Improved Deep Learning Technique to Detect Freezing of Gait in Parkinson’s Disease Based on Wearable Sensors

Bochen Li 1,2, Zhiming Yao 1,*, Jianguo Wang 1,2, Shaonan Wang 1,2, Xianjun Yang 1 and Yining Sun 1

1 Hefei Institutes of Physical Science, Chinese Academy of Sciences, Hefei 230031, China; bchli@mail.ustc.edu.cn (B.L.); jgw@mail.ustc.edu.cn (J.W.); wangsnnn@mail.ustc.edu.cn (S.W.); xjyang@iim.ac.cn (X.Y.); ynsun@iim.ac.cn (Y.S.)
2 Science Island Branch of Graduate School, University of Science and Technology of China, Hefei 230026, China
* Correspondence: zhmyao@iim.ac.cn; Tel.: +86-0551-6559-1181

Received: 19 October 2020; Accepted: 10 November 2020; Published: 14 November 2020

Abstract: Freezing of gait (FOG) is a paroxysmal dyskinesia, which is common in patients with advanced Parkinson’s disease (PD). It is an important cause of falls in PD patients and is associated with serious disability. In this study, we implemented a novel FOG detection system using deep learning technology. The system takes multi-channel acceleration signals as input, uses one-dimensional deep convolutional neural network to automatically learn feature representations, and uses recurrent neural network to model the temporal dependencies between feature activations. In order to improve the detection performance, we introduced squeeze-and-excitation blocks and attention mechanism into the system, and used data augmentation to eliminate the impact of imbalanced datasets on model training. Experimental results show that, compared with the previous best results, the sensitivity and specificity obtained in 10-fold cross-validation evaluation were increased by 0.017 and 0.045, respectively, and the equal error rate obtained in leave-one-subject-out cross-validation evaluation was decreased by 1.9%. The time for detection of a 256 data segment is only 0.52 ms. These results indicate that the proposed system has high operating efficiency and excellent detection performance, and is expected to be applied to FOG detection to improve the automation of Parkinson’s disease diagnosis and treatment.

Keywords: Parkinson’s disease; freezing of gait; detection; deep learning; convolutional neural networks; long short-term memory; attention mechanism; squeeze-and-excitation block; data augmentation

1. Introduction

Parkinson’s disease (PD) is a very common neurodegenerative disease with typical motor clinical symptoms such as bradykinesia, freezing of gait (FOG), rigidity, resting tremor and postural tremor [1]. These symptoms can interfere with patients’ daily activities, endanger their mental health, and cause their quality of life to decline. About 50% of PD patients have experienced FOG symptoms, which is the main cause of falls [2,3]. FOG is defined as a “brief, episodic absence or marked reduction in forward progression of the feet despite the intention to walk” [4]. Schaafsma et al. [5] defined five subtypes of FOG: start hesitation, turn hesitation, hesitation in tight quarters, destination hesitation, and open space hesitation. Generally, FOG is associated with a subjective feeling of “the feet being glued to the ground” [6]. The environment, medications, and anxiety all affect the frequency and duration of FOG [7].
Scales are commonly used to evaluate FOG in clinical practice, such as Unified Parkinson’s Disease Rating Scale (UPDRS), Activities of Daily Living (ADL) part 14 [8] and freezing of gait questionnaire (FOG-Q) [9]. However, scale-based assessment is a subjective method that depends on the doctor’s experience, the description of the patient or his caregiver, and the patient’s performance during the assessment. Due to the episodic and paroxysmal nature of FOG, it is extremely difficult to obtain accurate frequency and duration of FOG based on the scale method. Levodopa is currently the most effective treatment for PD, but the dosage schedule needs to be adjusted dynamically according to the condition [10]. Therefore, there is a need for an objective FOG detection method that can monitor freezing events in real time in daily life, give warnings before FOG to reduce the frequency of freezing, and intervene after FOG to reduce the duration of FOG and risk of falling. The computerized method for FOG detection is very helpful for quantifying diseases, formulating drug regimen, reducing the risk of falls and improving the quality of life.

2. Related Work

Basically, computerized FOG detection methods can be roughly divided into two groups according to the analyzed signal. The first group tries to figure out the differences in physiological signals between dyskinesia and normal walking to detect or predict FOG [11–14]. For example, Handojoseno et al. used the wavelet coefficients of electroencephalogram (EEG) signals as the input of the MultiLayer Perceptron Neural Network and k-Nearest Neighbor classifier, which can predict the transition from walking to FOG with 87% sensitivity and 73% accuracy [14]. The second group generally uses three-dimensional (3D) motion analysis systems [15–18], plantar pressure measurement systems [19–21] or inertial sensors (accelerometers, gyroscopes or magnetometers) [22–25] to obtain more intuitive gait kinematics or dynamics signals. Delval et al. used multiple cameras to capture the gait kinematics signals of patients with reflective markers attached to their bodies from different angles [17], and Okuno et al. used a plantar pressure measurement system (1.92 × 0.88 m) to record the soles of the patients walking [19]. Although the above-mentioned sensors all can be applied to FOG detection, the current FOG detection in the community environment is mainly based on inertial sensors. Inertial sensors have the characteristics of small size, low power consumption, low price, and wearability, making them the most suitable device for long-term monitoring of PD patients’ FOG in community environment.

Through frequency-domain analysis of the vertical acceleration signal of the left calf of PD patients, Moore et al. defined the ‘freeze’ band (3–8 Hz), ‘locomotor’ band (0.5–8 Hz) and freeze index (FI, the ratio of the power in the ‘freeze’ band to the power in the ‘locomotor’ band). Using the frozen threshold to detect FOG, 78% sensitivity and 80% specificity were achieved in a patient-independent model, and 89% sensitivity and 90% specificity were achieved in a patient-dependent model [22]. Several improved indicators [17,24–27] based on FI were further proposed to improve the accuracy of detection. For example, energy threshold [25], acceleration signal entropy [24] and cadency variation [27]. However, such threshold-based methods can only provide linear classification capabilities. FOG detection is a challenging task. Firstly, human motion signals are complex, and the start and end of FOG events are random. Secondly, the signals recorded by inertial sensors contain noise and irrelevant redundancy. Lastly, everyone has their own baseline health status, and the difference compared to the baseline can only indicate whether they deviate from the optimal state of health [28], which has a great impact on the performance of the classification model.

Deep learning (DL) refers to a method of representation learning for sample data based on a multi-layer artificial neural network. Recently, DL has achieved impressive success in the fields of computer vision [17,29], bioinformatics [30], and speech recognition [31]. Convolutional neural networks (CNNs) [32] are a type of deep neural network (DNN) that can automatically extract multi-level features of signals through convolution kernels of different sizes. These features usually have better distinguishing capabilities than handcrafted ones. Long short-term memory (LSTM) is a special recurrent neural network (RNN) which solves the problem of gradient vanishing and gradient exploding in the training process of long sequences by adding forget gates in RNN, and can be used for
modeling sequence data. The combination of CNNs and LSTM in a unified framework can capture the temporal correlation of features extracted by convolution operations, and has been successfully applied in fields such as natural language processing [33] and human activity recognition (HAR) [34,35].

Inspired by the excellent performance of DL in many classification and recognition tasks, some scholars [34,36–38] tried to use deep learning methods to analyze human inertial signals. For example, Xia Yi et al. designed a five-layer CNN to detect FOG, where the first three layers are used for feature extraction, the fourth layer is used for feature fusion, and the last layer is used for classification [37]. Guan et al. demonstrated that ensembles of deep LSTM learners outperform individual LSTM networks in human activity recognition using wearables [36]. Although the DL-based methods have achieved promising performance in HAR and FOG detection, these neural networks were not optimized for fusing multi-sensor/multi-channel signals and were designed without considering the problems of the gradient vanishing of LSTM when the sequence is too long. We designed a DL network for detecting FOG in PD patients based on CNN and LSTM, and improved the model’s performance by adding squeeze-and-excitation block (SE block) and attention layers. The experimental results show that the performance of the proposed model is better than previous models and has high efficiency.

3. Methods

In this section, we proposed a DL framework for detecting FOG of PD patients, and detailed the components and principles of the framework. The deep convolutional network in the framework is used to automatically extract signal features, and the attention-enhanced LSTM is used to model the temporal dependencies between features. In this study, acceleration sensors were used to collect patient motion signals.

3.1. Deep Convolutional Network

Since the acceleration signal is a time series signal, we choose a temporal convolutional network (TCN, a special CNN whose input is generally time series signal) for learning feature representations [39]. Considering a temporal convolutional network with \( L \) convolutional layers, and its input is a 1D signal, \( X \), where \( X_t \in \mathbb{R}^{F_0} \) is the input feature vector of length \( F_0 \) for time step \( t \) for \( 0 < t \leq T \). Note that we denote the number of time steps in each layer as \( T_l \). The filter duration for each layer is \( d \), and they are parameterized by tensor \( W^{(l)} \in \mathbb{R}^{d \times d \times F_{l-1}} \) and biases \( b^{(l)} \in \mathbb{R}^{F_l} \), where \( l \in \{1, \ldots, L\} \) is the layer index, \( F_l \) is the feature map number of the \( l \)-th layer. Then, for the \( l \)-th layer, the feature component \( F^{(l)}_{i,l} \in \mathbb{R}^{F_l} \) at the position of \( i \) is a function of the incoming activation tensor \( E^{(l-1)}_{i,l-1} \in \mathbb{R}^{F_{l-1} \times T_{l-1}} \) of the previous layer, such that,

\[
F^{(l)}_{i,l} = f_i^{(l)} + \sum_{j=1}^{d} W_{i,j}^{(l)} E^{(l-1)}_{i,d+j-l},
\]

for each time \( t \), where \( \langle \cdot \rangle \) denotes the inner product, and \( f(\cdot) \) is a rectified linear unit. Typically, a convolutional layer is followed by batch normalization [40].

Although the convolutional network has strong ability to learn features, it cannot make full use of the global information of each channel due to the equal importance of all channels. Therefore, Hu Jie et al. proposed the squeeze-and-excitation block (SE block) [41]. By introducing a weight for each channel to indicate its importance, the network can focus on the feature map of important channels. As shown in Figure 1a, the SE block is composed of squeeze and excitation. The squeeze operation refers to generating channel-wise statistics by using global average pooling. Given a computational unit \( U = [u_1, u_2, \ldots, u_C] \) with \( C \) channels, where \( U \in \mathbb{R}^{M \times N \times C} \). Formally, a statistic \( z \in \mathbb{R}^C \) generated by shrinking \( U \) through spatial dimensions \( M \times N \), where the \( c \)-th element of \( z \) is calculated by:

\[
z_c = F_{sq}(u_c) = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} u_c(i, j)
\]
where $F_{sq}(u_c)$ is the channel-wise global average over the spatial dimensions $M \times N$.

![Diagram](image_url)

**Figure 1.** (a) The architecture of 2D squeeze-and-excite block. $u_1, u_2, \ldots, u_C$ represents the $C$ channels of a computing unit $U$, and the feature map size of each channel is $M \times N$; (b) the computation of 1D squeeze-and-excite block used in this system.

Excite operation employs a simple gating mechanism with a sigmoid activation to capture the channel-wise dependencies, as follows:

$$s = F_{ex}(z, W) = \sigma(g(z, W)) = \sigma(W_2\delta(W_1z))$$  (3)

where $\sigma$ is the Sigmoid activation function, $\delta$ is the ReLU activation function, $W_1 \in \mathbb{R}^{C \times C}$, $W_2 \in \mathbb{R}^{C \times F}$ and $F_{ex}$ is parameterized as a neural network with two fully connected layers (a dimensionality-reduction layer with parameters $W_1$ with reduction ratio and a dimensionality-increasing layer with parameters $W_2$). $W_1$ and $W_2$ are used to limit model complexity and aid with generalization.

Finally, the output of the block is rescaled as follows:

$$\overline{x}_c = F_{scale}(u_c, s_c) = s_c \cdot u_c$$  (4)

where $\overline{x} = [\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_C]$ and $F_{scale}(u_c, s_c)$ denotes the channel-wise multiplication between the feature map $u_c \in \mathbb{R}^{M \times N}$ and the scale $s_c$.

### 3.2. Attention-Enhanced LSTM

LSTM is one of the most commonly used artificial neural network models in time series analysis [42]. As shown in Figure 2, a basic LSTM cell contains a hidden vector $h_t$ a memory state cell $c_t$ and three gate functions (input $i_t$, forget $f_t$ and output $o_t$). $g_t$ is the new cell state. The input gate controls which state of $g_t$ is used to update memory cell state. The forget gate controls what to be forgotten and what to be remembered by the memory cell, and the output gate lets the state of the memory cell impact the output at the current time step. Each vector in the LSTM cell can be computed as follows:

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f)$$  (5)

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i)$$  (6)

1. LSTM is one of the most commonly used artificial neural network models in time series analysis [42].
where $W_f$, $W_i$, $W_g$, $W_t$ are the recurrent weight matrices and $b_f$, $b_i$, $b_g$, $b_t$ are bias vectors. $\sigma$ is the Sigmoid activation function, $\odot$ is an element-wise multiplication, and $x_t$ is the input of LSTM cell unit, while $h_t$ is the hidden state vector of the $t$-th time.

$$g_t = \tanh(W_g[h_{t-1}, x_t] + b_g)$$
$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o)$$
$$c_t = f_t \odot c_{t-1} + i_t \odot g_t$$
$$h_t = o_t \odot \tanh(c_t)$$

The detection of FOG is a challenging task. The research of Guan et al. showed that the duration of FOG in reality is random [36]; however, the original acceleration signal is generally expressed as sequence data with equal time span during data preprocessing. This fixed window length with uniform distribution of sample weights will not naturally lead to ideal modeling [35], since not all observations at all time steps contribute equally to the model. Therefore, we introduced an attention mechanism to evaluate each time step observation with an importance score, and construct a hidden representation by integrating these scores to obtain better classification performance [43,44].

Generally, temporal context information is used to construct and learn such importance scores [43,44]. As an example, the attention mechanism proposed in [43] is shown in Figure 3. Given an input sequence $X = \{x_1, \ldots, x_T\}$ of length $T$, where $x_t$ represents the $t$-th element of the sequence, $0 < t \leq T$, the dimension of $x_t$ is $d$, and the attention weight is a scalar value $w_t$, which represents the importance of the $t$-th element in the entire sequence. The attention weight of the sequence can be expressed as $W = [w_1, \ldots, w_T]$. Usually, a set of linear layers or non-linear layers are used to calculate the attention weight of the sequence, which can map these $d$-dimensional vectors to one-dimensional score. These scores are then passed through the Softmax function to give the set of $T$ weights. The way in which each of these $T$ vectors is mapped to a one-dimensional fraction is architecture-specific.

![Figure 2](image_url)  
**Figure 2.** The structure of a basic LSTM cell, where $i_t$, $f_t$, $o_t$, $c$ respectively represent the input gate, forget gate, output gate and memory state cell. The update of each gate and state can be found in Equations (2)–(7).

![Figure 3](image_url)  
**Figure 3.** Illustration of attention mechanism. The input is a sequence $X$ with length $T$ and dimension $d$, which is mapped to a one-dimensional score by the linear layer, and these scores are then passed through the Softmax function to give the set of $T$ weights.
3.3. The Proposed Network for FOG Detection

In order to detect the FOG of PD patients, we present a novel DL framework with multi-channel input signals. The framework consists of three parts: data preprocessing, deep convolutional network, and attention-enhanced LSTM. Data preprocessing operations include filtering, segmentation and data augmentation. The original acceleration signals are preprocessed to obtain a multi-channel sequence dataset with equal time span. The deep convolutional network acts as a feature extractor, providing abstract representations of multi-channel sequence data in feature maps. The attention-enhanced LSTM models the temporal dynamics of the feature maps output by the deep convolutional network. Among them, the deep convolutional network is composed of temporal convolutional layers and an SE blocks. The SE block is located between adjacent convolutional layers and is used to fuse the global information of each channel during feature extraction. The proposed model consists of a deep convolutional network containing SE blocks and an attention-enhanced LSTM (ALSTM), which we refer to as SEC-ALSTM.

The structure of the proposed SEC-ALSTM is shown in Figure 4. The original signal is preprocessed to obtain a sequence dataset of length $T$ and number of channels $d$, and then fed to the deep convolutional network. The deep convolutional network has three time convolutional layers. In each convolutional layer, the convolution operation is followed by batch normalization (momentum = 0.99, epsilon = 0.001), and the batch normalization layer is succeeded by the non-linear activation function (e.g., ReLU). In addition, the first two convolutional layers conclude with a SE block. Note that the convolution kernel used in the convolution operation is one-dimensional instead of two-dimensional because the original signal is one-dimensional. The number of convolution kernels changes with convolution layers. Similar to the study by Karim et al. [45], the function of the SE block is only to fuse information from different channels without changing the size of the feature map and the number of channels. As shown in Figure 1b, Karim et al. extended the SE block to the case of one-dimensional signal models, which is different from the original 2D SE block during the squeeze operation. Assuming that the input is time series data $U$ with a time span of $T$ and a channel number of $C$, the input $U$ is shrunk through the temporal dimension $T$ to compute the channel-wise statistics, $z \in \mathbb{R}^C$. The $c$-th element of $z$ is then calculated by computing $F_{sq}(u_c)$, which is defined as:

$$z_c = F_{sq}(u_c) = \frac{1}{T} \sum_{t=1}^{T} u_c(t)$$

Subsequently, the outputs of the deep convolutional network are fed into the ALSTM. The ALSTM includes an LSTM network with $n$ hidden units and an attention layer. Among them, the hidden state vectors output by the LSTM cells are the input of the attention layer. Since not all observations in the time context have the same effect on classification, we use the attention mechanism to automatically determine the temporal context that is relevant for modeling activities and output the attention-weighted state vectors. Specifically, it is assumed that the LSTM has $T\tau$ hidden state vector sequences of length $n$. We apply a linear transformation to linearly transforming these hidden state vectors into new state vectors of equal size. The new state vector is multiplied by the last hidden state vector to obtain $T\tau$ values—that is, the scores of each hidden state vectors. These scores are then passed through the Softmax function to give the final weight set. These weights are used to calculate the weighted sum of all the $T\tau$ hidden states to give the attention-weighted state vector. The attention-weighted state vector is added to the last hidden state vector to give the final one-dimensional output vector. Finally, using the final one-dimensional vector for classification.
4. Experiments and Results

4.1. Experimental Settings

This section introduces the methods and metrics used to evaluate the performance of the proposed SEC-ALSTM. In order to compare with previous studies, we have done some experiments using two evaluation methods, 10-fold cross-validation (R10Fold) [37,46] and leave-one-subject-out (LOSO) cross-validation [34,38,47]. Among them, the specific implementation steps of R10Fold evaluation are as follows: first, we shuffle all samples and divide them into 10 folds as evenly as possible. Then, select one of them for testing, and the remaining nine folds for training. The process was repeated 10 times so that all samples have been tested. Finally, the average result of 10 evaluations is used as the final result of model evaluation. The specifics of LOSO cross-validation are: the samples used for testing and training come from different patients. Only samples from one of the patients are used for testing, and samples from the remaining patients are used for training. Repeat this process until all samples have been tested. Then, the average of each evaluation result is used as the final evaluation result of the model’s performance. In this study, according to the different sample sources, the R10Fold evaluation was divided into two types:

1. The samples used for training and testing were from the same patient, and the average of the evaluation results of all patients is taken as the final evaluation result;
2. The samples used for training and testing were from all patients.

In addition, we also evaluated the gains of data augmentation, SE block, and attention mechanisms to model performance. At the end of the experiment, we studied the influence of the sensor position and the sampling frequency on the performance of the FOG detection of PD patients.

Because the FOG dataset is imbalanced, it is unreasonable to use only common indicators such as accuracy to evaluate the detection performance of the model. In this study, we used sensitivity (true positive rate, the ratio of positives that are correctly identified), specificity (true negative rate, the ratio of negatives that are correctly identified), accuracy, F1-score, area under the curve (AUC) and equal error rate (EER) to evaluate the model.
As shown in Table 1, our hardware platform was configured with Intel(R) Core(TM) i5-9400 CPU@2.90 GHz, Nvidia GeForce RTX 2060 6 GB GPU, and 8 GB RAM.

| Central Processing Unit | Graphics Processing Unit | Graphics Memory | Computer Memory |
|-------------------------|---------------------------|-----------------|-----------------|
| Intel(R) Core(TM) i5-9400 | Nvidia GeForce RTX 2060   | 6 GB            | 8 GB            |

4.2. Description of Dataset

The Daphnet dataset used in this study was created by Bachlin et al. [25]. This public dataset contains the body acceleration signals of 10 PD patients (three females) during the walking task. The age of PD patients in this dataset ranged from 59 to 75 years (66.4 ± 4.8 years), and Hoehn and Yahr (H & Y) score in ON period ranged from 2 to 4 (2.6 ± 0.65). The acceleration sensors are placed on the patient’s back (above the patient’s hip joint), left thigh (just above the knee) and left calf (just above the ankle). Three directions of the human body—horizontally forward (perpendicular to the frontal plane), vertical (perpendicular to the transverse plane) and horizontally lateral (perpendicular to the sagittal plane)—correspond to the x, y, and z axes of the sensor, respectively. The sensor is sampled at a 64 Hz frequency and transmitted via Bluetooth to a wearable computer. The main attributes of the Daphnet dataset are listed in Table 2.

| Number of Patients | Age (Years) | H & Y in ON | Sampling Rate (Hz) | Test Duration (Mins) | Number of Freeze Events |
|--------------------|-------------|-------------|--------------------|----------------------|------------------------|
| 10 (3 females)     | 66.4 ± 4.8  | 2.6 ± 0.65  | 64                 | 500                  | 237                    |

According to the protocol, participants were asked to perform three walking tasks (10–15 min each) at a normal pace to simulate different circumstances of their daily walking. The walking tasks included:

- Walk straight back and forth along the corridor
- Walk or stop freely in the lobby
- Simulate daily activities

The physiotherapist marked the beginning and duration of FOG through playback of the experiment’s video. In total, 8 h 20 min of acceleration signals were collected, including 237 FOG events.

4.3. Data Preprocessing

In this study, we propose a comprehensive data preprocessing method based on the features of the original dataset, which includes data filtering, data segmentation and data augmentation.

4.3.1. Data Filtering

Since the original acceleration signal used in this study contains very large or very small abnormal values, we adopted the “3-sigma-rule” to detect and filter these outliers. Since the empirical distribution of acceleration signals is a unimodal function, which basically conforms to the Gaussian distribution, it is reasonable to use the “3-sigma-rule”, which means that the probability that the acceleration signal falls within the “3-sigma” interval is 95%. Similar to the study in [48], we use the median value of the entire time series instead of the average value to characterize outliers, because the signal may be affected by some particularly large outliers causing the overall average value to be too large. However, the median is basically not affected by outliers. After testing, we found that the effect of using 4-SDs is better than 3-SDs, because 3-SDs may delete some larger normal points by mistake.
Figure 5 shows the result of outlier processing. Among them, the original acceleration signal in the vertical direction at ankle position during walking is shown in Figure 5a, where the detected outliers are marked with circles. The signal after replacing the outliers with the median value of the entire signal is shown in Figure 5b. The power spectra of the signals in Figure 5a,b are shown in Figure 5e,f, respectively. It can be seen that the signal in Figure 5a has high energy in both the low-frequency band and the high-frequency band due to the existence of outliers. After removing the outliers, the energy of the signal in Figure 5b is concentrated in the “locomotor” band [0.5–3 Hz].

![Figure 5](image-url)

Figure 5. (b–d) are the signals in (a–c) after outlier processing, and (e–h) are the power spectra of the signal in (a–d).

The results of the outlier processing of the acceleration signal while standing are shown in Figure 5c,d,g,h, which are the original signal, the signal after outlier processing, the power spectra of original signal, and the power spectra of the signal after outlier processing. Due to the existence of the outstanding outliers, the total energy of the signal in Figure 5c is substantially higher than that of the signal in Figure 5d. Moreover, the energy of the signal in Figure 5c is approximately equally distributed over the whole frequency spectrum (white noise).

4.3.2. Data Segmentation

A sliding window of length T is used to divide the entire time series data into many overlapping data segments. The optimal length T of the sliding window is determined by the minimum duration of the FOG and the sampling frequency. When T is too large, the window can contain sufficient FOG features, but the model cannot identify short-duration FOG, and the detection time lag of the model deepens; when T is too small, the window cannot contain enough FOG features, and the accuracy is reduced accordingly. As for the step size L of the sliding window, it is obvious that when L is small, more data instances can be generated in this limited dataset, but it will cause more redundant information between two adjacent pieces of data. The next step is to add labels to the data segments. Some studies [38,49] pointed out that the sample label that appears most frequently in the window should be used as the label of the data segment, while other studies [34] think that it is more reasonable to use the label of the last frame sample as the label of the entire data segment. In this study, a method similar to the previous one is adopted (Figure 6)—that is, the label with the most occurrences in the
window is selected as the window label. In addition, since the physiotherapists do not distinguish between normal walking and standing when labeling data, in this study, an energy-thresholding approach was used to remove the standing part of the original dataset, similar to the method in [25].

![Figure 6. Data segmentation and labeling. The original signals are segmented by a sliding window of length T. The step size of the window is L. The activity label within each sequence is considered to be the ground truth label with the most occurrences of that window.](image)

### 4.3.3. Data Augmentation

The dataset used in this study is imbalanced. Due to the randomness of FOG in Parkinson’s disease patients, the number of samples of normal walking is much larger than that of FOG, which may cause the model to pay more attention to the normal walking during training. Data augmentation can be seen as a technique of using prior knowledge to expand the original data without changing the original data label. Augmented data can cover unexplored input space, prevent overfitting, balance the dataset, and improve the robustness of the model [50]. Data augmentation technology has been widely used in image recognition. Since minor changes such as jittering, scaling, cropping, warping and rotating may occur during actual observation, these processing methods will not change the category label of the original image, and are often applied to image data augmentation [51]. Similarly, when using an acceleration sensor to collect the gait signal of PD patients, turning the sensor upside down or tilting it at a certain angle will not change the gait category. Therefore, in this study, a method similar to that in [51] was used to augment the FOG data through arbitrary rotation, so that we can get a balanced training set for the model. Figure 7 compares the original three-axis acceleration signals of standing, walking and FOG with their augmented signals. Among them, the original signals are shown in Figure 7a–c, and the augmented signals are shown in Figure 7d–f. It can be seen that the waveform of the signal before and after the enhancement are similar.

### 4.4. Performance Evaluation and Comparison with Previous Work

#### 4.4.1. R10Fold Evaluation

In order to evaluate the performance of the proposed model, we conducted two patient-dependent experiments. The evaluation scheme used in the experiment was 10-fold cross-validation. In the first experiment, the samples used for training and testing were from the same patient. In order to obtain the best detection performance, we use a sliding window of 256 frames (4 s) to segment the acceleration signal. When the step size is 4, a total of 147,306 data instances are obtained, of which 12,135 data instances are No-FOG and 25,954 data instances are FOG. The specific distribution is listed in Table 3. By using the important hyperparameter settings listed in Table 4, the test results are recorded in Table 5. It can be seen that the overall detection accuracy and the detection accuracy of each patient are very high. In addition, several indicators suitable for imbalanced data, such as sensitivity, specificity and F1 score, also obtained excellent results. Among them, the best result (accuracy of 0.999, sensitivity of 0.999, specificity of 0.999, F1-score of 0.997) was obtained with patient #6.
4.4. Performance Evaluation and Comparison with Previous Work

4.4.1. R10Fold Evaluation

In order to evaluate the performance of the proposed model, we conducted two patient-dependent experiments. The evaluation scheme used in the experiment was 10-fold cross-validation. In the first experiment, the samples used for training and testing were from the same patient. In order to obtain the best detection performance, we used a sliding window of 256 frames (4 s) to segment the acceleration signal. When the step size is 4, a total of 147,306 data instances are obtained, of which 12,135 data instances are No-FOG and 25,954 data instances are FOG. The specific distribution is listed in Table 3. By using the important hyperparameter settings listed in Table 4, the test results are recorded in Table 5. It can be seen that the overall detection accuracy and the detection accuracy of each patient are very high. In addition, several indicators suitable for imbalanced data, such as sensitivity, specificity and F1 score, also obtained excellent results. Among them, the best result (accuracy of 0.999, sensitivity of 0.999, specificity of 0.999, F1-score of 0.997) was obtained with patient #6.

![Graphs showing acceleration signals](image)

**Figure 7.** (a–c) are the original three-axis acceleration signals of standing, walking and freezing of gait (FOG), (d–f) are the augmented signals.

| Class | Patient 1 | Patient 2 | Patient 3 | Patient 4 | Patient 5 | Patient 6 | Patient 7 | Patient 8 | Patient 9 | Patient 10 | Overall |
|-------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|---------|
| Walking | 16,110    | 11,310    | 8557      | 12,032    | 11,794    | 13,548    | 16,770    | 4915      | 9646      | 16,670    | 121,352 |
| FOG    | 1434      | 2856      | 4241      | 0         | 7514      | 1907      | 1053      | 2896      | 4053      | 0         | 25,954  |

**Table 3.** The class distribution of data instances for each patient. The data instance are obtained by segmenting the acceleration signals using a sliding window, where the length of the sliding window is 256 and the step size is 4.

**Table 4.** The hyper-parameters for the model proposed in this study.

| Learning Parameter | Value or Method |
|--------------------|-----------------|
| Batch size         | 64              |
| Regularization     | Batch Normalization |
| Learning rate      | 0.001           |
| Reduction ratio(r) | 8               |
| Number of Conv.Filters | 64            |
| Kernel size (k1)   | 7               |
| Kernel size (k2)   | 5               |
| Kernel size (k3)   | 3               |
| Hidden Units (n)   | 128             |
| Input Dimension (d)| 9               |
Table 5. The results of a 10-fold cross-validation (R10Fold) evaluation. The samples used for training and testing were from the same patient. TP, TN, FP, and FN are abbreviations of true positive, true negative, false positive and false negative, respectively. Subject #4 and #10 did not have FOG during this study.

| Patient | TP       | TN       | FP       | FN       | Sensitivity | Specificity | Accuracy | F1-Score |
|---------|----------|----------|----------|----------|-------------|-------------|----------|----------|
| 1       | 1424     | 16,097   | 13       | 10       | 0.993       | 0.999       | 0.999    | 0.992    |
| 2       | 2834     | 11,281   | 22       | 29       | 0.992       | 0.997       | 0.996    | 0.991    |
| 3       | 4198     | 8516     | 43       | 41       | 0.990       | 0.995       | 0.994    | 0.990    |
| 4       | 12,032   | 0        | 0        | 0        | -           | 1.000       | 1.000    | -        |
| 5       | 7406     | 11,728   | 66       | 108      | 0.986       | 0.994       | 0.991    | 0.988    |
| 6       | 1905     | 13,540   | 8        | 2        | 0.999       | 0.999       | 0.999    | 0.997    |
| 7       | 1032     | 16,757   | 13       | 21       | 0.980       | 0.999       | 0.998    | 0.984    |
| 8       | 2883     | 4950     | 10       | 13       | 0.996       | 0.998       | 0.997    | 0.996    |
| 9       | 4033     | 9610     | 36       | 20       | 0.995       | 0.996       | 0.996    | 0.993    |
| 10      | 16,670   | 0        | 0        | 0        | -           | 1.000       | 1.000    | -        |
| Overall | 25,710   | 121,141  | 211      | 244      | 0.991       | 0.998       | 0.997    | 0.991    |

In the second patient-dependent experiment, the samples used for training and testing were from all patients. In order to compare with the baseline, the sliding window length used in segmenting the acceleration signal is 256, and the step size is 64. Figure 8 shows specificity vs. sensitivity curves and AUC and EER values of the experimental results. Each curve represents the result of a cross-validation in this R10Fold evaluation. The overall results were sensitivity, specificity, accuracy, and F1-score of 0.951, 0.988, 0.981, and 0.947, respectively. This result is not as good as the first patient-dependent experiment. This phenomenon is reasonable. On the one hand, the gait features of each patient are different, which increases the difficulty of detection. On the other hand, because the step size of the sliding window becomes larger, the number of samples is reduced, which leads to insufficient generalization ability of the model.

4.4.2. LOSO Evaluation

Figure 9 shows the results of each patient when evaluated using LOSO, excluding patients #4 and #10. Patients #3 and #5 provided the worst results. As can be seen in Table 3, the FOG events of these two patients accounted for a relatively high proportion, indicating that the patient’s condition may be heavier than other patients, and therefore contains more special gait features. During the LOSO evaluation, different from R10Fold evaluation, this unique gait features did not appear sufficiently in the training set, which leads to poor results. Except for patients #3 and #5, the AUCs of the remaining
patients were all greater than 0.90, and the EERs were all less than 15%. This shows that the proposed model can also obtain good evaluation results in LOSO evaluation.

4.4.3. Comparison with Previous Work

Bachlin et al. was the first to use the Daphnet dataset to study the FOG detection model of PD patients [25]. They used two features extracted from acceleration signals: the freezing index and the energy in the frequency band of 0.5–8 Hz. A sensitivity of 0.731 and a specificity of 0.816 were achieved in the patient-independent experiment, and a sensitivity of 0.781 and a specificity of 0.869 were achieved in the patient-dependent experiment using the global threshold. The best results based on the Daphnet dataset were reported by Mazilu et al. [46]. Based on the research of Bachlin et al., they further extracted five features of signal mean, standard deviation, variance, frequency entropy, and energy. However, San-Segundo et al. [38] proved that the best results could be reproduced only with a significant overlap between the sliding windows after reproducing the study of Mazilu et al. Furthermore, in an experiment with a 4 s window and 75% overlap rate, the R10Fold evaluation achieved a sensitivity of 0.934 and a specificity of 0.939, and the LOSO evaluation achieved an AUC of 0.900 and an EER of 17.3%. San-Segundo et al. studied whether or not the inclusion of features from adjacent time windows would improve classification performance, and achieved the best detection performance when using three previous and three subsequent 4 s windows with a 75% overlap (AUC = 0.931, EER = 12.5%).

Tables 6 and 7 compare the results of the proposed model and several previous models. It can be seen that our proposed model is significantly better than other models in both R10Fold and LOSO evaluations. In addition, we also compared the performance of the proposed model with other neural networks, and the results are listed in Table 8. It can be seen that the performance of the proposed model is significantly better than the deep convolutional neural network and LSTM network.

Table 6. The performance comparison of different approaches on the same dataset using a R10Fold evaluation.

| Reference                          | Description                                      | Sensitivity | Specificity |
|------------------------------------|--------------------------------------------------|-------------|-------------|
| Bachlin et al. [25]                | FI index, energy in the frequency band 0.5–8 Hz  | 0.781       | 0.869       |
| San-Segundo et al. [38] (reproducing Mazilu et al.’s system [46]) | Signal mean, standard deviation, variance, frequency entropy, energy, FI index, etc. | 0.934       | 0.939       |
| The proposed                       | DL-based feature learning                        | 0.951       | 0.988       |
4.5. Evaluation of Several Modules of SEC-ALSTM

We evaluated the gains of data augmentation, SE block, and attention mechanisms to model performance. For this purpose, we built four different models: the complete SEC-ALSTM model, the SEC-ALSTM model without data augmentation modules (SEC-ALSTM without DA), the SEC-ALSTM model without SE blocks (CALSTM), and the SEC-ALSTM model without an attention mechanism (SEC-LSTM). For comparison, all models differ from the SEC-ALSTM model only in their missing parts. The experimental settings are: the length of the sliding window is 256, the step size is 64, and using a LOSO evaluation. Without loss of generality, here only the experimental results for patient #8 are reported. The results of this comparison experiment are listed in Table 9. Figure 10 shows the curve for additional comparison of sensitivity and specificity. According to Table 9, the complete SEC-ALSTM model obtains the best performance. This means that each component has a certain gain to the performance of the model. The data augmentation module has the best effect on improving the sensitivity of the model, and the sensitivity of the model without data augmentation modules is reduced by 0.152; the SE block has the best specific effect on the enhancement model, and the specificity of the model without the SE block is reduced by 0.073.

4.6. Evaluation of Sensor Position and Sampling Frequency

In order to analyze which part of the body’s acceleration signal can provide more useful information for FOG detection, the detection performance was investigated when using the signal from a single sensor (ankle, knee, and back accelerometers). The results of this comparative experiment are listed in Table 10, and Figure 11 shows the average curve for additional comparison of sensitivity and specificity. In general, the detection performances of single accelerometers are all slightly lower than that of using three accelerometers. The results of the ankle accelerometer were very similar to the case of using the three accelerometers, which indicates that the ankle motion signals are most suitable for detecting FOG events. The worst results were obtained when using only the back accelerometer, which may be due to the relatively weak back motion signal.
The frequency of the original acceleration signal is resampled to 16, 32, and 48 Hz. It can be seen from the results in Table 10 that as the sampling frequency decreases, the detection performance gradually declines, and this situation becomes more obvious when the sampling frequency is lower than 32 Hz. Compared with the sampling frequency of 64 Hz, when the sampling frequency is 8 Hz, the sensitivity is reduced by 0.111 and the specificity is reduced by 0.04.

In addition, we also compared the impact of different sampling frequencies of sensors on detection performance. The frequency of the original acceleration signal is resampled to 16, 32, and 48 Hz. It can be seen from the results in Table 10 that as the sampling frequency decreases, the detection performance gradually declines, and this situation becomes more obvious when the sampling frequency is lower than 32 Hz. Compared with the sampling frequency of 64 Hz, when the sampling frequency is 8 Hz, the sensitivity is reduced by 0.111 and the specificity is reduced by 0.04.
5. Discussion

We studied the detection of FOG in Parkinson’s patients based on deep learning technology. The FOG data used in the experiment come from the Daphnet dataset. The experimental results show that our proposed model achieves very good performance in R10Fold evaluation, with sensitivity and specificity higher than previous studies by 0.017 and 0.045 respectively. In addition, in the LOSO evaluation, the performance of the model is also very good: Compared with previous studies, the EER decreased from 12.5% to 10.6%, and the AUC increased from 0.931 to 0.945. Therefore, the proposed model can be used for FOG detection.

In the experiment, the result of R10Fold evaluation was better than that of LOSO evaluation. The reason for this may be that on the one hand, each patient has some unique gait features. When using R10Fold for evaluation, the samples used for training and testing are from all patients, and the model can learn the gait features of all patients. In the LOSO evaluation, the samples used for training and testing come from different patients, and the model cannot learn the unique gait features of the patients used for testing, so the detection performance declines. On the other hand, when using a sliding window to segment the acceleration signal, adjacent windows will partially overlap. When using R10Fold for evaluation, the sample will be shuffled and then divided into a test set and a training set, which will mean that the training set and test set may contain some of the same data. We compared the results of R10Fold evaluation when the sliding step length was 4, 16, 32, and 64, and found that as the sliding step size increases, the detection performance gradually declines, which confirms our second conjecture.

In order to improve the detection performance, the proposed model adds SE blocks to the convolutional network module and introduces an attention mechanism to the LSTM module. In addition, in the data preprocessing, the up-sampling method was used to balance the samples. The results of comparative experiments show that these methods can indeed improve the detection performance of the model. Although the proposed model is a more complicated approach, the entire model only contains 155,138 trainable parameters and 384 non-trainable parameters because the convolutional neural network used in the model is one-dimensional, which makes the model run at high efficiency. After a rough test, a data instance with a length of 256 only need 0.52 ms to complete the category detection.

We also compared the effects of different sensor positions and sampling frequencies on the detection performance. The results show that using only the ankle accelerometer can obtain detection performance similar to that of using three accelerometers at the same time, while the detection performance of using only the knee or back accelerometer is poor. This means that ankle motion signals can better distinguish between FOG and non-FOG. In addition, when the sampling frequency is between 32 and 64 Hz, as the sampling frequency decreases, the detection performance declines slightly. Once the sampling frequency is lower than 32 Hz, the detection performance declines significantly. The research of Bächlin et al. [25] proved that the frequency of human motion is mainly distributed within 30 Hz. Therefore, when the sampling frequency is lower than 32 Hz, some motion signal features will be lost, resulting in a significant decline in detection performance.

For future research, we still have a lot of work to do. First of all, the amount of data in the current FOG dataset is too small, and it is necessary to collect as many motion signals as possible from patients with FOG symptoms in different environments, so that the model can fully learn the features used to distinguish frozen and non-frozen events during training. This allows the trained model to be used not only in a laboratory setting, but also in a real patient’s daily life. In addition, it is recommended that in the subsequent data collection, only the acceleration signals with a frequency of 32–48 Hz at the left and right ankles should be collected because the higher sampling frequency will lead to greater power consumption and the use of multiple accelerometers will lead to cumbersome wearing. Secondly, the research of Yungher et al. [53] showed that the peak swing phase velocity of several strides prior to FOG successively decreased. Therefore, it is possible to develop an early warning system for FOG that can give an alarm before a freezing event occurs. Finally, many previous studies
have shown that proper visual or auditory stimuli can help patients overcome freeze and restart the gait. Therefore, the detection and early warning model can be integrated into the intervention system, so that the system can give certain visual and auditory stimulation to patients before or during freezing events, so as to achieve the purpose of reducing the risk of a freezing event, shortening the duration of FOG, and improving the quality of life of patients.

6. Conclusions

We have proposed a novel FOG detection system for Parkinson’s disease patients. The proposed system uses a wearable accelerometer to obtain the patient’s motion signal and use it as an input to detect FOG. A sliding window is used to segment the signal, and the 3-sigma method is used to remove outliers. In addition, in order to eliminate the impact of the imbalanced training set on the detection performance, arbitrary rotation is applied to the frozen gait data instance to expand the dataset. The deep convolutional network acts as a feature extractor, providing abstract representations of multi-channel sequence data in feature maps. The attention-enhanced LSTM models the temporal dynamics of the feature maps output by the deep convolutional network. In addition, in order to make full use of the global information of each channel of the network and temporal context information, we apply SE blocks and an attention mechanism in the system. Experimental results show that the proposed system can automatically learn the discriminative features used to distinguish between FOG and non-FOG, and a sensitivity of 0.951 and a specificity of 0.988 were achieved in the R10Fold evaluation, which are 0.017 and 0.045 higher than the previous results, respectively. In the LOSO evaluation, compared with previous studies, the EER decreased from 12.5% to 10.6%, and the AUC increased from 0.931 to 0.945. In terms of model running speed, a data instance with a length of 256 only needs 0.52 ms to complete the category detection. These results indicate that the proposed system has high operating efficiency and excellent detection performance, and can be used to detect FOG of Parkinson’s disease patients.

Author Contributions: All authors contributed to writing, reviewing, and editing the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This research is supported by National Natural Science Foundation of China (NSFC) for Youth (grant number 61701483); the National Key Research and Development Program of China (grant number 2016YFB1001300).

Acknowledgments: The authors would like to thank Bächlin et al. for the Daphnet dataset.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

We summarize our notation in Table A1.

| Symbol | Meaning | First Introduced |
|--------|---------|-----------------|
| ⟨⟩     | Inner product | 3.1 |
| f(·)   | Rectified Linear Unit | 3.1 |
| σ      | Sigmoid activation function | 3.1 |
| δ      | ReLU activation function | 3.1 |
| ∗      | Channel-wise multiplication between the feature map and the scale | 3.1 |
| ⊙      | Elementwise multiplication | 3.2 |

References
1. Harris, J.R. Protein Aggregation and Fibrillogenesis in Cerebral and Systemic Amyloid Disease; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2012; Volume 65.
2. Nutt, J.G.; Bloem, B.R.; Giladi, N.; Hallett, M.; Horak, F.B.; Nieuwboer, A. Freezing of gait: Moving forward on a mysterious clinical phenomenon. *Lancet Neurol.* **2011**, *10*, 734–744. [CrossRef]
3. Backer, J.H. The symptom experience of patients with Parkinson’s disease. J. Neurosci. Nurs. 2006, 38, 51. [CrossRef]

4. Giladi, N.; Nieuwboer, A. Understanding and treating freezing of gait in parkinsonism, proposed working definition, and setting the stage. Mov. Disord. 2008, 23, S423–S425. [CrossRef] [PubMed]

5. Schaafsma, J.D.; Balash, Y.; Gurevich, T.; Bartels, A.L.; Hausdorff, J.M.; Giladi, N. Characterization of freezing of gait subtypes and the response of each to levodopa in Parkinson’s disease. Eur. J. Neurol. 2003, 10, 391–398. [CrossRef] [PubMed]

6. Giladi, N.; Treves, T.; Simon, E.; Shabtai, H.; Orlov, Y.; Kandinv, B.; Palecuc, D.; Korczyn, A. Freezing of gait in patients with advanced Parkinson’s disease. J. Neural Transm. 2001, 108, 53–61. [CrossRef] [PubMed]

7. Ehgoetz Martens, K.A.; Ellard, C.G.; Almeida, Q.J. Does anxiety cause freezing of gait in Parkinson’s disease? PLoS ONE 2014, 9, e106561. [CrossRef] [PubMed]

8. Fahn, S. Recent developments in Parkinson’s disease. Macmillan Health Care Inf. 1987, 2, 293–304.

9. Giladi, N.; Tal, J.; Azulay, T.; Rascal, O.; Brooks, D.J.; Melamed, E.; Oertel, W.; Poewe, W.H.; Stocchi, F.; Tolosa, E. Validation of the freezing of gait questionnaire in patients with Parkinson’s disease. Mov. Disord. 2009, 24, 655–661. [CrossRef]

10. Pahwa, R.; Lyons, K.E. Levodopa-related wearing-off in Parkinson’s disease: Identification and management. Curr. Med. Res. Opin. 2009, 25, 841–849. [CrossRef]

11. Malik, O.A. Deep Autoencoder for Identification of Abnormal Gait Patterns Based on Multimodal Biosignals. Int. J. Comput. Digit. Syst. 2020, 10, 2–8.

12. Cole, B.T.; Roy, S.H.; Nawab, S.H. Detecting freezing-of-gait during unscripted and unconstrained activity. In Proceedings of the 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Boston, MA, USA, 30 August–3 September 2011; pp. 5649–5652.

13. Nieuwboer, A.; Dom, R.; De Weerd, W.; Desloovere, K.; Janssens, L.; Stijn, V. Electromyographic profiles of gait prior to onset of freezing episodes in patients with Parkinson’s disease. Brain 2004, 127, 1650–1660. [CrossRef] [PubMed]

14. Handojoseno, A.A.; Shine, J.M.; Nguyen, T.N.; Tran, Y.; Lewis, S.J.; Nguyen, H.T. Using EEG spatial correlation, cross frequency energy, and wavelet coefficients for the prediction of Freezing of Gait in Parkinson’s Disease patients. In Proceedings of the 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Osaka, Japan, 3–7 July 2013; pp. 4263–4266.

15. Alamri, A.; Gumaei, A.; Alrakhami, M.; Hassan, M.; Alhussein, M.; Fortino, G. An Effective Bio-signal-based Driver Behavior Monitoring System Using a Generalized Deep Learning Approach. IEEE Access 2020, 8, 135037–135049. [CrossRef]

16. Castro, F.; Marin-Jiménez, M.; Guiu, N.; Perez de la Blanca, N. Multimodal feature fusion for CNN-based gait recognition: An empirical comparison. Neural Comput. Appl. 2020, 32. [CrossRef]

17. Delval, A.; Snijders, A.H.; Weerdesteyn, V.; Duysens, J.E.; Defebvre, L.; Giladi, N.; Bloem, B.R. Objective detection of subtle freezing of gait episodes in Parkinson’s disease. Mov. Disord. 2010, 25, 1684–1693. [CrossRef]

18. Shah, S.A.; Tahir, A.; Ahmad, J.; Zahid, A.; Pervaiz, H.; Shah, S.Y.; Ashliebta, A.M.A.; Hasanali, A.; Khattak, S.; Abbasi, Q.H. Sensor fusion for identification of freezing of gait episodes using Wi-Fi and radar imaging. IEEE Sens. J. 2020, 20, 14410–14422. [CrossRef]

19. Okuno, R.; Fujimoto, S.; Akazawa, J.; Yokoe, M.; Sakoda, S.; Akazawa, K. Analysis of spatial temporal plantar pressure pattern during gait in Parkinson’s disease. In Proceedings of the 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vancouver, BC, Canada, 20–25 August 2008; pp. 1765–1768.

20. Hausdorff, J.; Schaafsma, J.; Balash, Y.; Bartels, A.; Gurevich, T.; Giladi, N. Impaired regulation of stride variability in Parkinson’s disease subjects with freezing of gait. Exp. Brain Res. 2003, 149, 187–194. [CrossRef]

21. Plotnik, M.; Giladi, N.; Balash, Y.; Peretz, C.; Hausdorff, J.M. Is freezing of gait in Parkinson’s disease related to asymmetric motor function? Ann. Neurol. 2005, 57, 656–663. [CrossRef]

22. Moore, S.T.; MacDougall, H.G.; Ondo, W.G. Ambulatory monitoring of freezing of gait in Parkinson’s disease. J. Neurosci. Methods 2008, 167, 340–348. [CrossRef]

23. Wu, Y.; Lin, Q.; Jia, H.; Hassan, M.; Hu, W. Auto-Key: Using Autoencoder to Speed Up Gait-based Key Generation in Body Area Networks. Proc. ACM Interact. Mobile Wearable Ubiquitous Technol. 2020, 4, 32. [CrossRef]
24. Tripoliti, E.E.; Tzallas, A.T.; Tsipouras, M.G.; Rigas, G.; Bougia, P.; Leontiou, M.; Konitsiotis, S.; Chondrogiorgi, M.; Tsoulis, S.; Fotiadis, D. Automatic detection of freezing of gait events in patients with Parkinson’s disease. *Comput. Methods Prog. Biomed.* 2013, 110, 12–26. [CrossRef]

25. Bachlin, M.; Plotnik, M.; Roggen, D.; Maidan, I.; Hausdorff, J.M.; Giladi, N.; Troster, G. Wearable assistant for Parkinson’s disease patients with the freezing of gait symptom. *IEEE Trans. Inf. Technol. Biomed.* 2010, 14, 436–446. [CrossRef] [PubMed]

26. Prado, A.; Kwei, K.; Vanegas-Arroyave, N.; Agrawal, S.K. Identification of Freezing of Gait in Parkinson’s Patients Using Instrumented Shoes and Artificial Neural Networks. In Proceedings of the 2020 8th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechatronics (BioRob), New York, NY, USA, 29 November–1 December 2020; pp. 68–73.

27. Pepa, L.; Ciabattoni, L.; Verdini, F.; Capecci, M.; Ceravolo, M. Smartphone based fuzzy logic freezing of gait detection in parkinson’s disease. In Proceedings of the 2014 IEEE/ASME 10th International Conference on Mechatronic and Embedded Systems and Applications (MESA), Senigallia, Italy, 10–12 September 2014; pp. 1–6.

28. Razavian, N.; Sontag, D. Temporal convolutional neural networks for diagnosis from lab tests. *arXiv Preprint 2015*, arXiv:1511.07938.

29. Drakopoulos, G.; Mylonas, P. Evaluating graph resilience with tensor stack networks: A Keras implementation. *Neural Comput. Appl.* 2020, 32, 4161–4176. [CrossRef]

30. Guan, Y.; Plötz, T. Ensembles of deep lstm learners for activity recognition using wearables. *Proc. ACM Interact. Mobile Wearable Ubiquitous Technol.* 2017, 1, 1–28. [CrossRef]

31. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv Preprint 2015*, arXiv:1502.03167.
44. Yang, Z.; Yang, D.; Dyer, C.; He, X.; Smola, A.; Hovy, E. Hierarchical attention networks for document classification. In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego, CA, USA, 12–17 June 2016; pp. 1480–1489.

45. Karim, F.; Majumdar, S.; Darabi, H.; Harford, S. Multivariate LSTM-FCNs for time series classification. Neural Netw. 2019, 116, 237–245. [CrossRef]

46. Mazilu, S.; Hardegger, M.; Zhu, Z.; Roggen, D.; Tröster, G.; Hausdorff, J.M. Online detection of freezing of gait with smartphones and machine learning techniques. In Proceedings of the 2012 6th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops, San Diego, CA, USA, 21–24 May 2012; pp. 123–130.

47. Samà, A.; Rodriguez-Martín, D.; Pérez-López, C.; Català, A.; Alcaine, S.; Mestre, B.; Prats, A.; Crespo, M.C.; Bayés, A. Determining the optimal features in freezing of gait detection through a single waist accelerometer in home environments. Pattern Recognit. Lett. 2018, 105, 135–143. [CrossRef]

48. Wu, Y.; Krishnan, S. Computer-aided analysis of gait rhythm fluctuations in amyotrophic lateral sclerosis. Med. Biol. Eng. Comput. 2009, 47, 1165–1171. [CrossRef] [PubMed]

49. Yang, J.B.; Nguyen, M.N.; San, P.P.; Li, X.L.; Krishnaswamy, S. Deep convolutional neural networks on multichannel time series for human activity recognition. In Proceedings of the 24th International Conference on Artificial Intelligence, Buenos Aires, Argentina, 25–31 July 2015; pp. 3995–4001.

50. Goodfellow, I.; Bengio, Y.; Courville, A.; Bengio, Y. Deep Learning; MIT press: Cambridge, MA, USA, 2016; Volume 1.

51. Um, T.T.; Pfister, F.M.; Pichler, D.; Endo, S.; Lang, M.; Hirche, S.; Fietzek, U.; Kulić, D. Data augmentation of wearable sensor data for parkinson’s disease monitoring using convolutional neural networks. In Proceedings of the 19th ACM International Conference on Multimodal Interaction, Glasgow, UK, 13–17 November 2017; pp. 216–220.

52. Ashour, A.S.; El-Attar, A.; Dey, N.; Abd El-Kader, H.; Abd El-Naby, M.M. Long short term memory based patient-dependent model for FOG detection in Parkinson’s disease. Pattern Recognit. Lett. 2020, 131, 23–29. [CrossRef]

53. Yungher, D.A.; Morris, T.R.; Valentina, D.; Shine, J.M.; Naismith, S.L.; Lewis, S.J.G.; Moore, S.T. Temporal Characteristics of High-Frequency Lower-Limb Oscillation during Freezing of Gait in Parkinson’s Disease. Parkinson’s Dis. 2014, 2014, 606427. [CrossRef]

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).