An Interpretable Deep Learning Optimized Wearable Daily Detection System for Parkinson's Disease

Min Chen, Zhanfang Sun, Tao Xin, Yan Chen, and Fei Su

***Abstract*— Walking detection in the daily life of patients with Parkinson's disease (PD) is of great significance for tracking the progress of the disease. This study aims to implement an accurate, objective, and passive detection algorithm optimized based on an interpretable deep learning architecture for the daily walking of patients with PD and to explore the most representative spatiotemporal motor features. Five inertial measurement units attached to the wrist, ankle, and waist are used to collect motion data from 100 subjects during a 10-meter walking test. The raw data of each sensor are subjected to the continuous wavelet transform to train the classification model of the constructed 6-channel convolutional neural network (CNN). The results show that the sensor located at the waist has the best classification performance with an accuracy of 98.01%±0.85% and the area under the receiver operating characteristic curve (AUC) of 0.9981±0.0017 under ten-fold cross-validation. The gradient-weighted class activation mapping shows that the feature points with greater contribution to PD were concentrated in the lower frequency band (0.5∼3Hz) compared with healthy controls. The visual maps of the 3D CNN show that only three out of the six time series have a greater contribution, which is used as a basis to further optimize the model input, greatly reducing the raw data processing costs (50%) while ensuring its performance (AUC=0.9929±0.0019). To the best of our knowledge, this is the first study to consider the visual interpretation-based optimization of an intelligent classification model in the intelligent diagnosis of PD.**

***Index Terms*— Parkinson's disease, wearable sensors, daily detection, deep learning, visual interpretation.**

I. Introduction

PARKINSON'S disease (PD), as a neurodegenerative disease, is affecting the health of more and more people, accompanied by movement symptoms such as bradykinesia, rigidity, tremor, and gait disorders [1], [2], [3]. It has become the fastest developing neurological disease in the world, posing a great challenge to the health of the elderly population because the prevalence rate increases significantly with age [4], [5]. Objective assessment of the clinical progress of PD plays an important role in formulating treatment strategies for patients [6], [7], [8]. However, there are some limitations in the evaluation of the motor performance of patients through the Unified Parkinson's Disease Rating Scale (UPDRS) in clinical practice: (1) strained medical resources due to the need for specialized neurologists, (2) subjectivity of clinical scoring and patient response, (3) fluctuating patient symptoms, and (4) unsustainability of the assessment [9], [10], [11]. Wearable sensor technology is becoming popular in the medical field, which can provide continuous objective longitudinal data, emerging as an effective approach for detecting and evaluating the severity of disease [12], [13], [14], [15].

Almost all relevant studies using wearable sensors to objectively assess PD patients have been implemented based on machine learning (ML) and deep learning (DL) algorithms [7], [16]. Regarding ML algorithms, Talitckii et al. [17] revealed three of the most effective exercises to assist doctors in diagnosing PD from 15 common motor tasks. Chen et al. [9] constructed a genetic algorithm optimized random forest classification model by collecting three upper limb tasks in the UPDRS-III, which can help clinicians identify subtle changes in the motor performance of patients with PD. However, ML based on handcrafted features is very time-consuming, and features determined by experience may not comprehensively describe the motion details of complex disease progression [18]. Comparatively, DL technology, which can automatically learn the spatial hierarchy of features from low to high levels, has gradually become the current development

Manuscript received 1 March 2023; revised 31 July 2023; accepted 31 August 2023. Date of publication 11 September 2023; date of current version 13 October 2023. This work was supported in part by the Innovation Project of Clinical Science and Technology of Jinan City under Grant 202134047 and in part by the Academic Promotion Program of Shandong First Medical University under Grant 2019LJ005. *(Corresponding authors: Fei Su; Yan Chen.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethics Committee of Shandong Provincial Hospital in Jinan, China, under Application No. 2020-600.

Min Chen and Fei Su are with the Department of Neurosurgery, The First Affiliated Hospital of Shandong First Medical University and Shandong Provincial Qianfoshan Hospital, Jinan 250014, China, and also with the Department of Radiology, Shandong First Medical University and Shandong Academy of Medical Sciences, Taian 271016, China (e-mail: sufei@sdau.edu.cn).

Zhanfang Sun is with the Department of Neurology, Provincial Hospital Affiliated to Shandong First Medical University, Jinan 250021, China.

Tao Xin is with the Department of Neurosurgery, The First Affiliated Hospital of Shandong First Medical University and Shandong Provincial Qianfoshan Hospital, Jinan 250014, China.

Yan Chen is with the Neurology Department, Shanghai Jiahui International Hospital, Shanghai 200233, China (e-mail: cyrz001@163.com).

This article has supplementary downloadable material available at https://doi.org/10.1109/TNSRE.2023.3314100, provided by the authors.

Digital Object Identifier 10.1109/TNSRE.2023.3314100

trend [18], [19]. Vásquez-Correa et al. [20] proposed a DL architecture that integrated multimodal information from speech, handwriting, and gait to evaluate patients at different disease stages. The above studies significantly improved the objective diagnosis of PD, but at the same time required patients to perform a series of tedious tests on their initiative, which not only makes them bear a large burden, but also fails to continuously monitor the exercise performance of patients.

The gait disorders are also important factors affecting the quality of life of PD patients, and walking occupies most of the exercise time in daily life [21], [22]. Therefore, for PD patients with gait disorders, the passive detection of motor performance has been proposed in many studies to achieve long-term monitoring [21], [23]. Although the body wearable sensor network can reflect the motor performance of PD patients more comprehensively, using multiple sensors increases the patient burden and may not be feasible for long-term monitoring of patients or management of the disease [24]. Therefore, current studies are moving toward a minimum number or even external detection of sensors [25], [26]. Daneault et al. showed that the use of multiple sensors on a single limb did not significantly improve clinical scores, indicating that one sensor per limb is sufficient [25]. The wrists and ankles have been consistently identified as the most prominent locations for capturing motion features, as demonstrated by their remarkable accuracy in previous studies [27]. Lonini et al. [26] collected movement data from 20 participants performing 13 common tasks using six sensors attached to the limbs, and demonstrated that a single sensor mounted on the hand was sufficient for the classification of bradykinesia and tremor. Mikos et al. [28] integrated the developed freezing of gait detection system into a single sensor node attached to the ankle, achieving an average sensitivity and specificity of 92.9%. Peraza et al. [29] proposed an automatic gait analysis pipeline based on DL algorithms, and the results demonstrated the potential feasibility of using a single wearable sensor attached to the foot to assess gait in free-living and home scenarios. Furthermore, due to its proximity to the center of gravity of the body, the waist serves as a typical location for capturing comprehensive movement patterns and is considered a standard sensor installation site [30], [31]. Juutinen et al. [30] utilized waist-mounted smartphones with built-in sensor to collect data from participants during a 20-step walking test for the purpose of identifying individuals with PD, and achieving an accuracy rate of up to 84.5%. In order to minimize the burden of wearing and managing devices for patients, some studies have made innovative explorations. Liu et al. [10] passively measured home gait speed by detecting and analyzing reflected radio waves from the human body using radio devices and used it as a key metric for characterizing and monitoring PD, enabling an objective assessment of disease severity, progression, and medication response at home. In this study, we will investigate how to determine the optimal sensor placement for PD patients during walking, with the goal of reducing subject burden and providing a theoretical foundation for future integrated device development.

The combination of DL technology and wearable sensor data brings great opportunities and potential for intelligent assessment and prediction of PD. However, the ensuing problem is that the uninterpretable nature of the DL model makes people unable to fully trust it, thus limiting its further clinical application [32]. The interpretability of DL model output is particularly important in healthcare fields such as PD assessment, where clinicians can more easily judge whether the model output is based on the relevant features of the input, thus to make better decisions and reduce misdiagnosis [33]. Besides, there are ethical and legal reasons for using interpretable methods, due to the high-risk nature of healthcare applications in which every decision made must be fully informed [34]. Therefore, the understanding and interpretation of DL models have gradually attracted attention of researchers, who try to explain the feature coding and decision-making mechanisms behind the models from different perspectives. Some interpretable methods have been proposed to explain the underlying decision mechanism of DL models, such as gradient-weighted class activation mapping (Grad-CAM) and Shapley additive explanations (SHAP) [35], [36]. Chu et al. [37] obtained characteristic frequency bands (high-delta (3.5–4.5 Hz) and low-alpha (7.5–11 Hz)) and spatial distribution features for early PD identification using EEG based on the interpretability of the model and visualization of the decision-making process, which enhances the reliability of early clinical diagnosis. Shahtalebi et al. [38] used the interpretation of the CNN model by Grad-CAM to compare the heat maps generated by the hand movements of patients with PD and essential tremor, in order to identify distinct frequency bands associated with hand features for each disease, providing insight into their differences. In our previous work, SHAP was introduced to explore the contribution of different motor tasks to the model and identify the primary sources of features that have the greatest impact on the model output, providing references for the clinical diagnosis of PD [9]. Previous studies have explained the model through interpretable methods, promoting the transparency and fairness of the decision-making process. However, there have been few studies on DL model visual interpretation in the field of intelligent PD assessment, and how to optimize models according to the visual interpretation of models is worthy of further exploration. In this study, we conducted research that will help to address these questions.

This study provides an objective detection algorithm that promises to enable long-term monitoring during the daily walking of patients, track the progression of PD, and reduce the dependence on clinicians. A total of 100 subjects were included in the experiment, in which each subject repeatedly collected 10-meter walking data using wearable sensors. The raw data are subjected to continuous wavelet transform (CWT) to obtain more detailed time-frequency information before training in the constructed DL architecture based on a 6-channel convolutional neural network (CNN). The classification performance of each sensor data-driven model was compared and interpretable visualized, with the waist sensor having the best performance. Finally, this work further visualizes the contribution of the data in each channel to the model performance and optimizes the model according to the analysis results. To the best of our knowledge, this is the first study

TABLE I
SUBJECT CHARACTERISTICS

Characteristics	HC (N = 50)	PD (N=50)
Age (Years, $\mu \pm \sigma$)	63.3 ± 8.6	63.6 ± 7.2
Gender (Male/Female)	22/28	22/28
Duration of disease (Years, $\mu \pm \sigma$)	N.A.	4.6 ± 3.2
Total score of UPDRS-III ($\mu \pm \sigma$)	N.A.	27.0 ± 13.3

μ = average, σ = standard deviation, N.A. = not applicable.

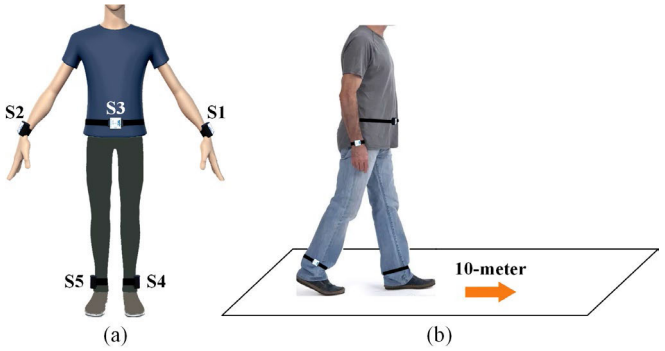

Fig. 1. Data acquisition. (a) The placements of sensors. (b) The 10-meter straight walking test.

to consider the visual interpretation-based optimization of an intelligent classification model in the intelligent diagnosis of PD.

II. MATERIALS AND METHODS

All data processing was conducted using Python 3.7, PyTorch 1.10, and other relevant Python libraries.

A. Data Acquisition

A total of 50 patients with varying severities of PD symptoms and 50 age-matched healthy controls (HC) participated in this study, excluding psychological, other neurological diseases, and movement disorders. All PD subjects were scored by the same neurologist according to the UPDRS-III, and other details of subject characteristics are shown in Table I. All subjects understood the purpose of the data collection and the significance of the study and signed an informed consent form. This study was approved by the Ethics Committee of Shandong Provincial Hospital in Jinan, China (NO.2020-600).

IMUs (BWT901CL, Wit-Motion Company) with the 3-axis accelerometer and 3-axis gyroscope were employed to collect motion data, which could be sent synchronously to the connected personal computer via Bluetooth. Five IMUs attached to the wrist, ankle, and waist using comfortable Velcro straps form a body sensor network, as shown in Fig. 1(a). The selection of these five representative positions enables comprehensive capture of the movement of subjects throughout the experiment while minimizing physiological strain [27]. This fixed manner ensured a consistent sensor position and orientation throughout the acquisition. From the perspective of the sensors, the x, y, and z axes were defined as the transverse direction (left is positive), anterior-posterior (backward is positive), and vertical direction (up is positive), respectively. Subjects deploying the body sensor network were instructed by the neurologist to perform a 10-meter straight walking test at their preferred speed, with each subject repeating the test three times, as shown in Fig. 1(b). All sensors collected gait data with a sampling frequency of 50 Hz, and the built-in accelerometer calibration function of the sensor eliminated human error caused by the different wearing angles of different subjects.

B. Data Pre-Processing

1) Data Screening and Filtering: All data were initially screened to discard data recorded owing to the substandard performance of the subject or non-compliant sensor placement. In addition, at the beginning and end of each walking test, redundant data that did not belong to the actual test owing to human error were removed. Aiming at the problem that a small part of the signal was lost due to unstable transmission during the data recording process, the linear interpolation method was used to supplement the data according to the position of the lost data determined by the data acquisition time interval.

Band-pass filtering is often performed on time series data to eliminate low-frequency and high-frequency noise to reduce external influences. However, DL models usually require minimal filtering or introduce noise into the input data, thereby preventing model overfitting and improving generalization performance and robustness [39], [40]. Therefore, the data were not filtered in this study.

2) Continuous Wavelet Transform: Compared with the fixed time-frequency resolution of the short-time Fourier transform (STFT) technique, CWT enables dynamic resolution analysis of non-stationary signals to capture more detailed time-frequency information [41]. In this study, the Morlet wavelet was used to perform the continuous transform on each time series (3-axis acceleration and angular velocity) of each sensor, providing richer input for the CNN model. The specific form is as follows:

$$CWT(\psi, a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(t)\psi^*\left(\frac{t-b}{a}\right) dt,$$
$$\psi(t) = \exp\left(-\frac{t^2}{2}\right) \cdot \cos(5t), \quad (1)$$

where ψ represents the Morlet mother wavelet, $\psi*$ represents the complex conjugate, a and b represent the scale and translation parameters, respectively.

3) Dataset Construction: To capture transient changes in characteristics during walking, all data were segmented into 2 seconds windows (100 samples) with 50% overlap. The entire dataset consisted of 1832 windows, including 975 for PD and 857 for HC. The training and test sets were randomly divided at 8:2 with the same PD and HC ratios. Each sample contains 6 channels, in which each channel represents a time-frequency map obtained after CWT of a time series component, with a height and width of 64 and 100, respectively.

For each time series of the sensor, the modulus range of the coefficients obtained from the CWT was different, so the data were normalized by scaling the coefficients between -1 and 1 to improve the convergence speed of the model. To prevent potential overfitting, only the training set was

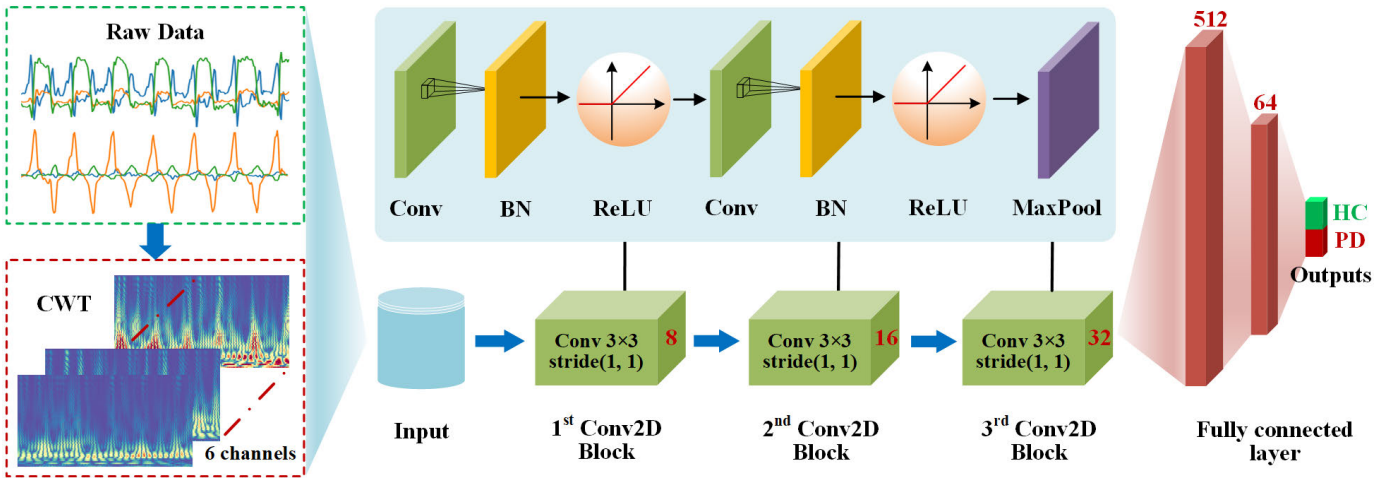

Fig. 2. The CNN Architecture. The entire architecture included four parts: the input layer, three convolutional blocks, two fully connected layers, and the output layer. Different convolution blocks have the same structure except for the number of filters.

used to calculate the normalization parameters, and the entire dataset was processed.

C. CNN Architecture and Training Strategy

CNN was widely used in classification, target detection, image segmentation, and other fields owing to its advantages in automatically learning complex features. However, training a deeper model requires a large amount of computing resources and data [39]. Considering the limitation of the input window sizes, this study constructed a 6-channel CNN architecture with only six convolutional layers according to the data characteristics to avoid overfitting, in which the channel represents the modulus obtained by the CWT on the time-series component of the sensor data.

As shown in Fig. 2, the "Raw Data" is the signal collected by a single sensor. The top represents the acceleration of the three axes, and the bottom represents the angular velocity of the three axes. These time-series data were directly used as input to the 6-channel CNN after CWT, and were not saved as the picture format. Each convolution block included two convolution layers with the same number of filters, and the kernel size was 3×3. The number of filters for the different blocks was 8, 16, and 32, respectively. To improve the convergence speed and robustness of the model, a batch normalization (BN) layer was added after each convolutional layer. The activation function adopted the rectified linear unit (ReLU) function. Each block ends with a max pooling layer of pool size 2×2 and stride 2×2 for downsampling to reduce information redundancy. The last convolutional block was followed by two fully connected layers with 512 and 64 neurons respectively.

The performance of the model was evaluated using 10-fold cross-validation, which divided the training set equally into 10 subsets. In the training process, nine subsets were used for training and the remaining one was used as validation data to test the performance of the model. This process was repeated 10 times to ensure that each subset could participate in training and validation. The hyperparameters were selected based on the average performance of the validation folds from 10-fold cross-validation, and the test set was used to evaluate the final performance of the model. The following settings and hyperparameters were selected for the model:

- Adaptive Moment Estimation (Adam) optimizer with initial learning rate = 0.001.
- Cross entropy loss function.
- Batch size of 8.
- Max epochs equal 100.

Moreover, the problem of model overfitting is often seen in CNNs, which perform extremely well on the training set but fail to generalize to the test set. Therefore, this study employs L2-norm regularization techniques to decay weights [42], combined with early stopping (stopping training when the validation set loss does not drop significantly within 30 epochs) to overcome overfitting.

D. Model Evaluation and Visualization

Accuracy, defined as the percentage of subjects who were correctly classified, was involved throughout the training and testing phases as a measure of performance [43]. Besides, the area under the receiver operating characteristic (ROC) curve (AUC) was introduced to comprehensively evaluate the performance of the proposed model [44]. The AUC represents the probability that the value of the positive sample is higher than that of the negative sample in a random sampling of a positive sample and a negative sample, that is, the larger the AUC value (no more than 1), the better the performance [37].

To make the output more understandable, Grad-CAM was used for visual interpretation of the model [35]. Grad-CAM can make existing state-of-the-art deep models interpretable, avoiding the trade-off between interpretability and accuracy [35]. Specifically, as described in [35]

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k}$$
$$L_{Grad-CAM}^c = ReLU\left(\sum_k \alpha_k^c A^k\right) \quad (2)$$

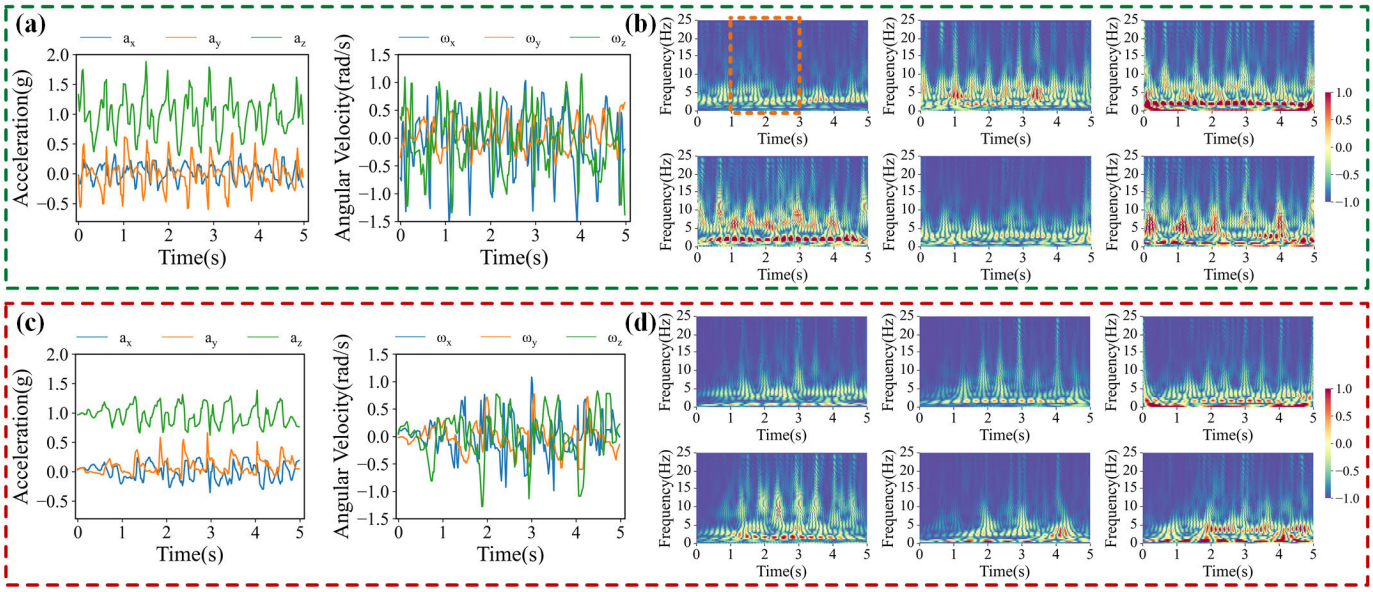

Fig. 3. The raw data and the time-frequency plot from waist sensor (sensor 3) of HC and PD. The green and red boxes represent HC and PD subjects, respectively. (a) and (c) are the acceleration and angular velocity curves. (b) and (d) are the normalized time-frequency plots corresponding to each component of acceleration (upper) and angular velocity (lower). The orange box represents the 2s time window.

where $\partial y^c/\partial Ak\ ij$ represents the gradient of the score y^c (before softmax) for class c with respect to neuron $Ak\ ij$ of the convolutional layer. The gradients obtained by backpropagation are subjected to global average pooling to obtain the importance weights $\alpha c\ k$ of feature map A^k.

Grad-CAM can visualize any layer of any structural CNN without modifying the network structure or retraining. In this study, the last layer of the feature maps was visually interpreted considering rich high-level semantic information and detailed spatial information. The obtained activation feature maps and weights can be used to visualize the basis for model judgment, which in turn helps us distinguish the frequency range and spatiotemporal features with the most significant differences between PD and HC.

III. Results

A. Results of Data Pre-Processing

Fig. 3 shows the raw data and time-frequency plot of each time-series component obtained by the CWT from the waist sensor (sensor 3) of randomly selected HC and PD subjects. The green and red boxes represent HC and PD subjects, respectively, where the acceleration and angular velocity curves are shown on the left ((a) and (c)), (b) and (d) are the normalized time-frequency plots corresponding to each component of acceleration (upper) and angular velocity (lower). From the raw data, HC subjects exhibited greater motion amplitude and frequency than PD subjects, while exhibiting larger values on the time-frequency map.

B. Classification Results

To determine the optimal sensor locations for daily walking detection, the classification performance of five different sensor data-driven CNN models, including accuracy and AUC, are compared in Fig. 4. Fig. 4(a) shows the average accuracy of the different models with 10-fold cross-validation on the test set, and the solid black lines represent their respective standard deviations. The results showed that the performance of the waist sensor (sensor 3) was the best, reaching 98.01% $\pm$ 0.85%, which is significantly better than the other results under one-way ANOVA ($p<0.005$). Fig. 4(b)-(f) depict the ROC and AUC of different models with 10-fold cross-validation on the test set, in which the blue solid line represents the average value of 10 solutions, the other lines in different colors represent the results of 10 calculations, and the gray shading represents the standard deviation. The results also showed that the waist sensor exhibited the best performance (AUC $=$ 0.9981 $\pm$ 0.0017). The results of the one-way ANOVA verified that the performance of the CNN model driven by the waist sensor was significantly better than those of the other sensors ($p<0.005$).

C. Model Visualization

The last convolutional layers of the CNN models that can capture higher-level visual constructs were visually interpreted using Grad-CAM [35], and visual maps of the same size as the input data were obtained through up sampling, as shown in Fig. 5. Four subjects (two HC and two PD) were randomly selected for presentation, where each column in the figure represents a subject and each row represents a sensor in Fig. 1(a). As shown, the feature points that significantly contributed to the discrimination of PD subjects were more concentrated in the lower frequency bands (0.5-3 Hz) relative to HC (4-6 Hz).

The best-performing waist sensor has better discrimination between HC and PD because the waist is closer to the center of gravity of the body and can better reflect the overall motion condition than the unilateral limb [30], [31]. Moore et al. [45] defined the locomotor band of walking as 0.5-3 Hz through the IMU attached on the left shank, while the sensor mounted

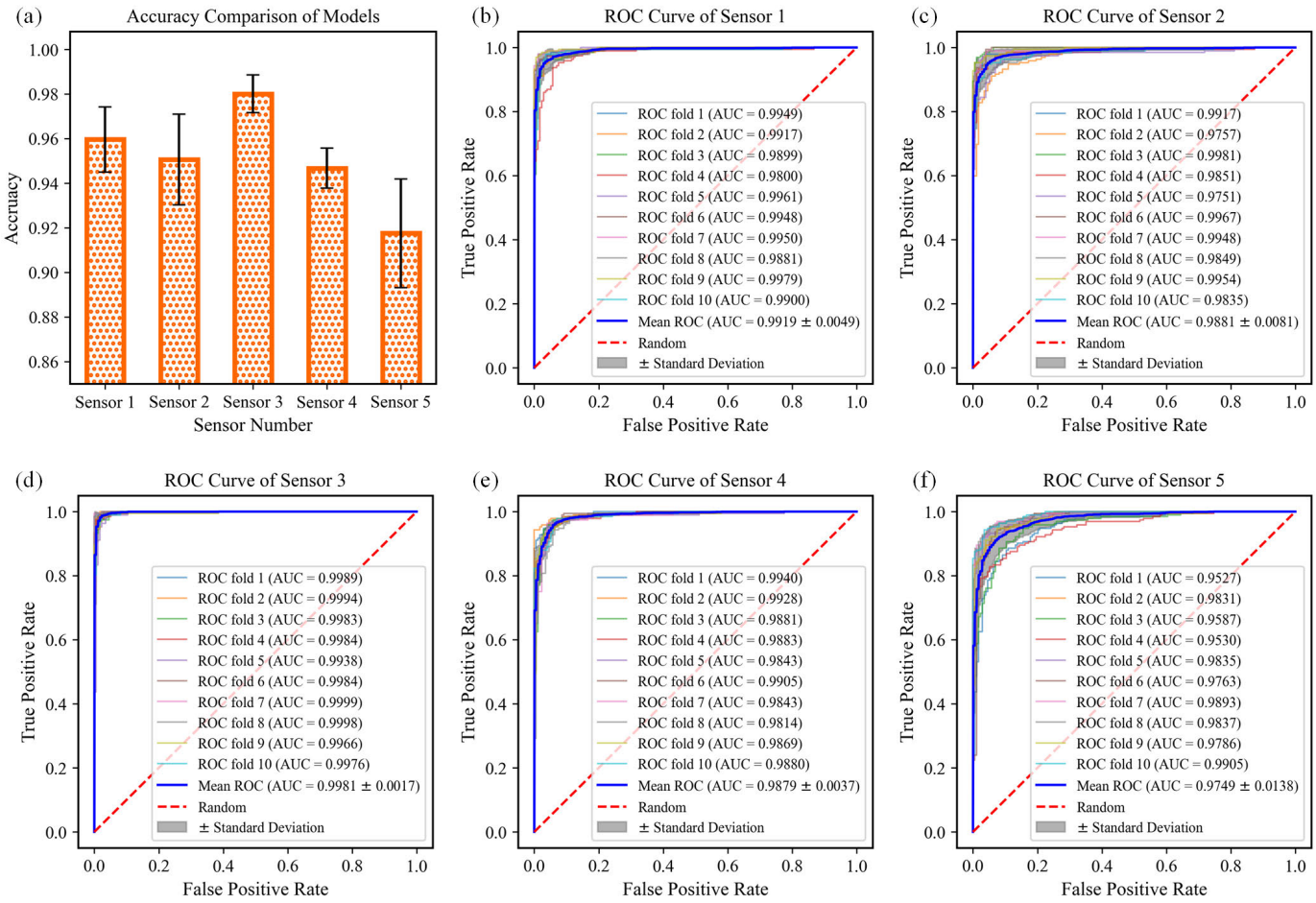

Fig. 4. Accuracy and AUC comparison of CNN models driven by 5 different sensor data. (a) The average accuracy of different models with 10-fold cross-validation on the test set, and the solid black lines represent their respective standard deviations. (b-f) The ROC and AUC of different models with 10-fold cross-validation on the test set. The blue solid line represents the average value of 10 solutions, the other lines in different colors represent the results of 10 calculations, and the gray shading represents the standard deviation.

on the waist in this work could display the activities of both body segments, thus the identified frequency band of HC was 4-6 Hz. In addition, the decision basis of the left-wrist sensor was also evident between the two classes, which was also reflected in the AUC results. Motor symptoms were more serious in the dominant hand [46], and the subjects included in this work were almost all right-handed, which was reflected in right-wrist sensor of PD with a wider range of frequency band (0.5-7 Hz, including tremor).

D. Model Optimization

Although the human body posture can be solved by IMU data, the limited time series components (e.g., only accelerometer data) of a particular movement may be sufficient to describe its motion characteristics [9], [47], thus reducing the burden of data acquisition. The above model visualization method explained the decision basis of each sensor and proved the advantage of waist sensor over the others, however it was unable to verify the contribution of its various time series components to classification. Therefore, the waist sensor-driven model with the best performance was further optimized to remove redundant data, reduce the data processing pressure, improve the generalization performance, and explore the interpretability of the model.

The dataset structure was reconstructed from a 6-channel 2D input ($6 \times 64 \times 100$) to a 1-channel 3D input ($1 \times 6 \times 64 \times 100$) to visualize the key features of each time series (3-axis acceleration and angular velocity). The proposed network was modified to a 3D CNN and retrained while keeping the basic architecture unchanged, only performing the waist sensor with the best performance. The contribution of each time-series component for the waist sensor to HC and PD classifications is visually interpreted in Fig. 6. As shown, PD focused on lower frequency band features in a_x and ω_z compared with HC, which means that the lateral movement of patients becomes difficult. Moreover, PD exhibited higher frequency features on a_z than HC because of the more stable gait in HC. The a_y component was not obvious in HC but was focused on the low-frequency band in PD. Interestingly, these two groups have the same interpretation of ω_y which represents side-to-side shaking. The inability to distinguish effectively may be due to the stable gait of HC and the slow movement of PD, both of which show low-frequency features. In addition, due to individual differences in patient symptoms, ω_x was confused on both. Additional sample statistics are shown in Supplementary Fig. S6-S11.

According to the above analysis, components that could not effectively distinguish HC from PD were removed,

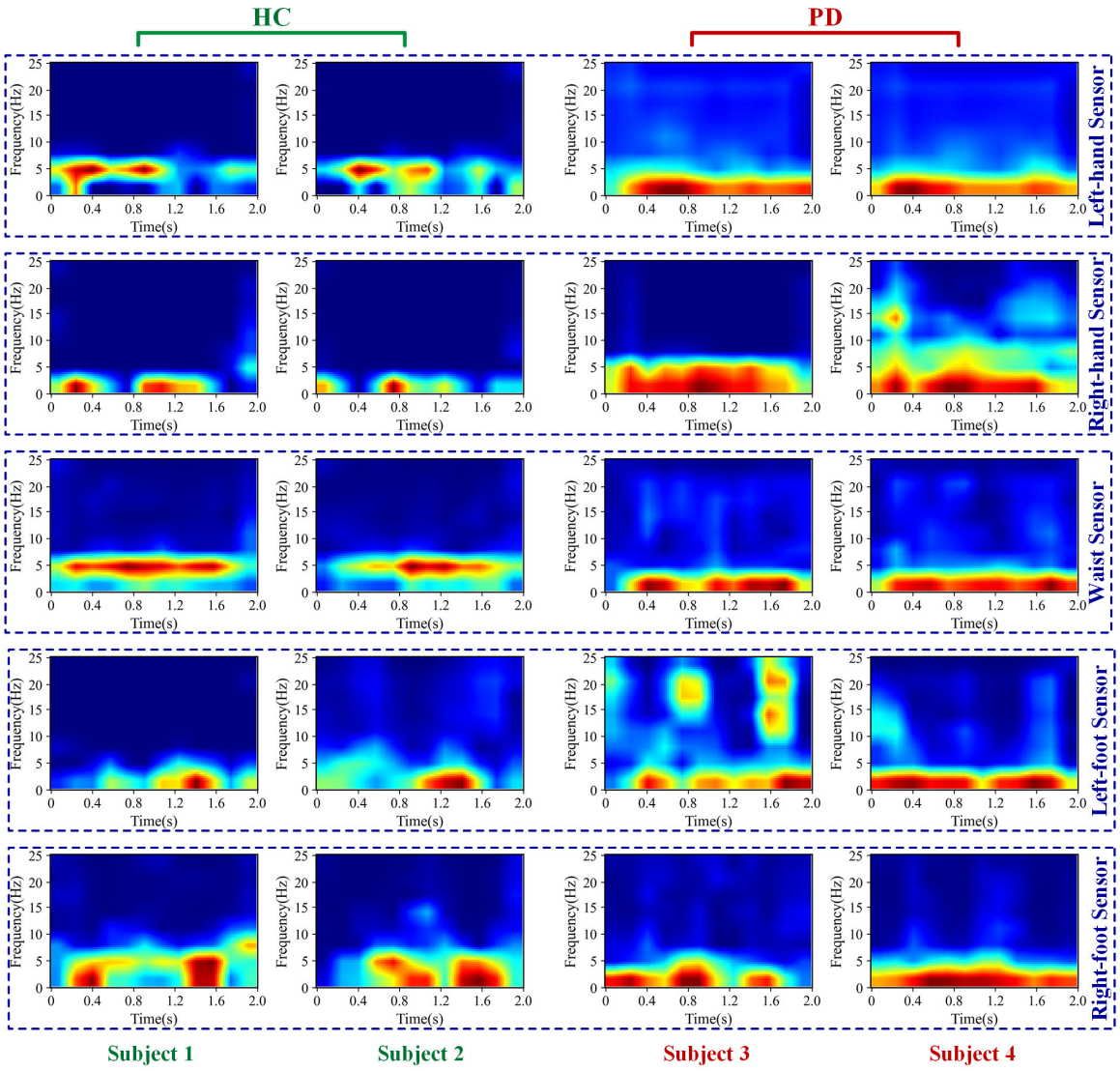

Fig. 5. The visual interpretation of the last convolutional layers. The *x*-axis and *y*-axis represent time and frequency, respectively. Four subjects (2 HC and 2 PD) were randomly selected for presentation, where each column represents a subject, and each row represents a sensor.

and only a_x, a_z, and ω_z were retained to train the new CNN model. Although the data component was reduced by half, the new model still showed good performance (AUC $= 0.9929 \pm 0.0019$). The optimized model focused on the important characteristics of walking in patients with PD. Although the performance of the optimized model was slightly lower than that of the original model, the raw data processing pressure was significantly reduced (50%), which provided the possibility for the development of a real-time monitoring system for daily gait.

IV. Discussion

In this study, we adopted DL technology and wearable sensors to monitor gait disorders in patients with PD for a long time. The original data collected by the IMU were one-dimensional time series. If it was directly input into the CNN model for one-dimensional convolution operation, only time-domain features could be obtained, while the equally important frequency domain features were ignored, resulting in limited model performance. Some studies used the fast Fourier transform method to extract the frequency domain features of data [9], [47], [48], but it could not guarantee the time-frequency information of the data at the same time. The STFT solved this problem to some extent but had a fixed time-frequency resolution [49], [50]. Therefore, CWT with dynamic resolution was used in this study to capture more detailed time-frequency features before input to the CNN. CNNs were designed to process data from multiple arrays, such as the color image consisting of three channels (RGB) [20]. According to the collected data format (3-axis acceleration and angular velocity), we regarded each time series component as a channel array to construct a customized 6-channel CNN. Considering the limited sample size and reduced computational cost, only six convolutional layers were designed for the network.

In order to mitigate potential patient burden and discomfort, current research is shifting towards employing a minimal number of wearable sensors [10], [25]. Hence, in this study,

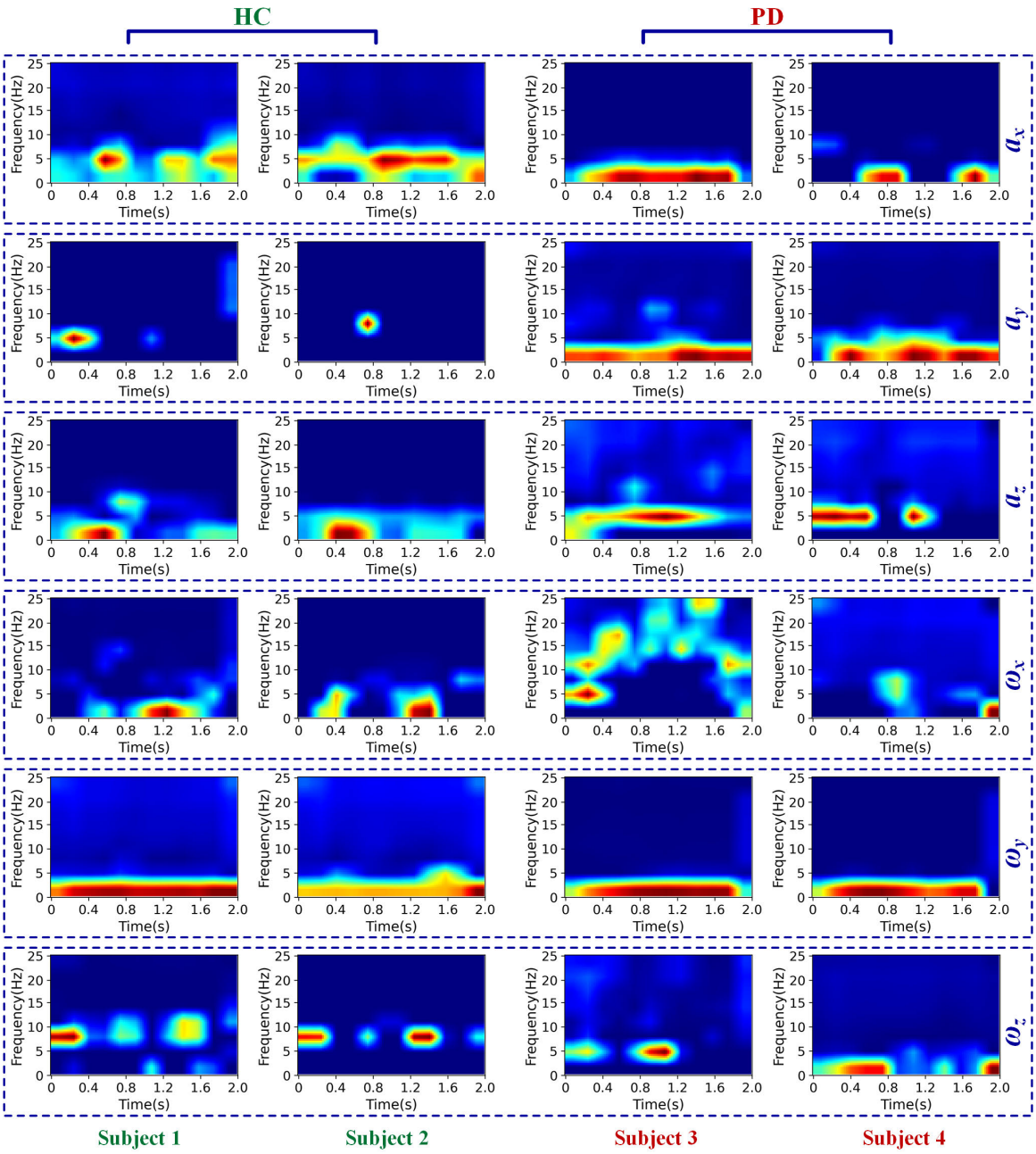

Fig. 6. The contribution of each time-series component for waist sensor driven 3D CNN model. Each column represents a subject, and each row represents a time-series component.

we constructed five CNN models based on the data of each sensor and compared their classification performance. The reason why the waist sensor (sensor 3) has better performance (AUC $=$ 0.9981 $\pm$ 0.0017) than the other sensors may be that its position is closer to the center of gravity of the body, reflecting the overall movement during walking [30], [31]. Supplementary Fig. S13 shows the training-validation learning curve of a waist sensor-driven model in 10-fold cross validation, presenting Loss and Accuracy metrics across the number of training epochs.

However, the black-box nature of CNN limits it in terms of high-risk decision-making in the medical field even with excellent performance [51]. To build trust in intelligent systems, we must construct "transparent" models that explain the decision-making mechanisms behind them [35]. Therefore, to further explore the decision-making mechanism of the proposed model, we used the Grad-CAM method to visually interpret the feature maps of the last layer of each model without changing the structure. The CNN model driven by the waist sensor showed better discrimination ability, and the feature points that make a significant contribution to PD are concentrated in the lower frequency band (0.5$\sim$3 Hz) compared with HC. Although the performances of the other four sensor-driven models were similar except for waist sensor, the decision basis was worse when attached to the ankle compared to the wrist. This was because the motor manifestations of

PD patients started distally in one upper limb [52], whereas this study involved a small number of early stage patients without motor disorders in the lower limb. The analysis results of the sensors attached to the ankles also showed some differences, but most of the features were concentrated in the lower frequency band determined by the locomotor band (0.5-3 Hz) of the unilateral limb [45], resulting in invalid differentiation. Another reason may be the heterogeneity of limb motor symptoms in PD patients. Due to the individual differences of patients, the sensors located in limbs cannot accurately measure the overall movement of patients, so the visual maps showed the disorder. Additional sample statistics are shown in Supplementary Fig. S1-S5.

In addition, not all windows were classified correctly, and a few incorrect cases were used to further reveal the underlying decision-making mechanism. Supplementary Fig. S12 shows the visual maps of different sensor-driven models in the case of decision-making errors, where the first two columns show that HC is misclassified as PD and the last two columns show that PD is misclassified as HC. As shown, misclassification occurred because a small number of HC windows exhibited disordered visual maps, and PD windows had an important decision basis in higher frequency bands (4-6 Hz). This was consistent with the above findings that the feature points that play an important role in the classification of PD were concentrated in the lower frequency bands relative to HC. Some patients had mild motor symptoms, indicating that gait impairment is not always present. Similarly, HC subjects may have gait fluctuations owing to the influence of the external environment, resulting in classification errors.

The ultimate goal of intelligent PD monitoring is to develop an integrated system [16] that combines data collection sensors and classification algorithms to monitor gait performance for a long time in the daily life of patients. However, the high computational cost of the CNN model is a great challenge for the microprocessor [53], even if a CNN with a simple architecture was used in this work. Therefore, we further optimized the waist sensor-driven model to reduce data processing pressure. The decision-making basis of each time series component was visualized by the reconstructed 3D CNN, and a new 3-channel CNN model was trained after removing the three poorly performing components. Although the performance of the optimized model is slightly reduced, the computational cost of the raw data is reduced by 50%, which makes it possible to develop a system for monitoring daily gait for a long time. Some researchers have also explored hardware implementations of DL algorithms. Mikos et al. [28] demonstrated that the freezing of a gait detection system built from neural networks capable of learning in real-time was integrated into a single sensor node based on field programmable gate arrays (FPGA), and could operate for more than 9 hours while providing auditory biofeedback cues. Langer et al. [54] reported the implementation of a temporal convolutional network trained to detect freezing of gait based on an FPGA, which can ensure sufficient time to trigger the cue in less than a millisecond to prevent the patient from falling. These will facilitate the design of dedicated hardware for PD.

However, it is important to acknowledge certain limitations. This study only involved straight walking while ignored turning, and could only distinguish PD patients from healthy subjects while ignoring the different severity degrees of PD patients. In the future, we will further explore the application of the proposed approach in these aspects.

V. Conclusion

Due to impaired motor function, PD patients showed symptoms such as slow walking and unsteady gait, and showed lower movement amplitude and frequency in the sensor data and CWT map. By comparing the performance of the CNN models driven by different sensors, the optimal sensor position was determined to be the waist with an AUC of 0.9981 ± 0.0017. Grad-CAM further demonstrated that the feature points with greater contribution to PD were concentrated in the lower frequency band (0.5∼3 Hz) compared with HC. Moreover, we combined Grad-CAM with 3D CNN to propose a model optimization scheme that visually interprets the contribution of each time series component to the model to remove components with insufficient contribution, which reduces the raw data processing pressure by 50% at a slight performance sacrifice. In the future, we plan to design an integrated dedicated hardware circuit that can be attached to the waist of patients to monitor gait performance for a long time and track the progress of PD. This is expected to guarantee the realization of personalized treatment and improve the treatment effect.

References

[1] A. Channa, N. Popescu, and V. Ciobanu, "Wearable solutions for patients with Parkinson's disease and neurocognitive disorder: A systematic review," *Sensors*, vol. 20, no. 9, p. 2713, May 2020.

[2] E. Angelopoulou, Y. N. Paudel, S. G. Papageorgiou, and C. Piperi, "Environmental impact on the epigenetic mechanisms underlying Parkinson's disease pathogenesis: A narrative review," *Brain Sci.*, vol. 12, no. 2, p. 175, Jan. 2022.

[3] E. Dorsey, T. Sherer, M. S. Okun, and B. R. Bloem, "The emerging evidence of the Parkinson pandemic," *J. Parkinsons Dis.*, vol. 8, no. s1, pp. S3–S8, Dec. 2018.

[4] G. Pagano, N. Ferrara, D. J. Brooks, and N. Pavese, "Age at onset and Parkinson disease phenotype," *Neurology*, vol. 86, no. 15, pp. 1400–1407, Feb. 2016.

[5] P. W.-L. Ho et al., "Age-dependent accumulation of oligomeric SNCA/α-synuclein from impaired degradation in mutant LRRK2 Knockin mouse model of Parkinson disease: Role for therapeutic activation of chaperone-mediated autophagy (CMA)," *Autophagy*, vol. 16, no. 2, pp. 347–370, Feb. 2020.

[6] C. H. Adler et al., "Low clinical diagnostic accuracy of early vs advanced Parkinson disease: Clinicopathologic study," *Neurology*, vol. 83, no. 5, pp. 406–412, Jul. 2014.

[7] Z. Ayaz, S. Naz, N. H. Khan, I. Razzak, and M. Imran, "Automated methods for diagnosis of Parkinson's disease and predicting severity level," *Neural Comput. Appl.*, vol. 35, pp. 14499–14534, Mar. 2022.

[8] M. Belić, V. Bobić, M. Badža, N. Šolaja, M. Durić-Jovičić, and V. S. Kostić, "Artificial intelligence for assisting diagnostics and assessment of Parkinson's disease—A review," *Clin. Neurol. Neurosurg.*, vol. 184, Sep. 2019, Art. no. 105442.

[9] M. Chen, Z. Sun, F. Su, Y. Chen, D. Bu, and Y. Lyu, "An auxiliary diagnostic system for Parkinson's disease based on wearable sensors and genetic algorithm optimized random forest," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 2254–2263, 2022.

[10] Y. Liu et al., "Monitoring gait at home with radio waves in Parkinson's disease: A marker of severity, progression, and medication response," *Sci. Transl. Med.*, vol. 14, no. 663, Sep. 2022, Art. no. eadc9669.

[11] W. Maetzler, J. Domingos, K. Srulijes, J. J. Ferreira, and B. R. Bloem, "Quantitative wearable sensors for objective assessment of Parkinson's disease," *Movement Disorders*, vol. 28, no. 12, pp. 1628–1637, Oct. 2013.

[12] A. C. Albán-Cadena, F. Villalba-Meneses, K. O. Pila-Varela, A. Moreno-Calvo, C. P. Villalba-Meneses, and D. A. Almeida-Galárraga, "Wearable sensors in the diagnosis and study of Parkinson's disease symptoms: A systematic review," *JME&T*, vol. 45, no. 7, pp. 532–545, Jun. 2021.

[13] E. Rovini, C. Maremmani, and F. Cavallo, "Automated systems based on wearable sensors for the management of Parkinson's disease at home: A systematic review," *Telemed. e-Health*, vol. 25, no. 3, pp. 167–183, Mar. 2019.

[14] R. Deb, G. Bhat, S. An, U. Ogras, and H. Shill, "A systematic survey of research trends in technology usage for Parkinson's disease," *Sensors*, vol. 22, no. 15, p. 5491, Jul. 2022.

[15] D. Johansson, K. Malmgren, and M. A. Murphy, "Wearable sensors for clinical applications in epilepsy, Parkinson's disease, and stroke: A mixed-methods systematic review," *J. Neurol.*, vol. 265, no. 3, pp. 1740–1752, 2018.

[16] S. Ancona et al., "Wearables in the home-based assessment of abnormal movements in Parkinson's disease: A systematic review of the literature," *J. Neurol.*, vol. 269, no. 1, pp. 100–110, Jan. 2022.

[17] A. Talitckii et al., "Defining optimal exercises for efficient detection of Parkinson's disease using machine learning and wearable sensors," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–10, 2021.

[18] S. K. Khare, V. Bajaj, and U. R. Acharya, "PDCNNet: An automatic framework for the detection of Parkinson's disease using EEG signals," *IEEE Sensors J.*, vol. 21, no. 15, pp. 17017–17024, Aug. 2021.

[19] N. Sharma, R. Sharma, and N. Jindal, "Machine learning and deep learning applications—A vision," *Global Transitions Proc.*, vol. 2, no. 1, pp. 24–28, Jun. 2021.

[20] J. C. Vásquez-Correa, T. Arias-Vergara, J. R. Orozco-Arroyave, B. Eskofier, J. Klucken, and E. Nöth, "Multimodal assessment of Parkinson's disease: A deep learning approach," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 4, pp. 1618–1630, Jul. 2019.

[21] A. Balakrishnan, J. Medikonda, and P. K. Namboothiri, "Role of wearable sensors with machine learning approaches in gait analysis for Parkinson's disease assessment: A review," *Eng. Sci.*, vol. 19, pp. 5–19, Jan. 2022.

[22] L. Brognara, P. Palumbo, B. Grimm, and L. Palmerini, "Assessing gait in Parkinson's disease using wearable motion sensors: A systematic review," *Diseases*, vol. 7, no. 1, p. 18, Feb. 2019.

[23] C. Chatzaki et al., "The smart-insole dataset: Gait analysis using wearable sensors with a focus on elderly and Parkinson's patients," *Sensors*, vol. 21, no. 8, p. 2821, Apr. 2021.

[24] J.-F. Daneault et al., "Accelerometer data collected with a minimum set of wearable sensors from subjects with Parkinson's disease," *Sci. Data*, vol. 8, no. 1, pp. 1–13, Feb. 2021.

[25] J.-F. Daneault et al., "Estimating bradykinesia in Parkinson's disease with a minimum number of wearable sensors," in *Proc. IEEE/ACM Int. Conf. Connected Health, Appl., Syst. Eng. Technol. (CHASE)*, Jul. 2017, pp. 264–265.

[26] L. Lonini et al., "Wearable sensors for Parkinson's disease: Which data are worth collecting for training symptom detection models," *NPJ Digit. Med.*, vol. 1, no. 1, pp. 1–8, Nov. 2018.

[27] M. Sica et al., "Continuous home monitoring of Parkinson's disease using inertial sensors: A systematic review," *PLoS ONE*, vol. 16, no. 2, Feb. 2021, Art. no. e0246528.

[28] V. Mikos et al., "A wearable, patient-adaptive freezing of gait detection system for biofeedback cueing in Parkinson's disease," *IEEE Trans. Biomed. Circuits Syst.*, vol. 13, no. 3, pp. 503–515, Jun. 2019.

[29] L. R. Peraza et al., "An automatic gait analysis pipeline for wearable sensors: A pilot study in Parkinson's disease," *Sensors*, vol. 21, no. 24, p. 8286, Dec. 2021.

[30] M. Juutinen et al., "Parkinson's disease detection from 20-step walking tests using inertial sensors of a smartphone: Machine learning approach based on an observational case-control study," *PLoS ONE*, vol. 15, no. 7, Jul. 2020, Art. no. e0236258.

[31] D. Rodríguez-Martín et al., "Home detection of freezing of gait using support vector machines through a single Waist–Worn triaxial accelerometer," *PLoS ONE*, vol. 12, no. 2, Feb. 2017, Art. no. e0171764.

[32] C. Janiesch, P. Zschech, and K. Heinrich, "Machine learning and deep learning," *Electron. Mark.*, vol. 31, no. 3, pp. 685–695, 2021.

[33] M. A. Ahmad, A. Teredesai, and C. Eckert, "Interpretable machine learning in healthcare," in *Proc. IEEE Int. Conf. Healthcare Informat. (ICHI)*, Jun. 2018, pp. 447–447.

[34] R. Katarya, P. Sharma, N. Soni, and P. Rath, "A review of interpretable deep learning for neurological disease classification," in *Proc. 8th Int. Conf. Adv. Comput. Commun. Syst. (ICACCS)*, vol. 1, Mar. 2022, pp. 900–906.

[35] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 618–626.

[36] S. M. Lundberg et al., "From local explanations to global understanding with explainable AI for trees," *Nature Mach. Intell.*, vol. 2, no. 1, pp. 56–67, Jan. 2020.

[37] C. Chu et al., "Deep learning reveals personalized spatial spectral abnormalities of high delta and low alpha bands in EEG of patients with early Parkinson's disease," *J. Neural Eng.*, vol. 18, no. 6, Dec. 2021, Art. no. 066036.

[38] S. Shahtalebi, S. F. Atashzar, R. V. Patel, M. S. Jog, and A. Mohammadi, "A deep explainable artificial intelligent framework for neurological disorders discrimination," *Sci. Rep.*, vol. 11, no. 1, p. 9630, May 2021.

[39] B. Shi, A. Tay, W. L. Au, D. M. L. Tan, N. S. Y. Chia, and S.-C. Yen, "Detection of freezing of gait using convolutional neural networks and data from lower limb motion sensors," *IEEE Trans. Biomed. Eng.*, vol. 69, no. 7, pp. 2256–2267, Jul. 2022.

[40] N. E. Khalifa, M. Loey, and S. Mirjalili, "A comprehensive survey of recent trends in deep learning for digital images augmentation," *Artif. Intell. Rev.*, vol. 55, pp. 1–27, Sep. 2021.

[41] T. Wang, C. Lu, Y. Sun, M. Yang, C. Liu, and C. Ou, "Automatic ECG classification using continuous wavelet transform and convolutional neural network," *Entropy*, vol. 23, no. 1, p. 119, Jan. 2021.

[42] A. Badola, V. P. Nair, and R. P. Lal, "An analysis of regularization methods in deep neural networks," in *Proc. IEEE 17th India Council Int. Conf. (INDICON)*, Dec. 2020, pp. 1–6.

[43] N. D. Cilia, T. D'Alessandro, C. De Stefano, F. Fontanella, and M. Molinara, "From online handwriting to synthetic images for Alzheimer's disease detection using a deep transfer learning approach," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 12, pp. 4243–4254, Dec. 2021.

[44] M. Nasseri et al., "Non-invasive wearable seizure detection using long–short-term memory networks with transfer learning," *J. Neural Eng.*, vol. 18, no. 5, Apr. 2021, Art. no. 056017.

[45] S. T. Moore, H. G. MacDougall, and W. G. Ondo, "Ambulatory monitoring of freezing of gait in Parkinson's disease," *J. Neurosci. Methods*, vol. 167, no. 2, pp. 340–348, Jan. 2008.

[46] A. van der Hoorn, H. Burger, K. L. Leenders, and B. M. de Jong, "Handedness correlates with the dominant Parkinson side: A systematic review and meta-analysis," *Movement Disorders*, vol. 27, no. 2, pp. 206–210, Feb. 2012.

[47] R. San-Segundo et al., "Parkinson's disease tremor detection in the wild using wearable accelerometers," *Sensors*, vol. 20, no. 20, p. 5817, Oct. 2020.

[48] M. Mancini et al., "Measuring freezing of gait during daily-life: An open-source, wearable sensors approach," *J. NeuroEng. Rehabil.*, vol. 18, no. 1, pp. 1–13, Jan. 2021.

[49] A. Talitckii et al., "Avoiding misdiagnosis of Parkinson's disease with the use of wearable sensors and artificial intelligence," *IEEE Sensors J.*, vol. 21, no. 3, pp. 3738–3747, Feb. 2021.

[50] J. L. Adams et al., "A real-world study of wearable sensors in Parkinson's disease," *NPJ Parkinson's Disease*, vol. 7, no. 1, pp. 1–8, Nov. 2021.

[51] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," *Nature Mach. Intell.*, vol. 1, no. 5, pp. 206–215, May 2019.

[52] M. H. G. Monje, Á. Sánchez-Ferro, J. A. Pineda-Pardo, L. Vela-Desojo, F. Alonso-Frech, and J. A. Obeso, "Motor onset topography and progression in Parkinson's disease: The upper limb is first," *Movement Disorders*, vol. 36, no. 4, pp. 905–915, Jan. 2021.

[53] A. Shawahna, S. M. Sait, and A. El-Maleh, "FPGA-based accelerators of deep learning networks for learning and classification: A review," *IEEE Access*, vol. 7, pp. 7823–7859, 2019.

[54] P. Langer, A. Haddadi Esfahani, Z. Dyka, and P. Langendörfer, "FPGA-based realtime detection of freezing of gait of Parkinson patients," in *Proc. EAI Int. Conf. Body Area Netw.* Cham, Switzerland: Springer, 2021, pp. 101–111.