

Wearable Sensor-Based Multi-modal Fusion Network for Automated Gait Dysfunction Assessment in Children with Cerebral Palsy

Lu Tang, Xiangrui Wang, Pengfei Lian, Zhiyuan Lu, Qibin Zheng,* Xilin Yang, Qianyuan Hu, and Hui Zheng

Gait, fundamental to human movement, becomes compromised in cerebral palsy (CP), a childhood-onset central nervous system motor disorder. Precise assessment of patients' gait is crucial for tailored rehabilitation interventions. Currently, clinical scales assessing CP gait dysfunction mostly, while valuable, rely on subjective clinician observations. To enhance objectivity and efficiency in CP diagnosis and rehabilitation, there is a need for more objective assessment procedures. This study introduces a multi-modal and multi-scale feature fusion (MMFF) framework, a new framework for automating gait dysfunction assessment in children with CP. By utilizing surface electromyography and acceleration signals recorded during children's walking, MMFF generates a feature vector enriched with adaptively refined feature maps, cross-mode correlations, and both local and global information. Validation of MMFF's effectiveness is evident through an accomplished classification accuracy of 99.13%. The mean values for precision, recall, and F1-score in Gross Motor Function Classification System (GMFCS)-1, GMFCS-2, and GMFCS-3, reaching 99.00%, 99.00%, and 98.33%, respectively, further reflect the accuracy of functional assessments at each level. This study underscores MMFF's potential as an objective, streamlined tool for clinicians, promising improved gait assessment and well-informed rehabilitation strategies for children with CP.

Addressing gait motor function rehabilitation in CP has always been brought into focus by healthcare professionals. Accurate assessment of patients' gait motor function is the important premise of appropriate rehabilitation treatment for them.^[4] Nowadays, clinical scale and motion sensing technology has been used to assess gait motor dysfunction of CP. Clinical scales, such as Gross Motor Function Measure (GMFM) scale and Gross Motor Function Classification System (GMFCS), have been internationally accepted and widely used to measure the gross motor function of children with CP in clinical practice. The assessment results based on gait part in the scale can track the development of gait motor function and assess the gait rehabilitation effect.^[5] However, GMFM and GMFCS scales are limited to the subjective observation of clinicians who assess patient's gross motor function empirically, which may lead to distinct results of the assessment given by different clinicians. Therefore, a more objective and efficient motor function assessment procedure

1. Introduction

Gait is the most basic of all human movements. Cerebral palsy (CP) is a central nervous system motor disorders syndrome developed in early childhood, which typically results in motor dysfunction, usually accompanied by gait dysfunction.^[1–3]

should be adopted for the clinical diagnosis and rehabilitation treatment of patients with CP.

Gait assessment based on motion sensing technology provides four types of gait data: spatial-temporal parameters, kinematics, kinetics, and surface electromyography (sEMG).^[6] More and more recent studies focus particularly on the acquisition of

L. Tang, X. Wang, Q. Zheng, X. Yang, Q. Hu, H. Zheng
School of Health Science and Engineering
University of Shanghai for Science and Technology
Shanghai 200093, China
E-mail: qbzheng@usst.edu.cn

P. Lian
School of Integrated Circuits
East China Normal University
Shanghai 200241, China

Z. Lu
School of Rehabilitation Science and Engineering
University of Health and Rehabilitation Sciences
Qingdao 266072, China

 The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/aisy.202300845>.

© 2024 The Authors. Advanced Intelligent Systems published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

DOI: 10.1002/aisy.202300845

sEMG signals in children with CP,^[7–10] probably due to the neuromuscular involvement of this disorder, indeed, strongly advised in CP.^[11] To capture the features of gait behavior more comprehensively, multisensory fusion technology has also been applied to gait function assessment in children with CP. Multisensory-based gait dysfunction assessment mainly focuses on calculating predefined coefficients and parameters to analyze gait synergetic neuromuscular control or lower limb muscle coactivation in patients with CP; some studies also aim to generate a gait assessment score reflecting the abnormal degree of gait.^[5,12–15] However, predefined features or coefficients may not fully model the non-stationary and non-linear sEMG signal measured in complex gait patterns, especially in patients with motor disorders.^[16]

To enhance the learning of multisensory information and address variations in abnormal gait patterns, studies focusing on abnormal lower limb motion analysis have employed machine learning (ML) approaches for developing automated rehabilitation devices. In the context of abnormal gait function in patients with CP, recent ML-based approaches primarily concentrate on gait pattern recognition, physical activity identification, and gait phase detection. ML-based studies typically utilize wearable sensors, such as the accelerometer, pressure sensor, and sEMG sensor, as the means to represent the motion function in children with CP.^[6,17–19] However, most ML-based approaches rely on manual feature extraction and further feature selection to identify the optimal feature subset for their tasks. Consequently, ML-based approaches may introduce uncertainty and subjectivity during the model design and training process.

Due to the significant advances in convolutional neural networks (CNNs) automating feature extraction through convolution on input data to derive high-level features, as compared to the traditional manual feature extraction, CNNs automatically update convolution kernel parameters using the gradient descent algorithm to extract semantic information from input data, thereby reducing subjectivity in the model training process,^[20] which have the potential to design an end-to-end gait assessment system. Currently, some research related to the abnormal gait analysis has shifted to CNNs. Chakraborty employed inertial measurement unit (IMU) sensors to detect the abnormal gait of CP and normal gait of typically developing (TD) children using CNN-based model.^[21] Gao et al. classified three abnormal gait patterns and two normal gait patterns using a combined long short-term memory and CNNs model with IMU sensor.^[22] Bajpai et al. further proposed a three-module CNNs model to identify nine CP gait patterns, with the model input being the knee angle collected by a motion capture system.^[23] Some CNN-based analyses of the abnormal gait function in patients with Parkinson's disease have also been carried out to identify patients' gait patterns.^[24–26] However, current CNN-based studies mentioned above focus more on the recognition of motor activity and gait events in patients with CP and have not conducted an objective assessment of gait dysfunction in patients with CP.

Considering the limitations of clinical scales and the motion and neuromuscular characteristic capture abilities of sensor technologies, this work proposed a novel approach for objectively grading abnormal gait function in patients with CP, specifically, an automatic CNN-based framework for grading gait dysfunction

in patients with CP, for the first time. In comparison to previous studies that employed only kinematic or physiological parameters, this study innovatively combined sEMG and acceleration (ACC) signals to construct a multi-modal CNN-based framework. This enabled the model to learn not only neuromuscular information during walking but also intuitive gait motion parameters. In summary, this approach only requires patients to measure their sEMG and ACC signals during walking to automatically identify the patient's gait dysfunction level, which allows clinicians to further develop patients' gait dysfunction rehabilitation schedules.

2. Experimental Section

2.1. Subjects

This experiment recruited 12 children with CP (4.38 ± 2.05 years) and 5 TD children (5.92 ± 1.54 years). **Table 1** shows the details of the 17 subjects. Inclusion criteria for the children with CP were: 1) confirmation of CP diagnosis, 2) absence of motor impairments unrelated to CP, and 3) assurance of the ability to engage in independent or assisted activities. GMFCS levels were assessed by three clinicians, and the evaluation results are averaged across the grades assigned by the three clinicians, enhancing the objectivity of the results. All patients' guardians signed an informed consent form before the experiment. Experiment protocol employed in this study underwent assessment and received approval from the Institutional Review

Table 1. Basic information of the subjects.

Subjects	Gender ^{a)}	Age (years)	Cadence [steps min ⁻¹]	Stride time ^{b)} [s]	Mode ^{c)}
01	F	3.0	29	2.07	GMFCS-1
02	M	3.7	34	1.76	GMFCS-1
03	M	1.5	21	2.86	GMFCS-1
04	M	6.0	41	1.46	GMFCS-1
05	F	2.0	29	2.07	GMFCS-1
06	M	5.5	27	2.22	GMFCS-2
07	M	6.0	29	2.07	GMFCS-2
08	F	4.5	22	2.73	GMFCS-3
09	M	2.9	18	3.33	GMFCS-3
10	M	4.0	21	2.86	GMFCS-3
11	M	4.4	25	2.40	GMFCS-3
12	F	9.0	31	1.94	GMFCS-3
13	F	4.4	49	1.22	TD
14	M	4.5	43	1.40	TD
15	M	6.0	57	1.05	TD
16	F	8.1	62	0.97	TD
17	F	6.6	54	1.11	TD

^{a)}The "F" and "M" in the "Gender" column represent female and male, respectively.

^{b)}Stride time: one gait cycle duration. ^{c)}Mode: GMFCS denotes the level of the Gross Motor Function Classification System and TD denotes the typically developing children.

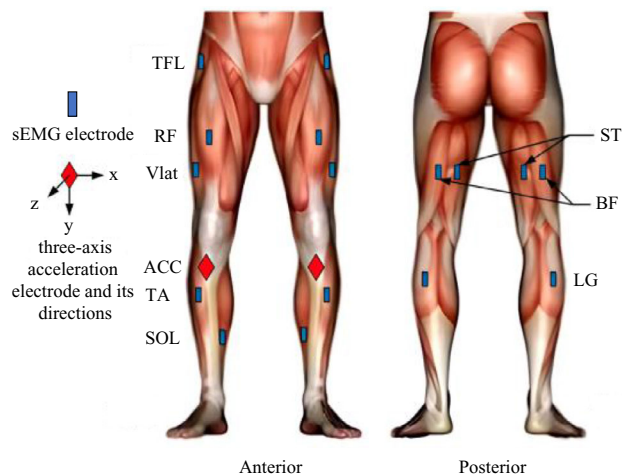


Figure 1. The placements of the sEMG and ACC electrodes. The x-axis of the ACC electrode is oriented horizontally from the subject's left leg to the right leg, the y-axis is oriented vertically downward, and the z-axis is oriented in the direction of the posterior side of the leg to the anterior side.

Board of the University of Health and Rehabilitation Sciences (Protocol Code: 2021-035).

2.2. Experimental Protocols

Subjects took at least 50 consecutive steps along a straight line at a walking speed that felt comfortable to them, and the sEMG and ACC signals were recorded simultaneously. The sEMG signals were recorded from 8 muscles on each side of the lower limb: tibialis anterior (TA), soleus (SO), lateral gastrocnemius (LG), vastus lateralis (VL), rectus femoris (RF), semitendinosus (SE), biceps femoris (BF), and tensor fasciae latae (TFL) for each leg. The placement of sEMG sensors and ACC sensors is shown in **Figure 1**. The sEMG signals were measured by the Ag-AgCl electrodes. The ACC signals were acquired using two identical 3-channel triaxial ACC sensor modules. Both sEMG and ACC signals were recorded by the self-developed sEMG acquisition device with 1000 Hz and 100 Hz sample rate for sEMG and ACC signal, respectively.

3. Methods

3.1. Data Preprocessing

The effective frequency of sEMG is mainly distributed in the range of 10–500 Hz,^[27] the noise frequency of motion artifacts is generally concentrated in the range of 0–20 Hz, and the noise of power frequency interference is concentrated at 50 Hz.^[28] Therefore, a 50th-order Butterworth bandpass filter from 50 to 450 Hz was employed to filter out the above noise.

Human walking movement is cyclical. In this study, a complete gait period is defined as the time elapsed from one heel strikes the ground to the same heel strikes the ground again on the same side. Considering a gait motor abnormality in children with CP in which the heel may not touch the ground,^[29] we defined the initial contact as a start of a gait cycle. The initial

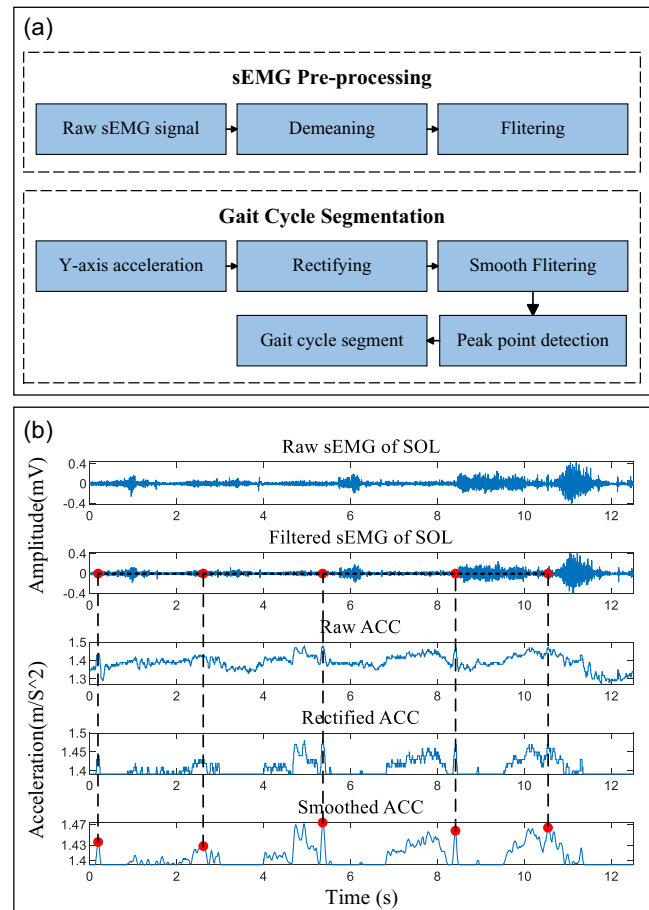


Figure 2. Signal preprocessing: a) presents the signal preprocessing and the gait cycle segmentation procedure and b) describes a partial result of (a). Note that the ACC signal in (b) is collected from y-axis.

contact causes the y-axis ACC to reach its maximum value. As shown in **Figure 2a**, the peak value of the y-axis ACC was detected and its corresponding time was recorded to segment the gait cycle. As shown in **Figure 2b**, the red dots presented in the smoothed ACC signal denote the locations of the peak values, and the interval between two adjacent red dots is a gait cycle. Since the peak point of ACC for some patients in GMFCS-3 level and GMFCS-2 level may be ambiguous, this study aims to enhance their visibility. Initially, the mean value of the ACC signal to be segmented was calculated, and then signal sections below the mean value were rectified to match the mean value (illustrated as ‘Rectified ACC’ in **Figure 2b**). Subsequently, a 50-ms sliding average smoothing filter was applied to smooth the ACC signal, reducing interference for peak detection.

3.2. Markov Transition Field

The prevalent methods for converting 1-D sEMG signals to 2-D include feature image and spectrogram approaches.^[30,31] Feature images involve manually selected features, while spectrograms sacrifice temporal/spatial resolution.^[32–34] Alternatively, Markov transition field (MTF) provides a time series coding

method that captures dynamic transition properties of sEMG signals. MTF generates a transition probability matrix by dividing signals into quantile bins and calculating transition probabilities between states.^[35] It avoids the trade-off between temporal and spatial resolution and enhances recognition capabilities of CNN models without subjective hand-crafted features.^[36,37] Consequently, MTF has found success in sEMG-based muscle fatigue visualization and emotion recognition research.^[38,39]

This study employed MTF to convert 1-D sEMG into its 2-D representation. The aim of the MTF transformation is to construct a matrix that represents the transition probability from one state to another over time.^[40] As shown in **Figure 3**, the MTF transformation contains 5 steps. Given a sEMG signal $X_{1 \times n} = [x_1, x_2, x_3, \dots, x_t, \dots, x_n]$, first, Q quantile bins of $X_{1 \times n}$ are identified, each bin represents a state mentioned above, and each x_t is in its corresponding bins q . Second, a $Q \times Q$ Markov transition matrix W is calculated by computing the transitions among the Q quantile bins in the way of a first-order Markov chain along the time axis. Each element w_{q_i, q_j} ($i, j \in [1, Q], w \in \mathbb{N}$) of the Markov transition matrix W (Equation (1)) represents bins q_i have w_{q_i, q_j} elements followed by an element located in q_j .

$$W = \begin{bmatrix} w_{q_1, q_1} & w_{q_1, q_2} & \dots & w_{q_1, q_Q} \\ w_{q_2, q_1} & w_{q_2, q_2} & \dots & w_{q_2, q_Q} \\ \vdots & \vdots & \ddots & \vdots \\ w_{q_Q, q_1} & w_{q_Q, q_2} & \dots & w_{q_Q, q_Q} \end{bmatrix} \quad (1)$$

Third, the probability $m_{i,j}$ that the next value of $x_t \in q_i$ is at q_j is calculated by Equation (2). And then, the MTF matrix M (Equation (3)) can be constructed by all the $m_{i,j}$ calculated along the time axis. By assigning transition probabilities of $q_i \rightarrow q_j$, the MTF matrix M encodes the multi-span transition probabilities of the sEMG signals.

$$m_{i,j} = \frac{w_{q_i, q_j}}{w_{q_1, q_j} + w_{q_2, q_j} + \dots + w_{q_i, q_j} + \dots + w_{q_Q, q_j}} \quad (2)$$

$$M = \begin{bmatrix} m_{i,j}|x_1 \in q_i, x_1 \in q_j & \dots & m_{i,j}|x_1 \in q_i, x_n \in q_j \\ m_{i,j}|x_2 \in q_i, x_1 \in q_j & \dots & m_{i,j}|x_2 \in q_i, x_n \in q_j \\ \vdots & \ddots & \vdots \\ m_{i,j}|x_n \in q_i, x_1 \in q_j & \dots & m_{i,j}|x_n \in q_i, x_n \in q_j \end{bmatrix} \quad (3)$$

To improve the computational efficiency, the dimensionality of M is reduced by averaging the pixels in each non-overlapping $k \times k$ patch of the M with a blurring kernel $\{\frac{1}{k^2}\}_{k \times k}$. Due to the unequal length of each gait cycle, the K in this study varied with the length of the gait cycle. The final size for one MTF matrix M after the blurring kernel processing was 48×48 .

The final dataset included the MTF matrices of sEMG and the corresponding ACC signals. Sixteen-channel sEMG was transformed into MTF matrices, respectively, to create the sEMG sub-dataset. ACC signals were also intercepted according to the gait cycle nodes, forming the ACC sub-dataset.

3.3. Automated Gait Dysfunction Assessment Framework

The proposed multi-modal and multi-scale feature fusion (MMFF) automated gait dysfunction assessment framework is shown in **Figure 4a**. The aims of the MMFF model are: 1) acquiring a feature map contains both the local and global information, 2) acquiring an adaptive-refined feature map, and 3) fusing feature maps with different modes, namely the sEMG-mode and the ACC-mode (Figure 4a). The inputs to the proposed model consist of the MTF matrices of the 16-channel sEMG signal and the tri-axial ACC signal. Regarding the ACC input format, this study employed bilinear interpolation to scale the original tri-axial ACC into the shape of $(48 \times 48 \times 1)$. Notably, although the ACC signals were processed for gait cycle segmentation, the inputted ACC signals were still the raw signal to retain the original movement information of gait. Thus, in the initial stage of the MMFF framework (Figure 4a), Convolutional Block Attention Module (CBAM) (Figure 4c) and ConvBN blocks (Figure 4b) were employed to obtain adaptive-refined feature maps with an extended receptive field. Then, ResNext architecture was implemented to extract higher-level feature maps containing global information of its inputs. Finally, the cross-attention module was used to selectively fuse the neuromuscular and motion information contained in the sEMG-mode and the ACC-mode. The details of the above-mentioned modules are described in the following subsections.

3.3.1. Convolutional Block Attention Module

To enhance the inter-channel learning capability of CNN model, recent research has focused on integrating attention mechanisms into CNNs to improve performance. Hu et al. proposed the squeeze-and-excitation (SE) module, which captures channel-wise attention weights to model inner-channel relationships.^[41] However, SE only considers interactions between channels. In contrast, Woo et al. introduced the CBAM, which computes attention weights from both channel and spatial axes. By incorporating CBAM into the CNN model, improved accuracy through adaptively refined feature maps resulting from the multiplication of the feature map with adaptive attention weights was achieved.^[42] Therefore, this study embedded CBAM modules in the MMFF framework.

As shown in Figure 4c, the input of the CBAM is an intermediate feature map $F \in \mathbb{R}^{C \times H \times W}$, the CBAM aggregate of the 1-D channel attention map $M_c \in \mathbb{R}^{C \times 1 \times 1}$, and the 2-D spatial

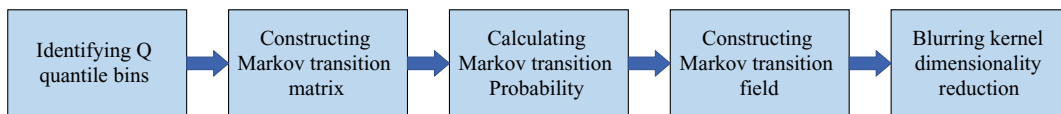


Figure 3. Calculation procedures of MTF.

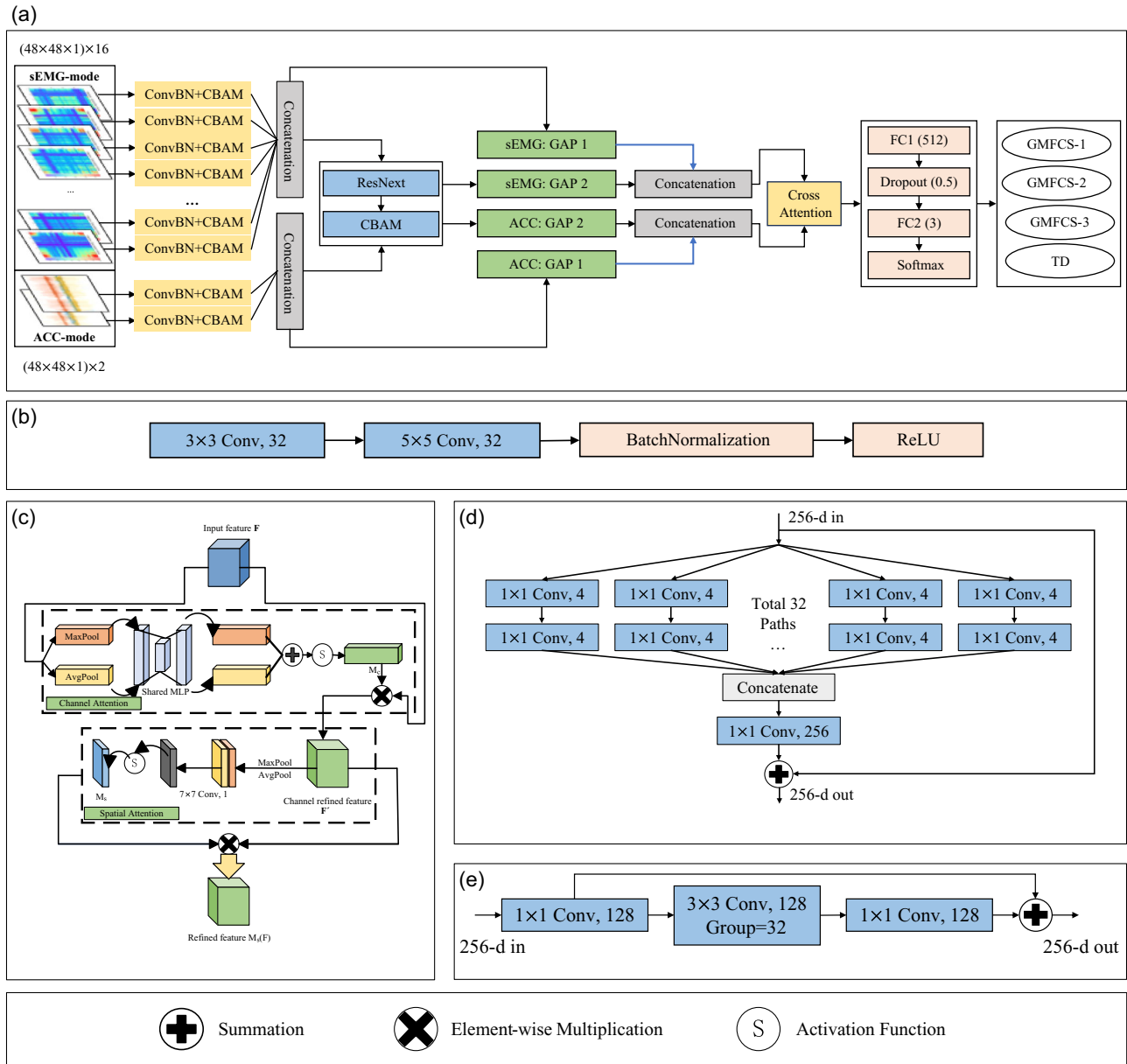


Figure 4. The MMFF framework. a) The pipeline of the MMFF framework, where the “FC” in (a) denotes the fully connected layers, the GAP denotes the global average pooling, sEMG inputs were the MTF matrices of the 16-channel sEMG signals, and ACC inputs were from the left and right leg ACC signals, b) ConvBN block, c) CBAM, d,e) two forms of the aggregated transformation, where the “256-d in” and “256-d out” denote the dimensionality of the input and output feature map. The format of the convolution hyperparameters was: $n \times n$ Conv, k , where $n \times n$ is the convolution kernel size and k is the number of convolution kernels.

attention map $M_s \in \mathbb{R}^{1 \times H \times W}$ to obtain the channel-spatial refined feature map. To obtain the channel attention map, first, max-pooling and average-pooling are performed on the channel axis of the input features F to calculate the pooled 1-D feature vectors $F_{\max}^c$ and F_{avg}^c . Then both the $F_{\max}^c$ and F_{avg}^c are forwarded to the shared multi-layer perceptron (MLP) to produce the channel attention map $M_c(F)$. The equation of the channel attention is shown as:

$$M_c(F) = \sigma(W_1(W_0(F_{\max}^c)) + W_1(W_0(F_{\text{avg}}^c))) \quad (4)$$

where σ denotes the sigmoid activation function, $W_0 \in \mathbb{R}^{C/r \times C}$, $W_1 \in \mathbb{R}^{C \times C/r}$ denotes the shared weights of MLP, and r is the reduction ratio. To obtain the spatial attention map, max-pooling and average-pooling are performed on the spatial axis of the input channel refined feature F' to calculate two 2-D channel information maps $F_{\max}^s \in \mathbb{R}^{1 \times H \times W}$ and $F_{\text{avg}}^s \in \mathbb{R}^{1 \times H \times W}$. The spatial attention map is computed as:

$$M_s(F) = \sigma(f^{7 \times 7}([F_{\max}^s; F_{\text{avg}}^s])) \quad (5)$$

where σ denotes the sigmoid activation function, and the $f^{7 \times 7}$ represents a convolutional filter with the size of 7×7 .

3.3.2. ResNext Module

This study applied the ResNext-50 architecture to extract higher-level feature map. The ResNext architecture in this study is shown in **Table 2**. ResNext was an upgraded version of the ResNet.^[43,44] Compared to the ResNet, ResNext increased the representational power and reduced the training parameters by replacing the bottleneck block of the ResNet with a parallel set of bottleneck blocks termed the aggregated transformation through the grouped convolution.

The aggregated transformation can be described as:

$$\mathcal{F}(x) = \sum_{i=1}^C \mathcal{T}_i(x) \quad (6)$$

where $\mathcal{T}_i(x)$ is the i -th path neuron (Figure 4d), and C is the number of the parallel set of bottleneck blocks termed cardinality, which is the “path” of Figure 4d and the “group” of Figure 4e.

Table 2. The model architecture of the ResNext module.

Aggregated transformation block	Model architecture ^{a)}
Block 1	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128, \\ 1 \times 1, 256 \end{bmatrix} C = 32 \times 3$
Block 2	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256, \\ 1 \times 1, 512 \end{bmatrix} C = 32 \times 4$
Block 3	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512, \\ 1 \times 1, 1024 \end{bmatrix} C = 32 \times 6$
Block 4	$\begin{bmatrix} 1 \times 1, 1024 \\ 3 \times 3, 1024, \\ 1 \times 1, 2048 \end{bmatrix} C = 32 \times 3$

^{a)}The brackets are the shape of each aggregated transformation block and “ $C = 32$ ” denotes the grouped convolutions with 32 groups.

C was set to 32 in this study, which was consistent with reference.^[39] The output of a parallel set of bottleneck block is

$$y = x + \sum_{i=1}^C \mathcal{T}_i(x) \quad (7)$$

where y is the output and x is the input, by summing the input x with the $\mathcal{F}(x)$, an aggregated transformation residual architecture formed.

3.3.3. Cross-Attention Module

Our proposed model takes sEMG and ACC signals as inputs, and their fusion approach significantly impacts the classification performance. Traditional fusion methods, such as concatenation and summation, have been commonly used for multi-mode fusion. Previous studies utilized concatenation and summation to fuse feature maps from sEMG and ACC signals, achieving successful results.^[45,46] However, for fusion involving multi-mode heterogeneous and information-imbalanced features, concatenation and summation are still inadequate.^[47] To address this, cross-attention mechanism, a powerful tool for selectively attending to relevant parts of each modality, was adopted.^[48] The cross-attention computes an attention score map representing the relevance between elements of different input branches. By employing cross-attention, the interaction between sEMG and ACC signals is enhanced, enabling the feature vector to capture information from both modalities.

This study performed the 1-D multi-head cross-attention on the sEMG and ACC data.^[49] As shown in **Figure 5**, the inputs of the cross-attention module are the feature map of the MTF matrix of the sEMG signals and the feature map of the ACC signals. The sEMG-branch is called the target, and the ACC-branch is called the source. According to Equation (8), q_i , k_i , and v_i are calculated, which denotes the vector of the query, key, and value, respectively:

$$q_i = x_{\text{target}} W^q, k_i = x_{\text{source}} W^k, v_i = x_{\text{source}} W^v \quad (8)$$

where W^q , W^k , and W^v are three learnable weight vectors, and $i \in \mathbb{R}$ denotes the index of each head. The number i was set to 2

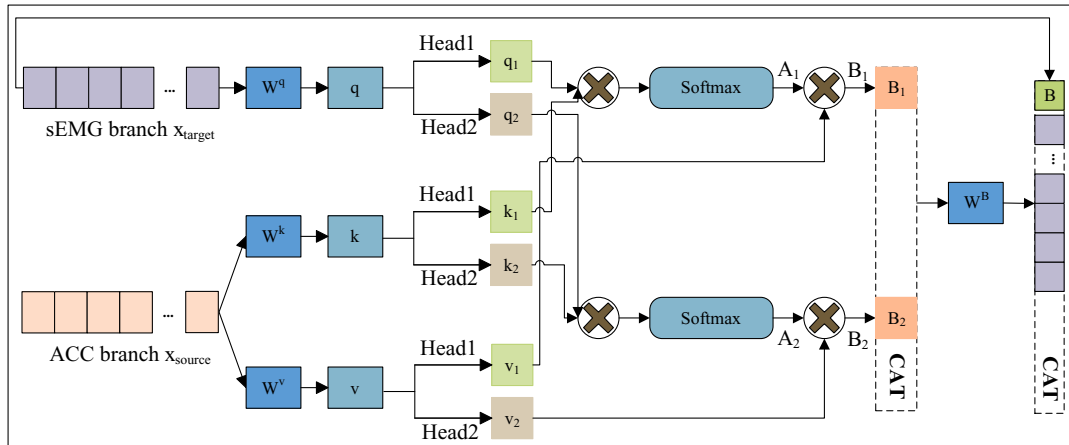


Figure 5. The cross-attention module. “CAT” denotes concatenation.

in this study. Then, the attention score vector of each head is computed through Equation (9), and the output of each head will be calculated as Equation (10):

$$A_i = \text{softmax} \left(\frac{q_i k_i^T}{\sqrt{d_{k_i}}} \right) \quad (9)$$

$$B_i = A_i v_i \quad (10)$$

where k_i^T is the transpose of k_i , h is the number of heads, and d_{k_i} is the dimensionality of k_i .

The output of the cross-attention is a concatenation of B and the x_{target} to form a residual shortcut. The mathematical results can be described as follows:

$$B = W_o([B_1 \| B_2]) \quad (11)$$

$$\text{CrossAttention} = [x_{\text{target}} \| B] \quad (12)$$

where W_o is a learnable weight vector and B_1, B_2 denotes the output of each head in Figure 5. By applying the cross-attention module, feature vectors with more representative and contextual representations of the input data can be learned.

4. Experiments and Results

4.1. Experimental Setup

Signal preprocessing was carried out on MATLAB 2018b, and MTF transformation and deep learning model training and testing process were carried on the hardware 16 vCPU Intel(R) Xeon(R) Platinum 8350C CPU @ 2.60 GHz and software Pycharm with the Keras 2.5.0 library. Especially, the deep learning model was computed on the RTX 3090 (24 GB). The batch size for the training and testing was set to 8 and 1, respectively. The training epoch was set to 100, and the Adam was chosen as the optimizer for the stage of training. The learning rate was initialized to 0.001 and scheduled to be reduced after 50, 65, 80, and 90 epochs. This study assigned 10 subjects (3 from GMFCS-1, 1 from GMFCS-2, 3 from GMFCS-3, and 3 from TD children) to the training set, 4 subjects (selecting one participant from each class who differed from the training and testing sets) to the validation set, and 4 subjects (selecting one participant from each

class who differed from the training and validation sets) to the testing set. Due to the limited number of subjects at GMFCS-2 level, only two individuals were available for this category. Therefore, data from one subject at GMFCS-2 level, collected from distinct trials on different dates, were assigned to the training and validation sets, respectively. Meanwhile, data from another subject at GMFCS-2 level were designated for the testing set, thus to avoid the problem from data leakage.

4.2. Automated Gait Dysfunction Assessment Results

The proposed MMFF framework achieved an overall accuracy of 99.13% on the testing set for automating gait dysfunction. To clearly depict the convergence in the model training process, Figure 6 illustrates the accuracy, loss curves, and the learning rate curve for both the training and validation sets. As depicted in Figure 6, after 60 epochs, the accuracy and loss curves for both the training and validation sets gradually stabilized, which indicated that our model had converged.

4.3. Comparison of Different Bins of the MTF

Different bins lead to different dynamical transition statistics representation. Wang et al. recommended larger bins since the larger bins provide a more detailed transition statistics representation.^[40] In this section, the bins were set to {4, 8, 16, 32}, so as to explore the best classification performance. As shown in Table 3, the increase of bins from 4 to 16 corresponded to an improvement in classification performance, rising from 97.69% to 99.13%. However, with a subsequent increase in the number of bins to 32, a discernible degradation in classification accuracy occurred, specifically resulting in a decline to

Table 3. Accuracies and loss values of the test dataset.

Bins	Accuracy [%]	Test loss
04	97.69	0.1118
08	98.27	0.1106
16	99.13^{a)}	0.0558^{a)}
32	95.95	0.2439

^{a)}Bold indicates the best result.

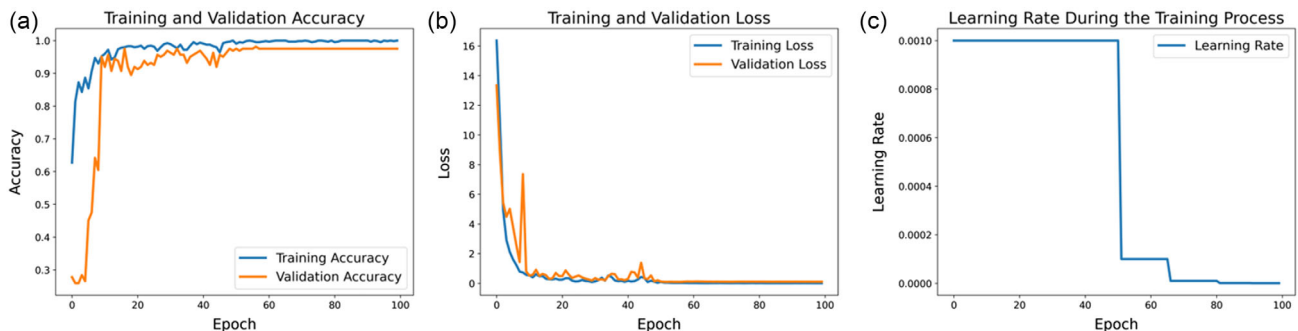


Figure 6. Convergence changes during the training process. a) Training and validation accuracy curve. b) Training and validation loss curve. c) Learning rate reduction curve.

Table 4. The evaluation metrics values of each bin.

	GMFCS-1				GMFCS-2				GMFCS-3				Typically developing			
Bins	04	08	16	32	04	08	16	32	04	08	16	32	04	08	16	32
Precision	0.96	0.97	0.98	0.96	0.98	0.97	1.00	0.95	0.97	0.99	0.99	0.93	1.00	1.00	1.00	0.99
Recall	0.99	0.99	1.00	0.96	0.97	0.98	0.98	0.95	0.95	0.96	0.98	0.92	1.00	1.00	1.00	1.00
F1-score	0.97	0.98	0.99	0.96	0.97	0.98	0.99	0.95	0.96	0.97	0.98	0.93	1.00	1.00	1.00	0.99
Average	0.97	0.98	0.99^{a)}	0.96	0.97	0.97	0.99^{a)}	0.95	0.96	0.97	0.98^{a)}	0.93	1.00	1.00	1.00	0.99

^{a)}Bold indicates the best result.

95.95%. To show the classification performance of each GMFCS level and TD children more intuitively, the precision, recall, and F1-score were employed as the evaluation metrics. As shown in **Table 4**, for the classification of the children with CP, Bins16 achieved the best performance, the averaged metrics of Bins16 were all the highest in each class. Moreover, for the evaluation metrics, values of the TD children were all greater than 0.99, whereas, each of the evaluation metrics of Bins32 was the lowest in each class. In general, Bins16 preformed the best among the four bins.

4.4. The Effectiveness of CBAM

As a lightweight channel-spatial attention module, CBAM produced an adaptive-refined feature map to improve the classification performance. The effectiveness of CBAM was examined by

omitting it from the MMFF model and replacing it with the SE module, facilitating a comparative analysis of their contributions. As shown in **Figure 7a**, the model with CBAM reached the best results among four evaluation metrics, namely the accuracy, precision, recall, and F1-score. When removing the CBAM, the accuracy was reduced from 99.13% to a minimum 94.22%, and the precision, recall, and F1-score were also declined from 99.25%, 99.00%, and 99.00% to 94.00%, 94.50%, and 94.25%, respectively. Additionally, when the CBAM was replaced by the SE module, the accuracy, precision, recall, and F1-score were decreased slightly to 97.98%, 98.00%, 98.00%, and 97.75%, respectively. We also applied the *t*-SNE visualization technique to visualize the features extracted by the latest FC layer.^[50]

Figure 8a–c depicts visualizations of the proposed MMFF model, the model incorporating the SE module, and the model with the CBAM module removed, respectively. As we can see the three

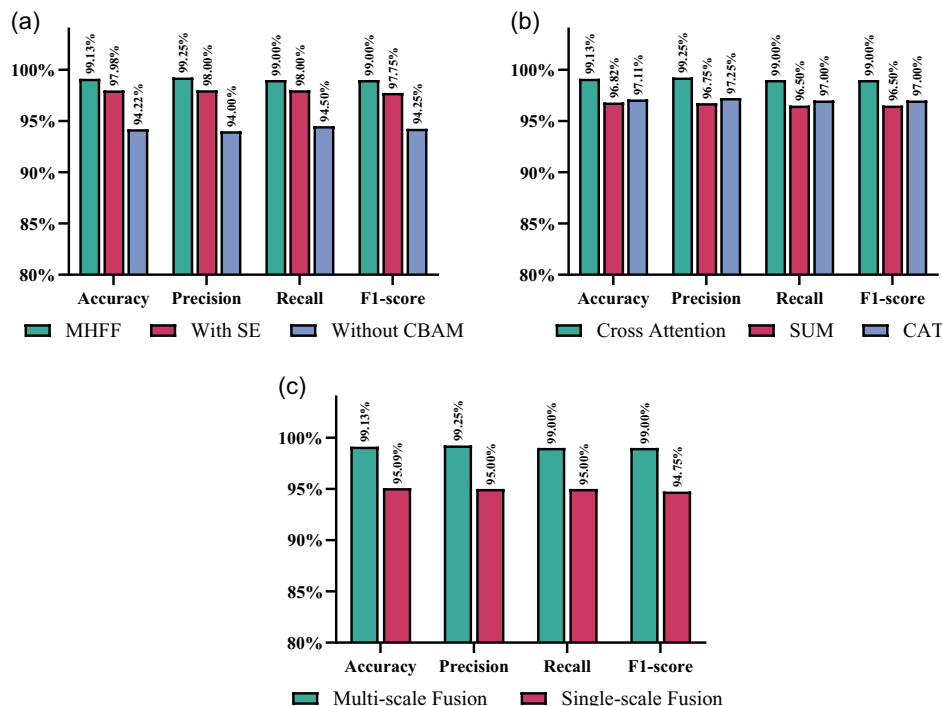


Figure 7. The evaluation metrics of the compared models. a) Comparison of the classification performance among the model with and without CBAM and the model with SE module. b) Performance comparison of the multi-mode fusion strategies. “SUM” denotes “summation” and “CAT” denotes “concatenation”. c) Performance of the multi-scale feature fusion and the single-scale feature fusion strategies. Precision, recall, and F1-score were averaged across the three classes.

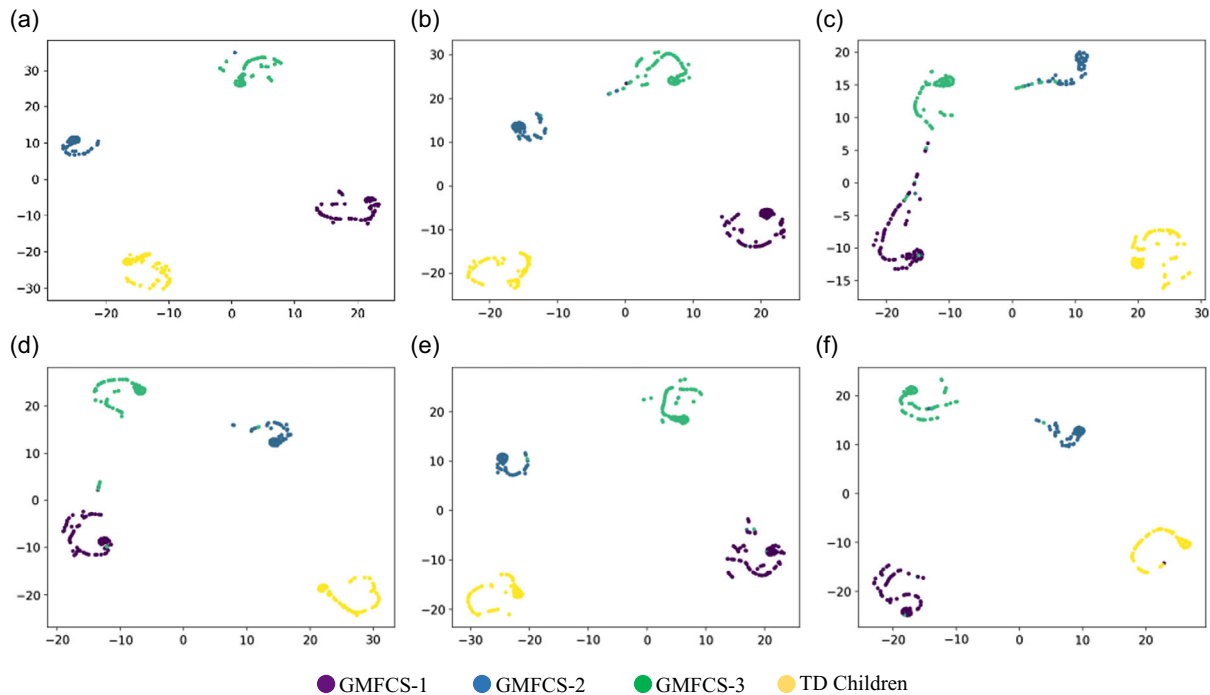


Figure 8. The *t*-SNE visualization of the compared models on the testing dataset: a) proposed MMFF model, b) model with SE module, c) model removed the CBAM, d) cross-attention module replaced by the summation fusion, e) cross-attention module replaced by the concatenation fusion, and f) single-scale feature fusion.

models all have good discrimination, but a bunch of the features of the GMFCS3 in the model without CBAM was misclassified into the GMFCS1 cluster. These results were consistent with tendency of Figure 7a.

4.5. Comparison of Different Fusion Strategies

Cross-attention mechanism is more suitable for the multi-mode information fusion tasks than the simple summation and concatenation. In Figure 7b, substituting cross-attention module in MHFF framework with summation led to a reduction in accuracy, precision, recall, and F1-score to 96.82%, 96.75%, 96.50%, and 96.50%, respectively. Meanwhile, adopting the concatenation approach resulted in a decrease to 97.11%, 97.25%, 97.00%, and 97.00%. These findings underscored the superior classification performance of cross-attention, while highlighting the comparatively poorer performance of summation and concatenation fusion strategies.

Moreover, to fuse feature vectors containing local and global information, before the cross-attention module, the global average pooling (GAP) was implemented to compute the 1-D representation of the 2-D feature map. As shown in Figure 4a, the GAP1 of both the sEMG-branch and the ACC-branch represented the feature vector with local information. After the feature maps pass through the ResNext module, the receptive field of the feature map increased. Thus, the GAP2 of both the sEMG-branch and the ACC-branch represented the feature vector with global information. This study fused the two vectors of each mode to improve the classification

performance. In Figure 7c, our proposed multi-scale feature fusion strategy underwent a comprehensive evaluation against the single-scale feature fusion approach. Comparative analysis of accuracy, precision, recall, and F1-score between multi-scale and single-scale feature fusion revealed a decrease in the latter to 95.09%, 95.00%, 95.00%, and 94.75%, respectively. These results clearly demonstrated that the multi-scale fusion strategy outperformed the single-scale fusion across all evaluation metrics, which suggested that our feature map contains richer information, contributing to enhanced performance.

5. Discussions

5.1. Discussions on the Related Work

This study automated the gait dysfunction assessment process in children with CP using a multi-modal CNN-based network. To highlight practical applications, related studies in patients with CP have been summarized in Table 5. In most studies, researchers primarily focused on gait activity recognition and gait phase detection.^[6,17–19] Regarding gait dysfunction severity evaluation, ref. [21] detected only abnormal and normal gait. In comparison, this study expanded the classes to include normal gait and specific gait dysfunction levels, namely GMFCS-1, GMFCS-2, and GMFCS-3. More detailed expansion aids clinicians in developing more precise gait dysfunction rehabilitation schedules for patients. Ref. [51] estimated gait deviation severities based on the Edinburgh Visual Scale (EVS) and estimated severities from Observation 1 to 7, which represented different gait pathological

Table 5. Comparing other wearable sensor-based research in patients with CP.

Authors	Application	Sensor	Number of subjects	ML method	Accuracy
Mancinelli ^[51]	Classifying the gait deviation severities	ActiveGait system with 15 insole sensors and one ankle angle sensor	11 children with CP	RF	>80%
Chakraborty ^[21]	Gait abnormality detection	Inertial sensors	9 CP diagnosed children and 9 healthy children	CNNs	96.4% and 90.97% for segment-wise and subject-wise
Hegde ^[17]	Physical activity recognition	5 FSR sensors and a 3-D accelerometer	10 children with CP and 11 healthy children	MLD	95.3% for children with CP and 96.2% for healthy children
Ahmadi ^[18]	Physical activity recognition	sEMG sensors	22 children and adolescents with CP	RF, SVM, BDT	SVM (82.0–89.0%), RF (82.6–88.8%), and BDT (76.1–86.2%)
Wei ^[19]	Gait phase recognition	sEMG sensors	10 children with CP	SVM	89.40%
Morbidoni ^[6]	Gait phase prediction	Foot-floor-contact and sEMG sensor	20 children with CP	SVM, RF, KNN, MLP	MAE, 14.8 ± 3.2 ms for HS and 17.6 ± 4.2 ms for TO
Ours	Automated gait dysfunction assessment	sEMG and ACC sensor	12 children with CP and 5 TD children	CNNs	99.13%

features, in the EVS, with accuracies exceeding 80%. Similarly, this study classified different gait abnormal severities. As shown in Table 4, results under Bins16 of the MTF matrix consistently exceeded 98.00% for each GMFCS level and TD control. These findings indicate that the proposed multimodal CNN-based framework has significant potential to contribute to auxiliary diagnostic devices for gait dysfunction assessment in children with CP.

5.2. Effect of the MTF Bins

The larger the number of bins, the more detailed the MTF contains the dynamic transition probability of the sEMG. However, as shown in Table 3, when the number of bins increased to 32, the classification accuracy decreased from 99.13% of the Bins16 to 95.95% of the Bins32, this was due to the increased sparsity of the input data. When the number of bins increased, the transition probability from one state to another may be 0 because some of the bins may have very small or no instances of transitions, which resulted in the increase of the sparsity of the MTF matrix, leading to instability during the model training process.

5.3. Approaches of the Multi-Modal Fusion

The concatenation and summation are the most commonly used approaches for the multi-mode fusion tasks, which fuse the output of the sEMG-branch and the ACC-branch (Figure 4a) into a single feature vector or matrix by a simple combination. In contrast, the cross-attention mechanism allows the model to selectively attend to the most relevant parts of each mode, which has more interactions between the sEMG and ACC signals when fusing the two different modes. As shown in Figure 7b, the results of the comparison of the cross-attention and the simple concatenation and summation were consistent with the discussions above. Specifically, the cross-attention reached the highest accuracy 99.13%, while the simple concatenation and summation reached 97.11% and 96.82%, respectively. Results indicated that

the cross-attention integrated the sEMG and ACC information in a more fine-grained way than concatenation and summation.

5.4. Effect of the Multi-Scale Fusion

In our investigation, feature maps with distinct information scopes were analyzed. Specifically, feature maps with local information focused on recognizing fine-grained patterns within smaller regions of the input images. On the other hand, feature maps with global information exhibited a broader receptive field, allowing them to capture larger-scale patterns contain in the input images. The proposed multi-scale fusion strategy fused the global and local information in one feature vector. Thus, the model learns to recognize a broader range of the patterns and features contained in the sEMG and ACC inputs by supplementing local information with global information. This work compared the multi-scale fusion and single-scale fusion strategy. After removing the blue arrows in Figure 4a, the multi-scale fusion was transformed into the single-scale fusion. As a result, the local information of the feature vector was reduced, and the classification accuracy decreased from 99.13% to 95.09% (Figure 7c). In short, a feature with both local and global information improves the classification accuracy.

6. Conclusion

This study introduced a novel MMFF framework for automating the assessment of gait dysfunction in children with CP, streamlining the GMFCS assessment process. Leveraging sEMG and ACC signals recorded during the children's walking, MMFF utilized CBAM to obtain the adaptively refined feature maps from both channel and spatial axes. Additionally, it incorporated a cross-attention module to acquire cross-mode correlations, so as to selectively obtain a feature vector containing both sEMG and ACC information. The MMFF framework further integrated a multi-scale fusion strategy to extract a feature vector enriched with both local and global information. The classification accuracy reached 99.13%, while the mean values for model evaluation

metrics, including precision, recall, and F1-score in GMFCS-1, GMFCS-2, and GMFCS-3, were 99.00%, 99.00%, and 98.33%, respectively. These results served as a validation for the effectiveness of MMFF. Results affirmed its potential as an objective, lightweight tool for clinicians to assess gait function and inform rehabilitation strategies for children with CP. While this work concentrated on gait motor dysfunction, future works aim to extend the automated assessment approach to lower limb motor dysfunction in children with CP.

Acknowledgements

This project was supported by a grant from National Key R&D Program of China (no. 2023YFF0719202). The authors also acknowledge all the volunteers and participants in this work for the signal acquisition.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Keywords

acceleration, cerebral palsy, gait dysfunction assessments, Markov transition fields, surface electromyography

Received: December 5, 2023

Revised: January 24, 2024

Published online:

- [1] M. Sadowska, B. Sarecka-Hujar, I. Kopyta, *Neuropsychiatr. Dis. Treat.* **2020**, *16*, 1505.
- [2] T. S. Pearson, R. Pons, R. Ghaoui, C. M. Sue, *Mov. Disord.* **2019**, *34*, 625.
- [3] S. Armand, G. Decoulon, A. Bonnefoy-Mazure, *EFORT Open Rev.* **2016**, *1*, 448.
- [4] L. Tang, X. Chen, S. Cao, D. Wu, G. Zhao, X. Zhang, *Front. Hum. Neurosci.* **2017**, *11*, 130.
- [5] X. Chen, Q. Wu, L. Tang, S. Cao, X. Zhang, X. Chen, *Med. Biol. Eng. Comput.* **2020**, *58*, 101.
- [6] C. Morbidoni, A. Cucchiarelli, V. Agostini, M. Knaflitz, S. Fioretti, F. D. Nardo, *IEEE Trans. Neural Syst. Rehabil. Eng.* **2021**, *29*, 819.
- [7] V. Agostini, A. Nascimbeni, A. Gaffuri, M. Knaflitz, *Clin. Biomech.* **2015**, *30*, 908.
- [8] D. M. Bojanic, B. D. Petrovacki-Balj, N. D. Jorgovanovic, V. R. Ilic, *J. Neurosci. Methods* **2011**, *198*, 325.
- [9] F. Di Nardo, A. Strazza, A. Mengarelli, S. Cardarelli, T. Andrea, V. Federica, N. Alberto, A. Valentina, K. Marco, S. Sandro, *Biosens.-Basel* **2019**, *9*, 82.
- [10] D. Patikas, S. Wolf, L. Döderlein, *Eur. J. Neurol.* **2005**, *12*, 691.
- [11] C. Frigo, P. Crenna, *Clin. Biomech.* **2009**, *24*, 236.
- [12] Y. Yu, X. Chen, S. Cao, D. Wu, X. Zhang, X. Chen, *J. Neurophysiol.* **2019**, *121*, 1680.
- [13] Y. Gagnat, S. M. Brændvik, K. Roeleveld, *Front. Neurol.* **2020**, *11*, 202.
- [14] C. Castagneri, V. Agostini, S. Rosati, G. Balestra, M. Knaflitz, *IEEE Trans. Neural Syst. Rehabil. Eng.* **2019**, *27*, 772.
- [15] R. Bajpai, D. Joshi, *IEEE Trans. Neural Syst. Rehabil. Eng.* **2021**, *29*, 2530.
- [16] J. Figueiredo, C. P. Santos, J. C. Moreno, *Med. Eng. Phys.* **2018**, *53*, 1.
- [17] N. Hegde, T. Zhang, G. Uswatte, E. Taub, J. Barman, S. Mckay, A. Taylor, D. M. Morris, A. Griffin, E. S. Sazonov, *IEEE Trans. Neural Syst. Rehabil. Eng.* **2017**, *26*, 477.
- [18] M. Ahmadi, M. Neil, M. Fragala-Pinkham, N. Lennon, S. Trost, *J. Neuroeng. Rehabil.* **2018**, *15*, 163180.
- [19] P. Wei, R. Xie, R. Tang, C. Li, J. Kim, M. Wu, *Ann. Biomed. Eng.* **2019**, *47*, 223.
- [20] X. Wang, L. Tang, Q. Zheng, X. Yang, Z. Lu, *Sensors* **2023**, *23*, 5775.
- [21] J. Chakraborty, A. Nandy, *Biomed. Signal Process. Control* **2020**, *62*, 102076.
- [22] J. Gao, P. Gu, Q. Ren, J. Zhang, X. Song, *IEEE Access* **2019**, *7*, 163180.
- [23] R. Bajpai, A. Tiwari, D. Joshi, R. Khataavkar, in *Int. Conf. Advanced Technologies*, Goa, India, January 2022.
- [24] L. I. Mihăilă, C. G. Cordoș, R. R. Ilcșan, P. Faragó, S. Hintea, in *Int. Conf. Telecommunications and Signal Processing*, Prague, Czech Republic, July 2022.
- [25] C. Fricke, J. Alizadeh, N. Zakhary, T. B. Woost, M. Bogdan, J. Classen, *Front. Neurol.* **2021**, *12*, 666458.
- [26] B. Vidya, P. Sasikumar, *Eng. Appl. Artif. Intell.* **2022**, *114*, 105099.
- [27] S. Shen, X. Wang, M. Wu, K. Wu, X. Chen, X. Geng, *IEEE Sens. J.* **2023**, *23*, 4052.
- [28] K. Li, J. Zhang, L. Wang, M. Zhang, J. Li, S. Bao, *Biomed. Signal Process. Control* **2020**, *62*, 102074.
- [29] J. Perry, J. R. Davids, *J. Pediatr. Orthop.* **1992**, *12*, 815.
- [30] Y. Yu, X. Chen, S. Cao, X. Zhang, X. Chen, *IEEE J. Biomed. Health Inf.* **2019**, *24*, 1310.
- [31] J. Fan, M. Jiang, C. Lin, G. Li, J. Fiaidhi, C. Ma, W. Wu, *Neural Comput. Appl.* **2021**, *35*, 16101.
- [32] T. Bao, S. Q. Xie, P. Yang, P. Yang, P. Zhou, Z. Zhang, *IEEE J. Biomed. Health Inf.* **2022**, *26*, 3822.
- [33] Y. Zhang, S. Chen, W. Cao, P. Guo, D. Gao, M. Wang, J. Zhou, T. Wang, *Expert Syst. Appl.* **2021**, *185*, 115639.
- [34] M. Sharma, H. Malhotra, N. Kumar, J. Yadav, in *Int. Conf. Computing, Communication and Power Technology*, New Delhi, India, September 2019.
- [35] X. Zhao, H. Sun, B. Lin, H. Zhao, Y. Niu, X. Zhong, Y. Wang, Y. Zhao, F. Meng, J. Ding, X. Zhang, L. Dong, S. Liang, *IEEE Sens. J.* **2021**, *22*, 3348.
- [36] N. Bahador, J. Kortelainen, *Biomed. Signal Process. Control* **2022**, *72*, 103300.
- [37] S. Yu, X. You, W. Ou, X. Jiang, K. Zhao, Z. Zhu, Y. Mou, X. Zhao, *Neurocomputing* **2016**, *204*, 211.
- [38] R. Li, Y. Wu, Q. Wu, N. Dey, R. G. Crespo, F. Shi, *Measurement* **2022**, *189*, 110470.
- [39] D. Sasidharan, G. Venugopal, S. Ramakrishnan, in *Annual Int. Conf. IEEE Engineering in Medicine and Biology Society*, IEEE, Glasgow, UK, July 2022.
- [40] Z. Wang, T. Oates, in *AAAI Conf. Artificial Intelligence*, TX, January 2015.
- [41] J. Hu, L. Shen, G. Sun, in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, IEEE, UT, June 2018.
- [42] S. Woo, J. Park, J. Y. Lee, I. S. Kweon, in *Proc. European Conf. Computer Vision*, Springer, Munich, Germany, July 2018.
- [43] S. Xie, R. Girshick, P. Dollár, Z. Tu, K. He, in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, IEEE, HI, July 2017.
- [44] K. He, X. Zhang, S. Ren, J. Sun, in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, IEEE, NV, June 2016.
- [45] S. Duan, L. Wu, B. Xue, A. Liu, R. Qian, X. Chen, *IEEE Sens. J.* **2023**, *23*, 2773.

- [46] S. Wang, L. Huang, D. Jiang, Y. Sun, G. Jiang, J. Li, C. Zou, H. Xiong, B. Chen, *Front. Bioeng. Biotechnol.* **2022**, 10, 909023.
- [47] Z. Xue, X. Tan, X. Yu, B. Liu, A. Yu, P. Zhang, *IEEE Trans. Image Process.* **2022**, 31, 3095.
- [48] C.-F.-R. Chen, Q. Fan, R. Panda, in *Proc. IEEE/CVF Int. Conf. Computer Vision*, IEEE, Canada, BC, October 2021.
- [49] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, *Adv. Neural Inf. Process. Syst.* **2017**, 30.
- [50] L. V. D. Maaten, G. Hinton, *J. Mach. Learn. Res.* **2008**, 9, 2579.
- [51] C. Mancinell, S. Patel, L. C. Deming, D. Nimec, J. J. Chu, J. Beckwith, R. Greenwald, P. Bonato, in *2012 Annual Int. Conf. IEEE Engineering in Medicine and Biology Society*, IEEE, San Diego, CA, August 2012.