Article

Implementation of Sequence-Based Classification Methods for Motion Assessment and Recognition in a Traditional Chinese Sport (Baduanjin)

Hai Li [1,2] , Selina Khoo [2] and Hwa Jen Yap [3,*]

1   Teaching and Research Section of Physical Education, College of Sport, Neijiang Normal University, Neijiang 641112, China; 10000861@njtc.edu.cn
2   Centre for Sport and Exercise Sciences, University Malaya, Kuala Lumpur 50603, Malaysia; selina@um.edu.my
3   Department of Mechanical Engineering, Faculty of Engineering, University Malaya, Kuala Lumpur 50603, Malaysia
*   Correspondence: hjyap737@um.edu.my; Tel.: +603-79675240

Abstract: This study aimed to assess the motion accuracy of Baduanjin and recognise the motions of Baduanjin based on sequence-based methods. Motion data of Baduanjin were measured by the inertial sensor measurement system (IMU). Fifty-four participants were recruited to capture motion data. Based on the motion data, various sequence-based methods, namely dynamic time warping (DTW) combined with classifiers, hidden Markov model (HMM), and recurrent neural networks (RNNs), were applied to assess motion accuracy and recognise the motions of Baduanjin. To assess motion accuracy, the scores for motion accuracies from teachers were used as the standard to train the models on the different sequence-based methods. The effectiveness of Baduanjin motion recognition with different sequence-based methods was verified. Among the methods, DTW + k-NN had the highest average accuracy (83.03%) and shortest average processing time (3.810 s) during assessing. In terms of motion reorganisation, three methods (DTW + k-NN, DTW + SVM, and HMM) had the highest accuracies (over 99%), which were not significantly different from each other. However, the processing time of DTW + k-NN was the shortest (3.823 s) compared to the other two methods. The results show that the motions of Baduanjin could be recognised, and the accuracy can be assessed through an appropriate sequence-based method with the motion data captured by IMU.

Keywords: inertial sensor measurement systems; motion accuracy; motion recognition; Baduanjin

1. Introduction

Traditional Chinese sports are an essential part of physical education (PE) in universities in China. The official document issued by the Ministry of Education requires that the PE curriculum of all universities in China include traditional Chinese sports [1]. An official report in 2012 showed that 76.7% of universities in China had chosen Chinese martial arts as part of their PE curriculum. Although traditional Chinese sports have been incorporated into university PE curriculums, there have been some problems in implementation. The most serious problem is the high student–teacher ratio in Chinese universities. In 1999, the government in China decided to increase the number of students in universities, which led to a rapid rise in the number of students [2]. According to the latest official report [3], the number of students in universities in China hit a record high of 40.02 million in 2019. Of this number, 30.31 million were undergraduates, and the average number of students per university was 15,176. The insufficient number of PE teachers in universities in China and the number of students in PE courses exceeding the recommended numbers have resulted in a high student–teacher ratio [4]. Due to the large number of students in Chinese martial arts classes, teachers can neither provide individual guidance to students nor can they correct the errors in the motions of students [5]. Moreover, the Ministry of Education requires

that the assessments in PE focus on formative assessment, instead of only using summative assessment [1]. However, PE teachers are unable to conduct formative assessment due to the high student–teacher ratio.

The application of motion capture (MoCap) in sports provides a way to solve the above problems in teaching traditional Chinese sports in PE in Chinese universities. In the past 10 years, researchers have developed various systems for different sports on motion data captured by MoCap [6,7]. These systems measure the motions of users in real-time on MoCap, analyse the captured motion data, detect errors in the motions, and give feedback to the users to help them recognise and correct their motions. In the study by Yamaoka et al. [7], a system-applied MoCap (Kinect device) for frisbee learners was developed. The 3D motion data of learners were captured using the Kinect device while throwing the flying disc. The system checked the positions in pre/exercise/post-exercise and gave information to learners. The results show that this system can effectively improve the motions of learners.

When considering the actual requirements of PE, it was found that not all MoCaps are suitable for PE courses. The photoelectric system (OMS) shows the highest measurement accuracy in MoCap, but the high cost is a barrier for using OMS in PE [8,9]. Electromagnetic systems (EMSs) are susceptible to interference in the electromagnetic environment [10]. The image processing system (IMS) has better accuracy than the EMS and an improved range compared to the OMS [11]. Many studies used a low-cost IMS (e.g., Kinect device) to capture the motion data to analyse motion in PE6 [12]. However, a low-cost IMS has some disadvantages: low accuracy, limited environment adaptability, and limited motion ranges led us to choose the inertial sensor measurement system (IMU) for this study.

Therefore, we aimed to develop a formative assessment system using IMU to assess the motion accuracy of Baduanjin to assist teachers and students identify errors in motions. Baduanjin is a popular traditional sport in China that consists of eight decomposition motions (Figure 1).

The formative assessment system needs to recognise and assess the motions of Baduanjin effectively. The motion sequences in Baduanjin are fixed (from Motion-1 to Motion-8) and cannot be changed. However, some students have problems practising the correct motions sequence while learning it. Therefore, the system needs to have the ability to recognise Baduanjin motions to evaluate whether the student is following the correct motion sequence. Recognising the motions of Baduanjin can be interpreted as a classification problem. During the classroom teaching session, teachers use the traditional manual grading method (vision) to assess the motion accuracy of Baduanjin. Therefore, the assessment of the motion accuracies of the Baduanjin motions is converted into a classification problem [14]. Commonly used classification methods can be divided into sample-based methods and sequence-based methods [14]. Few studies have applied sequence-based methods to assess and recognise motions of Baduanjin and other traditional Chinese sports, such as tai chi. Chen et al. [12] used DTW to assess the motions of tai chi. Sequence-based methods are used in this study, including dynamic time warping (DTW) combined with classifiers, hidden Markov model (HMM) [15], and recurrent neural networks (RNNs) [16,17].

Int. J. Environ. Res. Public Health **2022**, *19*, 1744

3 of 29

Motion 1: Holding the Hands High with Palms Up to Regulate the Internal Organ

Motion 2: Posing as an Archer Shooting Both Left- and Right-Handed

Motion 3: Holding One Arm Aloft to Regulate the Functions of the Spleen and Stomach

Motion 4: Looking Backwards to Prevent Sickness and Strain

Motion 5: Swinging the Head and Lowering the Body to Relieve Stress

Motion 6: Moving the Hands down the Back and Legs, and Touching the Feet to Strengthen the Kidneys

Motion 7: Thrusting the Fists and Making the Eyes Glare to Enhance Strength

Motion 8: Raising and Lowering the Heels to Cure Diseases

Figure 1. Eight standard motions of Baduanjin [13].

2. Materials and Methods

This study is comprised of three steps. The first step was to recruit volunteers and use IMU to capture the motion data of Baduanjin. Teachers and students were recruited from a university in Southwest China to participate in the study. The motion data of Baduanjin of all participants were captured using IMU. The second step involved extracting motion data and keyframes. The raw motion data needed to be converted into quaternion format for data analysis and the extracted keyframes, to prevent data redundancy and reduce processing time. The last step was to verify the effectiveness of sequence-based methods for assessing motion accuracy and recognising the motions of Baduanjin. In order to assess the motion accuracy of Baduanjin, Baduanjin experts were invited to score the motion accuracies of the captured motions and then use the scores to train the models for the different sequence-based methods (Figure 2).

2.1. Recruiting Volunteers and Capturing Motion Data

Undergraduate students or teachers with no clinical/mental illnesses or physical disabilities from a university in Southwest China were invited to participate in this study. All participants were informed and understood the purpose and scope of the data in this study and any associated privacy risks, and they accepted the researcher's commitment to protecting their privacy and the security of data. The study was conducted following the Declaration of Helsinki, and the protocol was approved by the University of Malaya Research Ethics Committee (UM.TNC2/UMREC-558).

The Baduanjin motions of the volunteers were captured with the commercial IMU "Perception Neuron 2.0", developed by Noitom [18]. Perception Neuron 2.0 has 17 inertial sensing units with a 3-axis gyroscope, 3-axis accelerometer, and 3-axis magnetometer [19]. Sers et al. verified the effectiveness the commercial IMU in measuring accuracy [20]. The captured motion data were the "output" in the Biovision Hierarchical Structure [21] motion file through the Axis Neuron software developed by Noitom.

2.2. Motion Data Conversion and Keyframes Extraction

Fifty-six participants were recruited and divided into two groups. The first group comprised 20 students and a martial arts teacher. The teacher has bachelor and master's degrees in traditional Chinese sports (martial arts) and more than 10 years of experience teaching Baduanjin. The IMU was used to capture the motions of these participants three times. The second group comprised 35 students, and IMU was used to measure the motions of these participants. Students were undergraduates without disabilities and no clinical or mental illnesses. Therefore, for each motion of Baduanjin, 98 "motion data" were captured. The entire dataset of motions included $98 \times 8 = 784$ motion data (760 motions of students and 24 motions of a teacher).

The BVH file output from Perception Neuron 2.0 used the rotation data on skeleton points represented by Euler angles. The data on Euler angles were converted to quaternions to avoid universal joint locking and singularity in the rotation data represented by Euler angles [22]. The quaternion is a four-dimensional super-complex number representing the three-dimensional vector space on the real number [23]. The symbol used to represent a quaternion is as follows:

$$q = [w, x, y, z] \tag{1}$$

In Equation (1), w is the scalar component, and x, y, z are the vectors. Set the rotation order of Euler angles as z, y, x, and set the rotation angles around the x, y, and z axes as α, β, γ, then the Euler angles can be converted to quaternions, as follows:

$$q = \begin{bmatrix} w \\ x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \cos(\gamma/2) \\ 0 \\ 0 \\ \sin(\gamma/2) \end{bmatrix} \begin{bmatrix} \cos(\beta/2) \\ 0 \\ \sin(\beta/2) \\ 0 \end{bmatrix} \begin{bmatrix} \cos(\alpha/2) \\ \sin(\alpha/2) \\ 0 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} \cos(\gamma/2)\cos(\beta/2)\cos(\alpha/2) + \sin(\gamma/2)\sin(\beta/2)\sin(\alpha/2) \\ \cos(\gamma/2)\cos(\beta/2)\sin(\alpha/2) - \sin(\gamma/2)\sin(\beta/2)\cos(\alpha/2) \\ \cos(\gamma/2)\sin(\beta/2)\cos(\alpha/2) + \sin(\gamma/2)\cos(\beta/2)\sin(\alpha/2) \\ \sin(\gamma/2)\cos(\beta/2)\sin(\alpha/2) - \cos(\gamma/2)\sin(\beta/2)\sin(\alpha/2) \end{bmatrix} \quad (2)$$

The motion data obtained from MoCap for a single motion is several thousand frames. Table 1 shows the mean duration and the mean number of frames of Baduanjin captured in the study.

Table 1. The mean duration and the mean number of frames of Baduanjin.

Motion	Mean Duration ($\pm$SD) [1]	Mean Number of Frames ($\pm$SD)
Motion 1	12.13 ± 2.80	1517 ± 350
Motion 2	21.72 ± 4.01	2715 ± 501
Motion 3	16.58 ± 3.78	2073 ± 472
Motion 4	15.09 ± 3.94	1887 ± 492
Motion 5	19.43 ± 4.67	2428 ± 583
Motion 6	16.40 ± 3.86	2050 ± 483
Motion 7	13.17 ± 3.50	1646 ± 438
Motion 8	2.94 ± 1.01	367 ± 126

[1] second.

Due to the limited storage space and bandwidth capacity available to users in actual applications, the large amount of collected motion data may limit its application [24]. The keyframe extraction technique, to extract a small number of representative keyframes from long motion sequences, has been widely used in motion analysis. Therefore, this technique was applied to extract keyframes to reduce the amount of motion data, to improve data storage and subsequent data analysis [24,25]. In order to extract the keyframes on the desired compression rate, a method that extracts the keyframes by a preset compression rate on k-means clustering was used in the study [26]. In this algorithm, Step 1 is to preset the value of k, which is the value of the preset keyframes. In this study, the value of k is determined by the preset compression rate of the extracted keyframes, as follows [27]:

$$k = C\_rate \times N \quad (3)$$

In Equation (3), k is the value of the preset keyframes, C_rate is the compression rate of the keyframes to be extracted, and N is the number of original frames. On the preset value of k, Step 2 uses k-means to extract k cluster centroids from the dataset of 3D coordinates ([x, y, z]) of the skeleton points of the original frame. In k-means, the distance is the applied Euclidean distance in this study. The skeleton model has 17 skeleton points. Therefore, a cluster centroid is composed of 51 (17×3) vectors. Based on these extracted cluster centroids, the keyframes were extracted by calculating the Euclidean distance of skeleton points between the cluster centroids and the original frames. The algorithm to extract the keyframe is shown in Figure 3.
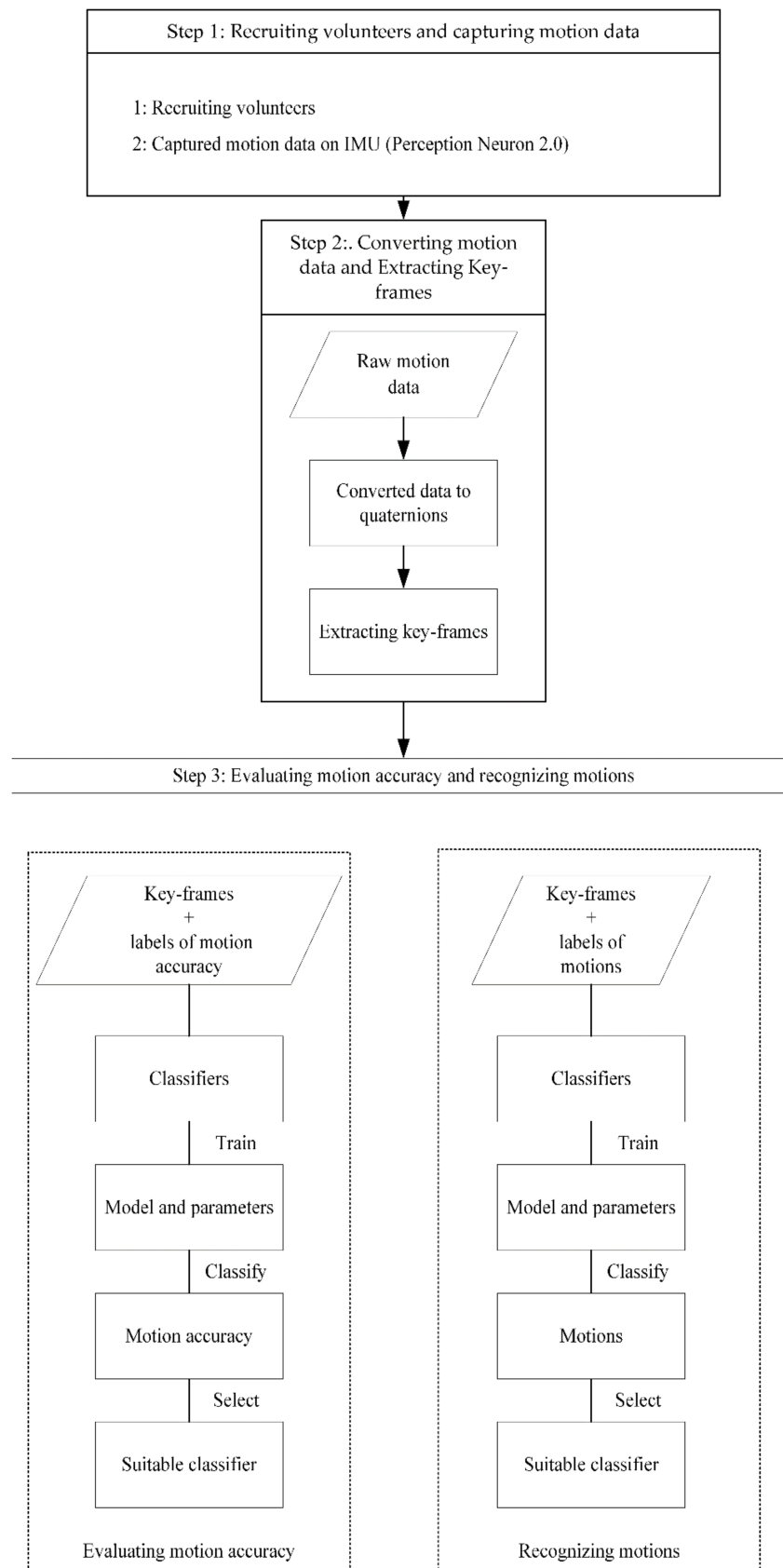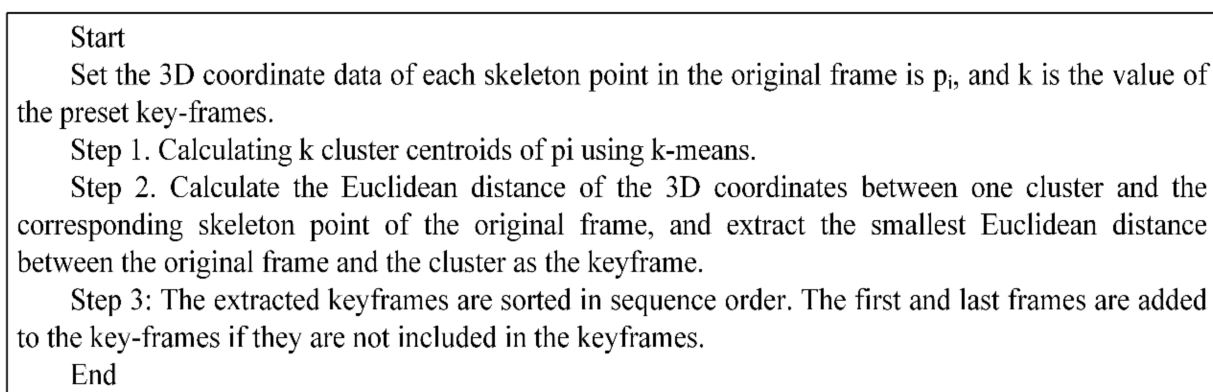
Figure 2. Flow diagram for the study.

> Start
> Set the 3D coordinate data of each skeleton point in the original frame is p_i, and k is the value of the preset key-frames.
> Step 1. Calculating k cluster centroids of pi using k-means.
> Step 2. Calculate the Euclidean distance of the 3D coordinates between one cluster and the corresponding skeleton point of the original frame, and extract the smallest Euclidean distance between the original frame and the cluster as the keyframe.
> Step 3: The extracted keyframes are sorted in sequence order. The first and last frames are added to the key-frames if they are not included in the keyframes.
> End

Figure 3. Algorithm to extract the keyframes from the original frames on the cluster centroids.

The reconstruction error between the reconstructed frames and the original frames was calculated to evaluate the effects on extracting keyframes in this study [28]. Motion reconstruction is based on reconstructing non-keyframes by interpolating adjacent keyframes, thereby rebuilding the same number of frames as the original frames. The interpolation method is: p_1 and p_2 are the positions of the skeleton points of adjacent keyframes at time t_1 and t_2, calculation of p_t (representing the position of the non-key frame point at time t) as follows [29]:

$$
\begin{aligned}
u(t) &= \frac{t_2 - t}{t_2 - t_1}, \\
p_t &= u(t)p_1 + (1 - u(t))p_2, \\
t_1 &< t < t_2
\end{aligned}
\tag{4}
$$

After reconstructing the motion by interpolation, the reconstruction error is calculated as follows:

$$
\begin{aligned}
\text{Dis}(p_1^i - p_2^i) &= \sum_{j=1}^{n} \left\| p_{1,j}^i - p_{2,j}^i \right\|^2, \\
\text{Error}(m_1, m_2) &= \frac{1}{N} \sum_{i=1}^{N} \text{Dis}(p_1^i - p_2^i)
\end{aligned}
\tag{5}
$$

In Equation (5), n represents the number of skeleton points, in the study m = 17; $p_{1,j}^i$ is the coordinate of j skeleton point for i frame in original frames, and $p_{2,j}^i$ is the coordinate of the j skeleton point for the coordinate frames in reconstructed frames. $\text{Dis}(p_1^i - p_2^i)$ represents the calculated distance of the human posture in the i frame between the original frames and reconstructed frames. m_1 is the original frames, m_2 is the reconstruction frames. N represents the number of frames for reconstructed frames or original frames. Error (m_1, m_2) is the reconstructed errors calculated by the distance of human posture.

Five different compression ratios (5%, 10%, 15%, 20%, 25%) are chosen to extract keyframes and calculate the reconstruction errors of the corresponding keyframes under the different compression ratios (Figure 4). This figure shows the average reconstruction error between the extracted keyframes on the corresponding compression rate and the original frames. The Y-axis represents the reconstruction error, and the X-axis represents the five compression ratios (5%, 10%, 15%, 20%, 25%) to the extracted keyframes. The reconstruction errors corresponding to the eight motions of Baduanjin under different compression ratios are represented by different symbols and line segments (as shown in the legend in the figure).

It can be seen from Figure 4 that as the compression rate decreases, the error of motion reconstruction increases. When the compression ratio increases from 15% to 25%, the reconstruction error does not change much. However, the reconstruction error increases significantly when the compression ratio is between 5% and 15%. Therefore, keyframes on the compression ratio (15%) are extracted to ensure that the compression ratio and reconstruction error are reasonable.

Int. J. Environ. Res. Public Health **2022**, 19, 1744

8 of 29
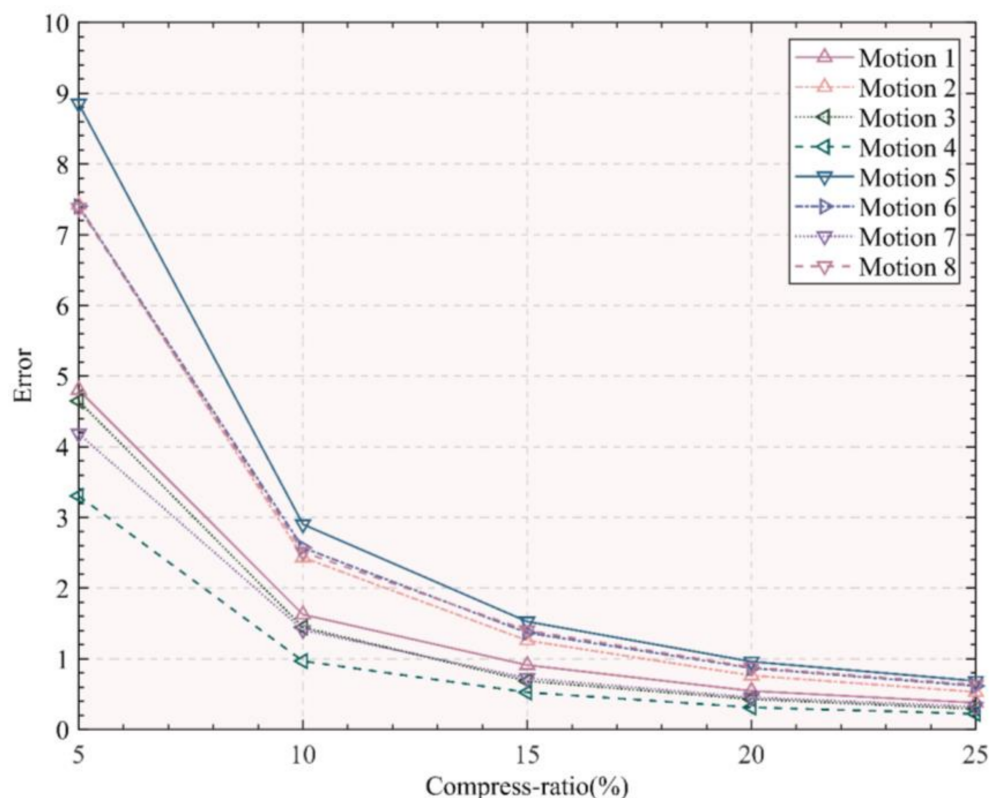
Figure 4. The average reconstruction errors of the corresponding keyframes under the different compression ratios.

### 2.3. Traditional Manual Assessment of Baduanjin

Motion recognition can be regarded as a classification issue of time-varying data, where the test motion is matched with a pre-calibrated motion representing typical motion [30]. Therefore, assessing motion accuracy is similar to an issue for classifying motions. In this study, two martial arts teachers who have more than 10 years of experience teaching Baduanjin from a university in Southwest China were invited to assess the motion accuracy for each student's motions. According to the grading scale of Baduanjin, the two teachers assess the motion accuracy of the students' motions into three grades: fail, pass, and good. The Kendall correlation coefficient test is used to calculate the correlation between the scores results from two teachers to ensure the consistency of the evaluation.

Table 2 shows that the captured motions of students have three different grades (fail, pass, and good) except for Motion-8, with only two grades (good and pass). The two teachers explained that Motion-8 is a simple motion in Baduanjin, which the students easily pass (Figure 5). The Kendall coefficient values between the scores of the two teachers are higher than 0.7, indicating that the two teachers have a high degree of consistency in assessing motion accuracy in Baduanjin. Based on the motion data and the scores of the two teachers, the three sequence-based methods were applied to assess motion accuracy: dynamic time warping (DTW) combined with classifiers, hidden Markov model (HMM) and recurrent neural network (RNN).

Int. J. Environ. Res. Public Health **2022**, 19, 1744

9 of 29

Table 2. The scores of two teachers and Kendall values of the scores.

Motion	Teacher A			Teacher B			Kendall Value
	Good	Pass	Fail	Good	Pass	Fail	
Motion-1	16	57	22	15	55	25	0.941
Motion-2	21	53	21	24	50	21	0.882
Motion-3	26	58	11	23	49	23	0.831
Motion-4	22	58	15	19	49	27	0.824
Motion-5	20	57	18	20	56	19	0.838
Motion-6	23	55	17	20	54	21	0.907
Motion-7	29	59	7	26	62	7	0.944
Motion-8	61	34	0	61	34	0	0.862

Start   Heel up   End
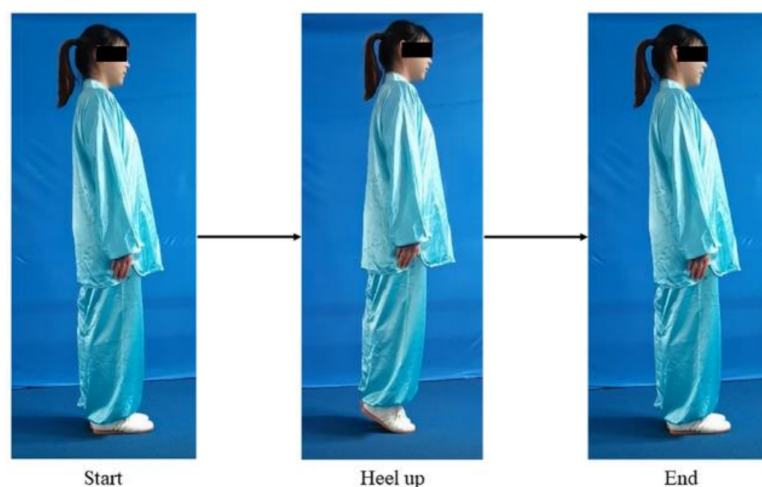
Figure 5. Motion-8.

2.3.1. Dynamic Time Warping (DTW) Combined with Classifiers

In previous studies [12,31], one method assessed the motion accuracy on the differences between the motions of the students and the teacher. The duration of the captured motions was different. DTW, the method used for calculating differences for different time series was applied [32]. Although the difference between motions calculated by DTW can be used to recognise motions, it is difficult to classify the motions based on the difference to assess the motion accuracy without classifiers. Therefore, classifiers were used to classify the motions after calculating the difference between the students' motions and the teacher's motions through DTW. The specific steps are as follows:

- Step 1: to calculate the minimum cumulative distance between corresponding skeleton points between two motions of students and teacher with DTW [13,32]. In order to prevent DTW from mismatching by excessive time warping, the global warping window was limited in this study, which was set to 10% of the entire window range: $0.1 \times \max(n, m)$. n and m are, respectively, the frame length of the two motions. Since the warping path of each corresponding skeleton point of the two motions may be different, the average value of the cumulative distance of the warping path was used to indicate the differences between the points:

$$\text{dis}(q_{stu}^{i}, q_{tea}^{i}) = \frac{\text{DTW}(q_{stu}^{i}, q_{tea}^{i})}{\text{length}(\text{path}^{i})}, i = 1, 2, \ldots, n \tag{6}$$

In Equation (6), stu represents the students' motions; tea represents the teacher's motions; qi is the vectors of the quaternion of the i skeleton point in the two motions, and the total number of points is n. length(path^{i}) is the length of the warping path on the i skeleton point.

Int. J. Environ. Res. Public Health **2022**, *19*, 1744

10 of 29

- Step 2: because the human posture is composed of multiple skeleton points, the overall difference between the two motions from students and the teacher was calculated as distance:

$$D(m_{stu}, m_{tea}) = \frac{\sum\limits_{i=1}^{n} dis(q_{stu}^i, q_{tea}^i)}{n} \qquad (7)$$

- Step 3: after calculating the difference between the teacher's motion and the student's motion, the classifier is trained with the scores of the two teachers to classify motion-for-motion accuracy. A variety of classifiers were selected in the study, namely *k*-nearest neighbour (*k*-NN) [33], support vector machine (SVM) [34], naive Bayes (NB) [35], logistic regression [36], decision tree (DT) [37], and artificial neural network (ANN). Due to the diversity of ANNs, two commonly used ANNs: back propagation neural network (BPNN) and radial basis function neural network (RBFNN) [38], were chosen as classifiers (Figure 6).
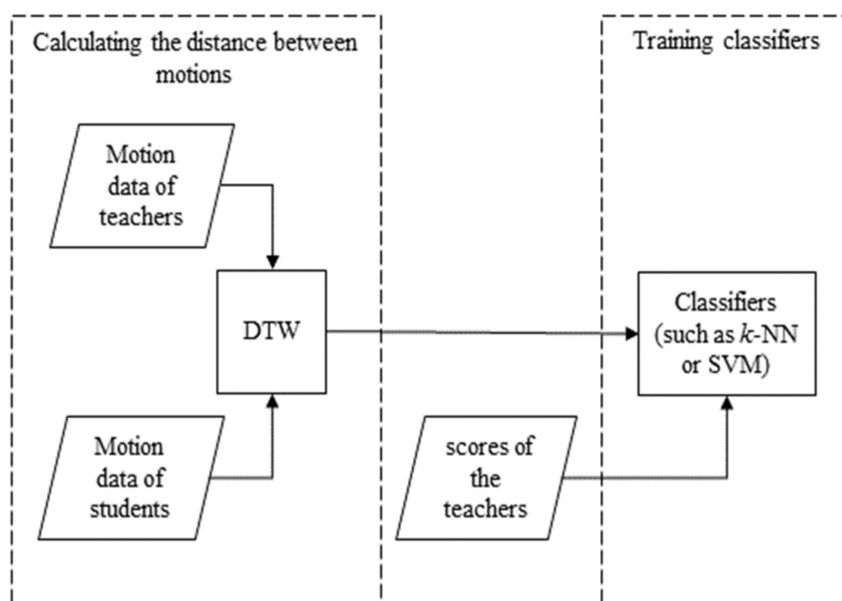
Figure 6. Flow diagram for DTW combined with classifiers.

BPNN is a multi-layer feedforward network trained according to the error backpropagation algorithm. In this study, BPNN was constructed with three layers. The first layer was the input layer. The number of neurons was equivalent to the dimension of the feature vectors; the second layer was the hidden layer, the tangent sigmoid equation was applied as the activation equation:

$$f(x) = \frac{2}{(1 + e^{-2x})^{-1}} \qquad (8)$$

The third layer is the output layer.

The RBFNN is a kind of feedforward network trained using a supervised training algorithm, but the calculation and processing time is lower than that of BPNN. The main advantage of the RBFNN is that it has only one hidden layer and uses the radial basis equation as the activation equation [38]. Unlike BPNN, the Gaussian equation is used as the basic equation in RBFNN, as follows:

$$f(x) = \sum_{i=1}^{M} w_i \exp(\frac{\|x - ci\|^2}{d^2}) \qquad (9)$$

2.3.2. Hidden Markov Model (HMM)

HMM is a sequence-based method that has been successfully applied in recognising human motion [39,40]. HMM is a double random process consisting of a hidden Markov chain and a set of explicit discrete probabilities or probability density functions [40]. HMM can be expressed as follows:

$$\lambda = [N, M, \pi, A, B]$$ (10)

In Equation (8), N presents the set of states in the HMM model, which is the set of possible states of the event. M is the set of possible observation events corresponding to each state, which is the observation set. π is the a priori probability of state occupancy. A is the state transition probability matrix. B is associated with the observations from the various states, which describes the probability of each observation event in each state. HMM can solve three kinds of problems: evaluation, decoding, and learning. In this study, recognising or classifying motions is an evaluation problem [41]. The method supposes the output probability in one observation sequence: $O = o_1, o_2, ..., o_r$ of a given HMM(λ) is: $P(O|\lambda)$. The state with the highest probability in $P(O|\lambda)$, the result of recognition or classification. $P(O|\lambda)$ can be calculated by the forward algorithm, a dynamic programming algorithm that uses a table to store intermediate values when establishing the probability of an observation sequence. $\alpha_t(j)$ in the forward trellis represents the forward probability of the given HMM(λ), which the observed sequence $o_1, o_2, ..., o_r$ at time t and the state is j:

$$\alpha_t(j) = P(o_1, o_2, \ldots, o_t, q_t = j | \lambda)$$ (11)

$\alpha_t(j)$ is calculated by superimposing all possible paths:

$$\alpha_1(j) = \pi_j b_j(o_1), j = (1, 2, \ldots, N)$$ (12)

Initialisation:

$$\alpha_t(j) = \left[\sum_{i=1}^{N} \alpha_{t-1}(i) a_{ij} \right] b_j(o_t), t = (1, 2, \ldots, T), j = (1, 2, \ldots, N)$$ (13)

In Equations (10) and (11), N represents the number of paths from $t - 1$ to t when the state at time t is j. Then:

$$P(O|\lambda) = \sum_{i=1}^{N} \alpha_T(j)$$ (14)

Corresponding to the forward algorithm is the backward algorithm, which is used to train the model. $\beta_t(i)$ represents the backward probability of the given HMM(λ), which the observed sequence $o_{t+1}, o_{t+2}, ..., o_T$ at time t and the state is i:

$$\beta_t(i) = P(o_{t+1}, o_{t+2}, \ldots, o_T | q_t = i, \lambda)$$ (15)

Initialisation:

$$\beta_T(i) = 1 i = (1, 2, \ldots, N)$$ (16)

Recursion:

$$\beta_t(i) = \sum_{j=1}^{N} a_{ij} b_j(o_{t+1}) \beta_{t+1}(j), t = (1, 2, \ldots, T), i = (1, 2, \ldots, N)$$ (17)

2.3.3. Recurrent Neural Network (RNN)

Motion data of humans is a type of sequence data on time. A RNN is a recursive neural network that takes sequence data as input, recurses in the evolution direction of the sequence, and all nodes (recurrent units) are connected in a chain network [42]. Trained by the end-to-end training method, RNN can solve sequence labelling problems with

Int. J. Environ. Res. Public Health **2022**, *19*, 1744

12 of 29

unknown input–output alignment [43]. However, traditional RNNs have the problem of vanishing gradients or exploding gradients. Vanishing gradients cause the parameters to be affected by short-term information, and the earlier information decays with the increase of the time steps. Exploding gradients cause long-term dependence, which refers to the state of the current system that may be affected by the previous state of the system that can lead to information overload [42]. Long short-term memory (LSTM), a type of RNN, is a neural network especially designed to reduce the problem of vanishing or exploding gradients by using gates to obtain relevant information and forget irrelevant information [44]. Therefore, LSTM was used as a method of RNN to assess the motion accuracy in this study.

Figure 7 shows the architecture of LSTM applied in the study. In the figure, X_t is input; h_t is hidden states; Y_t is output; U_t is the cell state, and $\mathfrak{S}$ is the activation function. There are three gates in the hidden state of LSTM: forget, input, and output. The equations in Figure 7 are:

$$
\begin{aligned}
&f_t = \sigma(W_f \bullet [h_{t-1}, x_t] + b_f) && \text{(Forget gate)}\\
&i_t = o_t = \sigma(W_o \bullet [h_{t-1}, x_t] + b_o) && \text{(Input gate)}\\
&\sigma(W_i \bullet [h_{t-1}, x_t] + b_i) && \text{(Output gate)}\\
&\tilde{C}_t = \tanh(W_C \bullet [h_{t-1}, x_t] + b_C)\\
&C_t = f_t * C_{t-1} + i_t * \tilde{C}_t\\
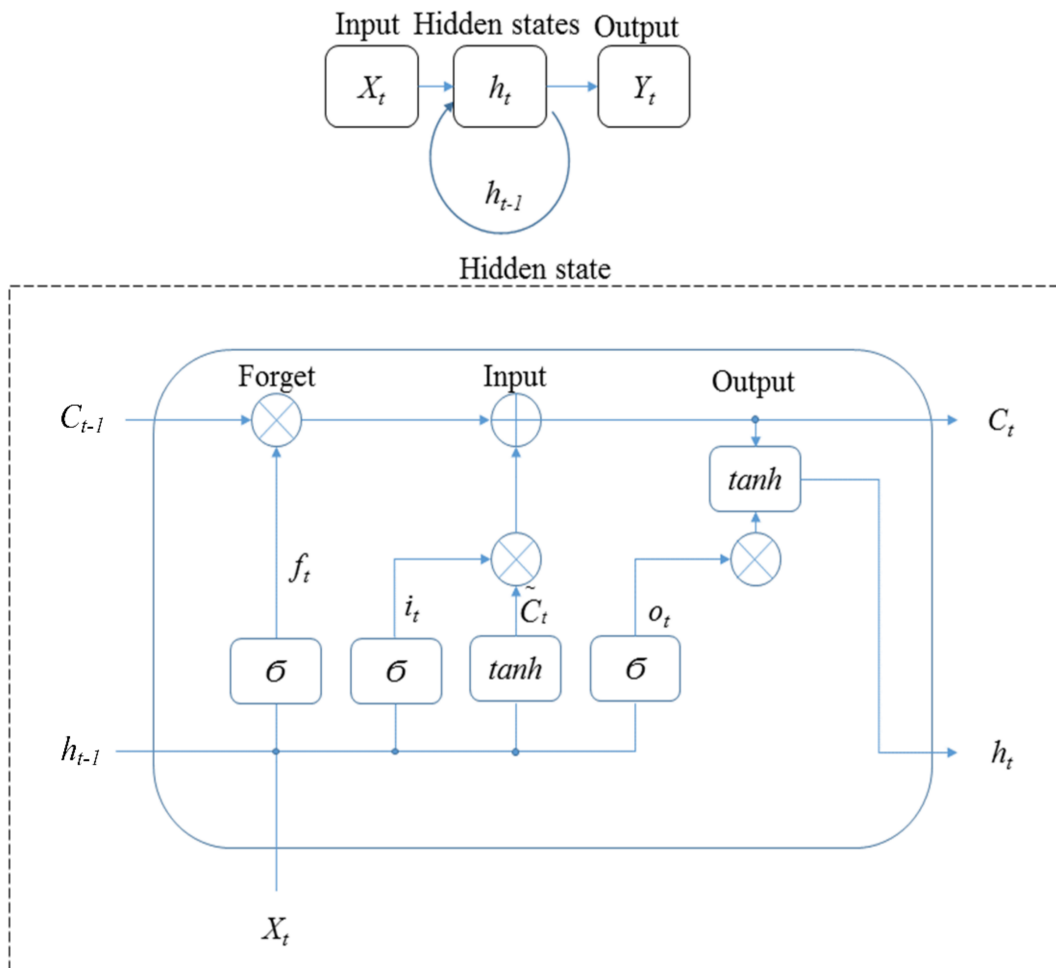&h_t = o_t * \tanh(C_t)
\end{aligned}
\tag{18}
$$

Figure 7. The architecture of LSTM.

Int. J. Environ. Res. Public Health **2022**, *19*, 1744

13 of 29

LSTM remembers and forgets the long sequence input data using the gates. However, LSTM also has problems: whether the gates included in the architecture of LSTM already provide good predictions and whether additional data training is needed to further improve the predictions [45]. Bidirectional LSTM (BiLSTM) is an improvement to this problem of LSTM, which enables additional training by training the input data twice. Based on the traditional LSTM for positive training on a time series from front to back, BiLSTM adds negative training on a time series from back to front [45]. The architecture of BiLSTM is shown in Figure 8.
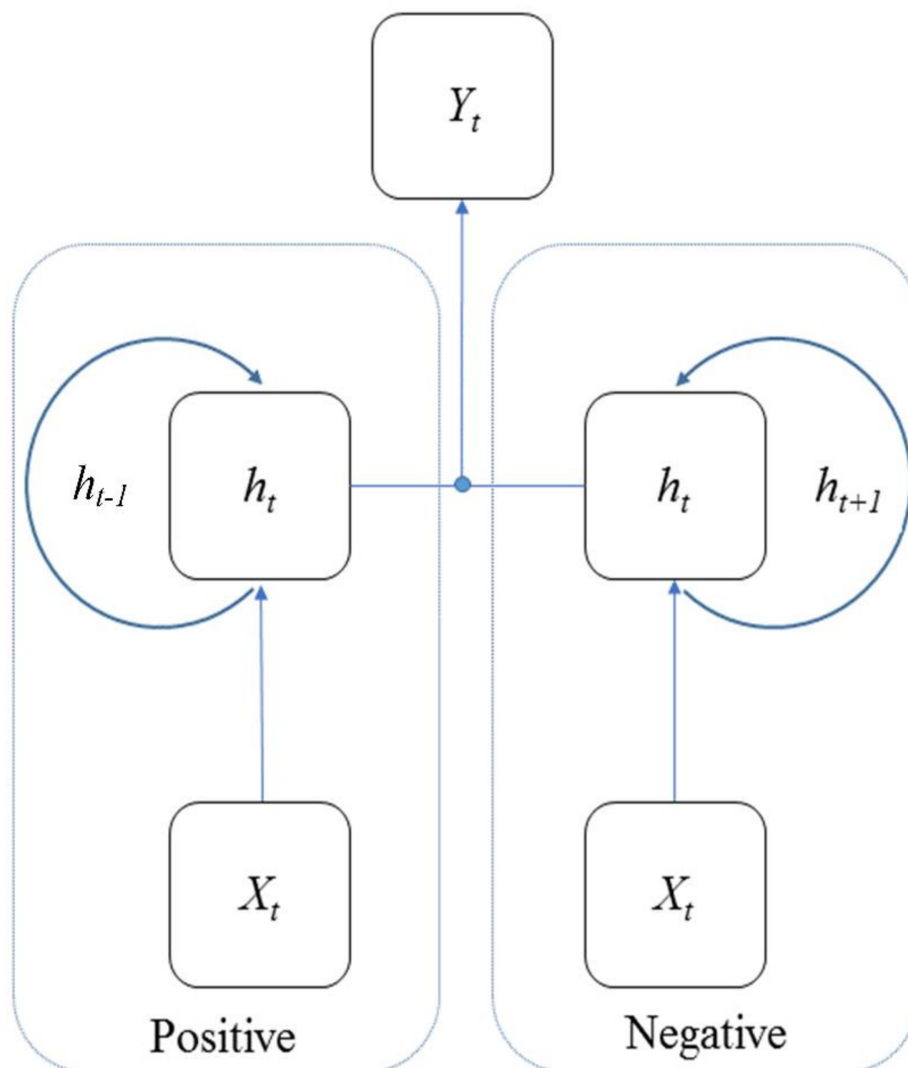
Figure 8. The architecture of BiLSTM. In the figure, the hidden units in BiLSTM are the same as in LSTM (see Figure 7).

Cho et al., in 2014, proposed a model similar to LSTM but with simpler calculations and implementations, namely the gated recurrent unit (GRU) [46]. Similar to LSTM, GRU is a positive training model based on time series [47]. However, the architecture of the hidden state in GRU is different from that of LSTM. The architecture of the hidden state in GRU applied in this study is shown in Figure 9.

In Figure 9, X_t is input; h_t is hidden states; Y_t is output, and δ is activation function. There are two gates in the hidden state of LSTM. The equations in Figure 9 are:

$$
\begin{aligned}
r_t &= \sigma(W_r \bullet [h_{t-1}, x_t] + b_r) & \text{(Reset gate)}\\
z_t &= \sigma(W_z \bullet [h_{t-1}, x_t] + b_z) & \text{(Update gate)}\\
\widetilde{h}_t &= \tanh(W_{\widetilde{h}} \bullet [r_t * h_{t-1}, x_t] + b_{\widetilde{h}})\\
h_t &= (1 - z_t) * h_{t-1} + z_t * \widetilde{h}_t
\end{aligned}
\tag{19}
$$

From Figure 7 to Figure 9, the hidden state of LSTM and GRU is similar, in that the output at time t is calculated using the hidden state at time $t - 1$ and the value of input at time t. In addition, the equation of the forget gate for LSTM is similar to the equation of the reset gate for GRU. The difference is that LSTM has three gates, whereas GRU has two gates. There is no output gate in GRU as in LSTM, which means GRU directly transmitted the memory to the next cell, but LSTM selects whether to transmit memory by the output gate.
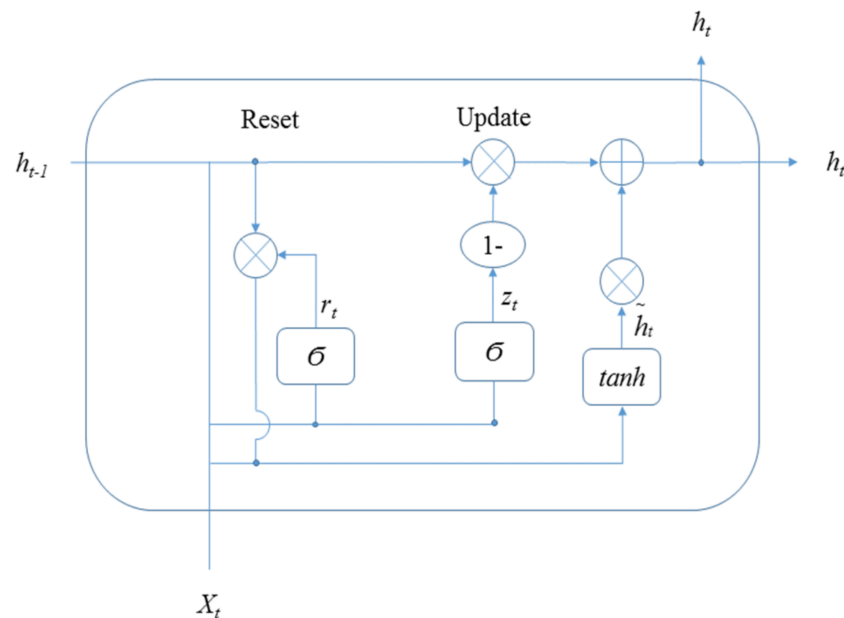
Figure 9. The typical architecture of the hidden state in GRU.

2.3.4. Evaluating the Effectiveness of Methods

In this study, cross-validation is applied to evaluate the effectiveness of different sequence-based methods for assessing motion accuracy. In cross-validation, the dataset is divided into a training set and a test set: (1) the training set is used to train the model and (2) the test set is used to test the trained model. The 10-fold cross-validation was used in the study. The confusion matrix and the assessment accuracy rate were used to express the effectiveness of different sequence-based methods. The confusion matrix shows the degree of confusion in the classification between different motions. The assessment accuracy is calculated as:

$$
\text{Assessment Accuracy} = \frac{\text{Number of correctly classified}}{\text{Overall sample size}}
\tag{20}
$$

3. Results

The assessment of motion accuracy used captured motion data and the traditional manual scoring results from the two teachers. With 10-fold cross-validation, the models on the sequence-based methods were trained. Using the scoring results from Teacher A as a standard, the accuracies of different sequence-based methods (on assessing motion

accuracy) are shown in Table 3. Figure 10 shows an example of the confusion matrices for the sequence-based methods for Motion-1.

From Table 3, the results show that none of the methods had the highest accuracy for all eight motions of Baduanjin. The DTW + k-NN method had the highest accuracy in assessing five motions (Motion-1, Motion-2, Motion-5, Motion-7, and Motion-8). The highest accuracy for Motion-3 was the LSTM method, reaching 80.00%, the highest accuracy for Motion-4 was HMM, and BiLSTM with 90.53%. Lastly, the highest accuracy of Motion-6 was LSTM, with 84.21%. The highest average accuracy was the DTW + k-NN with 84.21%. From Figure 10, it can be seen from the confusion matrix of Motion-1 that, for most methods, the errors in classifying occur mainly in pass and fail motions. DTW + SVM and DTW + logistic regression classified all fail motions as pass motions.

Table 3. The accuracies of assessing the motion accuracy of different sequence-based methods using the scoring results from Teacher A.

Methods	Accuracy (%)								
	Motion-1	Motion-2	Motion-3	Motion-4	Motion-5	Motion-6	Motion-7	Motion-8	Mean
DTW + k-NN	94.74 [1]	86.32 [1]	77.90	80.00	84.21 [1]	77.90	87.37 [1]	85.26 [1]	84.21 [1]
DTW + SVM	66.32	62.11	69.47	74.74	63.16	65.26	69.47	78.95	68.68
DTW + NB	77.90	72.63	74.74	84.21	65.26	70.53	70.53	74.74	73.82
DTW + Logistic regression	66.32	63.16	67.37	73.68	63.16	63.16	66.32	74.74	67.24
DTW + DT	69.47	63.16	82.11	70.53	67.37	68.42	74.74	69.47	70.66
DTW + BPNN	85.26	71.58	71.58	73.68	66.32	67.37	69.47	84.21	73.68
DTW + RBFNN	89.47	84.21	72.63	75.79	80.00	81.05	82.11	83.16	81.05
HMM	84.21	80.00	78.95	90.53 [1]	76.84	78.95	83.16	77.90	81.32
LSTM	75.79	77.90	82.11 [1]	84.21	72.63	84.21 [1]	78.95	78.95	79.34
BiLSTM	84.21	80.00	78.95	90.53	76.84	78.95	83.16	77.90	81.32
GRU	80.00	75.79	67.37	83.16	74.74	81.05	82.11	72.63	77.11

[1] The highest accuracy.

When using the scoring results from Teacher B as a standard, the accuracy of different sequence-based methods on assessing motion accuracy is shown in Table 4. The confusion matrices for the sequence-based methods with Motion-1 is used as an example in Figure 11.

Table 4. The accuracies of assessing the motion accuracy on different sequence-based methods using the scoring result from Teacher B.

Methods	Accuracy (%)								
	Motion-1	Motion-2	Motion-3	Motion-4	Motion-5	Motion-6	Motion-7	Motion-8	Mean
DTW + k-NN	92.63 [1]	77.89 [1]	77.90	80.00 [1]	83.16	83.16 [1]	86.32 [1]	83.16	83.03 [1]
DTW + SVM	66.31	60.00	69.47	69.47	66.32	65.26	72.63	77.90	68.42
DTW + NB	75.79	73.68	71.58	71.58	74.74	75.79	74.74	67.37	73.16
DTW + Logistic regression	64.21	61.05	61.05	68.42	61.05	62.11	70.53	71.58	65.00
DTW + DT	67.37	62.11	60.00	70.53	76.84	76.84	73.68	81.05	71.05
DTW + BPNN	68.42	62.11	66.32	65.26	66.32	54.74	71.58	78.95	66.71
DTW + RBFNN	78.95	74.74	76.84	62.11	86.32 [1]	78.95	86.32 [1]	83.16	78.42
HMM	83.16	73.68	80.00 [1]	80.00[1]	77.90	76.84	82.11	85.26	79.87
LSTM	76.84	71.58	77.90	75.79	76.84	82.11	84.21	82.11	78.42
BiLSTM	82.11	75.79	78.95	74.74	76.84	82.11	83.16	85.26	79.87
GRU	76.84	67.37	69.47	71.58	75.79	78.95	77.90	88.42 [1]	75.79

[1] The highest accuracy.

From Table 4, the results show that, in assessing motion accuracy, not one method had the highest accuracy for all eight motions of Baduanjin. However, DTW + k-NN had the highest accuracy in assessing five motions (Motion-1, Motion-2, Motion-4, Motion-6, and Motion-7), with its mean accuracy reaching 83.03%, which is the highest compared to other methods. The highest accuracy for Motion-3 was HMM, at 80.00%, and the highest accuracy for motion-5 was DTW + RBFNN at 86.32%. The highest accuracy of Motion-8 was the GRU at 88.42%. On the confusion matrix of Motion-1 (Figure 11), the result is similar to Figure 10 when using Teacher A as the standard. The classification errors occur mainly in pass and fail, especially DTW + SVM and DTW + logistic regression.
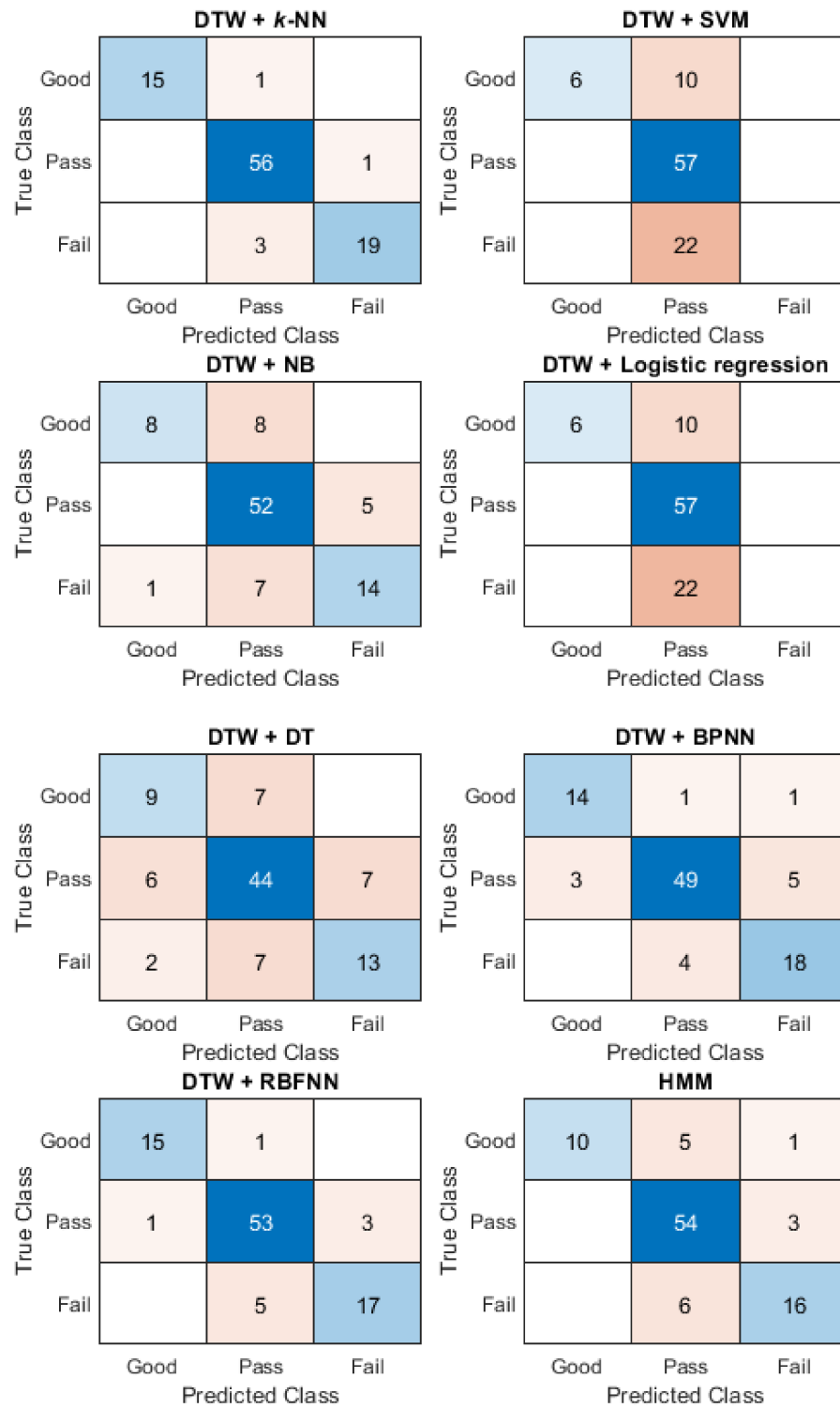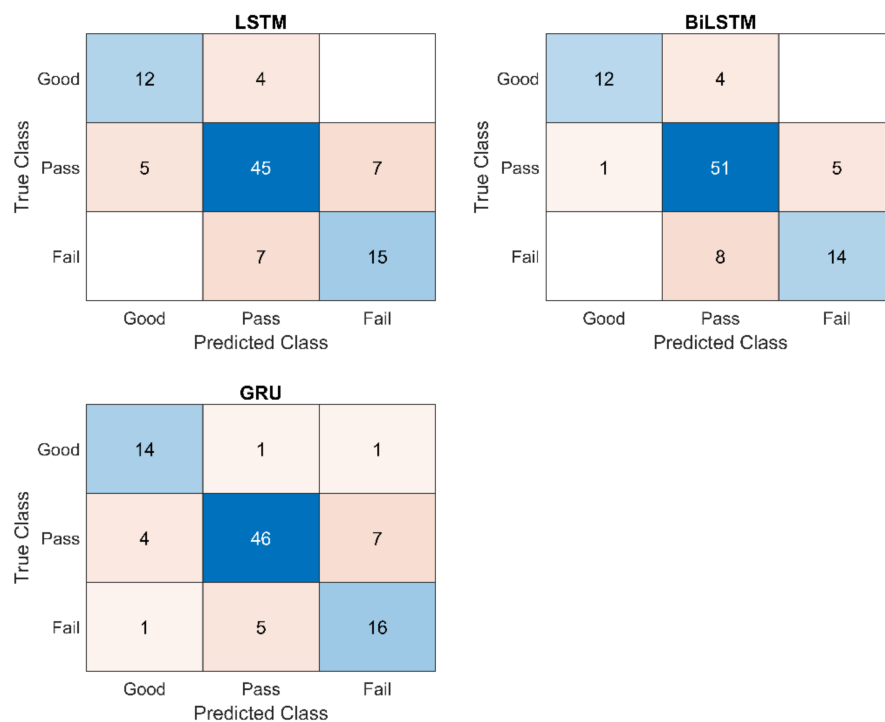
Figure 10. *Cont.*

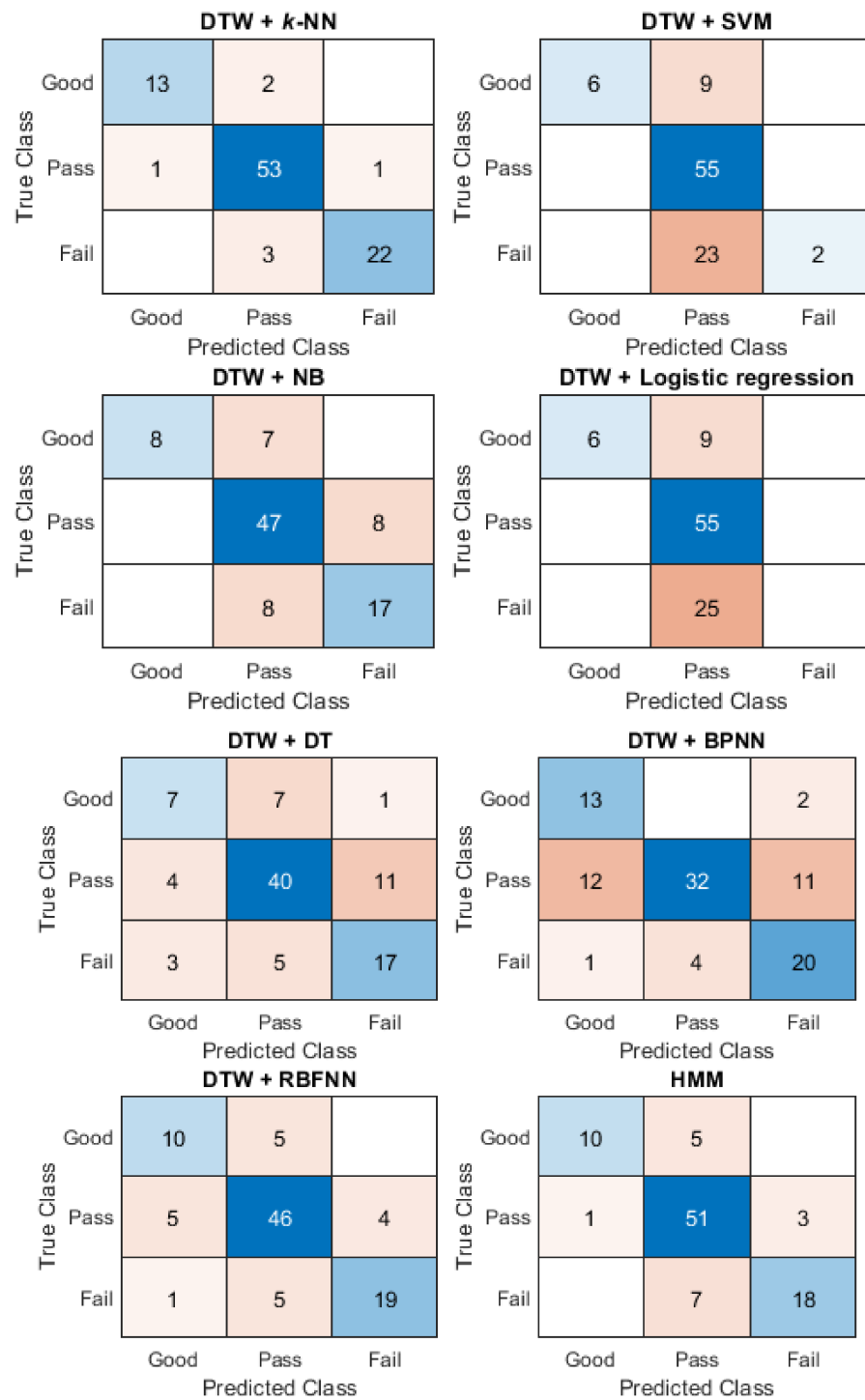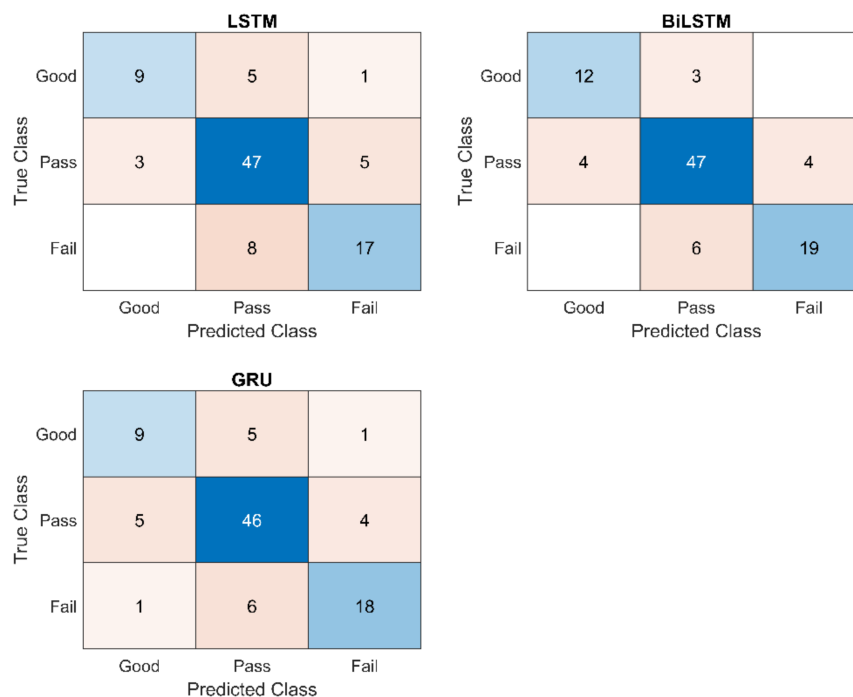*Int. J. Environ. Res. Public Health* **2022**, *19*, 1744

17 of 29

Figure 10. The confusion matrices of the different sequence-based methods for Motion-1 using the scoring results from Teacher A.

The processing time (training and classifying) of each method is presented in Table 5.

Table 5. Processing times (training and classifying) of the different sequence-based methods for assessing motion accuracy.

Methods	Processing Time (Seconds)
DTW + *k*-NN	3.810 [1]
DTW + SVM	4.119
DTW + NB	4.057
DTW + Logistic regression	4.382
DTW + DT	3.947
DTW + BPNN	14.830
DTW + RBFNN	3.898
HMM	4.119
LSTM	14.132
BiLSTM	27.995
GRU	11.943

[1] Minimum processing time.

The results show that the processing time of DTW + *k*-NN was the shortest (3.810 s). The processing times of the three recurrent neural network methods (LSTM, BiLSTM, and GRU) were much longer, all exceeding 15 s.

Figure 11. *Cont.*

Int. J. Environ. Res. Public Health **2022**, *19*, 1744

19 of 29

Figure 11. The confusion matrices of the different sequence-based methods for Motion-1 using the scoring result from Teacher B.

Recognising Motions of Baduanjin

For recognising motion with 10-fold cross-validation, the accuracy and processing time of the sequence-based methods are shown in Table 6. The confusion matrices of different sequence-based methods for motion recognition are shown in Figure 12.
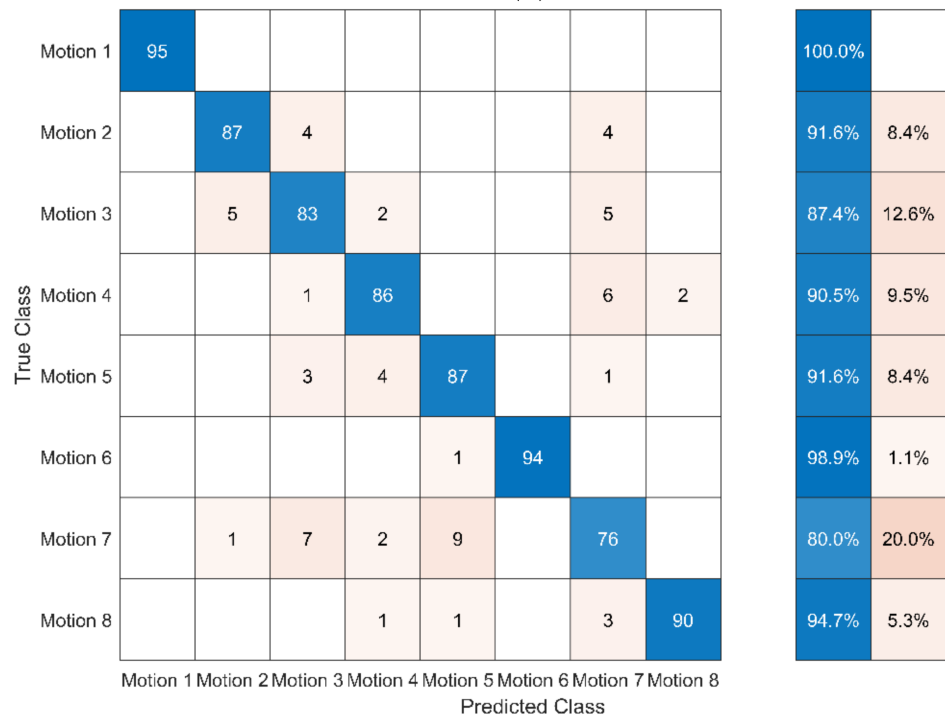

(**a**)

Figure 12. *Cont*.

Int. J. Environ. Res. Public Health **2022**, *19*, 1744

20 of 29

(**b**)

(**c**)

Figure 12. *Cont.*

Int. J. Environ. Res. Public Health **2022**, 19, 1744

21 of 29

(**d**)

(**e**)

Figure 12. *Cont.*

Int. J. Environ. Res. Public Health **2022**, *19*, 1744

22 of 29

(**f**)

(**g**)

Figure 12. *Cont.*

Int. J. Environ. Res. Public Health **2022**, *19*, 1744

23 of 29

(**h**)

(**i**)

Figure 12. *Cont*.

Int. J. Environ. Res. Public Health **2022**, *19*, 1744

24 of 29

(j)

(k)

Figure 12. The confusion matrices for recognizing motion on different sequence-based methods: (**a**) DTW + *k*-NN; (**b**) DTW + SVM; (**c**) DTW + NB; (**d**) DTW + Logistic regression; I DTW + DT; (**f**) DTW + BPNN; (**g**) DTW + RBFNN; (**h**) HMM; (**i**) LSTM; (**j**) BiLSTM; (**k**) GRU.

Table 6. Accuracy and processing time (training and classifying) of different sequence-based methods for recognising motion.

Methods	Accuracy (%)	Processing Time (Seconds)
DTW + *k*-NN	99.47	3.823 [2]
DTW + SVM	99.61 [1]	6.909
DTW + NB	91.84	6.757
DTW + Logistic regression	94.21	10.163
DTW + DT	93.68	4.809
DTW + BPNN	91.05	24.665
DTW + RBFNN	75.79	5.439
HMM	99.08	61.144
LSTM	96.45	123.477
BiLSTM	97.37	239.190
GRU	97.50	106.513

[1] The highest accuracy. [2] Minimum processing time.

The results show that, by using sequence-based methods to recognise the motions of Baduanjin, the DTW method combined with the classifier will improve their accuracy (except for DTW + RBFNN). It can be seen for the classifier *k*-NN and SVM, the accuracy of both methods exceeded 99.00% (DTW + *k*-NN reached 99.47% and DTW + SVM reached 99.61%). It was the highest accuracy among all verification methods tested in this paper. However, the accuracy of DTW + RBFNN was not satisfactory, only reaching 75.79%. Besides DTW + BFNN, the processing time of DTW combined with several classifiers was between 3.823 and 10.163 s. The processing time for DTW + *k*-NN was the minimum of all methods at 3.823 s. The accuracies of HMM and RNN (LSTM, BiLSTM, and GRU) also exceeded 96%. However, the processing times of these four methods were relatively high (between 61.144 and 239.190 s), especially LSTM, BiLSTM, and GRU. These three methods had much higher processing times compared to DTW combined with classifiers. BiLSTM, with the maximum processing time, was many times longer than DTW + *k*-NN with the minimum processing time.

The Chi-Square Test of the High Accuracy Methods on Recognising Motions

From the confusion matrices, it was found that the accuracy of DTW + *k*-NN, DTW + SVM, and HMM were all higher than 99%. The chi-square test on the number of correct and incorrect judged motions by the three methods were obtained, and the results are shown in Tables 7 and 8 (calculation on SPSS 23.0).

Table 7. The number of correct and incorrect judged motions by DTW + *k*-NN, DTW + SVM, and HMM.

Methods	Recognised Motions		Total
	Correct	Incorrect	
DTW + *k*-NN	756	4	760
DTW + SVM	757	3	760
HMM	753	7	760

Table 8. The chi-square test results of DTW + *k*-NN, DTW + SVM, and HMM.

	Value	Degree of Freedom	Asymptotic Significance (2-Sided)	Exact Sig (2-Sided)
Pearson chi-square	1.869 [1]	2	0.393	0.497
Likelihood ratio	1.804	2	0.406	0.497
Fisher's exact test	1.722	–	–	0.497

[1] 3 cells (50.0%) have expected count less than 5. The minimum expected count is 4.67.

From the results of the chi-square test, it is found that the three methods have no significant difference in the accuracy of recognising motions. Therefore, the effectiveness of the three methods on accuracy can be considered the same.

Second, the incorrect motions recognised by the three methods are not the same. For example, for DTW + SVM, 3 of the 706 motions were misclassified, including Motion-3, Motion-5, and Motion-7. All three motions were incorrectly recognised as Motion-4. However, for HMM, the problem was in recognising the motions in Motion-7 and Motion-8 as other motions.

4. Discussion

This research used IMU to capture motion data to assess motion accuracy and recognise Baduanjin motions that can be used to assist students in knowing their errors during practice. To assess motion accuracy, previous studies have suggested and verified that motion accuracy could be assessed by comparing the motions to be assessed with the standard motions [12,31]. For example, in a study involving assessing the motion accuracy of tai chi, DTW was used to calculate the differences between the motions of teachers and students to assess the motion accuracies of the students' motions [12]. The correlation between the assessment results of this method and the assessment results between teachers and students reached 80%. This method assesses the motion accuracy by calculating the difference or similarity between the motions, in which this method needs a quantitative assessment for the accuracy of motions. However, in teaching and learning, the traditional assessment for the accuracy of motions is usually a manually graded assessment, a qualitative assessment, such as the existing method in the course of Baduanjin in universities. Therefore, in this study, the motion accuracy of Baduanjin was assessed as a classification problem.

In this study, three types of sequence-based methods were applied, and the traditional manual assessment results of teachers were used as the classification criteria to train the classifiers. It is proven that the results of some classifiers have the ability to approach teacher assessment scoring results. However, none of the classifiers have the highest accuracies in assessing the motion accuracies of all eight motions of Baduanjin. Among the classifiers, with Teacher A as the classification criteria, DTW + *k*-NN reached the highest accuracies in five motions (Motion-1, Motion-2, Motion-5, Motion-7, and Motion-8). However, there were some differences in the results using Teacher B as the classification criteria. Among the classifiers, DTW + *k*-NN reached the highest accuracies in five motions (Motion-1, Motion-2, Motion-4, Motion-6, and Motion-7). Although no selected classifier had the best accuracy for all eight motions of Baduanjin, the highest average accuracy was DTW + *k*-NN, which also had the lowest average processing time among the selected classifiers.

In addition, the motion recognition of Baduanjin was also studied. In actual applications, the motion can be recognised, and the motion accuracy assessed using the corresponding optimal classifier. The results of motion recognition show that the selected classifiers reached high accuracy (all above 90%), except for DTW + RBFNN (75.79%). The accuracy of the three classifiers (DTW + *k*-NN, DTW + SVM, and HMM) exceeded 99%. From the confusion matrix of the three classifiers and the chi-square test performed thereon, there was no significant difference in the accuracy of the three types of classifiers, indicating that the accuracies of the three classifiers for motion recognition under the existing experimental data were the same. Therefore, as the shortest processing time among the three classifiers, DTW + *k*-NN is concluded as the best choice for motion recognition of Baduanjin.

There is a limitation in this study: the captured motions used for assessing motion accuracies were relatively small, which could be why the accuracies of the three types of RNN (LSTM, BiLSTM, and GRU) were not high in assessing the motion accuracy. It is expected that by increasing the motion dataset in further research, the accuracy of RNN will be increased [48].

5. Conclusions

This study shows that, on the motion data captured by IMU, the sequence-based method to classify the Baduanjin motions can assess the motion accuracies of the Baduanjin motions and recognise the motions with high accuracy and short processing times. The methods verified in this study could be used to assess the motions of Baduanjin in PE. Based on the verified sequence-based method in this study, a formative assessment system for Baduanjin in PE can be developed.

References

1. Ministry of Education of People's Republic of China. The Guidelines of Physical Education in Colleges and Universities. Available online: http://www.moe.gov.cn/s78/A10/moe_918/tnull_8465.html (accessed on 6 August 2020).
2. Ministry of Education of People's Republic of China. 21st Century Action Plan for Invigorating Education. Available online: http://old.moe.gov.cn/publicfiles/business/htmlfiles/moe/s6986/200407/2487.html (accessed on 11 October 2020).
3. Ministry of Education of the People's Republic of China. 2019 National Education Development Statistical Bulletin. Available online: http://www.gov.cn/xinwen/2020-05/20/content_5513250.htm (accessed on 11 October 2020).
4. Department of Education of Jiangsu. Notice of Provincial Education Department on the Results of the Fifth Batch of Public Physical Education Courses in Colleges and Universities. Available online: https://wenku.baidu.com/view/bb4c0d8f876fb8 4ae45c3b3567ec102de2bddf83.html (accessed on 7 August 2020).
5. Zhan, Y.Y. Exploring a New System of Martial Arts Teaching Content in Common Universities in Shanghai. Master's thesis, East China Normal University, Shanghai, China, 2015.
6. Elaoud, A.; Barhoumi, W.; Zagrouba, E.; Agrebi, B. Skeleton-based comparison of throwing motion for handball players. *J. Ambient. Intell. Humaniz. Comput.* **2019**, *11*, 419–431. [CrossRef]
7. Yamaoka, K.; Uehara, M.; Shima, T.; Tamura, Y. Feedback of flying disc throw with Kinect and its evaluation. *Procedia Comput. Sci.* **2013**, *22*, 912–920. [CrossRef]
8. Spörri, J.; Schiefermüller, C.; Müller, E. Collecting kinematic data on a ski track with optoelectronic stereophotogrammetry: A methodological study assessing the feasibility of bringing the biomechanics lab to the field. *PLoS ONE* **2016**, *11*, e0161757. [CrossRef]
9. Thomsen, A.S.S.; Bach-Holm, D.; Kjærbo, H.; Højgaard-Olsen, K.; Subhi, Y.; Saleh, G.M.; Park, Y.S.; la Cour, M.; Konge, L. Operating room performance improves after proficiency-based virtual reality cataract surgery training. *Ophthalmology* **2017**, *124*, 524–531. [CrossRef] [PubMed]
10. Schuler, N.; Bey, M.; Shearn, J.; Butler, D. Evaluation of an electromagnetic position tracking device for measuring in vivo, dynamic joint kinematics. *J. Biomech.* **2005**, *38*, 2113–2117. [CrossRef]
11. Van der Kruk, E.; Reijne, M.M. Accuracy of human motion capture systems for sport applications; state-of-the-art review. *Eur. J. Sport Sci.* **2018**, *18*, 806–819. [CrossRef]
12. Chen, X.M.; Chen, Z.B.; Li, Y.; He, T.Y.; Hou, J.H.; Liu, S.; He, Y. ImmerTai: Immersive motion learning in VR environments. *J. Vis. Commun. Image Represent.* **2019**, *58*, 416–427. [CrossRef]

Int. J. Environ. Res. Public Health **2022**, *19*, 1744

28 of 29

13. Li, H.; Selina, K.; Yap, H.J. Differences in Motion Accuracy of Baduanjin between Novice and Senior Students on Inertial Sensor Measurement Systems. *Sensors* **2020**, *20*, 6258. [CrossRef] [PubMed]

14. Mannini, A.; Sabatini, A.M. Machine Learning Methods for Classifying Human Physical Activity from On-Body Accelerometers. *Sensors* **2010**, *10*, 1154–1175. [CrossRef] [PubMed]

15. Carmona, J.M.; Climent, J. A Performance Evaluation of HMM and DTW for Gesture Recognition. In Proceedings of the Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications, Buenos Aires, Argentina, 3–6 September 2012; pp. 236–243.

16. Wang, X.; Gao, L.; Song, J.; Shen, H. Beyond Frame-level CNN: Saliency-Aware 3-D CNN With LSTM for Video Action Recognition. *IEEE Signal Process. Lett.* **2017**, *24*, 510–514. [CrossRef]

17. Wang, H.; Wang, L. Learning content and style: Joint action recognition and person identification from human skeletons. *Pattern Recognit.* **2018**, *81*, 23–35. [CrossRef]

18. Noitom Technology. *Perception Neuron 2.0.*; Noitom Technology: Beijing, China. Available online: https://www.noitom.com.cn/perception-neuron-2-0.html (accessed on 20 August 2020).

19. Li, H.; Yap, H.J.; Selina, K. Motion Classification and Features Recognition of a Traditional Chinese Sport (Baduanjin) Using Sampled-Based Methods. *Appl. Sci.* **2021**, *11*, 7630. [CrossRef]

20. Sers, R.; Forrester, S.E.; Moss, E.; Ward, S.; Zecca, M. Validity of the Perception Neuron inertial motion capture system for upper body motion analysis. *Measurement* **2019**, *149*, 107024. [CrossRef]

21. Srivastava, R.; Sinha, P. Hand Movements and Gestures Characterization Using Quaternion Dynamic Time Warping Technique. *IEEE Sens. J.* **2016**, *16*, 1333–1341. [CrossRef]

22. Yap, H.J.; Taha, Z.; Dawal, S.Z.M.A. A generic approach of integrating 3D models into virtual manufacturing. *J. Zhejiang Univ. -SCIENCE C (Comput. Electron.)* **2012**, *13*, 22–30. [CrossRef]

23. Mukundan, R. Quaternions: From classical mechanics to computer graphics, and beyond. In Proceedings of the The 7th Asian Technology Conference in Mathematics, Melaka, Malaysia, 17–21 December 2002; pp. 97–105.

24. Kim, M.H.; Chau, L.P.; Siu, W.C. Keyframe selection for motion capture using motion activity analysis. In Proceedings of the 2012 IEEE International Symposium on Circuits and Systems (ISCAS), Seoul, Korea, 20–23 May 2012; pp. 612–615.

25. Yan, C.; Qiang, W.; He, X.J. *Multimedia Analysis, Processing and Communications*; Springer:Berlin/Heidelberg, Germany, 2011; pp. 535–542.

26. Shi, X.B.; Liu, S.P.; Zhang, D.Y. Human action recognition method based on key frames. *J. Syst. Simul.* **2015**, *27*, 2401–2408. [CrossRef]

27. Zhang, Y.; Cao, J. 3D Human Motion Key-Frames Extraction Based on Asynchronous Learning Factor PSO. In Proceedings of the 2015 Fifth International Conference on Instrumentation and Measurement, Computer, Communication and Control (IMCCC), Qinhuangdao, China, 18–20 September 2015; pp. 1617–1620.

28. Li, S.Y.; Hou, J.; Gan, L.Y. Extraction of motion key-frame based on inter-frame pitch. *Comput. Eng.* **2015**, *41*, 242–247. [CrossRef]

29. Liu, X.M.; Hao, A.M.; Zhao, D. Optimization-based key frame extraction for motion capture animation. *Vis. Comput.* **2013**, *29*, 85–95. [CrossRef]

30. Cai, M. *3D Hunam Motion Analysis and Action Recognition*; Central South University: Changsha, China, 2013.

31. Alexiadis, D.S.; Daras, P. Quaternionic signal processing techniques for automatic evaluation of dance performances from MoCap data. *IEEE Trans. Multimed.* **2014**, *16*, 1391–1406. [CrossRef]

32. Keogh, E.; Ratanamahatana, C.A. Exact indexing of dynamic time warping. *Knowl. Inf. Syst.* **2005**, *7*, 358–386. [CrossRef]

33. Altun, K.; Barshan, B.; Tunel, O. Comparative study on classifying human activities with miniature inertial and magnetic sensors. *Pattern Recognit.* **2010**, *43*, 3605–3620. [CrossRef]

34. Jegham, I.; Khalifa, A.B.; Alouani, I.; Mahjoub, M.A. Vision-based human action recognition: An overview and real world challenges. *Digit. Investig.* **2020**, *32*, 200901. [CrossRef]

35. Iglesias, J.A. Creating Evolving User Behavior Profiles Automatically. *IEEE Trans. Knowl. Data Eng.* **2012**, *24*, 854–867. [CrossRef]

36. Edgar, T.W.; Manz, D.O. *Research Methods for Cyber Security*; Syngress: Houston, TX, USA, 2017.

37. Lin, T. Code comment analysis for improving software quality. In *The Art and Science of Analyzing Software Data*; Bird, C., Menzies, T., Zimmermann, T., Eds.; Morgan Kaufmann: San Mateo, CA, USA, 2015; pp. 493–517.

38. Satapathy, S.K.; Dehuri, S.; Jagadev, A.K.; Mishra, S. EEG Signal Classification Using RBF Neural Network Trained With Improved PSO Algorithm for Epilepsy Identification. In *EEG Brain Signal Classification for Epileptic Seizure Disorder Detection*; Satapathy, S.K., Dehuri, S., Jagadev, A.K., Mishra, S., Eds.; Academic Press: Cambridge, MA, USA, 2019; pp. 67–89.

39. Li, S.; Ferraro, M.; Caelli, T.; Pathirana, P.N.J.I.J.o.B.; Informatics, H. A Syntactic Two-Component Encoding Model for the Trajectories of Human Actions. *IEEE J. Biomed. Health Inform.* **2014**, *18*, 1903–1914. [CrossRef]

40. Mannini, A.; Sabatini, A.M. Gait phase detection and discrimination between walking-jogging activities using hidden Markov models applied to foot motion data from a gyroscope. *Gait Posture* **2012**, *36*, 657–661. [CrossRef] [PubMed]

41. Jurafsky, D.S.; Martin, J.H. *Speech and Language Processing*; Chapman & Hall: London, UK, 2000.

42. Goodfellow, I.; Bengio, Y.; Courville, A. *Regularization for Deep Learning*; MIT Press: Cambridge, MA, USA, 2016; Volume 1.

43. Graves, A.; Mohamed, A.-r.; Hinton, G. Speech Recognition with Deep Recurrent Neural Networks. In Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal, Vancouver, BC, Canada, 26–31 May 2013; pp. 6645–6649.

Int. J. Environ. Res. Public Health **2022**, *19*, 1744

29 of 29

44. Donahue, J.; Hendricks, L.A.; Rohrbach, M.; Venugopalan, S.; Guadarrama, S.; Saenko, K.; Darrell, T. Long-term Recurrent Convolutional Networks for Visual Recognition and Description. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 677–691.

45. Siami-Namini, S.; Tavakoli, N.; Namin, A.S. The Performance of LSTM and BiLSTM in Forecasting Time Series. In Proceedings of the 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 9–12 December 2019; pp. 3285–3292.

46. Cho, K.; Merrienboer, B.V.; Gulcehre, C.; Ba Hdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. *Comput. Sci.* **2014**, *1*. [CrossRef]

47. Rui, F.; Zuo, Z.; Li, L. Using LSTM and GRU neural network methods for traffic flow prediction. In Proceedings of the 2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC), Wuhan, China, 11–13 November 2016; pp. 324–328.

48. Rashedi, N.; Sun, Y.; Vaze, V.; Shah, P.; Paradis, N.A.J.M.M. Early Detection of Hypotension Using a Multivariate Machine Learning Approach. *Mil. Med.* **2021**, *186*, 440–444. [CrossRef] [PubMed]