temporal dynamics of gait signal sequences recorded by wearable devices.

In order to explore intra-signal temporal dependencies related to PD severity levels from gait movement with limited computational overhead, we propose a data-driven PD detection model using an attention-based temporal network to model patients' gait dynamics. With the sequential VGRF signal data, the self-attention mechanism of the Transformer network is suitable for temporal feature excavation on account of its flexibility for changing series and parallel computing capability to capture PD-related gait information, which can improve the detecting sensitivity and carry on to the diversity and complexity of movement behavior objective analysis. The experimental results illustrate that the ability of our proposed method to assess the PD severity scale is superior to the above state-of-art approaches.

### 3. Methods

The proposed attention-based temporal network (ATN) framework is shown in **Fig. 1** which gets the VGRF signals as input and outputs the detection result of PD severity rating to assist doctors in decision-making. ATN model is comprised of three components: a data preprocessing module mapping raw data into a uniform dimension, a temporal feature extraction based on a self-attention mechanism to capture discriminative temporal features, and a classifier making a final decision.

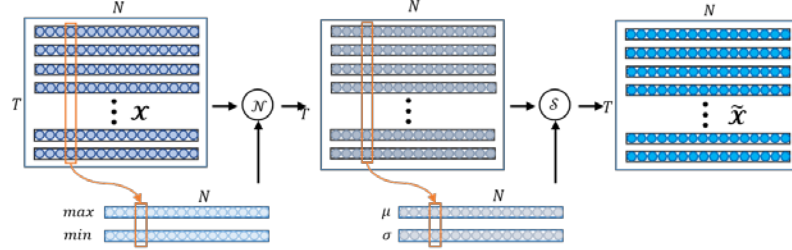


**Fig. 1.** The architecture of ATN model. There are three modules in this model: a preprocessing module for data normalization, a temporal feature extraction module for discriminative feature learning, and a softmax classifier for decision making.

#### 3.1 VGRF data preprocessing

The VGRF data collected is in the form of variable-length signal sequences, and each of them has a label indicating the category assigned to it which is associated with the subject's PD severity whom it is from. In order to successfully identify the severity level, we first split it into multiple partially overlapping small segments. Each small segment is a subset of the

original sequence containing rich gait information about the subject while further expanding the scale of data.



**Fig. 2.** Workflow of data preprocessing. The raw data is  $\mathcal{X} \in \mathbb{R}^{T \times N}$  from a time period of length  $T$ . Function  $\mathcal{N}$  and  $\mathcal{S}$  normalize the values of  $N$  features of the data  $\mathcal{X}$  in turn. The output  $\tilde{\mathcal{X}} \in \mathbb{R}^{T \times N}$  can be entered into  $T$  tokens in the temporal feature extraction module.

As shown in **Fig. 2**, a VGRF signal subsequence obtained by intercepting can be defined as  $\mathcal{X} = \{x_i \in \mathbb{R}^N \mid i = 1, 2, \dots, T\}$  where  $T$  represents the subsequence's length.  $x_i = \{\alpha_1, \alpha_2, \dots, \alpha_N\}$  is a one-dimensional vector of the number of collected signals  $N$  at each time instant, and its value is related to the scale of the sensor system currently working.  $\alpha_j$  is the VGRF signal value measured by the  $j$ -th sensor in the system at time step  $i$ . Taking into account uncertainty in measurement magnitude as well as the possibility of outliers, the original data  $\mathcal{X}$  needs to be preprocessed to normalize the values of its elements into the partition  $[0,1]$  and make it fit the normal distribution to increase the stability of predictions. The preprocessing process is formulated as

$$\begin{aligned} \tilde{\mathcal{X}} &= \mathcal{S}(\mathcal{N}(\mathcal{X})) \\ \mathcal{N}(A) &= \frac{A - A_{min}}{A_{max} - A_{min}} \\ \mathcal{S}(A) &= \frac{A - \mu}{\sigma}, \quad A = [a_{i,j}]_{n \times m} \end{aligned} \quad (1)$$

where the function  $\mathcal{N}$  and  $\mathcal{S}$  describe the computation process of normalization and standardization method respectively, and  $A$  denotes any two-dimensional matrix of size  $n \times m$ .  $\mu$  and  $\sigma$  describe the distribution of all elements in matrix  $A$  as their mean and variance.

So far, we have get preprocessed input data  $\tilde{\mathcal{X}} = [\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_N]^T$  whose shape is  $T \times N$  consistent with that of original data  $\mathcal{X}$ . Before being sent to the next module for effective PD-related temporal feature extraction, the data  $\tilde{\mathcal{X}}$  needs to go through a feature mapping layer consisting of multi-layer perceptron (MLP) which maps low-dimensional original data into a uniform high-dimensional space dimension  $P$ .

$$\tilde{F} = \tilde{\mathcal{X}} \cdot \mathcal{W} \quad (2)$$

As shown in (2),  $\tilde{F} \in \mathbb{R}^{T \times P}$  is the high-dimensional features transformed through matrices multiplication with weight matrix  $\mathcal{W} \in \mathbb{R}^{N \times P}$ . The process is a transformation in the feature dimension from  $N$  to  $P$ , in which parameters at every instant are shared with each other and there is nothing changed in the time direction.

### 3.2 Temporal feature extraction

The purpose of this section is to explore the data in temporal direction  $T$  and then output discriminative gait feature  $F_t \in \mathbb{R}^R$  related to PD severity levels from 2-dimensional primary feature tensor  $\tilde{F} \in \mathbb{R}^{T \times P}$ . The temporal feature extraction module designed here follows the computational principles of the Transformer with the attention mechanism as the core, which has the capability to mine temporal global dependency information from the input data in parallel.

The feature extractor is stacked by multiple Transformer blocks as depicted in its architecture, and each of them undertakes the function of feature extraction layer by layer. In order to meet the input specification of Transformer blocks, we first need to adjust input  $\tilde{F}_t$  to divide tokens along the dimension  $T$  and add location information describing the temporal relationship between tokens.

$$F = \mathcal{P} + [\tilde{f}_0 | \tilde{F}]^T$$

$$\mathcal{P}_{i,j} = \begin{cases} \sin(i \times \omega_j), & j \% 2 = 1 \\ \cos(i \times \omega_j), & j \% 2 = 0 \end{cases} \quad (3)$$

As shown in (3), input feature metric  $\tilde{F} = [\tilde{f}_1, \tilde{f}_2, \dots, \tilde{f}_N]^T \in \mathbb{R}^{T \times R}$  is divided in  $T$  tokens along the primary dimension and each one represents a feature vector at a time step. And an extra learnable token  $\tilde{f}_0$  as CLASS Token is concatenated with  $\tilde{F}$  for the final classification task. As an attention mechanism, the kernel of this extractor discovers the inter-token dependency relationships via similarity calculation computation with additional needs of sequence information which is embedded in the inherent architecture of convolutional and recurrent networks, the positional encoding tensor  $P \in \mathbb{R}^{(T+1) \times R}$  is added into  $T + 1$  characteristic tokens through element-wise addition operations as the input of Transformer blocks, the basic unit of the module. In (3),  $i$  is the position index of  $T + 1$  time steps,  $\omega_j$  is the handcraft frequency for each dimension  $j$  and  $\%$  represents the modulo operation. The complete tokens  $F = \{f_0, f_1, f_2, \dots, f_T\}$  containing both characteristic and temporal position information are calculated and will be sent into stacked Transformer blocks for multi-layer temporal feature aggregation. The structure of the kernel attention module is shown in Fig. 3.

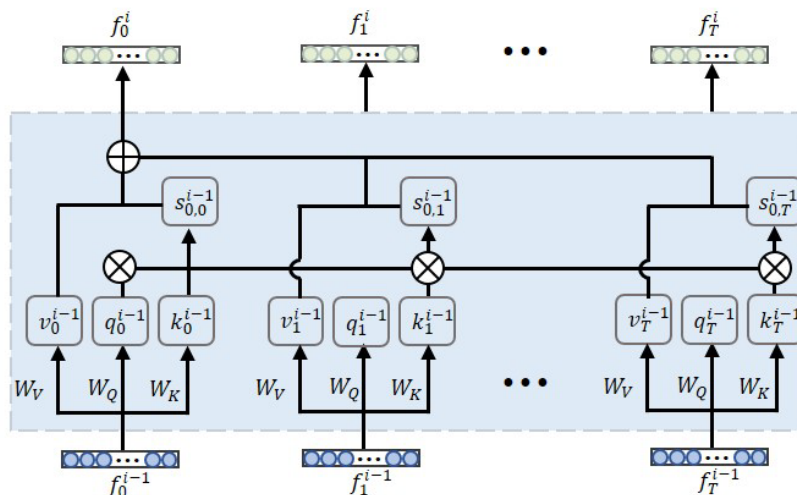


Fig. 3. Structure of attention module.



The  $i$ -th attention module gets features  $\{f_0^{i-1}, f_1^{i-1}, \dots, f_T^{i-1}\}$  composed of  $T + 1$  tokens from layer  $(i - 1)$  as input, then make calculation and output  $i$ -th layer's features  $\{f_0^i, f_1^i, \dots, f_T^i\}$  which will go to the next attention module. This module has two basic cells based on a softmax function  $\psi$  and matrix multiplication. The calculation equations are as follows:

$$\begin{aligned} f_j^i &= \sum_{k=0}^T v_j^{i-1} s_{j,k}^{i-1} \\ s_{j,k}^{i-1} &= \psi \left( \frac{q_j^{i-1} k_j^{i-1}}{\sqrt{d}} \right) \\ v_j^{i-1} &= \xi(f_j^{i-1} W_V) \\ q_j^{i-1} &= \xi(f_j^{i-1} W_Q) \\ k_j^{i-1} &= \xi(f_j^{i-1} W_K) \end{aligned} \quad (4)$$

where  $W_V, W_Q, W_K \in \mathbb{P}^P$  are weight metrics of value vector  $v_j^{i-1}$ , query vector  $q_j^{i-1}$  and key vector  $k_j^{i-1}$  corresponding to  $j$ -th feature vector  $f_j^{i-1}$ , and  $d$  is the feature dimension to perform a scale operation.  $\xi$  is an activation function providing nonlinear transformation. These weight metrics are all learnable, which makes the attention module independently learn temporal characteristics via inter-token scaled dot-product attention computation.

When passing the last transformer block, the class token  $F_t \in \mathbb{R}^P$  with index 0 is output by the whole temporal feature extractor as the final discriminative feature as shown in the model framework. In the next subsection, the classifier will make the final decision using feature  $F_t$ .

### 3.3 Classifier

The classifier is a multi-class decision-making module composed of a full-connection layer and a softmax function  $\psi$ . The weight metric  $W \in \mathbb{R}^{P \times C}$  maps the input feature  $F_t$  of length  $P$  into the dimension of the number of severity classes  $C$ . Then the probability that this sample belongs to each class  $\mathbf{1}, \mathbf{2}, \dots, \mathbf{C}$  is calculated by the softmax function  $\psi$ .  $\hat{y}$  is the model's final prediction result determined depending on the maximum of the calculated conditional probability of each class.

$$\begin{aligned} \hat{y} &= \psi(F_t W) \\ \psi(Z) &= \arg \max_i \frac{e^{z_i}}{\sum_{j=1}^C e^{z_j}} \end{aligned} \quad (5)$$

## 4. Experiment

In this section, we conduct experiments using the Pytorch and Python libraries to implement our proposed method for the PD stage rating task on a public dataset PDgait, which records sequences of plantar pressure signals from subjects with varying severity of Parkinson's disease. Details about experimental settings and result analysis will be presented as follows.

## 4.1 Experimental setup

### 4.1.1 Experimental data

In this work, the dataset we utilize to verify the effectiveness of our proposed method is the timing gait signal dataset PDgait from the Physionet databank. The data within PDgait is time series pressure signals of variable length. While gathering data, the organizers attached force-sensitive resistors to the soles of the subject's feet to the vertical ground force (VGRF) between feet and ground. There were 8 sensors under each sole, and the ones on the left and right foot were symmetrically distributed. The positions and relative coordinates of a total of 16 sensors are presented in Fig. 4. And they synchronously recorded VGRF signals at a sampling frequency of 100 Hz. Hence, VGRF signals obtained follow a  $16 \times t \times 100$ , where 16 denotes the dimension of gait signals at each sampling instant,  $t$  is the length of this gait sequence in seconds, and 100 corresponds to the sampling number per unit of time. Hence,  $t \times 100$  equals to time steps of a signal sequence.

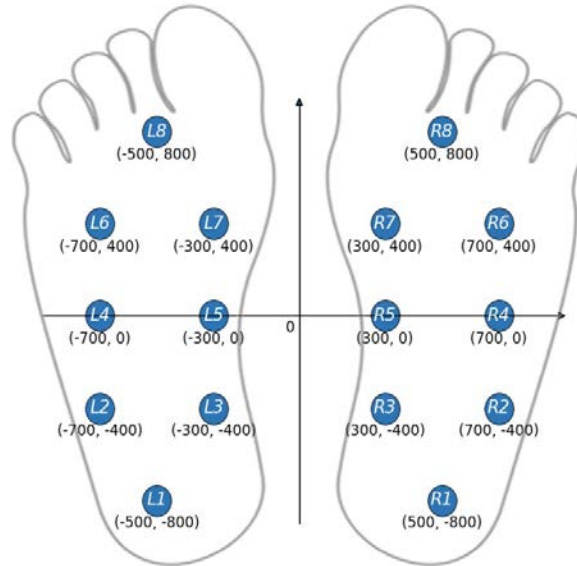


Fig. 4. The position coordinate of sensors measuring VGRF signals in rectangular coordinate system.

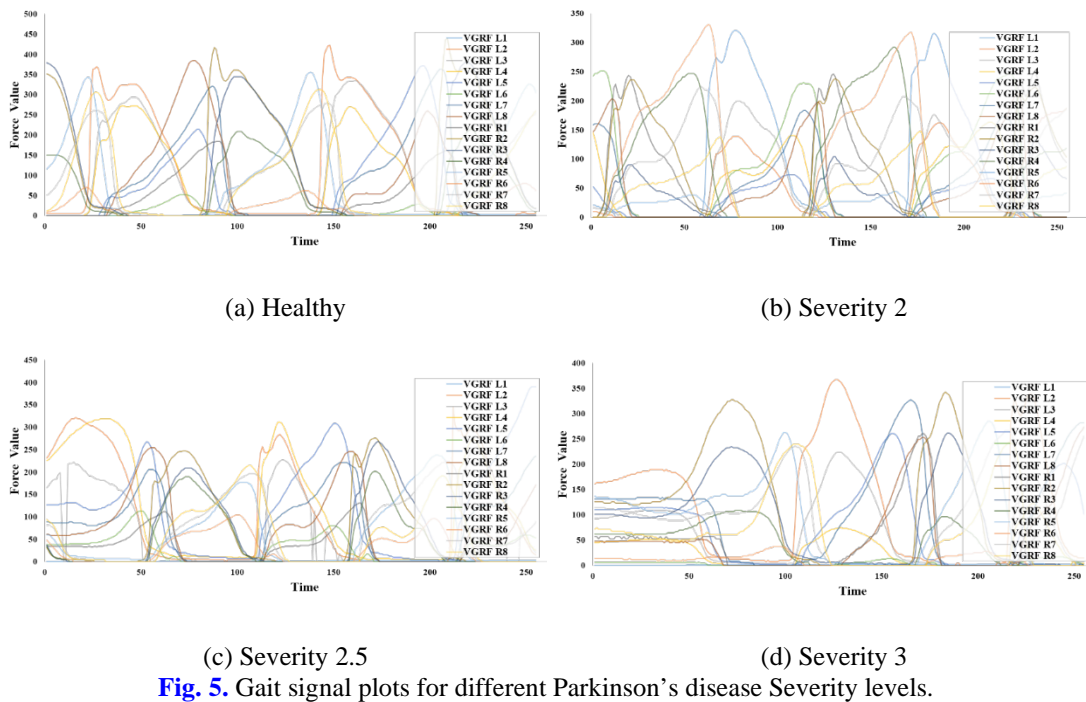
Table 1. Demographic information of volunteers collected VGRF signal data in PDgait

Dataset	Category	Subject		Age (year)		Weight (kg)	Height (m)
		Female	Male	< 70	≥ 70		
Ga [28]	PD subject	9	20	13	16	$73.1 \pm 11.3$	$1.67 \pm 0.06$
	Healthy Co	5	10	10	8	$74.2 \pm 12.6$	$1.68 \pm 0.09$
Si [29]	PD subject	13	22	28	7	$70.3 \pm 8.35$	$1.66 \pm 0.08$
	Healthy Co	11	18	27	2	$71.5 \pm 11.03$	$1.60 \pm 0.07$
Ju [30]	PD subject	13	16	19	10	$75.1 \pm 16.9$	$1.97 \pm 0.16$
	Healthy Co	14	12	20	6	$66.8 \pm 11.08$	$1.83 \pm 0.07$

Dataset PDgait collected VGRF records from 166 volunteers including 93 patients with Parkinson's disease and 73 healthy controls. The demographic information of these subjects is given in Table 1. The patients all had Parkinson's disease of different severity, and their severity levels had been marked according to the H&Y scale which contains a total of eight



PD severity ratings: Healthy status 0, Severity 1, 1.5, 2, 2.5, 3, 4, and 5. And our dataset involves four of these stages: Healthy, Severity 2, Severity 2.5, and Severity 3. This dataset is composed of three small datasets contributed by Frenkel-Toledo *et al.* [28], Yogev *et al.* [29], and Hausdorff *et al.* [30]. To maintain clarity and brevity, these datasets are referred to as Si, Ga, and Ju, based on the lead authors' names, indicating the study from which the data originated. When collecting gait signals, they were required to perform test tasks with sensors: walking directly on level ground, treadmill walking, or walking with rhythmic auditory stimulation. Fig. 5 presents line graphs depicting gait signal patterns recorded during walking for different severity levels of Parkinson's disease which illustrate the signal changes and the progression of PD symptoms from Healthy stage to Severity 3. From (a) to (d) in Fig. 5, we can find that patients' walking shows more pronounced stagnation as the disease progresses, although the difference is little clear in Fig. 5 (a) and Fig. 5 (b). And Table 2 gives the composition of the PD severity level of subjects who participated in walking tests in three datasets, and we use them as truth values to realize the effectiveness evaluation of the proposed method in this section.



**Table 2.** The composition of severity rating truth of subjects based on H&Y Scale in three datasets

Severity	Ga [28]	Si [29]	Ju [30]
Healthy	18	29	26
Severity 2	15	29	12
Severity 2.5	8	6	13
Severity 3	6	0	4

### 4.1.2 Implementation details

The experiment of this model is implemented in the system of GTX 3080 GPU, i7-11700 CPU, and 12-GB RAM. The dataset is divided into three parts according to proportion for experiments: 60% as the training set, 20% as the validation set, and 20% for final testing. The partition process is carried out based on the truth labels to ensure a balanced data distribution. For PDgait, the shape of a sample is  $100 \times 18$  in which 100 is the temporal length  $T$  set based on the sensor's sampling frequency 100 Hz, and input feature dimension 18 contains 16 VGRF signals from two feet and 2 statistics. The hidden layer feature dimension of this model is set to 32 with data dimension and computation time considered. The training process is iterated 200 times with a learning rate of 0.01.

### 4.2 Performance metrics

We use several performance metrics to ensure an objective evaluation of the model, including accuracy (Acc), Precision, Recall, and F1-score. Their calculation formulas are as follows:

$$\begin{aligned}
 Acc(\%) &= \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \\
 Precision(\%) &= \frac{TN}{TP+FP} \times 100\% \\
 Recall(\%) &= \frac{TP}{TP+FN} \times 100\% \\
 F1-score(\%) &= \frac{2 \times Precision \times Recall}{Precision + Recall} \times 100\%
 \end{aligned} \tag{6}$$

in which  $TP$  (true positives) is the number of samples whose labels are positive and prediction results are true,  $TN$  is the number of true negatives,  $FP$  represents the number of false positives, and  $FN$  donates the number of false negatives. Among these performance metrics, Acc describes the proportion of correct classification, Precision discovers the cost of false positives, and Recall discovers the cost of false negatives. F1-score combines both precision and recall using the harmonic mean of them.

### 4.3 Performance analysis

In this subsection, we evaluate the proposed attention-based temporal network (ATN) on PDgait Dataset for Parkinson's disease detection and severity rating and analyze its classification performance. As shown in [Table 2](#), PDgait dataset is made up of three subdatasets, and the subdataset Ga [\[28\]](#) and Ju [\[30\]](#) both involve four classes related to PD severity, and the subdataset Si [\[29\]](#) contain the first three severity levels within them. Because of the difference in their data composition, we separately conduct experiments on two datasets Ga and Si, and take both of them as a bigger dataset Ga-Si to verify the effectiveness of the proposed method. Their testing performances are shown in [Fig. 5](#), in which [Fig. 5 \(a\)](#), [\(b\)](#), and [\(c\)](#) are the confusion matrixes of datasets Ga-Si, Si, and Ga, and the classification accuracy of their test sets respectively are 98.06 %, 98.90%, and 98.86%. The proposed method can achieve great recognition ability on these datasets. What's more, an additional experiment is conducted on the dataset Ga without going through the preprocessing of normalization whose classification result is presented in [Fig. 6 \(d\)](#) with 97.61% accuracy, of which 99.30% for healthy controls, 96.72%, 94.53%, and 100% respectively for PD severity score 2, 2.5, and 3. Compared with that, the proposed ATN establishes better classification performance and has a more balanced discrimination capability in the face of PD patients with various severity levels and healthy individuals because of the raw data's uneven distribution.

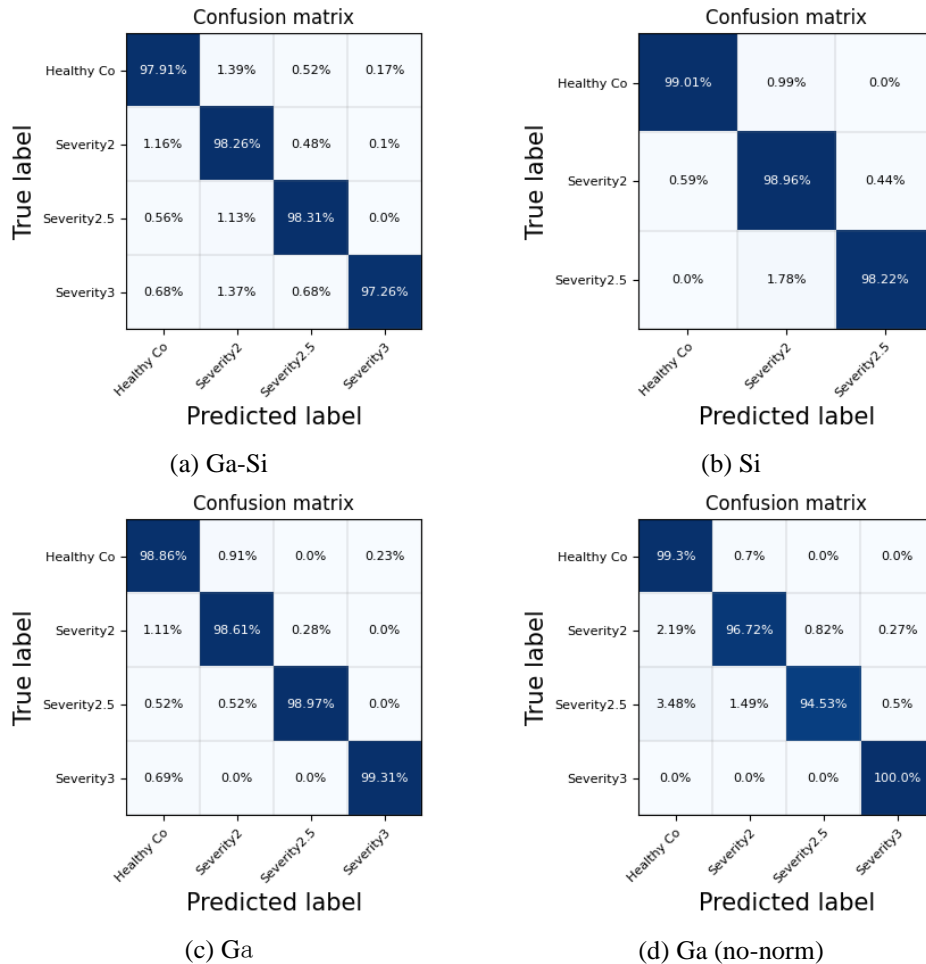
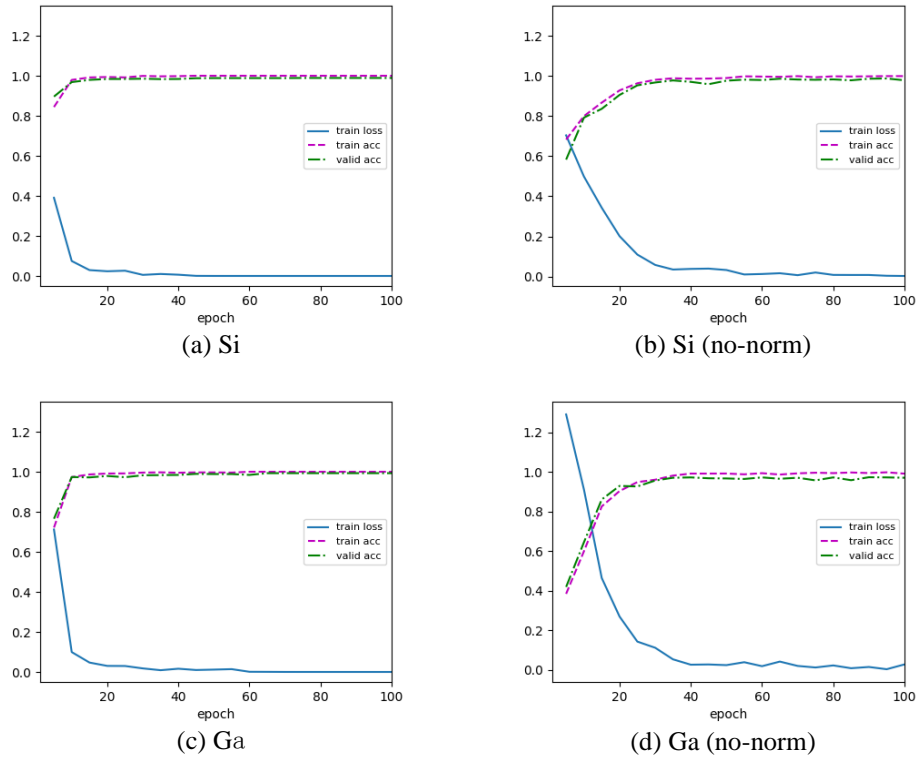


Fig. 6. Confusion matrixes of the classification results on PDgait dataset.

Loss function plots in Fig. 7 illustrate the training process of classifier models on two subdatasets Si and Ga through the trends of their training loss and classification accuracy on the training and validation sets. The classification models involve the proposed ATN corresponding to Fig. 7 (a), (c) and the classifier consisting of its remaining modules without preprocessing matching Fig. 7 (b), (d), and they are set with the same hyper-parameters and parameter initializer. The loss function plots highlight that the preprocessing module involving normalization and standardization techniques scales data and reduces the data distribution difference between the training set and validation set, which can alleviate overfitting issue and accelerate the model convergence to a certain extent.



**Fig. 7.** Loss function plots of model training on PDgait dataset.

**Table 3.** Performance of the proposed compared with existing methods

Method	Classification	Acc(%)	Precision(%)	Recall(%)	F1 (%)
SVM [18]	Two class	92.68	92.22	96.14	94.14
RF [8]	Two class	90.90	88.40	85.36	86.87
RNN	Multi-class	91.01	91.08	91.01	91.00
LSTM [24]	Two class	83.90	84.90	88.74	86.78
Dual_channel LSTM [6]	Multi-class	94.51	94.59	94.80	94.60
GRU	Multi-class	97.75	97.74	97.75	97.70
Bi-GRU	Multi-class	97.86	97.86	97.86	97.90
CNN [26]	Multi-class	96.34	96.34	96.34	96.31
Proposed (Transformer)	Multi-class	98.86	98.87	98.86	98.90

In addition, we further compared our proposed method with several previous approaches by conducting short experiments on the same dataset Ga, a small dataset in PDgait dataset, i.e., SVM, RF, RNN, LSTM, GRU, Bi-GRU and CNN models according to performance metrics introduced in Sec. 4.2, including binary classification and multi-classification. The quantitative metrics are presented in Table 3. Ref. [18], [8], [24] processed the labels of PDgait data as two classes to distinguish healthy individuals and PD patients. Remaining approaches respectively utilized RNN family variants and CNN network to perform multi-classification among patients with varying severity levels of Parkinson's disease coupled with healthy controls. Deep neural networks are better suitable for time series data processing than traditional classification methods. LSTM and GRU methods have greater ability to handle long sequence data benefiting from the design of memory gating units, and their performance is

significantly improved over RNN. Bi-GRU adopts a bidirectional structure to learn the change features of the forward and backward directions improved from the original GRU network. Our proposed method reveals better classification ability against them. The promising results illustrated the ability of the attention mechanism from the Transformer Network in dealing with sequential data and excavating global temporal dependencies.

In our experiments, we demonstrated that deep learning algorithms seem to particularly suit disease feature discovery and their applications in similar fields have been explored, such as hand movements and speech. On the basis of the survey, most studies focus on PD diagnosis according to single-modal motor data which is easily accessible, and only a few pay attention to the study of the combination of multi-modal motor or no-motor data with the limitations of the data source. These researches require the support of equipment for monitoring and collecting patients' clinical symptoms and signs with severity labels. We will carry out data collection from PD patients' daily lives and fuse multi-modal data for PD diagnosis and severity rating.

## 5. Conclusion

In this paper, we reported our research on predicting the severity of Parkinson's disease using an attention-based temporal network (ATN) with gait pressure data. We proposed ATN. The proposed method can effectively capture gait features from sequential signals and assess the severity of Parkinson's disease. Faced with gait signals from PD patients and healthy individuals, the attention mechanism of the Transformer can deal with variable length sequences more flexibly than one-dimensional convolution operation and has the ability to cope with variable length sequences with higher computational efficiency in comparison with recurrent structural networks. The model was evaluated and the best classification results were calculated.

Since the difficulty and great challenges in the PD diagnosis and severity evaluation are from its complex etiology and diverse symptoms, the current work has certain limitations because of its design only with the ability to process single-mode gait data. In the further, a large set of multi-modal data will be valid for more comprehensive and reliable information analysis relevant to Parkinson's disease. With these data, we will employ a variety of motion and non-motion features, involving lower limb movements, upper limb movements, and vocal cord ability to combine multi-feature and carry out comprehensive auxiliary diagnosis.

## Acknowledgement

This research was supported in part by National Natural Science Foundation of China under Grant No.62106117, China Postdoctoral Science Foundation under Grant No.2022M711741, and Natural Science Foundation of Shandong Province under Grant No.ZR2021QF084.

## References

- [1] J Pereira, C.R., Pereira, D.R., Weber, S.A., Hook, C., De Albuquerque, V.H.C., Papa, J.P., "A survey on computer-assisted parkinson's disease diagnosis," *Artificial intelligence in medicine*, vol. 95, pp. 48-63, Apr. 1995. [Article \(CrossRef Link\)](#)
- [2] W. P. Risk, G. S. Kino, and H. J. Shaw, "Fiber-optic frequency shifter using a surface acoustic wave incident at an oblique angle," *Opt. Lett.*, vol. 11, no. 2, pp. 115–117, Feb. 1986. [Article \(CrossRef Link\)](#)

- [3] Sepas-Moghaddam, A., Etemad, A., "Deep gait recognition: A survey," *IEEE transactions on pattern analysis and machine intelligence*, vol. 45, no. 1, pp. 264-284, Jan. 2023. [Article \(CrossRef Link\)](#)
- [4] Tölgyessy, M., Dekan, M., Chovanec, L., Hubinský, P., "Evaluation of the azure kinect and its comparison to kinect v1 and kinect v2," *Sensors*, vol. 21, no. 2, pp. 413, Jan. 2021. [Article \(CrossRef Link\)](#)
- [5] Marsico, M.D., Mecca, A., "A survey on gait recognition via wearable sensors," *ACM Computing Surveys (CSUR)*, vol. 52, no. 4, pp. 1-39, Aug. 2019. [Article \(CrossRef Link\)](#)
- [6] Zhao, A., Qi, L., Dong, J., Yu, H., "Dual channel lstm based multi-feature extraction in gait for diagnosis of neurodegenerative diseases," *Knowledge-Based Systems*, vol. 145, pp. 91-97, Apr. 2018. [Article \(CrossRef Link\)](#)
- [7] Zhong, C., Ng, W.W., "A robust frequency-domain-based graph adaptive network for parkinson's disease detection from gait data," *IEEE Transactions on Multimedia*, Oct. 2022. [Article \(CrossRef Link\)](#)
- [8] Khoury, N., Attal, F., Amirat, Y., Oukhellou, L., Mohammed, S., "Data-driven based approach to aid parkinson's disease diagnosis," *Sensors*, vol. 19, no. 2, pp. 242, Jan. 2019. [Article \(CrossRef Link\)](#)
- [9] Vidya, B., Sasikumar, P., "Gait based parkinson's disease diagnosis and severity rating using multi-class support vector machine," *Applied Soft Computing*, vol. 113, pp. 107939, Dec. 2021. [Article \(CrossRef Link\)](#)
- [10] Balaji, E., Brindha, D., Elumalai, V.K., Umesh, K., "Data-driven gait analysis for diagnosis and severity rating of parkinson's disease," *Medical Engineering & Physics*, vol. 91, pp. 54-64, May. 2021. [Article \(CrossRef Link\)](#)
- [11] Er, M.B., Isik, E., Isik, I., "Parkinson's detection based on combined cnn and lstm using enhanced speech signals with variational mode decomposition," *Biomedical Signal Processing and Control*, vol. 70, pp. 103006, Sept. 2021. [Article \(CrossRef Link\)](#)
- [12] Wingate, J., Kollia, I., Bidaut, L., Kollias, S., "Unified deep learning approach for prediction of parkinson's disease," *IET Image Processing*, vol. 14, no. 10, pp. 1980-1989, Jun. 2020. [Article \(CrossRef Link\)](#)
- [13] Flagg, C., Frieder, O., MacAvaney, S., Motamedi, G., "Real-time streaming of gait assessment for parkinson's disease," in *Proc. of the 14th ACM International Conference on Web Search and Data Mining*, New York, NY, USA, pp. 1081-1084, 2021. [Article \(CrossRef Link\)](#)
- [14] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I., "Attention is all you need," in *Proc. of NIPS2017*, 2017. [Article \(CrossRef Link\)](#)
- [15] Bhidayasiri, R., Tarsy, D., "Parkinson's disease: Hoehn and yahr scale," in *Movement Disorders: A Video Atlas*, 2012, pp. 4-5. [Article \(CrossRef Link\)](#)
- [16] Goetz, C.G., Poewe, W., Rascol, O., Sampaio, C., Stebbins, G.T., Counsell, C., Giladi, N., Holloway, R.G., Moore, C.G., Wenning, G.K., 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*, vol. 19, no. 9, pp. 1020-1028, Aug. 2004. [Article \(CrossRef Link\)](#)
- [17] Agostini, V., Balestra, G., Knaflitz, M. "Segmentation and classification of gait cycles," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 22, no. 5, pp. 946-952, Sep. 2014. [Article \(CrossRef Link\)](#)
- [18] Abdulhay, E., Arunkumar, N., Narasimhan, K., Vellaiappan, E., Venkatraman, V., "Gait and tremor investigation using machine learning techniques for the diagnosis of parkinson disease," *Future Generation Computer Systems*, vol. 83, pp. 366-373, Jun. 2018. [Article \(CrossRef Link\)](#)
- [19] E., B., D., B., R., B., "Supervised machine learning based gait classification system for early detection and stage classification of parkinson's disease," *Applied Soft Computing*, vol. 94, pp. 106494, Sept. 2020. [Article \(CrossRef Link\)](#)
- [20] Das, A., Mohapatra, S.K., Mohanty, M.N., "Design of deep ensemble classifier with fuzzy decision method for biomedical image classification," *Applied Soft Computing*, vol. 115, pp. 108178, Jan. 2022. [Article \(CrossRef Link\)](#)



- [21] Yujing, H., Ishtiaq, A., Lin, S., KyungHi, C., "SVM-based drone sound recognition using the combination of HLA and WPT techniques in practical noisy environment," *KSII Transactions on Internet and Information Systems (TIIS)*, pp. 5078-5094, October. 2019. [Article \(CrossRef Link\)](#)
- [22] Valdiviezo-Diaz, P., Ortega, F., Cobos, E., Lara-Cabrera, R., "A collaborative filtering approach based on naïve bayes classifier," *IEEE Access*, vol. 7, pp. 108581-108592, Aug. 2019. [Article \(CrossRef Link\)](#)
- [23] Charbuty, B., Abdulazeez, A., "Classification based on decision tree algorithm for machine learning," *Journal of Applied Science and Technology Trends*, vol. 2, no. 01, pp. 20-28, Nov. 2021. [Article \(CrossRef Link\)](#)
- [24] Ashour, A.S., El-Attar, A., Dey, N., Abd El-Kader, H., Abd El-Naby, M.M., "Long short term memory based patient-dependent model for fog detection in parkinson's disease," *Pattern recognition letters*, vol. 131, pp. 23-29, Mar. 2020. [Article \(CrossRef Link\)](#)
- [25] Kaur, S., Aggarwal, H., Rani, R., "Diagnosis of parkinson's disease using deep cnn with transfer learning and data augmentation," *Multimedia Tools and Applications*, vol. 80, pp. 10113-10139, Nov. 2021. [Article \(CrossRef Link\)](#)
- [26] El Maachi, I., Bilodeau, G.-A., Bouachir, W., "Deep 1d-convnet for accurate parkinson disease detection and severity prediction from gait," *Expert Systems with Applications*, vol. 143, pp. 113075, Apr. 2020. [Article \(CrossRef Link\)](#)
- [27] Goldberger, A.L., Amaral, L.A., Glass, L., Hausdorff, J.M., Ivanov, P.C., Mark, R.G., Mietus, J.E., Moody, G.B., Peng, C.-K., Stanley, H.E., "Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals," *circulation*, vol. 101, no. 23, pp. 215-220, Jun. 2000. [Article \(CrossRef Link\)](#)
- [28] Yogeve, G., Giladi, N., Peretz, C., Springer, S., Simon, E.S., Hausdorff, J.M., "Dual tasking, gait rhythmicity, and parkinson's disease: which aspects of gait are attention demanding?," *European journal of neuroscience*, vol. 22, no. 5, pp. 1248-1256, Sept. 2005. [Article \(CrossRef Link\)](#)
- [29] Frenkel-Toledo, S., Giladi, N., Peretz, C., Herman, T., Gruendlinger, L., Hausdorff, J.M., "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*, vol. 20, no. 9, pp. 1109-1114, May. 2005. [Article \(CrossRef Link\)](#)
- [30] Hausdorff, J.M., Lowenthal, J., Herman, T., Gruendlinger, L., Peretz, C., Giladi, N., "Rhythmic auditory stimulation modulates gait variability in parkinson's disease," *European Journal of Neuroscience*, vol. 26, no. 8, pp. 2369-2375, Oct. 2007. [Article \(CrossRef Link\)](#)



**Huimin Wu** received the bachelor's degree in computer science and technology from Nanjing University of Finance and Economics in June 2021. She is currently studying for the master's degree in Qingdao University, China, majoring in computer science and technology. Her research interests include computer vision, pattern recognition, and machine learning.



**Yongcan Liu** received the Bachelor's degree in computer science and technology from Qingdao Institute of Technology in 2021. He is currently pursuing the master's degree in the College of Computer Science & Technology in Qingdao University. His research interests include deep learning, pattern recognition and multimodal fusion.



**Haozhe Yang** received the Bachelor's degree in Computer Science and Technology from Qingdao University in June 2023. He is currently a postgraduate student majoring in Computer Science and technology in the School of Computer Science and Engineering at Northeastern University. His main research interests include computer vision, mixed reality technology, data processing and analysis, and image processing.



**Zhongxiang Xie** received the bachelor's degree in computer science from Qingdao University, Qingdao, China, in 2023. He is currently a postgraduate majoring in computer technology in Qingdao University. His research interests include computer version, pattern recognition, machine learning, data analysis and face recognition.



**Xianchao Chen** received the bachelor's degree in computer science and technology from Qingdao University, Qingdao, China, in 2023. He is now studying computer technology at Ocean University of China, Qingdao, China.



**Mingzhi Wen** received the bachelor's degree in computer science and technology from Qingdao University, Qingdao, China, in 2023. He is now studying software engineering at Southeast University of China, Nanjing, China.



**Aite Zhao** received her Ph.D. degree in June 2020 in the College of Information Science and Engineering in Ocean University of China. She is a visiting Ph.D. researcher in the School of Informatics, University of Leicester, Leicester, U.K. She is currently a Lecturer of College of Computer Science and Technology in Qingdao University. Her research interests include computer vision, pattern recognition, machine learning, data analysis and robotics.