



Assessment of Parkinson's Disease Severity Using Gait Data: A Deep Learning-Based Multimodal Approach

Nabid Faiem¹ , Tunc Asuroglu¹ , Koray Acici² , Antti Kallonen¹ ,
and Mark van Gils¹ 

¹ Faculty of Medicine and Health Technology, Tampere University, 33720 Tampere, Finland
nabid.faiem@tuni.fi

² Department of Artificial Intelligence and Data Engineering, Ankara
University, 06830 Ankara, Turkey

Abstract. The ability to regularly assess Parkinson's disease (PD) symptoms outside of complex laboratories supports remote monitoring and better treatment management. Multimodal sensors are beneficial for sensing different motor and non-motor symptoms, but simultaneous analysis is difficult due to complex dependencies between different modalities and their different format and data properties. Multimodal machine learning models can analyze such diverse modalities together, thereby enhancing holistic understanding of the data and overall patient state. The Unified Parkinson's Disease Rating Scale (UPDRS) is commonly used for PD symptoms severity assessment. This study proposes a Perceiver-based multimodal machine learning framework to predict UPDRS scores.

We selected a gait dataset of 93 PD patients and 73 control subjects from the PhysioNet repository. This dataset includes two-minute walks from each participant using 16 Ground Reaction Force (GRF) sensors, placing eight on each foot. This experiment used both raw gait timeseries signals and extracted features from these GRF sensors. The Perceiver architecture's hyperparameters were selected manually and through Genetic Algorithms (GA). The performance of the framework was evaluated using Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and linear Correlation Coefficient (CC).

Our multimodal approach achieved a MAE of 2.23 ± 1.31 , a RMSE of 5.75 ± 4.16 and CC of 0.93 ± 0.08 in predicting UPDRS scores, outperforming previous studies in terms of MAE and CC.

This multimodal framework effectively integrates different data modalities, in this case illustrating by predicting UPDRS scores using sensor data. It can be applied to diverse decision support applications of similar natures where multimodal analysis is needed.

Keywords: Multimodal machine learning model · Perceiver · Gait analysis · GRF sensors

1 Introduction

1.1 Parkinson's Disease (PD)

PD is the fastest growing neurological disorder according to the Global Burden of Diseases, Injuries, and Risk Factors (GBD) studies [1–3]. The World Health Organization (WHO) estimated that about 8.5 million individuals were living with PD worldwide in 2019 [4]. In the last thirty years, there has been a significant increase in prevalence and mortality rates of PD. Several components contributed to this upward trend, such as a growing elderly population, environmental and social factors, and extended duration of the disease. If the current trend persists, it is projected that the number of individuals with PD could exceed 17 million by the year 2040 [5], which will pose enormous challenges for any healthcare system.

Diagnosis of PD is typically performed by neurologists specialized in movement disorders and involves different neurological tests and patient interviews. However, the diagnosis of PD remains a difficult task due to its overlapping characteristics with other neurodegenerative diseases and the subjectivity in short-term assessment. A global shortage of specialized neurologists increases the risk of misdiagnosis, potentially preventing targeted treatment and increasing disease severity. PD symptoms are typically evaluated using a rating scale called Unified Parkinson's Disease Rating Scale (UPDRS), which is a widely accepted score. This PD severity rating scale has four distinct components that include both non-motor and motor parts [6]. Early diagnosis and frequent assessment of PD symptoms is required for more targeted medical intervention and to support remote monitoring for better treatment management, thereby improving the quality of life of PD patients.

1.2 Gait Analysis

Gait analysis can be a useful tool for measuring gait abnormalities as gait worsens with the disease progression. The analysis is conducted in a specialized laboratory equipped with video systems, motion-capturing cameras, floor-based force sensors, and electromyography (EMG) systems [7]. Although this complex laboratory setup provides accurate results, the availability of these systems is limited due to expensive infrastructures and lack of skilled personnel, particularly in developing countries and remote places. Ground Reaction Force (GRF) non-invasive wearable sensors can offer a cost-effective and accessible tool for gait analysis to provide a comprehensive overview of gait pattern. These sensors are designed to capture joint movements and muscle activities effectively [8]. They are small and typically placed in the insole or underneath shoes to measure kinetic force, temporal, and spatial characteristics of gait variability.

Gait abnormalities in PD patients under cognitive load become more severe with the disease progression [9]. Dynamic changes in gait can be detected through regular monitoring of daily activities, medications, social interactions, or environmental conditions outside of the artificial laboratory in real-life settings. Therefore, there is a need for inexpensive monitoring methods that can be used not only during healthcare encounters but also to improve treatment intervention and management throughout the patient's lifetime [10].

The gait signals obtained from the controlled laboratory settings are more structured, as the participants follow a specific protocol in a strictly controlled environment [11]. Gait analysis has traditionally focused on temporal or frequency domain analysis with standard hypothesis testing. This approach was mainly used because the datasets were simpler and more structured [12]. However, real-world gait signals are unstructured and noisy, and may require multimodal sensing for more comprehensive analysis taking into account a wider context. Therefore, the GRF based gait signal alone may not be enough for estimating PD symptoms severity in remote monitoring applications. Analyzing other gait or tremor signals with non-motor symptoms combined could give us a better understanding about the progression of PD symptoms in real-life scenarios.

1.3 Multimodal in Decision Support

Analyzing multimodal data from diverse sources is a challenging task due to possible complex, non-linear relationships, and temporal dependencies between modalities [13]. If these modalities are analyzed separately with own methods (signal processing and other approaches), and their results are combined and processed thereafter, there is a risk of losing potential joint information between them. Integrating these modalities requires harmonization and standardization prior to analysis in a computational model [14]. Multimodal machine learning is an evolving field of machine learning where multiple modalities can be combined simultaneously to support or aid each other in enhancing the predictive performance of the model.

Recently, DeepMind's Perceiver architecture has shown promising results in processing different data modalities [15]. This architecture is designed on top of Transformer networks, it is capable of processing different modalities including time series, images, and other signals. The core concept behind the Perceiver architecture is the use of an iterative attention mechanism [15]. This mechanism allows the model to concentrate on distinct parts of signals to capture the underlying pattern. By integrating multiple modalities, the model may generate robust predictions as it can observe patterns from different modalities and identify the relationship between input modalities and output.

A Perceiver architecture-based multimodal machine learning framework could be an effective solution for simultaneously analyzing both raw gait timeseries signals and extracted hand-crafted features from GRF sensors, allowing them to complement each other to improve the predictive ability for PD diagnosis and PD symptoms severity estimation. The iterative nature of the Perceiver architecture, along with the weight sharing strategy can enhance the predictive performance by efficiently reusing the same input multiple times [15].

Optimizing hyperparameters is a challenging task in any deep learning model, particularly when parameters are selected manually. To overcome this issue, state-of-the-art optimization techniques like Random search (RS), Grid search (GS), Bayesian optimization (BO), and Genetic Algorithms (GA) techniques explored in earlier studies [16]. GA has the advantage of simulating different hyperparameter settings to find the optimal configuration to achieve better prediction performance. This algorithm has been used for hyperparameter optimization in the detection of femoral neck fracture [17] and the diagnosis of nutritional anemia [18].

1.4 Goal of the Study

The main advantage of using raw gait signals is that it eliminates the manual processing steps and simplifies analysis prior to computational model. However, to incorporate explicitly expert-based knowledge in the form of well-defined features, and to evaluate the capability of the computational model for analyzing multimodal data, we utilized the extracted features from gait timeseries signal as separate modality. This is because expert-derived features represent interpretable biomechanical metrics, while the gait timeseries signal provides sensor data covering variation in spatiotemporal domains. In addition, the effect of GA optimization performance is assessed in this approach.

This research proposes a multimodal machine learning framework based on the Perceiver architecture for predicting the severity of PD symptoms. The study also examines the framework's performance for the diagnosis of PD. The major contributions of this paper can be summarized as follows:

1. Study if UPDRS can be predicted with gait timeseries signals from GRF sensors and compare framework's performance with other studies,
2. Compare performance of multimodal vs single model approaches,
3. Study GA optimization performance.

The rest of the paper follows as outlined below: Sect. 2 discusses state-of-the-art data analysis methods related to PD diagnosis and severity assessment. Studies related to applying the Perceiver architecture in disease diagnosis for other diseases than PD are also presented in this section. Section 3 describes the materials and explanation of each component of the proposed framework. Section 4 presents the results of the performance evaluation of the proposed approach. Section 5 discusses the results comparing different modalities, as well as the limitations, challenges, and future scope of this research. Section 6 concludes the paper.

2 Related Work

Despite the need for regular assessment of PD symptoms to improve treatment management, most existing studies focused on diagnostic solutions for detecting PD. Fewer studies have addressed estimating the exact symptoms severity of PD as regression problem. This study includes references to PD diagnosis research to provide a comprehensive overview of how wearable sensors and data analysis can improve diagnosis. The majority of these studies explored single modality data whereas only a small number of studies have used a multimodal approach. These multimodal studies are mostly based on small populations and/or imbalanced datasets [19]. Prior to machine learning, standard hypothesis statistical tests like t-tests, Mann-Whitney U test, and ANOVAs were employed for PD detection from gait data [12]. Recent literature shows successful implementation of machine learning and deep learning techniques for PD detection. Machine learning techniques like Random Forests (RF) [20] and Support Vector Machines (SVM) [21] have been explored for PD diagnosis from gait data. Convolutional Neural Network (CNN) and long short-term memory (LSTM) deep learning algorithms are mostly used in research for PD diagnosis [22]. Most studies use a single modality analysis for PD diagnosis and the PhysioNet Gait database is most used in these studies for gait analysis

[22]. This PhysioNet Gait dataset contains original UPDRS to reflect the severity of PD. The total UPDRS score ranges from 0 to 199, encompassing both motor and non-motor components, with 199 representing severe disability and 0 indicating the healthy state. The maximum score from the motor part of the scale is 108 [6, 23].

2.1 PD Severity Estimation

Aşuroğlu et al. [24] proposed a hybrid deep learning regression approach emphasizing on local pattern recognition for predicting PD symptoms severity from the gait signal. Their proposed framework is based on the combination of CNN and locally Weighted Random Forest (LWRF) that use multi-channel gait data to predict exact UPDRS scores. The convolutional part of their framework extracts local characteristics from the extracted time and frequency domain features and the LWRF part exploits the local relationships from these characteristics. Their proposed model achieved a state-of-the-art performance and outperformed the previous study. Aşuroğlu et al. [25] conducted a prior study focused on the same regression problem for estimating PD symptoms severity. In this study they used a decision tree-based supervised machine learning model. This study was the first one that utilizes multichannel GRF wearable sensors-based gait data in general. They utilized the same time and frequency domain features as [24] for the prediction. Their developed ML model exploited the local patterns from these features to better predict UPDRS scores.

2.2 PD Diagnosis

In this section, studies that used the PhysioNet gait dataset for PD diagnosis are discussed to maintain a consistent comparison. El Maachi et al. [26] used raw gait timeseries data for PD detection using a deep learning model based on the 1D CNN (1D-Convnet). This model was designed to simultaneously process 18 one-dimensional signals obtained from 16 GRF foot sensors and the total force from each foot, eliminating the need for manual feature extraction. The first part of their model used 18 parallel 1D-CNN signals for local spatial information extraction. Following this, a fully connected layer that integrates the relevant CNN spatial features for PD diagnosis. Their proposed algorithm achieved an accuracy of 98.7%, a sensitivity of 98.1% and a specificity of 100.0%. Alharthi et al. [27] used deep CNN architecture for PD diagnosis. They transformed the raw GRF sensor signal into a 3D matrix to provide input to the model. Their approach was designed to learn the spatiotemporal GRF signals without manual feature extraction. Their proposed model was robust against noise and effectively addressed the variability of human movement between individuals.

Pham et al. [28] focused on a single GRF sensor from the gait data for PD detection. From the timeseries signal, they extracted time-frequency and time-space features. With the extracted features they trained bi-LSTM (bi-LSTM). Their results showed better performance compared to conventional LSTM and other prior studies in terms of accuracy (100.0%), sensitivity (100.0%), specificity (100.0%) and F1 score (1.0). They also reported that their model was more efficient in terms of computational power and processing time as it used only one sensor. Balaji et al. [29] introduced a deep learning approach using LSTM for PD detection and severity classification, eliminating the need

for handcrafted features. They passed gait cycles from GRF sensors to train the LSTM network. Their network consists of four LSTM layers, four dropout layers followed by a fully connected layer and a SoftMax layer. Their approach achieved accuracy of 98.6% for PD detection.

Vidya et al. [30] used a hybrid CNN-LSTM model to explore the spatial and temporal dependence of GRF timeseries signals to differentiate between healthy subjects and different PD severity levels. They selected the optimal number of GRF sensors using a variability analysis. They applied the empirical mode decomposition (EMD) technique to extract the significant intrinsic mode functions (IMFs) through power spectral analysis to capture the non-linear and non-stationary characteristics of the timeseries signal. The dominant IMFs from the optimal GRF signals were then used to train the hybrid model for PD stage classification. They reported that their proposed hybrid model achieved better performance than other studies that used gait analysis to classify healthy subjects and PD severity stages.

Nguyen et al. [31] introduced a Transformer-based deep learning model that emphasized both temporal and spatial characteristics of gait signals to differentiate between healthy control subjects and PD patients. They applied one temporal Transformers for each gait sensor, and the dimensionally reduced outputs from these temporal Transformers were concatenated before being fed into a spatial Transformer. This spatially encoded feature set was then passed to two fully connected layers and an output layer for the final classification. Their model achieved accuracy of 95.2%, a sensitivity of 98.1% and a specificity of 86.8%.

2.3 Multimodal Data Analysis Using Perceiver

Although the Perceiver architecture is a recent development, it has already been implemented in other studies focusing on disease diagnosis other than PD.

Josef et al. [32] estimated the speed of human motion using IMU-based wearable sensors. They evaluated the performance of different deep learning methods, including the Perceiver architecture. In their experiment, they collected IMU data from a single foot, shin, and thigh. The Perceiver architecture, along with other deep learning techniques, outperformed conventional feature-based methods in estimating speed. Aadam et al. [33] evaluated the performance of the Perceiver architecture for classification of emotion from raw EEG signals. In this experiment, they used EEG signal from DEAP [34] dataset, and they used two modalities for the analysis. The first modality consisted of EEG signal from all channels as 1D vector and the spatial locations of electrodes as the second modality. They found that the Perceiver model performed better for multimodal configuration compared to single modality.

3 Materials and Methods

This section introduces a Perceiver architecture-based multimodal machine framework for PD symptoms severity estimation. The proposed framework can simultaneously process both raw gait timeseries signals and extracted features from GRF sensors as

multimodal input to predict UPDRS score. This framework can also process each modality input separately. Figure 1 depicts the workflow of the proposed framework which includes the Perceiver model and hyperparameter optimization using GA.

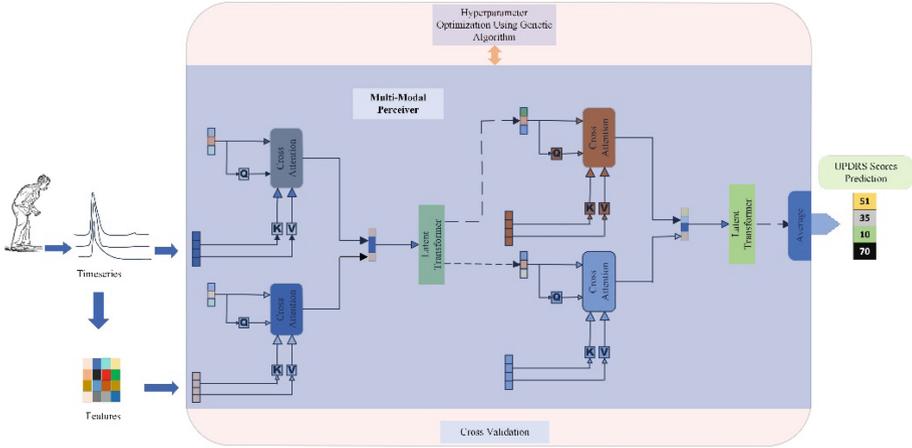


Fig. 1. Workflow of the proposed Perceiver architecture-based multimodal machine learning framework for UPDRS score prediction. The framework simultaneously processes raw gait time-series signals and extracted features from these signals to complement each other through the iterative process of Perceiver network. GA selects hyperparameters of the Perceiver architecture in the first cross validation (CV) and then applies them to subsequent CVs.

3.1 Dataset Description

The performance of the proposed architecture was evaluated using the dataset of PhysioNet [35] that includes walking sequences from 93 idiopathic PD patients and 73 healthy control subjects. The mean age of the PD patients was 66.3 years, and 63% of the patients were male. The average age of control subjects was 63.7 years among which 55% were men. This dataset was collected by three independent research groups (Yogev et al. [36], Hausdorff et al. [37], and Silvi Frenkel-Toledo et al. [38]) at the Laboratory for Gait & Neurodynamics, Movement Disorders Unit of the Tel Aviv Sourasky Medical Center.

Gait patterns were measured from each participant for two minutes by placing eight GRF sensors under each foot. Participants were asked to walk in two different scenarios: normal walking at a self-selected speed and dual-task walking. In the dual-task protocol, subjects were instructed to perform arithmetic tasks by serially subtracting seven from a pre-defined number.

These GRF sensors measure force (in Newton) with a sampling rate of 100 Hz and the force distribution of these sensors could be used to measure the gait impairment of subjects. The GRF sensors-based measurement system provides better foot distribution compared to force-sensitive resistors (FSR) due to their larger size and sensing area [39].

The dataset also includes demographic information about the participants, such as their gender, age, height, weight, and PD severity values as UPDRS scores. In this dataset, the mean UPDRS score of PD patients was 32, with a minimum score of 13 and a maximum of 70.

3.2 Pre-processing and Feature Extraction

The first 20 s and the last 10 s from each sensor data segment were excluded to minimize the start and end effects. After that, a median filter with a length of 3 was applied to remove outliers or large spikes. This filter reduces sudden walking fluctuations, and the small filter size preserves maximum force signal without distortion.

Gait cycles are repetitive in nature therefore sensor timeseries signals may contain redundant information. Hand-crafted features that reflect the spatiotemporal and frequency characteristics of timeseries signals may be useful for PD detection or PD severity estimation. Therefore, we extracted the frequency and time domain features from the median filtered signal of each GRF sensor. These seven-frequency domain and sixteen-time domain extracted features, presented in Table 1, effectively capture relevant information that has been demonstrated in previous studies [24, 25]. Extracted features were then standardized by subtracting the mean and scaling to unit variance, as defined in Eq. 1. In this equation, ‘ μ ’ represents a mean of a specific feature that is calculated from all subjects. ‘ σ ’ represents the standard deviation of that feature also calculated from all subjects, x represents feature value for a single subject and z is the standardized value of x .

$$z = \frac{x - \mu}{\sigma} \quad (1)$$

Table 1. Time and Frequency based features extracted from each sensor of all participants.

Feature Domain	Features calculated from each GRF sensor
Time	mean, harmonic mean, median, range, interquartile range (IQR), mean absolute deviation, maximum amplitude and minimum/maximum spread, skewness, kurtosis, root mean square (RMS), energy, power, and entropy
Frequency	mean, minimum, maximum, normalized, energy, power, and phase

3.3 Perceiver Architecture

The Perceiver architecture has the potential of combining multiple data modalities to improve a model’s predictive capability. The Perceiver leverages an iterative attention mechanism to scale high-dimensional multimodal data without making any domain specific assumptions. This architecture employs scalable Fourier features-based position encoding to preserve the temporal, spatial or spatiotemporal characteristics of the input data. These encoded features are then concatenated with input data before processing

in the main architecture. As illustrated in Fig. 2, the Perceiver architecture is composed of two main components: the cross-attention module and the latent Transformer. The cross-attention module reduces the dimensionality of the input data into a lower dimensional latent bottleneck. The latent Transformer then processes further to learn complex patterns in the data. This concept allows the construction of large networks of multiple cross-attention and the latent Transformer blocks for processing complex and high-dimensional multimodal data without changing the underlying architecture. The weight sharing strategy across each block can also improve predictive performance by efficiently reusing the same input multiple times. Finally, the model generates predictions by averaging the output of the final latent Transformer over the index dimension depending on the classification or regression task [15].

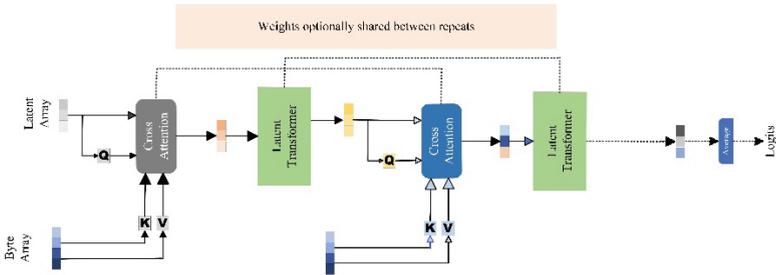


Fig. 2. The Perceiver architecture uses an iterative approach that alternates between cross-attention layers and several latent self-attention blocks. By alternating between these two types of blocks, the model can receptively attend to the input data and find patterns between the input and output layers more effectively. This mechanism allows the model to reduce the dimensionality of the input data while preserving the most critical information. The weight sharing strategy of the architecture can also improve predictive performance.

A potential disadvantage of the Perceiver architecture is that while the size of the latent array facilitates detail mapping of the input data, the bottleneck effect may limit the extent of detail. The use of multiple cross-attention layers can improve the precision of information extraction from the input data. However, this comes at the cost of increased computational resources that can lead to longer processing times [15].

3.4 Proposed Framework

The proposed framework uses the Perceiver architecture to identify the relationships between GRF signals with corresponding targets (either for classification or regression purposes). This work uses a multimodal version of the perceiver architecture that is capable of processing both unimodal and multimodal data in a single run [40]. This multimodal architecture leverages the attention mechanism to dynamically focus on distinct parts of these modalities to enhance the overall performance.

After the pre-processing phase, the timeseries signal of each participant is reshaped to $9025 \text{ samples} \times 16 \text{ sensors}$ and the features set is reshaped to 368×1 from $23 \text{ features} \times 16 \text{ sensors}$. This conversion is necessary as the multimodal version [40] requires the

inputs in specific format for processing. These reshaped datasets are then converted to a one-dimensional tensor vector to feed into the Perceiver architecture.

To address the class imbalance during training, we use the weighted random sampler algorithm from PyTorch without oversampling. This means that each training batch includes samples from both classes (0 = healthy and 1 = PD) proportional to their class weights, until all samples from the minority class have been utilized. After that, training continues with samples from the majority class. The training batch size for these experiments is set to six. Validation is conducted in a single batch, where samples are randomly selected without considering class balance to simulate the real-world scenario.

3.5 Experimental Setup

The Perceiver model's hyperparameters are initially selected in a non-optimized manner using a trial-and-error method, where the optimal selection is determined based on the lowest prediction error. We first begin with predetermined hyperparameters and then adjust them to achieve the lowest possible prediction error. Through this process, we then select a network depth of 4, a cross-attention layer of 6 and latent-attention layer of 6 and the weighting sharing between the cross-attention and the latent self-attention layers. For optimization, we use Adam optimizer with a learning rate of 10^{-4} .

The proposed framework is implemented in Python with the PyTorch library on the JupyterLab development environment. This experiment is conducted on a computer with an AMD Ryzen 5 5600X 6-Core Processor, 16 GB RAM, and a 12 GB NVIDIA GeForce RTX 3060 graphics card. CUDA library of GeForce is used to utilize graphics card. Typical training sessions for processing either multimodal data or timeseries signals lasts approximately 80 h. The duration of training session extends when GA optimization is employed.

3.6 Evaluation

The accuracy of the proposed framework for predicting PD severity symptoms depends on the error between actual and predicted UPDRS scores. We use Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Correlation Coefficient (CC) to evaluate the performance of the proposed framework. MAE measures the mean difference between two continuous variables (mean difference between actual and predicted UPDRS scores).

$$MAE = \frac{|p_1 - a_1| + \dots + |p_n - a_n|}{n} \quad (2)$$

$$RMSE = \sqrt{\frac{(p_1 - a_1)^2 + \dots + (p_n - a_n)^2}{n}} \quad (3)$$

Here n is the size of the sample, p and a are the target and estimated value (output of the algorithm), respectively.

Initially gait data shorter than two minutes are discarded to maintain consistency across samples. This process resulted in a comprehensive dataset of 189 samples from

126 participants (68 PD patients and 58 Healthy control subjects). Gait timeseries data and the extracted features from each sample are structured according to the requirements of the framework to train either a single or a multimodal Perceiver model, individually or together. In the framework, we divide the structured dataset into ten randomly equal-sized subsamples using the ten-fold CV method. For each iteration, one subsample representing 10% of the dataset is reserved for validation and the remaining nine subsamples are reserved for training. This process is then repeated ten times. During training in each CV, samples are divided in proportion to preserve class balance as indicated in the Proposed Framework.

The final performance of each model is evaluated using the mean and standard deviation (SD) of MAE, RMSE and CC obtained from the ten-fold CV. The best performing model is identified by the lowest mean and SD in both MAE and RMSE, along with the highest mean and lowest SD in CC. The model is then benchmarked against the outcomes of previous similar studies. Similarly, we compare the mean and SD of these metrics between the multimodal and single model approaches to determine the best model. The performance of these models is visually illustrated using scatter plots of actual versus predicted UPDRS scores.

3.7 Hyperparameter Optimization Using GA

Hyperparameters define the complexity of the Perceiver architecture and its learning behavior [16]. Hyperparameters are difficult to optimize while improving model performance and reducing complexity of the architecture. This work attempts to optimize hyperparameters like the network depth, the number of cross-attention and self-attention blocks to minimize the prediction error of the framework. The manual tuning of these hyperparameters is time-consuming, so after an initial effort, we use GA for hyperparameter optimization. The main mechanism behind GA is given below [16]:

- First, initialize the population's equivalent of chromosomes and genes, randomly. These parameters represent the search space, hyperparameter and hyperparameter values, respectively.
- A fitness value of each member of the current generation is evaluated using a fitness function. The objective of the fitness function is to minimize the prediction error from the optimized hyperparameter settings. We use MAE as the fitness value for regression tasks and accuracy as the fitness value for classification tasks.
- The termination criterion for the regression task is set at a MAE of 2.5, while for the classification task it is set at an accuracy of 100%. These values are set high to achieve better performance by exploring more generations and combinations. If there is no change in the accuracy or the MAE score for two consecutive generations, then the evolution process will stop. Otherwise, if the termination criterion is not met, then proceed with the following steps:
 - Select parents from the mating pool.
 - Perform crossover and mutation operations on the chromosomes to produce the next generation population.
 - Evaluate the fitness of each child in the new generation.

To implement this algorithm, we use the TorchGA open-source library, which has implemented GA using PyTorch library [41]. The hyperparameters of this algorithm are presented in Table 2. In this study, we use the population size of 10 and repeat the algorithm until 10 generations have passed if the termination condition is not met. We configure three hyperparameters such as net depth, the number of cross-attention and self-attention layers and represent them as chromosomes. Their corresponding values, as shown in Table 2, are specified as genes. Parents are randomly selected from the mating pool, and a crossover operation is performed after the selection. We do not include mutation operations in this task.

We use ten-fold CV to evaluate PD diagnosis and symptoms severity estimation performance, applying both single and multimodal model with the same structured gait data and extracted features. The class balance is also maintained during the training. The optimal hyperparameters are identified through GA optimization using the fitness function from the first CV. The resultant hyperparameters from the first CV are then used in the remaining CVs.

Table 2. GA hyperparameters

Hyperparameters	Values
Population Size	10
Generation	10
Chromosomes (Genes)	Net depth (4,8), Cross-attention (1,2,4), Latent attention (2,4,8)
Parent Selection Type	Random
Crossover Type	Uniform

4 Results

4.1 Empirical Results

Table 3 presents ten-fold CV results of models with manually selected hyperparameters. The model that incorporated both Features and Timeseries as input modalities together is referred to as the Multimodal Data model. This model outperformed the other two models where Timeseries or Features were used as separate input modalities. From Table 3, it is observed that the Multimodal Data model achieved the highest performance among all models with an MAE of 2.23 ± 1.31 , RMSE of 5.75 ± 4.16 and CC of 0.93 ± 0.08 , indicating that it enhances prediction with a certain degree of variability. The model that utilized the Features modality demonstrated slightly lower performance with an MAE of 2.72 ± 1.57 , RMSE of 6.79 ± 4.42 and CC of 0.91 ± 0.09 . The model used the Timeseries modality has the highest errors, as indicated by an MAE of 3.18 ± 1.60 , RMSE of 7.56 ± 3.89 and CC of 0.90 ± 0.08 . The MAE scores of each model demonstrate relatively

low means and small SD. The RMSE scores of each model exhibits high means and larger SD, particularly due to the mispredictions of higher UPDRS scores.

Table 3. Performance comparison of different input modalities with manually selected hyperparameters (Network Depth: 4, Cross-Attention Layer: 6, Latent-Attention Layer: 6)

Input Modality	MAE Mean \pm SD	RMSE Mean \pm SD	CC Mean \pm SD
Features	2.72 \pm 1.57	6.79 \pm 4.42	0.91 \pm 0.09
Timeseries	3.18 \pm 1.60	7.56 \pm 3.89	0.90 \pm 0.08
Multimodal Data	2.23 \pm 1.31	5.75 \pm 4.16	0.93 \pm 0.08

Figure 3 indicates the relationship between predicted and actual UPDRS scores for three models with manually selected hyperparameters. The blue reference line in the plot represents the ideal scenario where predicted scores would perfectly align with actual scores. Figure 3 illustrates that the Multimodal Data model performed better than other models, as most samples were comparatively closer to the reference line than others.

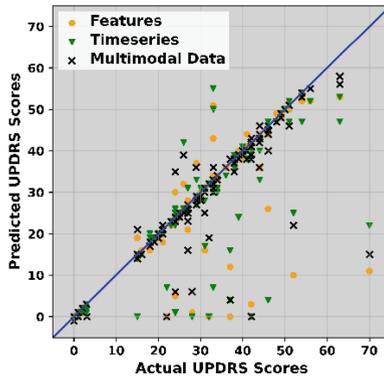


Fig. 3. Scatter plot showing the predicted against the actual UPDRS scores from Multimodal Data (black cross), Timeseries (green down-pointing triangle) and Features (yellow circles). Scores closely align with the $y = x$ reference line (blue) indicate a strong correlation between predicted and actual UPDRS scores.

Table 4 presents the performance of three models with GA selected hyperparameters. It selects identical hyperparameters for the Timeseries and Multimodal Data modalities but selected a larger network for the Feature modality. Despite having a larger network depth, the Feature modality performed better than Timeseries modality. All models with GA selected hyperparameters showed poor performance. This suggests that suboptimal hyperparameters selected in the first CV could reduce GA optimization performance. However, the Multimodal Data outperformed the other two models in MAE, RMSE and CC.

Table 4. Performance comparison of different input modalities for the selected hyperparameters (D: Depth, C: Cross-attention layer and L: Latent-attention layer) using GA.

Input Modality	Hyperparameter (D, C, L)	MAE Mean \pm SD	RMSE Mean \pm SD	CC Mean \pm SD
Features	8,1,2	3.64 \pm 3.26	6.92 \pm 5.54	0.92 \pm 0.08
Timeseries	4,1,8	4.04 \pm 1.82	7.79 \pm 2.98	0.91 \pm 0.07
Multimodal Data	4,1,8	2.58 \pm 1.39	6.12 \pm 3.46	0.93 \pm 0.06

To compare the performance of models using these modalities with and without GA selected hyperparameters, we used scatter plot in Fig. 4. The plot visually compares the alignment between the predicted and actual UPDRS scores for each modality. In Fig. 4a, the model using the Features modality showed a decrease in performance when predicting UPDRS scores for healthy control subjects. Similarly in Fig. 4b, the model with the Timeseries modality showed a decline in performance for predicting UPDRS scores for both healthy control subjects and PD patients. The variation slightly improved when predicting using GA selected hyperparameters. However, the Multimodal Data model, as demonstrated in Fig. 4c, accurately predicted UPDRS scores for healthy control subjects in both cases. Similarly, the prediction UPDRS scores were relatively close to the reference line for PD patients in both scenarios.

In Table 5, we compared the performance of the Multimodal Data model with previous studies that predicted UPDRS scores from GRF signals. The Multimodal Data model, using manually selected hyperparameters, showed better performance in terms of MAE and CC compared to referenced studies. RMSE performance of the Multimodal Data model was slightly higher compared to one of the previous studies.

Table 6 demonstrates the classification evaluation metrics for three modalities with and without GA optimization. As can be seen from Table 6, the Multimodal Data model demonstrated the highest performance, with an accuracy of 97.3%, AUC of 0.98, sensitivity of 96%, and specificity of 100%. However, the performance slightly decreased with GA optimized hyperparameters. The models using the Features input modality also showed a relatively good performance, both with and without GA optimization, while the models that used Timeseries input modality had slightly lower performance in comparison. Overall, integrating multiple data modalities improved the predictive performance for PD diagnosis.

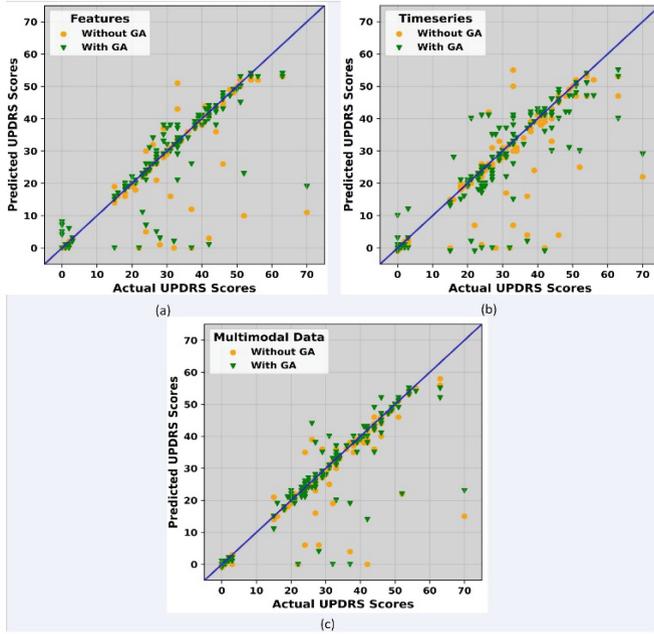


Fig. 4. Scatter plots comparing predicted and actual UPDRS scores for (a) Features Modality, (b) Timeseries Modality, and (c) Multimodal Data Modality with and without GA Optimization.

Table 5. Comparison with previous studies on PD severity estimation.

Authors	MAE	RMSE	CC
Aşuroğlu et al. [24]	3.01	4.56	0.90
Aşuroğlu et al. [25]	4.46	7.38	0.90
Present Method (Multimodal)	2.23 ± 1.31	5.75 ± 4.16	0.93 ± 0.08

5 Discussion

Currently used gait analysis methods are mostly limited to complex laboratories, but the advancement of wearable technologies demand increased support for remote monitoring. Regular assessment of PD symptoms severity can support and benefit all stakeholders, from patients to healthcare professionals, for better treatment management. Early diagnosis and monitoring may benefit individuals who are either developing motor symptoms or transitioning from non-motor to motor symptoms. This approach can also assist individuals who have manifested motor symptoms but have not received clinical diagnosis and clinical supervision.

Table 6. Performance Evaluation of PD Classifier for Different Input Modalities with and without GA Optimization.

Input Modality	Optimization	Hyperparameter (D, C, L)	AUC	Sensitivity (%)	Specificity (%)	Accuracy (%)
Features	Without GA	4,6,6	0.968	93.6	100	95.7
	With GA	4,1,2	0.972	94.4	100	96.3
Timeseries	Without GA	4,6,6	0.956	92.8	98.4	94.7
	With GA	4,1,8	0.949	94.4	95.3	94.7
Multimodal Data	Without GA	4,6,6	0.980	96.0	100	97.3
	With GA	4,1,8	0.956	92.8	98.4	94.7

The proposed framework successfully predicted UPDRS scores using the multimodal data, outperforming single model approaches in terms of MAE, RMSE and CC. Additionally, it outperformed previous studies in terms of MAE and CC. This multimodal approach might offer promising solution for combining different sensor modalities that capture the characteristics of motor and non-motor symptoms characteristics influenced by daily activities and treatment interventions. This approach would benefit from mutual support or co-learning from these modalities. However, multimodal machine learning is still a developing field, particularly in the biomedical sector. Special considerations are required for challenges like data linkage between modalities and dealing with noise and missing data. Training a Perceiver model on multimodal data requires extensive computational power due to large size of the data. Complexity and training time increase with larger batch sizes and large number of parameters of the Perceiver architecture. To effectively process multimodal data and reduce training time, a small batch size, optimized hyperparameters tailored to the data, and an advanced GPU are required.

The proposed framework predicted UPDRS scores using the gait timeseries signal, with promising results in terms of MAE, RMSE and CC. This approach could minimize the preprocessing steps required for automatic prediction in free living conditions. In this dataset, we observed higher gait variability in PD patients compared to healthy controls, particularly during dual task activities. Analyzing gait variability directly from the time-series signal is challenging, as gait characteristics of PD patients and healthy subjects are influenced by several factors. Nonetheless, this framework effectively captures these variabilities to estimate PD symptoms severity.

In this study, we only used GA to optimize the hyperparameters of the first cross-fold due to computational limitations because training the Perceiver model is time-consuming task. These selected hyperparameters may not be sufficient for subsequent folds. In addition, although GA automates the hyperparameter tuning, this approach has limitations. This algorithm introduces additional hyperparameter configuration like population size, generation number, crossover, and mutation rate. In addition, the time complexity of this algorithm is considerably high [16].

In this study, we used a CV strategy but did not reserve any data for testing because of the relatively small sample size. As a result, performance of the model might be biased

towards this dataset. During training, class balancing was maintained until all samples from the minority class were used to ensure balanced learning. Without this approach, the model might have biased or overfitted towards the majority class.

This dataset has a limited population, particularly PD patients with high UPDRS scores. Patients with high severity typically exhibit motor disabilities even during normal walking. They are often excluded from experiments because of the risk of falling during dual-task activities. This exclusion can result in a misrepresentation of the actual PD population, leading to a high bias in the dataset. A complex model such as Perceiver, with its numerous parameters, could potentially reduce this bias and variation with extensive training at the risk of overfitting.

For future direction of this study, we aim to use a large and diverse dataset that would capture free-living gait characteristics from smart shoe and smartwatch, combined with non-motor symptoms collected via smartphone, could enable continuous assessment of PD symptoms.

6 Conclusion

Frequent assessment of PD symptoms may open new opportunities for personalized remote monitoring and effective treatment intervention in free-living environments. Free-living conditions introduce uncertain variables like different activity levels and environmental factors can complicate gait patterns. These complexities, combined with non-motor symptoms and medication effects, make it significantly challenging and impossible for any single sensor. A comprehensive monitoring system, combining gait sensors for motor symptoms and smartphone data for non-motor symptoms can capture the whole spectrum. However, analyzing multimodal data from these diverse sources in free-living conditions is a challenging task due to several factors such as data source integration and potential biases. The Perceiver architecture has shown promising results in combining data from different modalities effectively. Our study demonstrates the combination GRF based gait timeseries signal with extracted features predicted UPDRS scores accurately. This attention-based architecture outperforms previous studies for predicting PD symptoms severities. This architecture has potential to be used as multimodal machine learning framework for decision support solution in personalized treatment management, remote care, and digital twin-based applications.

References

1. Rocca, W.A.: The burden of Parkinson's disease: a worldwide perspective. *The Lancet Neurol.* **17**, 928–929 (2018). [https://doi.org/10.1016/S1474-4422\(18\)30355-7](https://doi.org/10.1016/S1474-4422(18)30355-7)
2. Feigin, V.L., et al.: GBD 2015 neurological disorders collaborator group: global, regional, and national burden of neurological disorders during 1990–2015: a systematic analysis for the global burden of disease study 2015. *The Lancet Neurol.* **16**, 877–897 (2017). [https://doi.org/10.1016/S1474-4422\(17\)30299-5](https://doi.org/10.1016/S1474-4422(17)30299-5)
3. Dorsey, E.R., et al.: GBD 2016 Parkinson's disease collaborators: global, regional, and national burden of Parkinson's disease, 1990–2016: a systematic analysis for the global burden of disease study 2016. *The Lancet Neurol.* **17**, 939–953 (2018). [https://doi.org/10.1016/S1474-4422\(18\)30295-3](https://doi.org/10.1016/S1474-4422(18)30295-3)

4. Launch of WHO's Parkinson disease technical brief. <https://www.who.int/news/item/14-06-2022-launch-of-who-s-parkinson-disease-technical-brief>. Accessed 01 Jun 2023
5. Dorsey, E.R., Sherer, T., Okun, M.S., Bloem, B.R.: The emerging evidence of the Parkinson pandemic. *JPD*. **8**, S3–S8 (2018). <https://doi.org/10.3233/JPD-181474>
6. Goetz, C.G., Poewe, W., Rascol, O.: Movement disorder society task force on rating scales for Parkinson's disease: the unified Parkinson's disease rating scale (UPDRS): status and recommendations. *Mov. Disord.* **18**, 738–750 (2003). <https://doi.org/10.1002/mds.10473>
7. Chen, P.-H., Wang, R.-L., Liou, D.-J., Shaw, J.-S.: Gait disorders in Parkinson's disease: assessment and management. *Int. J. Gerontol.* **7**, 189–193 (2013). <https://doi.org/10.1016/j.ijge.2013.03.005>
8. Tong, J., Zhang, J., Dong, E., Du, S.: Severity classification of Parkinson's disease based on permutation-variable importance and persistent entropy. *Appl. Sci.* **11**, 1834 (2021). <https://doi.org/10.3390/app11041834>
9. Mirelman, A., et al.: Gait impairments in Parkinson's disease. *The Lancet Neurology*. **18**, 697–708 (2019). [https://doi.org/10.1016/S1474-4422\(19\)30044-4](https://doi.org/10.1016/S1474-4422(19)30044-4)
10. Papapetropoulos, S., Mitsi, G., Espay, A.J.: Digital health revolution: is it time for affordable remote monitoring for Parkinson's Disease? *Front. Neurol.* **6**, 126864 (2015). <https://doi.org/10.3389/fneur.2015.00034>
11. Del Din, S., Godfrey, A., Mazzà, C., Lord, S., Rochester, L.: Free-living monitoring of Parkinson's disease: lessons from the field: wearable technology for Parkinson's disease. *Mov. Disord.* **31**, 1293–1313 (2016). <https://doi.org/10.1002/mds.26718>
12. Chandrabhatla, A.S., Jonathan Pomeranec, I., Ksendzovsky, A.: Co-evolution of machine learning and digital technologies to improve monitoring of Parkinson's disease motor symptoms. *NPJ Digit. Med.* **5**(1), 32 (2022). <https://doi.org/10.1038/s41746-022-00568-y>
13. Chau, T.: A review of analytical techniques for gait data. Part 1: fuzzy, statistical and fractal methods. *Gait Posture* **13**, 49–66 (2001). [https://doi.org/10.1016/S0966-6362\(00\)00094-1](https://doi.org/10.1016/S0966-6362(00)00094-1)
14. Acosta, J.N., Falcone, G.J., Rajpurkar, P., Topol, E.J.: Multimodal biomedical AI. *Nat. Med.* **28**, 1773–1784 (2022). <https://doi.org/10.1038/s41591-022-01981-2>
15. Jaegle, A., et al.: Perceiver: general perception with iterative attention (2021). <http://arxiv.org/abs/2103.03206>
16. Yang, L., Shami, A.: On hyperparameter optimization of machine learning algorithms: theory and practice. *Neurocomputing* **415**, 295–316 (2020). <https://doi.org/10.1016/j.neucom.2020.07.061>
17. Beyaz, S.: Femoral neck fracture detection in X-ray images using deep learning and genetic algorithm approaches. *Jt Dis Relat Surg.* **31**, 175–183 (2020). <https://doi.org/10.5606/ehc.2020.72163>
18. Kilicarslan, S., Celik, M., Sahin, Ş: Hybrid models based on genetic algorithm and deep learning algorithms for nutritional Anemia disease classification. *Biomed. Signal Process. Control* **63**, 102231 (2021). <https://doi.org/10.1016/j.bspc.2020.102231>
19. Skaramagkas, V., Pentari, A., Kefalopoulou, Z., Tsiknakis, M.: Multi-modal deep learning diagnosis of Parkinson's disease—a systematic review. *IEEE Trans. Neural Syst. Rehabil. Eng.* **31**, 2399–2423 (2023). <https://doi.org/10.1109/TNSRE.2023.3277749>
20. Açıcı, K., Erdaş, Ç.B., Aşuroğlu, T., Toprak, M.K., Erdem, H., Oğul, H.: A random forest method to detect Parkinson's disease via gait analysis. In: Boracchi, G., Iliadis, L., Jayne, C., Likas, A. (eds.) *Engineering Applications of Neural Networks*, pp. 609–619. Springer International Publishing, Cham (2017). https://doi.org/10.1007/978-3-319-65172-9_51
21. Daliri, M.R.: Chi-square distance kernel of the gaits for the diagnosis of Parkinson's disease. *Biomed. Signal Process. Control* **8**, 66–70 (2013). <https://doi.org/10.1016/j.bspc.2012.04.007>
22. Sigcha, L., et al.: Deep learning and wearable sensors for the diagnosis and monitoring of Parkinson's disease: a systematic review. *Expert Syst. Appl.* **229**, 120541 (2023). <https://doi.org/10.1016/j.eswa.2023.120541>

23. Tsanas, A., Little, M., McSharry, P., Ramig, L.: Accurate telemonitoring of Parkinson's disease progression by non-invasive speech tests. *Nat Prec.* (2009). <https://doi.org/10.1038/npre.2009.3920.1>
24. Aşuroğlu, T., Oğul, H.: A deep learning approach for Parkinson's disease severity assessment. *Health Technol.* **12**, 943–953 (2022). <https://doi.org/10.1007/s12553-022-00698-z>
25. Aşuroğlu, T., et al.: Parkinson's disease monitoring from gait analysis via foot-worn sensors. *Biocybernetics Biomed. Eng.* **38**, 760–772 (2018). <https://doi.org/10.1016/j.bbe.2018.06.002>
26. El Maachi, I., Bilodeau, G.-A., Bouachir, W.: Deep 1D-Convnet for accurate Parkinson disease detection and severity prediction from gait. *Expert Syst. Appl.* **143**, 113075 (2020). <https://doi.org/10.1016/j.eswa.2019.113075>
27. Alharthi, A.S., Casson, A.J., Ozanyan, K.B.: Gait spatiotemporal signal analysis for Parkinson's disease detection and severity rating. *IEEE Sensors J.* **21**, 1838–1848 (2021). <https://doi.org/10.1109/JSEN.2020.3018262>
28. Pham, T.D.: Time–frequency time–space LSTM for robust classification of physiological signals. *Sci. Rep.* **11**, 6936 (2021). <https://doi.org/10.1038/s41598-021-86432-7>
29. Balaji, E., Brindha, D., Elumalai, V.K., Vikrama, R.: Automatic and non-invasive Parkinson's disease diagnosis and severity rating using LSTM network. *Appl. Soft Comput.* **108**, 107463 (2021). <https://doi.org/10.1016/j.asoc.2021.107463>
30. Vidya, B., Sasikumar, P.: Parkinson's disease diagnosis and stage prediction based on gait signal analysis using EMD and CNN–LSTM network. *Eng. Appl. Artif. Intell.* **114**, 105099 (2022). <https://doi.org/10.1016/j.engappai.2022.105099>
31. Nguyen, D.M.D., Miah, M., Bilodeau, G.-A., Bouachir, W.: Transformers for 1D signals in Parkinson's disease detection from gait (2022). <http://arxiv.org/abs/2204.00423>, <https://doi.org/10.48550/arXiv.2204.00423>
32. Justa, J., Šmídl, V., Hamáček, A.: Deep learning methods for speed estimation of bipedal motion from wearable IMU sensors. *Sensors* **22**, 3865 (2022). <https://doi.org/10.3390/s22103865>
33. Aadam, et al.: EmoPercept: EEG-based emotion classification through perceiver. *Soft. Comput.* **26**, 10563–10570 (2022). <https://doi.org/10.1007/s00500-021-06578-4>
34. Koelstra, S., et al.: DEAP: a database for emotion analysis; using physiological signals. *IEEE Trans. Affective Comput.* **3**, 18–31 (2012). <https://doi.org/10.1109/T-AFFC.2011.15>
35. Hausdorff, J.M.: Gait in Parkinson's disease (2008). <https://physionet.org/content/gaitpdb/>, <https://doi.org/10.13026/C24H3N>
36. Yogev, G., et al.: Dual tasking, gait rhythmicity, and Parkinson's disease: which aspects of gait are attention demanding? *Eur. J. Neurosci.* **22**, 1248–1256 (2005). <https://doi.org/10.1111/j.1460-9568.2005.04298.x>
37. Hausdorff, J.M., et al.: Rhythmic auditory stimulation modulates gait variability in Parkinson's disease: effects of RAS on gait variability in PD. *Eur. J. Neurosci.* **26**, 2369–2375 (2007). <https://doi.org/10.1111/j.1460-9568.2007.05810.x>
38. Frenkel-Toledo, S., et al.: Treadmill walking as an external pacemaker to improve gait rhythm and stability in Parkinson's disease. *Mov. Disord.* **20**, 1109–1114 (2005). <https://doi.org/10.1002/mds.20507>
39. Bae, J., Tomizuka, M.: A tele-monitoring system for gait rehabilitation with an inertial measurement unit and a shoe-type ground reaction force sensor. *Mechatronics* **23**, 646–651 (2013). <https://doi.org/10.1016/j.mechatronics.2013.06.007>
40. Campagne, F.: Multi modality perceiver - pytorch (2021). <https://github.com/fac2003/perceiver-multi-modality-pytorch>
41. Gad, A.F.: PyGAD: An intuitive genetic algorithm python library (2021). <http://arxiv.org/abs/2106.06158>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

