Article:
Sheng, B., Chen, X., Cheng, J. orcid.org/0000-0003-0673-928X et al. (4 more authors) (2024) A novel scoring approach for the Wolf Motor Function Test in stroke survivors using motion-sensing technology and machine learning: A preliminary study. Computer Methods and Programs in Biomedicine, 243. 107887. ISSN 0169-2607

https://doi.org/10.1016/j.cmpb.2023.107887

A novel scoring approach for the Wolf Motor Function Test in stroke survivors using motion-sensing technology and machine learning: a preliminary study

Bo Sheng, Xiaohui Chen, Jian Cheng, Yanxin Zhang, Shane (Sheng Quan) Xie, Jing Tao, and Chaoqun Duan

ABSTRACT

***Background and objective*: Human-administered clinical scales, such as the Functional Ability Scale of the Wolf Motor Function Test (WMFT-FAS), are widely utilized to evaluate upper-limb motor function in stroke survivors. However, these scales are generally subjective and labor-intensive. To end this, we proposed a novel scoring approach for the motor function assessment.**

***Methods*: The proposed novel scoring approach mainly contained one Microsoft Kinect v2, one customized motion tracking system, and one customized intelligent scoring system. Specifically, the Kinect v2 was used to capture stroke survivors' functional movements, the motion tracking system was developed for recording the gathered movement data, and the intelligent scoring system (kernel: feed-forward neural network, FFNN) was developed to evaluate movement quality and provide corresponding WMFT-FAS scores. Several methods have been applied to enhance the approach's usability, such as singular spectrum analysis and multi-ReliefF method.**

***Results*: Sixteen stroke survivors and ten healthy subjects were recruited for validation. Inspiring results of the proposed approach were achieved when compared with the clinical scores provided by a physiotherapist: 0.924 ± 0.027 for accuracy, 0.875 ± 0.063 for F1-score, 0.915 ± 0.051 for sensitivity, 0.969 ± 0.013 for specificity, 0.952 ± 0.038 for AUC, 0.098 ± 0.037 for mean absolute error, and 0.214 ± 0.078 for root mean squared error.**

***Conclusions*: The results indicate that the proposed novel scoring approach can provide objective and accurate assessment scores, which can help physiotherapists make individualized treatment decisions.**

***Keywords*—Stroke, Kinect v2, Intelligent scoring system, WMFT-FAS, Motor function assessment.**

1. Introduction

Motor function assessment is essential for stroke rehabilitation as it can assist in developing clinical interventions to maximize patients' independence and motor function [1]. Clinical scales like WMFT-FAS and Fugl-Meyer Assessment (FMA) have been developed to evaluate motor function in hospital settings. Their inter- and intra-rater reliability has been extensively validated [2]–[4]. However, being a manual procedure, it inherently has certain limitations. Firstly, visual inspection involves some subjectivity, potentially resulting in varying clinical scores for the same patient across different physiotherapists [5]. Secondly, conducting motor functional assessments through clinical scales is labor-intensive. For example, administering the entire WMFT-FAS scale for a patient takes approximately 30 minutes, escalating medical costs, even for healthcare providers. To solve these limitations, experts have proposed biomarker-based approaches for assessing motor function in stroke survivors.

Biomarkers, encompassing diverse biological data like EMG, EEG, and fMRI, serve as key analytical sources for effective rehabilitation assessment [6]. EMG, measuring muscle electrical signals, can elucidate muscle contraction intensity and frequency in stroke survivors [7]. This information aids doctors in comprehending muscle damage extent post-stroke and tailoring rehabilitation tasks across different stages of recovery [8]. EEG, focusing on brain electrical activity, allows researchers to utilize functional connectivity measures to assess motor function in stroke survivors [9]. Automation of patient motor function evaluation, combined with neural network algorithms [10] such as Convolutional Neural Networks (CNNs) and Residual Neural Networks (RNNs) [11], is also feasible using EEG data. In the realm of rehabilitation, researchers have employed the Electroencephalographic Phase Synchrony Index as a biomarker to assess the recovery status of post-stroke survivors [12]. Some have used specific motor imagery EEG models (e.g., elbow extension and flexion) to classify EEG related to various motor imagery activities (e.g., opening a drawer) for rehabilitation applications [13]. Compared to EEG, fMRI offers a clear view of the damaged brain region after stroke and alterations in its connections with other brain regions [14]. This deeper insight contributes to a better understanding of the rehabilitation mechanism in stroke survivors. While biological data markers offer an intelligent approach to stroke rehabilitation assessment, the necessity for patients to wear specialized devices poses challenges in promoting these

B. Sheng is with the School of Mechatronic Engineering and Automation, Shanghai University, Shanghai, 200444, China (e-mail: shengbo@shu.edu.cn).

X. Chen is with the School of Mechatronic Engineering and Automation, Shanghai University, Shanghai, 200444, China (e-mail: chen15838022398@shu.edu.cn).

J. Cheng is with the School of Mechatronic Engineering and Automation, Shanghai University, Shanghai, 200444, China (e-mail: Cheng_9801@shu.edu.cn).

Y. Zhang is with the Department of Exercise Sciences, The University of Auckland, Auckland, 1010, New Zealand (e-mail: yanxin.zhang@auckland.ac.nz).

S. Xie is with the School of Electronic and Electrical Engineering, University of Leeds, Leeds, LS2 9JT, United Kingdom (e-mail: S.Q.Xie@leeds.ac.uk).

J. Tao is with the School of Mechatronic Engineering and Automation, Shanghai University, Shanghai, 200444, China (e-mail: taojing1016@shu.edu.cn).

C. Duan is with the School of Mechatronic Engineering and Automation, Shanghai University, Shanghai, 200444, China (e-mail: chaoqun.duan@hotmail.com, the corresponding author).

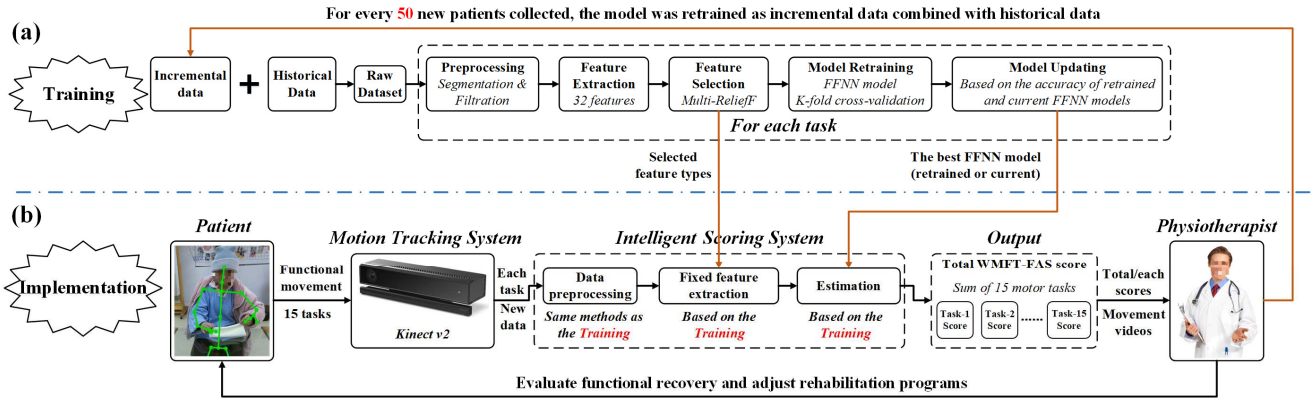

Fig. 1. The workflow of the proposed novel scoring approach based on the WMFT-FAS scale: (a) the training process; and (b) the implementation process.

approaches in community and primary healthcare settings. Hence, there is a growing focus on markerless-based intelligent rehabilitation assessment methods.

Methods based on image processing, such as Kinect [15], smartphones [16], and OpenPose [17], offer the advantages of low cost, portability, and minimal devices for patients to wear (if necessary). These methods hold promise for practical applications. Researchers have employed Kinect-V1 and data gloves to collect patient motion data, achieving an accuracy of around 82% using the Support Vector Machine (SVM) algorithm [18]. Expanding on this, some researchers have introduced rule-based binary logic classification algorithms to overcome the challenge of small sample sizes. They combined Kinect-V2 and force-sensitive sensors for motion data collection, ultimately achieving an accuracy of approximately 90% [19]. Another study proposed an automated method involving performing the FMA scale's movements with a handheld smartphone, extracting features from smartphone motion data. A decision tree algorithm was used to score 20/33 items from the FMA scale, resulting in an average accuracy of around 85% [16]. Other researchers proposed a method for real-time recognition of rehabilitation actions of stroke survivors based on the OpenPose and the full convolutional network (FCN). A one-dimensional full convolutional network was utilized to extract spatiotemporal features and classify actions, achieving an accuracy of 85.6% [20]. However, the existing markerless-based intelligent assessment systems have the potential for improvement: 1) the existing systems with advanced new hardware require a relatively large amount of computational resources (e.g., the new Azure Kinect needs a graphics card of RTX 3070 and above), which might not be readily available in communities; 2) physiotherapists may struggle to comprehend and explain the scores provided by the existing systems without detailed motor function results (e.g., movement trajectory, shoulder range); 3) To the best of the authors' knowledge, the WMFT-FAS has not been fully digitized by existing studies [21]–[26].

Therefore, this study aims to devise an innovative and cost-effective scoring approach for the intelligent and objective assessment of motor function in stroke survivors. Specifically, to enable practical applications in clinical settings, we designed a customized motion tracking system utilizing an inexpensive yet reliable depth-sensing device, Kinect v2, to capture motion data. Subsequently, we developed an intelligent scoring system employing classification-based scoring principles (feed-forward neural networks, FFNN) to construct estimation models for each motion task. Through the integration of these two systems, we were able to objectively evaluate motor function in stroke survivors and provide comprehensive WMFT-FAS scores.

The main contributions of this work were:

1) We first proposed a novel scoring approach (one Microsoft Kinect v2, one customized motion tracking system, and one customized intelligent scoring system) for stroke survivors' motor function assessment based on the full WMFT-FAS scale.

2) Experimental results (16 stroke survivors and 10 healthy subjects were recruited) showed that the proposed approach can perform the whole WMFT-FAS with an accuracy of 0.924 ± 0.027, and the developed multi-ReliefF method can dynamically select suitable features and increase the accuracy by around 2.9%.

2. METHODS

2.1 Workflow

Fig. 1(a) presents the workflow of the training process, consisting of the following steps: 1) get incremental data from the implementation process and build raw datasets by combining historical data (every 50 newly acquired patient data is treated as a new set of incremental data.); 2) preprocess raw datasets by the procedures of data classification, segmentation, and filtration; 3) extract 32 features from the preprocessed datasets based on different kinematics (e.g., endpoint, angular); 4) rank and select features by the multi-ReliefF method according to their sensitivity for distinguishing scores (e.g., scores 0 to 5) within each motor task; 5) use the K-fold cross-validation method to train the FFNN (feed-forward neural network) algorithm, employing the appropriate types of features selected in step 4. This involves randomly dividing the original dataset into K-equal (almost) subsets by subjects. In each round, one subset is designated as the testing set, while the remaining K-1 subsets are merged to form the training set. This process is repeated for a total of K rounds of training. 6) finally, update the estimation model in the implementation process if the retrained model achieves better accuracy.

Fig. 1(b) presents the workflow of the implementation process. Firstly, stroke survivors begin by performing

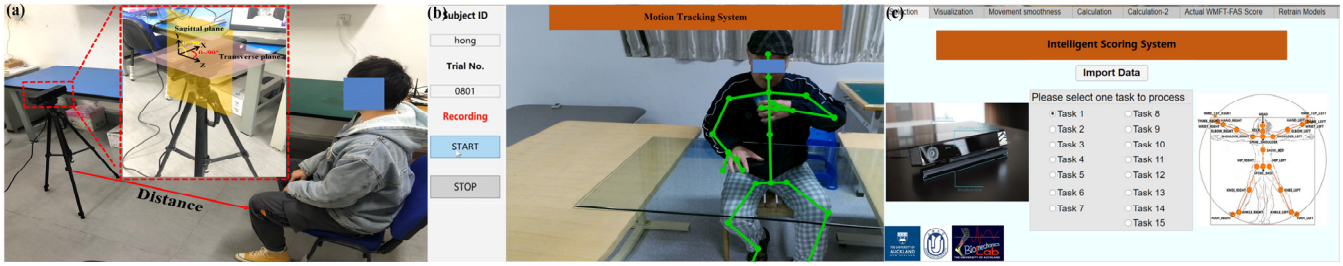

Fig. 2. Application development: (a) the hardware; (b) the motion tracking system; and (c) the intelligent scoring system.

upper-limb functional movements related to the WMFT-FAS. Their movement data are measured via the customized motion tracking system based on a Kinect v2 (Microsoft, Redmond, WA, USA). Next, new data are processed via the customized intelligent scoring system to get the estimated WMFT-FAS scores, including data preprocessing, feature extraction, and estimation. Specifically, the data preprocessing contains the same procedures conducted in the training process. The feature extraction extracts features from the preprocessed data based on the fixed types of features selected by the multi-ReliefF method executed in the training process. The score estimation employs the trained FFNN algorithm to estimate clinical scores that correspond to the movement quality of tasks. Finally, according to the estimated scores (total/each score) and movement videos, physiotherapists assess the functional recovery of stroke survivors and adjust rehabilitation programs.

The main differences between the two workflows are that the training process is responsible for determining and updating the parameters (e.g., the suitable types of features) and estimation models, which will be performed on a fixed schedule. The implementation process is only responsible for gathering new patient data and sending estimated scores movement videos to physiotherapists. The following sections will mainly introduce the training process workflow when gathering a raw dataset.

2.2 WMFT-FAS Motor Tasks for Stroke Survivors

The WMFT-FAS is a 6-point (0-5) ordinal rating scale for assessing the movement functional ability of stroke survivors rather than their completion time [4]. The WMFT-FAS contains 15 motor tasks (TABLE I) that progress from simple (e.g., forearm to table) to complex (e.g., flip cards). For each task, the score-0 implies no use of the affected side, while the score-5 implies the full ability to perform the task [27].

TABLE I
THE SET OF MOTOR TASKS FROM THE WMFT-FAS

Task	Description	Task	Description
T1	Forearm to table (side)	T9	Lift pencil (front)
T2	Forearm to box (side)	T10	Lift paper clip (front)
T3	Extend elbow (side)	T11	Stack checkers (front)
T4	Extend elbow with weight (side)	T12	Flip cards (front)
T5	Hand to table (front)	T13	Turn key in lock (front)
T6	Hand to box (front)	T14	Fold towel (front)
T7	Reach and retrieve with weight (front)	T15	Lift basket with weight (front)
T8	Lift can (front)		

We used the WMFT-FAS as the targeted scale because it has the following characteristics: 1) the scale mainly focuses on gross movements rather than fine movements (e.g., finger mass flexion/extension tasks in the FMA), which can be measured via the Kinect v2 (for the tasks of "flip cards" and "fold towel tasks", patients typically perform these tasks with compensatory strategies by using trunk and forearm); and 2) the scoring criteria include speed, movement smoothness, joint coordination, and the presence of pathological synergies and compensatory strategies, which can be characterized as features extracted from movement data. By comparison, motion-sensing technology cannot measure the strength criterion of grasp tasks in the FMA.

2.3 Application Development

The hardware of this study was the Kinect v2 (Fig. 2(a)) developed by Microsoft as a motion-sensing input for the game console. Subsequently, it was redesigned for various applications, such as healthcare and robotics. The Kinect v2 can measure the 3D coordinate data of 25 anatomical landmarks (Fig. 2(b)) for each body with a frame rate of 30 fps (frames per second) [28]. These landmarks can be gathered under standard lighting and fitted clothing conditions (a similar requirement for performing the WMFT-FAS assessment) without any actual markers, making the Kinect v2 suitable for clinical settings. Accordingly, a customized motion tracking system (Fig. 2(b)) was developed to save the gathered video, coordinate data, and time frames. A customized intelligent scoring system (Fig. 2(c)) was also developed to estimate motor tasks' scores and the total WMFT-FAS score. The motion tracking system was programmed based on the Kinect SDK 2.0 and Microsoft Visual Studio 2016 (Microsoft, Redmond, WA, USA). The intelligent scoring system was reprogrammed based on the 2021b Matlab's App Designer (MathWorks, Natick, MA, USA).

2.4 Data Processing and Feature Extraction

A total of 10 joints' movement data were used to evaluate the functional performance of the affected upper limbs, including shoulder/mid of spine, left/right of shoulder, elbow, wrist, and hand. The raw datasets were firstly classified according to each motor task and segmented into individual repetitions. They were then filtered via a singular spectrum analysis (SSA) algorithm to reduce the effects of noise [29]. The SSA algorithm performs better for smoothing raw movement data than other commonly used filtering methods, such as Butterworth filters and discrete wavelet transform [30]. Fig. 3 presents the results of the SSA algorithm (applied to a healthy subject performed Task-1 with three repetitions). As expected, the speed is approximately zero during the waiting for and after the reaching motion (the retrieving movement was not used for assessment according to the guideline of the WMFT-FAS), and the outline of the speed is well bell-shaped.

Four critical steps of the SSA are described as follows [31].

Step 1 (Embedding): the primary aim of the 1st step is to transform the observed one-dimensional time series $Y_T = (y_1,...y_T)$ to the corresponding trajectory matrix X:

$$\mathrm{X} = [X_1 : \ldots : X_K] = (y_{ij})_{i,j=1}^{L,K} = \begin{pmatrix} y_1 & y_2 & y_3 & \cdots & y_K \\ y_2 & y_3 & y_4 & \cdots & y_{K+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_L & y_{L+1} & y_{L+3} & \cdots & y_T \end{pmatrix} \quad (1)$$

Where L ($1<L<T$) is the window length and $K=T-L+1$.

Step 2 (Singular Value Decomposition (SVD)): the primary aim of the 2nd step is to perform the SVD of the trajectory matrix X. Set $S = XX^T$, and the eigenvalues ($\lambda_1 \geq \ldots \geq \lambda_L \geq 0$) and the corresponding eigenvectors ($U_1,\ldots, U_L$) of the matrix S will be processed by the SVD.

Step 3 (Eigentriple grouping): the primary aim of the 3rd step is to partition the set of indices $\{1,\ldots,d\}$ into the m disjoint subsets ($I_1,\ldots,I_m$). The grouped SVD expansion of the matrix X can then be written as

$$\mathrm{X} = X_{I_1} + \ldots + X_{I_m} \quad (2)$$

Step 4 (Diagonal averaging): the primary aim of the 4th step is to get the decomposition results, which are the main results of the SSA algorithm. Specifically, each matrix X_{Ij} of the grouped decomposition will be hankelized to obtain the Hankel matrix.

The Hankel matrix will then be transformed into a new series of length N processed by diagonal averaging. Finally, the initial series $y_1,...,y_T$ can be decomposed into a sum of m reconstructed subseries:

$$y_n = \sum_{k=1}^{m} \tilde{y}_n^{(k)} \; (n = 1,2,\ldots,N) \quad (3)$$

The key parameters were set as the same as our pilot study (e.g., 5 for the window of length L) [5].

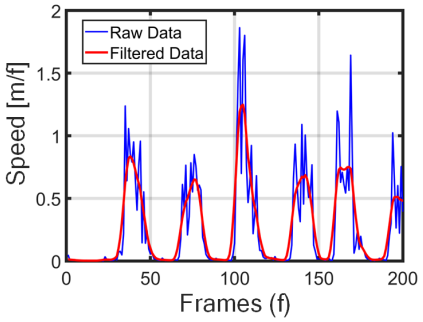

Fig. 3. The performance of the SSA algorithm.

For feature extraction, two principles are commonly used by existing automated systems: 1) extract specific features based on the characteristics of each motor task and the extracted features will not be easily updated further [16], [19]; 2) extract all features and then dynamically select suitable features based on the current training dataset, and the extracted features will be updated during the system maintenance [25]. This study chose the second principle for continuously improving the proposed approach via gathering movement data of stroke survivors in practical applications, even though the fixed number of participants in this preliminary study might not highlight the suitability of the second principle.

TABLE II
THE DETAILED INFORMATION ON THE EXTRACTED FEATURES

Position	Type	Name (number)	Brief Description	Ref
Endpoint	Speed	V_{max} (1)	Max velocity	N/A
		V_{mean} (2)	Mean velocity	N/A
		Mov_{time} (3)	Movement time	[32] [33]
	Control strategy	$Peak_{time}$ (6)	Time to velocity peak	N/A
		$Length_{ept}$ (7)	Path length of the endpoint	N/A
	Efficiency	IC_{ept} (8)	Index of curvature	[34]
	Movement coordination	$IJCI_{ept}$ (9)	Inter-joint coordination index	[32] [28]
	Smoothness	SAL_{ept} (10)	Spectral arc-length	[35]
		V_{ratio} (11)	Ratio of max to mean velocity	N/A
		$Peak_{no}$ (12)	Number of velocity peaks	N/A
		$Arrest_{ratio}$ (13)	Mean arrest period ratio	[36]
		$Jerk_{ept}$ (14)	Normalized mean absolute jerk	[37] [38]
Shoulder angle (Flexion, Adduction, Internal)	Rotational speed	$AngV_{maxF}$ (18), $AngV_{maxA}$ (19), $AngV_{maxI}$ (20)	Max angular velocity of shoulder	[39]
		$AngV_{meanF}$ (21), $AngV_{meanA}$ (22), $AngV_{meanI}$ (23)	Mean angular velocity of shoulder	[39]
	Smoothness	$AngV_{ratioF}$ (24), $AngV_{ratioA}$ (25), $AngV_{ratioI}$ (26)	Ratio of max to mean angular velocity	N/A
		$AngD_{ratioF}$ (15), $AngD_{ratioA}$ (16), $AngD_{ratioI}$ (17)	Ratio of angle difference	[40]
		$Jerk_F$ (27), $Jerk_A$ (28), $Jerk_I$ (29)	Normalized mean absolute jerk	[37] [38]
		SAL_F (30), SAL_A (31), SAL_I (32)	Spectral arc-length	[35]
Others		$Trunk_{mean}$ (4)	Mean trunk displacement	[41]
		$Trunk_{max}$ (5)	Max trunk displacement	[41]

A total of 32 features were therefore extracted and divided into three main aspects (TABLE II): 1) endpoint (12 features, e.g., max/mean velocity, index of curvature); 2) shoulder angle (18 features, e.g., max/mean angular velocity, normalized mean absolute jerk); and 3) others (2 features, max/mean trunk displacement). A detailed explanation of the features is provided as the supplementary material. The details of these extracted features can be found in our previous studies [5][40]. Their validity and reliability have been evaluated through clinical trials, which have the discriminatory ability to distinguish functional motor performance in specific aspects (e.g., smoothness and control strategy) [1]. It is crucial to emphasize that the term "Smoothness" presented in TABLE II serves as an evaluation index for assessing the functional motor

ability of patients. This evaluation metric is distinct from the concept of "smoothness" associated with the data smoothing process employing the SSA algorithm.

2.5 Feature Selection

Feature selection was conducted by the customized multi-ReliefF method in the training process only, which includes iterative ranking and filtration steps. The flowchart of the multi-ReliefF method is shown in Fig. A1. Firstly, a feature ranking algorithm named ReliefF was used to calculate the importance weights of raw features. The raw features were then ranked in descending order according to the values of their importance weights (Fig. A1, the blue bar chart). Secondly, the feature filtration process was conducted to select appropriate features. The features with negative importance weights were deleted, and the remaining features were sent to the ReliefF algorithm for the next processing cycle. The feature selection was completed until all features had positive importance weights; that is, no features with negative importance weights were found by the ReliefF algorithm in the first step. It should be noted that in order to achieve the automation of clinical scoring in practical applications, we did not use the Davies-Bouldin index (DBI) [42] to select top-ranked features since the optimal cutoff point has to be revised according to different training datasets [25].

The ReliefF algorithm, used in the ranking step, is extended based on the original Relief algorithm (a supervised feature selection method). It is not limited to two-class problems and can process noisy, incomplete, and skewed data [43]. The pseudo-code of the ReliefF algorithm is listed as follows [43]–[45].

ReliefF algorithm

Input: All features and classes (scores) of trials in the training dataset.

Output: A relevance index vector W, which contains the calculated importance weights of each feature A.

1. set all importance weights $W[A]$: =0.0;
2. **for** i: =1 **to** m **do begin**
3. randomly select an instance R_i;
4. find k nearest hits H_j;
5. **for** each class $C \neq$ class (R_i) **do**
6. from class C find k nearest misses $M_j(C)$;
7. **for** A: =1 **to** #all_features **do**
8. $$W[A] := W[A] - \sum_{j=1}^{k} \frac{diff(A, R_i, H_j)}{m \square k} + \sum_{C \neq class(R_i)} \left[\frac{\frac{P(C)}{1 - P(class(R_i))} \sum_{j=1}^{k} diff(A, R_i, M_j(C))}{m \square k} \right]$$
9. **end**;
10. **end**;

The ReliefF algorithm will randomly select an instance R_i (line 3) and find k nearest neighbors of the same class, named nearest hits H_j (line 4). It will then search k nearest neighbors for each of the other classes, named nearest misses $M_j(C)$ (lines 5 and 6). Finally, it will update the relevance estimation $W[A]$ for all features A based on their values for R_i, hits H_j, and misses $M_j(C)$ (lines 7 and 8). Here, the function $diff(A, I_1, I_2)$ calculates the difference between the values of feature A for two trials I_1, and I_2, and averages the contribution of all the hits and all the misses. The $P(C)$ is the prior probability of class C, which is used for weighting the contribution for each class of the misses. The $P(C)$'s value is estimated from the training set. The factor $1\text{-}P(class(R_i))$ represents the sum of probabilities for the misses' classes. The user-defined parameters m and k control the repetition times and the locality of the relevance estimates, respectively. In this study, the value of m was set as the default value (the number of trials), and the value of k was set to 10 according to previous studies [25][44].

2.6 Model Development

The estimation model is the critical component of the intelligent scoring system, which should be designed to provide accurate estimation results and be easily maintained. The feed-forward neural network (FFNN) algorithm was selected as the estimation model's kernel, a simple but robust artificial neural network. In Fig. 4, the FFNN algorithm's inputs were the filtered features, and its outputs were the occurrence possibility (OP) of the WMFT-FAS motor task's each score (score 0 to score 5). The structure of the FFNN algorithm consists of three layers: the input layer, the hidden layer, and the output layer. Specifically, the input layer contained dynamic numbers of neurons associated with the filtered features (32 features for the maximum). The hidden layer contained 20 neurons recommended by previous studies [46], [47]. The output layer contained 6 neurons associated with the scores of each WMFT-FAS motor task. The input data travel in the forward direction, from the input layer, through the hidden layer, to the output layer. To enhance the performance of the FFNN algorithm, the backpropagation algorithm was also used to calculate the error contribution of neurons. The detailed workflow of the FFNN algorithm can be found in [48].

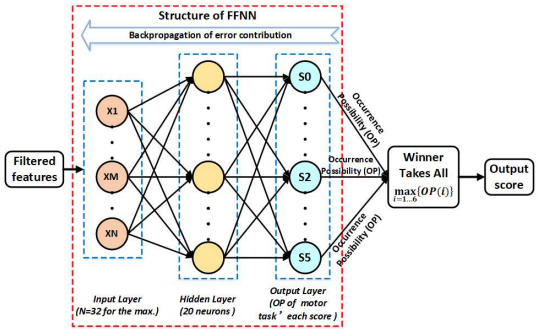

Fig. 4. The structure of the proposed estimation model.

The structure of the proposed estimation model is shown in Fig. 4. The approach employed the "Winner Takes All" criterion [49]–[51] for decision-making. In essence, this means selecting the highest OP (Occurrence Possibility), representing the FFNN algorithm's outputs for scores ranging from 0 to 5, as the output score for the motor task. For example, for task-1, the occurrence possibilities of scores 0 to 5 were 0%, 6%, 6%, 11%, 67%, and 10%, respectively. The final output score was four since score 4 had the highest occurrence possibility (67%).

A K-fold cross-validation method [52] was applied to train (retrain) and evaluate the proposed FFNN algorithm, one of the most frequently used evaluation methods in machine learning

[53]. The cross-validation procedure was applied on a subject-by-subject basis according to the structure of the prepared datasets, namely the K=26. Furthermore, during the training process (Fig. 1), the accuracy of the trained model would be calculated based on the raw dataset via the cross-validation method. The current model (the old model in the implementation process) would be replaced if the trained model (the new model) achieved higher accuracy.

3 Experiments and Results

Experimental validation was conducted in collaboration with the Rehabilitation Hospital of Fujian Province in China. Fujian University of Traditional Chinese Medicine Human Participants Ethics Committee approved all the experimental procedures and clinical data access with the reference 2018KY-019-02.

3.1 Participants

A total of 16 stroke survivors (4 females, 12 males, age: 54.2 ± 18.1 years old, 5.0 ± 6.1 months post stroke), all without mild cognitive impairment (Montreal Cognitive Assessment (MoCA) ≥ 23 [54][55]), participated in the study. Additionally, 10 healthy subjects (5 females, 5 males, age: 21.2 ± 1.2 years old), without upper-limb diseases or known nervous system disorders, were recruited for experimental validation. The exclusion criteria were: 1) participants with severe impairments; 2) participants with all kinds of infectious diseases; 3) participants with medical-tape allergy; and 4) other personal reasons. TABLE III presents the characteristics of the recruited stroke survivors. TABLE A1 offers the detailed WMFT-FAS scores of all stroke survivors.

TABLE III
THE CHARACTERISTICS OF STROKE SURVIVORS

ID	Age/ Sex	M[b]	S[c]	Stroke Location & Etiology	MoCA[d]	WMFT -FAS
P1	48/M	12	L	Frontal lobe, hemorrhage	30	48
P2[a]	56/F	1	L	Thalamus and basal ganglia, hemorrhage	25	23
P3	51/M	24	R	Unclear	30	60
P4	75/M	1	L	Basal ganglia, infarction	26	35
P5	44/M	5	L	Subarachnoid space, hemorrhage	27	43
P6	63/M	1	R	Unclear	29	67
P7	79/M	2	L	Basal ganglia, hemorrhage	28	57
P8	61/M	1	R	Frontal lobe and centrum ovale, infarction	27	52
P9	63/M	1	R	Basal ganglia, hemorrhage	28	38
P10	54/M	12	R	Basal ganglia, hemorrhage	26	46
P11	50/M	5	L	Subarachnoid space, hemorrhage	26	57
P12	41/M	2	R	Basal ganglia, infarction	30	60
P13	22/F	4	L	Subarachnoid space, hemorrhage	30	40
P14	12/F	1	L	Parietal lobe, infarction	28	60
P15	79/F	6	R	Pons, hemorrhage	29	42
P16	69/M	1	L	Frontal lobe, hemorrhage	30	58

[a]P2 only completed Tasks 1, 5-7, 9-12; [b]M = Monthes post stroke; [c]S = Stroke hemisphere; [d]MoCA = Montreal Cognitive Assessment, measured before the validation.

The healthy subjects were recruited to provide full-mark trials because stroke survivors with nearly full marks were hard to find in the rehabilitation hospital due to their financial and personal issues. Similar recruitment protocols have been reported by studies [56][57], although age-matched healthy subjects could provide more comparable full-mark trials. It should be noted that this study is not clinical research; therefore, the sample size of the recruited stroke survivors was not calculated by the G*Power software (University of Kent, Canterbury, UK).

3.2 Testing Protocol

Before clinical trials, verbal and video instructions were given to all participants, and the signed participant consent form and participant information sheet were collected if they agreed with the procedures. All participants' upper-limb functional motor performance was evaluated by an experienced physiotherapist using the WMFT-FAS. Participants were verbally encouraged to follow an instructional video repeated three times for each motor task during the assessment. Meanwhile, in front of participants (Fig. 2(a)), the Kinect v2 was placed on a tripod to record movement data. To enhance the accuracy of raw data, the distance and direction of the Kinect v2 were also adjusted for tasks according to the report [58] and our experience (TABLE IV).

TABLE IV
DETAIL PARAMETERS OF TASK ADJUSTMENTS

Task	Distance (cm)	Direction ([a]S/[b]T, degree)	Task	Distance (cm)	Direction (S/T, degree)
T1	150	15/5	T9	142	20/5
T2	150	15/5	T10	142	20/5
T3	150	15/5	T11	142	20/5
T4	150	15/5	T12	142	20/5
T5	145	18/0	T13	140	20/15
T6	145	18/0	T14	140	20/5
T7	145	18/0	T15	150	20/5
T8	142	20/5			

[a]S=Sagittal plane angle; [b]T = Transversal plane angle. The absolute values of parameters were taken.

These two parameters were fixed for different participants with the same task. Furthermore, a big glass platform was set for tracking participants' legs in several motor tasks (inaccurate leg tracking results might affect the tracking accuracy of upper limbs, Fig. 2(b). The video recordings of the assessment were also saved for later analysis. According to the previous work [25] and the limited number of recruited participants, we assumed that the repetitions of each participant were independent, and each repetition was considered a trial. Therefore, for each of the 15 tasks, the raw dataset encompassed 78 trials, each comprising up to 32 features, along with 1 clinical score, resulting in a maximum total of 66,924 data points. However, patient-2 did not complete Tasks 2-4, 8, 13-15, and the datasets of these tasks contained 75 trials.

3.3 Performance Evaluation

The performance of the proposed novel WMFT-FAS scoring approach was evaluated in three aspects: the whole approach, the multi-ReliefF method, and the FFNN kernel. For the first one, the estimated scores were compared with the actual WMFT-FAS scores provided by the experienced physiotherapist. For the second one, the proposed multi-ReliefF method was compared with the suggested ReliefF-DBI method [25] and the original features (without selection). For the third one, 10 commonly used machine learning algorithms were used to build respective estimation models with the same datasets for comparison. TABLE V shows the selected machine learning algorithms and their specifications. Specifically, among the

five mentioned neural networks—FFNN, NN, MNN, CFNN, and RNN—FFNN and CFNN are built based on specifications 1, 2, and 3. On the other hand, NN and MNN are developed relying on specifications 1, 2, 4, and 5, while RNN is constructed using specifications 1, 2, 4, and 6.

TABLE V
DETAILED INFORMATION ON THE SELECTED CLASSIFICATION ALGORITHMS

Model	Kernel	Specification
[a]FFNN (current)	Script: Feedforwardnet (one-layer)	1.HiddenSize=20 2.Maxepochs=5000
[b]NN	Script: Patternnet (one-layer)	3.TrainFcn=Levenberg-Marquardt 4.TrainFcn=Scaled Conjugate
[c]MNN	Script: Patternnet (three-layer)	Gradient 5.PerformFcn=Cross-Entropy
[d]CFNN	Script: Cascadeforwardnet (one-layer)	Loss 6.LayerDelays=1:2 for RNN
[e]RNN	Script: Layrecent (one-layer)	
[f]DT	Coarse Tree	1.Maximum number of splits=4 2.Split criterion: Gini's diversity index
[g]DA	Linear Discriminant	1.Covariance structure: Full
[h]NB	Kernel Naive Bayes	1.Kernel type: Gaussian
[i]SVM	Quadratic SVM	1.Kernel scale=Automatic 2.Box constraint level=1 3.Multiclass method: One-vs.-One
[j]KNN	Fine KNN	1.Number of neighbours=1 2.Distance metric: Euclidean 3.Distance weight: Equal
[k]EM	Subspace with Discriminant learner	1.Number of learner=30 2.Subspace dimension=16

[a]FFNN = Feed-Forward Neural Network, [b]NN = Neural Network created using patternnet, [c]MNN = Neural Network with Multiple Layers, [d]CFNN = Cascade-Forward Neural Network, [e]RNN = Recurrent Neural Network, [f]DT = Decision Tree, [g]DA = Discriminant Analysis, [h]NB = Naive Bayes, [i]SVM = Support Vector Machine, [j]KNN = K-Nearest Neighbour, [k]EM = Ensemble Method.

Seven metrics [59] were used to evaluate performance and make comparisons: accuracy, mean absolute error (MAE), root mean squared error (RMSE), sensitivity (namely recall), specificity, F1-score, and AUC (area under a curve). For this study (multi-class estimation/classification issues), accuracy, F1-score, sensitivity, specificity, and AUC are often used as important evaluation metrics, so we ranked these metrics first [60]. MAE and RMSE are used to measure average errors, which are commonly used in control strategies. However, they are still important metrics to evaluate the model performance. So, we put these two values last. The details of MAE and RMSE can be found in [25], and the details of other metrics can be found in our previous research [5][40]. The significant difference was analyzed via SPSS version 20 (IBM Corporation, NY, USA). The level of statistical significance was set to $p \leq 0.05$.

3.4 Results

Fig. 5 and TABLE A2 show the performance of the whole approach: the mean values for accuracy, F1-score, sensitivity, specificity, AUC, MAE, and RMSE are 0.924, 0.875, 0.915, 0.969, 0.952, 0.098, and 0.214, respectively. The values of standard deviation are from 0.013 to 0.078. Specifically, for accuracy and F1-score, Task-14 receives the highest value while Task-10 receives the lowest value (0.966 versus 0.879, 0.967 versus 0.741, respectively); for sensitivity, Task-3 receives the highest value while Task-12 receives the lowest value (0.980 versus 0.803); for specificity, Task-3 and Task-14 receive the highest value while Task-7 receives the lowest value (0.984 versus 0.947); for AUC, Task-14 still receives the highest value while Task-12 still receives the lowest value (1.000 versus 0.871); for MAE, Task-14 receives the lowest error value while Task-13 receives the highest error value (0.040 versus 0.157); for RMSE, the same trend is found (0.063 for Task-14 versus 0.330 for Task-13).

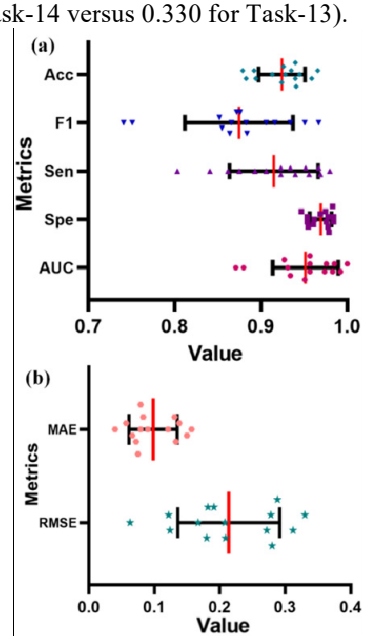

Fig. 5. Performance of the proposed approach (the red vertical line represents the mean: (a) the values of Accuracy, F1-score, Sensitivity, Specificity, and AUC; (b) the values of MAE and RMSE.

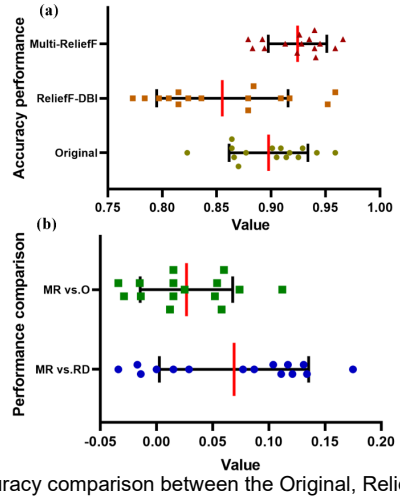

Fig. 6. Accuracy comparison between the Original, ReliefF-DBI, and Multi-ReliefF methods (the red vertical line represents the mean): (a) the accuracy performance; (b) the difference values. MR = Multi-ReliefF, RD = ReliefF-DBI, and O = Original.

Fig. 6 and TABLE A3 show the performance of the multi-ReliefF method: compared with the original features, around 75% of motor tasks' accuracy is improved through the proposed multi-ReliefF method, and only 4 motor tasks' accuracy is slightly decreased (1.5%~3.7% reduction). Compared with the ReliefF-DBI method, only 3 motor tasks' accuracy is slightly decreased (1.5%~3.7% reduction). In terms of the mean improvement of accuracy, the proposed multi-ReliefF method presents 2.9% (0.027 ± 0.040) and 8.1% (0.069 ± 0.064) increments compared with the original features and ReliefF-DBI method, respectively.

Meanwhile, we focused on analyzing the accuracy results (TABLE A3, the first three columns, tasks 1-15) obtained using the original method, ReliefF-DBI method, and Multi-ReliefF method. The Kolmogorov-Smirnov test was utilized to verify that the accuracy results from these three methods adhere to a normal distribution. Subsequently, we conducted an independent sample t-test to discern any statistically significant differences in the accuracy results across the various methods. Detailed results of the independent sample t-test are presented in TABLE VI.

TABLE VI
INDEPENDENT SAMPLE T-TEST RESULTS OF MULTI-RELIEFF VS. ORIGINAL AND MULTI-RELIEFF VS. RELIEFF-DBI

	[d]df	*p*-value	Cohen's d	95%CIs (Lower)	95%CIs (Upper)
[a]MR vs. [b]O	28	0.030	0.781	-0.051	-0.003
MR vs. [c]RD	19.343	0.001	1.195	-0.105	-0.334

[a]MR = Multi-ReliefF, [b]O = Original, [c]RD = ReliefF-DBI, [d]df = degrees of freedom.

Specifically, we observed statistically significant differences among the original, Relief-DBI, and multi-ReliefF methods. The p-value for the difference between the multi-ReliefF and original methods is 0.03, indicating a statistically significant difference in accuracy between these two methods, as it is less than 0.05. Additionally, the p-value for the difference between the multi-ReliefF and Relief-DBI methods is 0.001, signifying a highly significant accuracy difference between these two methods, given that it is much smaller than 0.05.

Regarding effect size, we utilized Cohen's d as a metric to quantify the effect size. Cohen's d is employed to characterize the magnitude of the mean difference between the two groups, where d=0.2 signifies a small effect size, d=0.5 indicates a moderate effect size, and d=0.8 denotes a large effect size. As indicated in TABLE VI, Cohen's d value for the difference between the multi-ReliefF and original methods is 0.781 (0.5<d<0.8), signifying a substantial difference between these two methods. Similarly, Cohen's d value for the difference between the multi-ReliefF and Relief-DBI methods is 1.195 (d>0.8), indicating a more pronounced difference between these two methods.

The confidence interval offers a possible range for parameter estimation. When the 95% confidence interval excludes 0, the parameter is typically considered statistically significant. As depicted in Table VI, the 95% confidence interval for the difference between the multi-ReliefF and original methods ranges from -0.003 to -0.051, with all values below 0. This implies that the multi-ReliefF method exhibits a statistically significant improvement in accuracy compared to the original method. Similarly, for the difference between the multi-ReliefF and Relief-DBI methods, the 95% confidence interval spans from -0.105 to -0.334, with both lower and upper limits significantly below 0. This strongly suggests that the multi-ReliefF method yields a substantially superior improvement in accuracy compared to the Relief-DBI method.

All of these statistical findings consistently demonstrate a significant enhancement in accuracy achieved by the proposed multi-ReliefF method in comparison to both the original method and the Relief-DBI method.

TABLE VII shows each task's best and deleted (worst) features. Specifically, features 12 (number of velocity peaks), 7 (path length of the endpoint), and 10 (spectral arc-length) exerted the most positive influences on estimation performance among all motor tasks, while features 9 (inter-joint coordination index), 13 (mean arrest period ratio), and 23 (mean internal angular velocity of the shoulder) presented opposite results. Meanwhile, the best and worst features were similar for task-5, task-7, and task-13: $Peak_{no}$, $Length_{ept}$ for the best, and SAL_I, $Arrest_{ratio}$ for the worst. One interesting finding is that for task-12 (flip cards), movement time was the best feature, and $trunk_{mean}/trunk_{max}$ were the worst features, along with a situation that the FFNN algorithm dropped out of the top three best algorithms (TABLE A4).

TABLE VII
THE BEST AND DELETED FEATURES

Task	Best	Deleted	Task	Best	Deleted
T1	16,14,7	11	T9	12,7,21	9,13,17
T2	12,16,7	9,4,13	T10	12,7,25	23,9,11
T3	12,7,14	9,13,17-23, 25,26,30	T11	12,7,3	11,21-23,8,4
T4	7,12,5	9,15,17,22	T12	3,10,24	4,31,5,8
T5	12,7,14	32,23,13	T13	12,7,5	32,15,13,9-11,17
T6	12,10,28	4,13,17	T14	3,13,6	8,9,15,20,22, 23,31
T7	12,7,10	13,32	T15	31,1,2	18-21,23,24, 30,32
T8	15,10,9	4,5,13,21,2 3,30-32			

Fig. 7 and TABLE A4 show the top three algorithms for each motor task with respect to accuracy. The FFNN algorithm receives promising comprehensive performance: 5 times for the best algorithm (T1, T3, T4, T8, and T14), 6 times for the second-best algorithm (T2, T6, T7, T11, T13, and T15), and 3 times for the third-best algorithm (T5, T9, and T10). For T12, the FFNN algorithm ranks as the fourth-best algorithm. For other evaluation metrics, the FFNN algorithm also presents the best comprehensive performance. The detailed results of different algorithms for each task can be found in TABLE A5-TABLE A19.

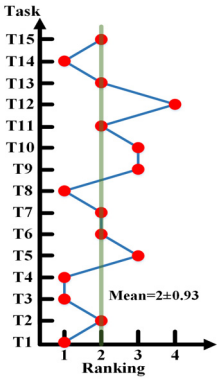

Fig. 7. Accuracy performance of FFNN algorithm in 15 tasks.

4 DISCUSSION

The main findings of this study were: 1) the proposed novel scoring approach can automate the entire WMFT-FAS with a mean accuracy of 0.924 and mean RMSE of 0.214/5; 2) the proposed multi-ReliefF method enhanced the final accuracy: the increment of 0.027 ± 0.041 (2.9% increment) with the

p-value of 0.030 was achieved when compared with the original features, and the increment of 0.069 ± 0.066 (8.1% increment) with the *p*-value of 0.001 was achieved when compared with the ReliefF-DBI method.

According to the research [18][19], a Kinect-based system might be an effective and accurate way to assess clinical scales automatically, and it is indeed a widely used device for scoring FMA [61], BBT [62], and RPS [63] scales. In this study, the movement data were collected via one Kinect v2, which might be the first time Kinect was used for the WMFT-FAS scoring [64]. In our previous work [28], Kinect v2 measured angular waveforms for elbow and shoulder flexion/extension with the inter-device coefficient of multiple correlation (CMC) > 0.87. It also showed relatively high test-retest reliability for most angular waveforms of upper limbs with CMC=0.75-0.99 (the Vicon system was used as the gold standard, and the perfect value was 1.00). Some articles reported that the tracking accuracy of Kinect v2 was not good enough for clinical applications, such as research [65]. The main reasons for making such conclusions were: 1) using relatively low-end computers to track movements (the measured fps is too low to represent the actual movement); 2) using the constant distance and direction of Kinect v2 to gather movement data (these two critical parameters should be modified according to different types of motor tasks, TABLE IV); and 3) using raw movement data to extract features (advanced algorithms are suggested to enhance the accuracy and robustness of preprocessing). Although the Kinect v2 is currently out of the market, the proposed approach can still be used for new practical applications in clinical scenarios. Specifically, the tracking skeleton of the proposed motion tracking system (Fig. 2(b)) can be revised to adapt the new anatomical landmarks provided by new devices (e.g., Azure Kinect, a new Kinect device released by Microsoft in 2019). As an affordable and portable assessment tool that can be used in hospital and community settings, this study used the Kinect v2 rather than other new devices to collect clinical data due to its low cost and friendly hardware requirement. Furthermore, according to the report [66], the Kinect v2 achieved better tracking results than the Azure Kinect in the mid and upper body regions, especially in the upper extremities. In fact, the most significant disadvantage of the Kinect v2 is that it does not have the ability to track fine movements (the same issue as the Azure Kinect), such as finger flexion/extension. Therefore, for this study, we have to use gross movement' features to evaluate specific motor tasks containing fine movements. This is why task-10 (lift paper clip) and task-12 (flip cards) received the worst assessment results. One interesting finding was that task-14 (fold towel) received the highest results in 6/7 evaluation metrics. The possible reason might be that task-14 requires a larger range of motion than task-10 and task-12, even though all of them contain fine movements. Therefore, patients have to use their whole upper bodies (e.g., shoulder, elbow, trunk) to finish task-14 due to synergistic patterns and muscle contractures.

To the best of the authors' knowledge, this is the first automated approach for scoring the entire WMFT-FAS. Compared with similar WMFT-FAS scoring systems that might be expensive or inconvenient/impossible for practical applications [21]–[26], the proposed approach is cheap, accurate, portable, and easy to operate. The experimental performance of these similar systems was promising (0.667 for the worst) but still far from clinical applications due to their hardware (e.g., a set of IMUs, which often requires precise sensor placement and therefore causes clothing restrictions) and specific assumptions (e.g., all of the recruited patients had relatively high-level motor capacity). The proposed approach does not contain such limitations and other time-consuming procedures: it requires only around eight minutes to set up an experiment (e.g., lighting, platform, Kinect's angle of view for each task, please see Fig. 2), and participants are required to perform tasks in front of the Kinect, making it potentially more accessible to be used in clinical settings. However, the primary users of the proposed approach are stroke survivors with moderate impairments (scores 2-4 of the WMFT-FAS). It is why this study did not recruit participants with severe impairments. That is, the proposed motion tracking system (Fig. 2(b)) cannot deal with stroke survivors in their early stages (e.g., scores 0-1 of the WMFT-FAS, score 0 of the FMA) since they can hardly move their bodies to perform the motor functional assessment. This situation has been confirmed by the report [64] that the usability of automated assessment systems will have a niche bounded by the level of patients' affection. Surface electromyography (sEMG) could be combined with the proposed approach in the future to cover stroke survivors in all stages since it can monitor muscle activity in their early stages [67].

The performance of our approach can be seen in TABLE A2: the mean accuracy is 0.924 (ranging from 0.879 to 0.966) with a low standard deviation (0.026). Similar values can be found in other evaluation metrics in TABLE A2. These results suggested that the proposed approach might have the potential to be used in clinical applications with further improvements. In addition, several methods have been applied to enhance the performance of the proposed approach. Firstly, the distance and direction of Kinect v2 were adjusted according to motor tasks' characteristics (approximately 20 seconds for each adjustment, TABLE IV). For example, for task-6 (hand to box-front, gross movements), the Kinect was set far from participants to track their whole bodies, and the parameters of distance and direction were set as 145 cm and 18/0 degrees, respectively. For task-14 (fold towel-front, contains fine movements), the Kinect was placed close to participants to track their whole upper bodies and a part of their lower bodies. The parameters of distance and direction were set as 140 cm and 20/5 degrees, respectively. Secondly, powerful preprocessing algorithms were applied to reduce the effects of noise on raw movement data suggested by the research [65]. For example, SSA (singular spectrum analysis) was used instead of Butterworth filters. Lastly, two customized algorithms were developed to enhance the overall performance.

The first customized algorithm is the multi-ReliefF method. It can be concluded from TABLE A3 that the proposed method can enhance accuracy with respect to the original features and the ReliefF-DBI method (*p*-value = 0.030, and 0.001, respectively). Specifically, around 2.9% increment has been achieved compared with the original features, and approximately 8.1% increment has been achieved compared with the suggested ReliefF-DBI method [25]. The possible reason for the ReliefF-DBI method's low performance in our study could be that the rule of the fixed feature selection (e.g.,

select the fixed top-N-ranked features) cannot deal with a large number of features (a total of 32 features for each trial). Furthermore, the features selected and deleted (please see TABLE VII) for task-12 (flip cards) might partly explain the low rank of the FFNN algorithm. The possible reason might be the deletion of trunk features ($Trunk_{mean}$, $Trunk_{max}$) since gross movements have been used to assess tasks with fine motor skills. The trunk features are essential for detecting synergistic patterns of stroke survivors and can also be used to assess the movement quality of flip cards (stroke survivors have to complete task-12 with a posture of the anteverted pelvis with an extended and rotated trunk). It should be noted that the influence of features is dependent on datasets. It means that for new training datasets built in later practical applications, the effect of each feature would be dynamically changed. However, as a novel scoring approach, the feature selection has to be conducted by the proposed approach itself, and this is why we did not use the suggested ReliefF-DBI in this study.

The other customized algorithm is the FFNN kernel, and its overall performance has been listed in TABLE A4. Compared with other machine learning algorithms, the FFNN kernel proposed promising performance in accuracy (around 2.9% increment). Although the FFNN algorithm did not rank first for all motor tasks, the most important for practical applications is that the utilized scoring algorithm should be able to handle various types of gathered movement data. The FFNN algorithm was therefore selected based on its three advantages. Firstly, the FFNN algorithm has a simple artificial neural network structure that can be easily built and maintained during practical applications. Secondly, as mentioned above (TABLE A2), the FFNN algorithm achieves balanced performance for all seven evaluation metrics, making it possible to evaluate new trials of stroke survivors in practical applications. Lastly, the FFNN algorithm is more suitable for processing discrete datasets of possible values to transfer to related classes than non-artificial neural network algorithms. The possible reason might be that we used larger numbers of features (at least 20 features) to create datasets for training and validation, which is beyond their abilities [68] (we used original versions of the selected algorithms for performance comparison). For example, the decision tree will create an over-complex tree due to various features, which might be unstable and easily overfit.

In terms of clinical applications, this study simplifies the process for patients by requiring them to complete a series of upper limb functional movements based on prompts. The system then automatically processes the data, extracts features, and generates clinical scores. Physiotherapists can use these scores, combined with their expertise, to develop customized treatment plans. The proposed method offers several advantages in clinical application. Firstly, in comparison to biomarker technologies like EMG and EEG, this method is cost-effective, easy to operate, and highly suitable for widespread adoption in clinical practice, particularly in community settings. Secondly, it alleviates the need for participants to wear any devices, thereby overcoming the limitations associated with wearable technologies. Thirdly, the proposed method continuously enhances scoring accuracy by utilizing an AI model retrained with additional movement data from patients, thereby ensuring ongoing improvement and precision in the scoring approach.

Regarding model selection, for the ordinal scoring system (0-5 points) of this study, the regression model may theoretically be a more intuitive choice, as it can naturally handle ordinal data and is more conducive to processing the ordinality between scores. However, existing research has employed classification methods in the automated motor function assessment system for stroke survivors [18][19]. A classification model was therefore utilized in this study. Based on Figs. 5 and 6, the actual performance of the proposed model demonstrates its robustness in scoring WMFT-FAS motor tasks, providing valuable insights for the motor function assessment of stroke survivors.

Three major limitations should be highlighted: 1) in line with many studies [14][36], the quantity and quality of the recruited participants might affect the performance of the proposed approach, such as the study did not include age-matched healthy and stroke participants; 2) the gold standard of this study is linked to the personal feelings of physiotherapists, which is an unavoidable aspect of existing intelligent scoring systems; and 3) a classification model was adopted in the ordinal scoring system, but a regression model might theoretically be a more suitable choice, especially for model training. Future work will focus on overcoming these limitations: 1) continual efforts will be made to recruit more stroke survivors and healthy subjects with similar ethnicity, gender, and mean age to improve the study's robustness; 2) two or more experienced physiotherapists will be invited to score simultaneously, and averaged scores will be used as the gold standard to enhance objectivity; 3) consider using regression models and further explore and compare the applicability of regression models and classification models for the ordinal scoring system; and 4) new devices, such as the Ultraleap Stereo IR170, will be employed to improve scoring precision by leveraging their new features, particularly hand movements.

The main improvements of the current study when compared with our previous work [5] can be concluded as follows: 1) the current study automated the entire WMFT-FAS scale (15 tasks) rather than four tasks in our previous work; and 2) the current study can dynamically select suitable features to enhance the classification/estimation performance by using the customized multi-ReliefF method (Table A3, the accuracy increased by around 2.9%).

5 Conclusion

This paper first proposed a novel scoring approach for the motor function assessment of stroke survivors based on the WMFT-FAS. The approach mainly contained one Microsoft Kinect v2, one customized motion tracking system, and one customized intelligent scoring system. Sixteen stroke survivors and ten healthy subjects were recruited for data collection and validation. Promising experimental results were achieved, indicating the developed approach has the potential for clinical applications as an affordable and portable assessment tool to help physiotherapists make treatment decisions, releasing the burden of healthcare resources. The proposed novel scoring approach can also monitor the progress and outcomes of stroke survivors' rehabilitation interventions in community-based settings.

APPENDIX

TABLE A1 THE WMFT-FAS SCORES OF ALL STROKE SURVIVORS

ID	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16
T1	4	3	4	3	3	5	4	4	3	3	4	4	3	4	3	4
T2	4	0	4	3	3	5	4	3	2	3	4	4	3	4	3	4
T3	4	0	4	3	3	5	4	3	3	3	4	4	3	4	3	4
T4	4	0	4	3	3	4	4	3	3	4	4	4	3	4	3	4
T5	4	4	4	3	3	5	4	4	3	3	4	4	3	4	3	4
T6	3	4	4	2	3	4	4	4	3	3	4	4	3	4	3	4
T7	3	3	4	2	3	4	4	4	3	3	4	4	3	4	3	4
T8	2	0	4	2	3	4	3	3	3	3	4	4	2	4	2	4
T9	3	3	4	2	3	4	4	3	3	3	4	4	3	4	3	4
T10	3	2	4	2	3	5	4	4	2	3	4	4	3	4	3	4
T11	3	2	4	2	3	5	4	4	3	3	4	4	1	4	3	4
T12	2	2	4	2	3	5	3	3	2	2	3	4	2	4	2	4
T13	3	0	4	2	2	4	4	3	2	4	3	4	3	4	3	2
T14	3	0	4	1	2	4	3	3	1	3	4	4	2	4	2	4
T15	3	0	4	3	3	4	4	4	2	3	3	4	3	4	3	4
Total	48	23	60	35	43	67	57	52	38	46	57	60	40	60	42	58

TABLE A2
PERFORMANCE OF THE PROPOSED APPROACH

Task	[a]Acc	[b]F1	[c]Sen	[d]Spe	[e]AUC	[f]MAE	[g]RMSE
T1	0.924	0.872	0.862	0.961	0.934	0.079	0.124
T2	0.941	0.866	0.923	0.978	0.957	0.072	0.208
T3	0.959	0.951	0.980	0.984	0.957	0.057	0.122
T4	0.928	0.906	0.934	0.962	0.951	0.121	0.280
T5	0.945	0.916	0.923	0.969	0.992	0.066	0.182
T6	0.926	0.864	0.875	0.978	0.931	0.075	0.191
T7	0.894	0.852	0.893	0.947	0.927	0.133	0.288
T8	0.940	0.883	0.956	0.983	0.985	0.090	0.209
T9	0.940	0.934	0.940	0.975	0.976	0.079	0.167
T10	0.879	0.741	0.841	0.954	0.957	0.138	0.278
T11	0.935	0.876	0.952	0.979	0.974	0.083	0.180
T12	0.883	0.751	0.803	0.955	0.871	0.150	0.312
T13	0.913	0.884	0.969	0.976	0.983	0.157	0.330
T14	0.966	0.967	0.967	0.984	1.000	0.040	0.063
T15	0.892	0.855	0.906	0.954	0.880	0.131	0.272
Mean	0.924	0.875	0.915	0.969	0.952	0.098	0.214
[h]SD	0.027	0.063	0.051	0.013	0.038	0.037	0.078

[a]Acc = accuracy; [b]F1 = F1-score; [c]Sen = Sensitivity; [d]Spe = Specificity; [e]AUC = AUC; [f]MAE = mean absolute error; [g]RMSE = root mean squared error; [h]SD = standard deviation.

TABLE A3
ACCURACY COMPARISON BETWEEN THE ORIGINAL, RELIEFF-DBI, AND MULTI-RELIEFF METHODS

Task	[a]O	[b]RD	[c]MR	Ovs.MR	RDvs.MR
T1	0.864	0.909	0.924	0.060	0.015
T2	0.929	0.824	0.941	0.012	0.117
T3	0.905	0.784	0.959	0.054	0.175
T4	0.870	0.797	0.928	0.058	0.131
T5	0.959	0.959	0.945	-0.014	-0.014
T6	0.901	0.815	0.926	0.025	0.111
T7	0.909	0.773	0.894	-0.015	0.121
T8	0.866	0.806	0.940	0.074	0.134
T9	0.925	0.836	0.940	0.015	0.104
T10	0.864	0.879	0.879	0.015	0.000
T11	0.823	0.952	0.935	0.112	-0.017
T12	0.917	0.917	0.883	-0.034	-0.034
T13	0.942	0.884	0.913	-0.029	0.029
T14	0.914	0.879	0.966	0.052	0.087
T15	0.877	0.815	0.892	0.015	0.077
Mean	0.898	0.855	0.924	0.027	0.069
[d]SD	0.036	0.060	0.027	0.041	0.066

[a]O = Original; [b]RD = ReliefF-DBI; [c]MR = Multi-ReliefF; [d]SD = standard deviation;

TABLE A4
TOP THREE ACCURACY PERFORMANCE OF DIFFERENT ALGORITHMS FOR EACH MOTOR TASK

Task	Top1	Top2	Top3
T1	FFNN	MNN	CFNN
T2	NN	FFNN	MNN, RNN
T3	FFNN	RNN	CFNN
T4	FFNN, RNN	NN, MNN, KNN	CFNN
T5	MNN	CFNN	FFNN, RNN
T6	RNN	FFNN, CFNN	NN
T7	MNN, CFNN, RNN	FFNN	NN
T8	FFNN	MNN	CFNN
T9	NN	RNN	FFNN
T10	RNN	MNN	FFNN, NN
T11	NN	FFNN	RNN
T12	RNN	MNN	CFNN
T13	CFNN	FFNN	MNN
T14	FFNN	NN, MNN, RNN	CFNN
T15	NN, CFNN	FFNN	RNN

TABLE A5 THE PERFORMANCE OF DIFFERENT ALGORITHMS FOR TASK-1

Name	[a]Acc	[b]F1	[c]Spe	[d]Sen	[e]AUC
FFNN	0.924	0.872	0.961	0.862	0.934
NN	0.864	0.810	0.937	0.855	0.941
MNN	0.909	0.870	0.954	0.883	0.935
CFNN	0.894	0.879	0.938	0.899	0.959
RNN	0.879	0.850	0.932	0.870	0.949
DT	0.636	0.528	0.795	0.535	0.717
DA	0.727	0.660	0.818	0.656	0.780
NB	0.636	0.498	0.789	0.507	0.710
SVM	0.788	0.664	0.858	0.643	0.860
KNN	0.818	0.722	0.864	0.688	0.777
EM	0.742	0.637	0.818	0.635	0.860

[a]Acc = accuracy; [b]F1 = F1-score; [c]Spe = Specificity; [d]Sen = Sensitivity; [e]AUC = AUC.

TABLE A6 THE PERFORMANCE OF DIFFERENT ALGORITHMS FOR TASK-2

Name	[*]Acc	[*]F1	[*]Spe	[*]Sen	[*]AUC
FFNN	0.941	0.866	0.978	0.923	0.957
NN	0.976	0.970	0.988	0.983	0.968
MNN	0.918	0.822	0.961	0.829	0.981
CFNN	0.871	0.701	0.939	0.693	0.933
RNN	0.918	0.894	0.963	0.912	0.925
DT	0.753	0.544	0.888	0.527	0.645
DA	0.706	[#]N/A	0.854	0.418	0.585
NB	0.718	0.561	0.898	0.545	0.720
SVM	0.824	0.618	0.925	0.620	0.885
KNN	0.729	0.558	0.843	0.597	0.720
EM	0.788	0.585	0.877	0.569	0.855

[*]Same abbreviation as Table A2, [#]N/A = Not Applicable.

TABLE A7 THE PERFORMANCE OF DIFFERENT ALGORITHMS FOR TASK-3

Name	*Acc	*F1	*Spe	*Sen	*AUC
FFNN	0.959	0.951	0.984	0.980	0.957
NN	0.865	0.804	0.928	0.851	0.951
MNN	0.851	0.799	0.914	0.829	0.910
CFNN	0.892	0.869	0.928	0.868	0.969
RNN	0.905	0.869	0.954	0.887	0.944
DT	0.689	0.609	0.803	0.606	0.727
DA	0.824	0.740	0.874	0.715	0.807
NB	0.703	0.644	0.824	0.679	0.840
SVM	0.784	0.673	0.849	0.642	0.913
KNN	0.716	0.524	0.754	0.491	0.877
EM	0.811	0.715	0.861	0.682	0.877

*Same abbreviation as Table A2.

TABLE A8 THE PERFORMANCE OF DIFFERENT ALGORITHMS FOR TASK-4

Name	*Acc	*F1	*Spe	*Sen	*AUC
FFNN	0.928	0.906	0.962	0.934	0.951
NN	0.899	0.873	0.950	0.903	0.953
MNN	0.899	0.876	0.938	0.872	0.976
CFNN	0.870	0.833	0.925	0.842	0.960
RNN	0.928	0.922	0.955	0.936	0.980
DT	0.652	0.516	0.745	0.490	0.647
DA	0.696	0.615	0.811	0.605	0.797
NB	0.696	0.617	0.827	0.643	0.867
SVM	0.812	0.751	0.867	0.730	0.877
KNN	0.899	0.832	0.944	0.817	0.863
EM	0.826	0.780	0.880	0.777	0.890

*Same abbreviation as Table A2.

TABLE A9 THE PERFORMANCE OF DIFFERENT ALGORITHMS FOR TASK-5

Name	*Acc	*F1	*Spe	*Sen	*AUC
FFNN	0.945	0.916	0.969	0.923	0.992
NN	0.918	0.909	0.956	0.941	0.989
MNN	0.986	0.968	0.995	0.980	1.000
CFNN	0.959	0.929	0.984	0.945	0.981
RNN	0.945	0.912	0.970	0.948	0.996
DT	0.740	0.640	0.811	0.647	0.767
DA	0.753	0.620	0.849	0.612	0.827
NB	0.753	0.718	0.831	0.710	0.893
SVM	0.836	0.756	0.858	0.702	0.923
KNN	0.877	0.786	0.891	0.729	0.810
EM	0.863	0.799	0.877	0.750	0.950

*Same abbreviation as Table A2.

TABLE A10 THE PERFORMANCE OF DIFFERENT ALGORITHMS FOR TASK-6

Name	*Acc	*F1	*Spe	*Sen	*AUC
FFNN	0.926	0.864	0.978	0.875	0.931
NN	0.889	0.853	0.955	0.882	0.959
MNN	0.864	#N/A	0.938	N/A	0.923
CFNN	0.926	N/A	0.972	N/A	0.733
RNN	0.951	N/A	0.985	N/A	0.981
DT	0.765	N/A	0.896	0.549	0.825
DA	N/A	N/A	N/A	N/A	N/A
NB	0.679	N/A	0.878	0.376	0.533
SVM	0.877	N/A	0.936	0.576	0.695
KNN	0.864	N/A	0.927	0.540	0.733
EM	0.815	N/A	0.900	0.544	0.678

*Same abbreviation as Table A2, #N/A = Not Applicable.

TABLE A11 THE PERFORMANCE OF DIFFERENT ALGORITHMS FOR TASK-7

Name	*Acc	*F1	*Spe	*Sen	*AUC
FFNN	0.894	0.852	0.947	0.893	0.927
NN	0.864	0.812	0.924	0.839	0.974
MNN	0.924	0.879	0.970	0.909	0.964
CFNN	0.924	0.906	0.944	0.877	0.982
RNN	0.924	0.871	0.961	0.893	0.959
DT	0.682	0.573	0.812	0.589	0.747
DA	0.773	0.681	0.873	0.684	0.803
NB	0.682	0.598	0.812	0.621	0.747
SVM	0.773	0.693	0.840	0.667	0.877
KNN	0.758	0.665	0.817	0.630	0.727
EM	0.773	0.678	0.807	0.635	0.860

*Same abbreviation as Table A2.

TABLE A12 THE PERFORMANCE OF DIFFERENT ALGORITHMS FOR TASK-8

Name	*Acc	*F1	*Spe	*Sen	*AUC
FFNN	0.940	0.883	0.983	0.956	0.985
NN	0.821	0.671	0.951	0.853	0.921
MNN	0.896	0.895	0.955	0.920	0.970
CFNN	0.866	0.783	0.942	0.783	0.950
RNN	0.821	0.672	0.939	0.856	0.955
DT	0.716	#N/A	0.886	0.459	0.800
DA	0.702	0.598	0.873	0.606	0.743
NB	0.657	0.490	0.850	0.492	0.578
SVM	0.851	0.729	0.927	0.720	0.890
KNN	0.821	0.619	0.912	0.592	0.753
EM	0.776	0.595	0.862	0.574	0.793

*Same abbreviation as Table A2, #N/A = Not Applicable.

TABLE A13 THE PERFORMANCE OF DIFFERENT ALGORITHMS FOR TASK-9

Name	*Acc	*F1	*Spe	*Sen	*AUC
FFNN	0.940	0.934	0.975	0.940	0.976
NN	0.970	0.963	0.991	0.975	1.000
MNN	0.866	#N/A	0.940	N/A	0.933
CFNN	0.910	0.888	0.966	0.910	0.980
RNN	0.955	N/A	0.987	N/A	0.980
DT	0.731	N/A	0.893	0.484	0.830
DA	N/A	N/A	N/A	N/A	N/A
NB	0.746	N/A	0.867	0.445	0.803
SVM	0.791	N/A	0.893	0.484	0.885
KNN	0.761	N/A	0.876	0.451	0.663
EM	0.851	N/A	0.923	0.556	0.688

*Same abbreviation as Table A2, #N/A = Not Applicable.

TABLE A14 THE PERFORMANCE OF DIFFERENT ALGORITHMS FOR TASK-10

Name	*Acc	*F1	*Spe	*Sen	*AUC
FFNN	0.879	0.741	0.954	0.841	0.957
NN	0.879	0.795	0.952	0.800	0.947
MNN	0.924	0.912	0.965	0.892	0.965
CFNN	0.894	0.752	0.969	0.781	0.927
RNN	0.955	0.967	0.980	0.971	0.982
DT	0.758	#N/A	0.911	0.419	0.863
DA	0.758	N/A	0.893	0.533	0.675
NB	0.621	N/A	0.853	0.336	0.583
SVM	0.803	N/A	0.902	0.464	0.838
KNN	0.742	N/A	0.873	0.420	0.648
EM	0.773	0.544	0.901	0.536	0.890

*Same abbreviation as Table A2, #N/A = Not Applicable.

TABLE A15 THE PERFORMANCE OF DIFFERENT ALGORITHMS FOR TASK-11

Name	*Acc	*F1	*Spe	*Sen	*AUC
FFNN	0.935	0.876	0.979	0.952	0.979
NN	0.968	0.964	0.993	0.991	0.999
MNN	0.855	N/A	0.961	N/A	0.927
CFNN	0.903	N/A	0.972	N/A	0.956
RNN	0.919	0.875	0.965	0.919	0.989
DT	0.855	N/A	0.960	0.524	0.886
DA	N/A	N/A	N/A	N/A	N/A
NB	0.790	N/A	0.937	0.352	0.598
SVM	0.855	N/A	0.941	0.550	0.690
KNN	0.758	N/A	0.905	0.498	0.700
EM	0.823	N/A	0.945	0.362	0.912

*Same abbreviation as Table A2, #N/A = Not Applicable.

TABLE A16 THE PERFORMANCE OF DIFFERENT ALGORITHMS FOR TASK-12

Name	*Acc	*F1	*Spe	*Sen	*AUC
FFNN	0.883	0.751	0.955	0.803	0.871
NN	0.867	0.744	0.931	0.733	0.941
MNN	0.917	0.792	0.967	0.807	0.984
CFNN	0.900	0.800	0.951	0.816	0.946
RNN	0.983	0.953	0.995	0.950	1.000
DT	0.767	0.503	0.885	0.510	0.803
DA	0.700	0.517	0.867	0.519	0.705
NB	0.800	0.578	0.924	0.593	0.878
SVM	0.817	0.664	0.889	0.646	0.845
KNN	0.783	0.544	0.900	0.521	0.713
EM	0.833	0.711	0.934	0.727	0.918

*Same abbreviation as Table A2.

TABLE A17 THE PERFORMANCE OF DIFFERENT ALGORITHMS FOR TASK-13

Name	*Acc	*F1	*Spe	*Sen	*AUC
FFNN	0.913	0.884	0.976	0.969	0.983
NN	0.870	0.812	0.951	0.857	0.977
MNN	0.899	0.874	0.959	0.899	0.941
CFNN	0.928	0.914	0.969	0.912	0.984
RNN	0.855	0.826	0.951	0.904	0.929
DT	0.652	0.468	0.839	0.510	0.670
DA	0.768	0.776	0.886	0.773	0.875
NB	0.768	0.721	0.901	0.704	0.823
SVM	0.783	0.709	0.885	0.645	0.915
KNN	0.841	0.779	0.918	0.755	0.838
EM	0.855	0.839	0.917	0.810	0.918

*Same abbreviation as Table A2.

TABLE A18 THE PERFORMANCE OF DIFFERENT ALGORITHMS FOR TASK-14

Name	*Acc	*F1	*Spe	*Sen	*AUC
FFNN	0.966	0.967	0.984	0.967	1.000
NN	0.931	0.911	0.986	0.983	0.983
MNN	0.931	0.852	0.976	0.907	0.988
CFNN	0.862	0.738	0.946	0.839	0.915
RNN	0.931	0.850	0.985	0.947	0.991
DT	0.724	#N/A	0.905	0.417	0.622
DA	0.759	0.629	0.894	0.612	0.704
NB	0.724	N/A	0.870	0.286	0.404
SVM	0.776	N/A	0.890	0.324	0.894
KNN	0.793	N/A	0.903	0.398	0.650
EM	0.862	0.664	0.944	0.686	0.910

*Same abbreviation as Table A2, #N/A = Not Applicable.

TABLE A19 THE PERFORMANCE OF DIFFERENT ALGORITHMS FOR TASK-15

Name	*Acc	*F1	*Spe	*Sen	*AUC
FFNN	0.892	0.855	0.954	0.906	0.880
NN	0.923	0.832	0.978	0.845	0.976
MNN	0.815	0.713	0.917	0.789	0.925
CFNN	0.923	0.879	0.970	0.942	0.930
RNN	0.877	0.838	0.939	0.848	0.917
DT	0.631	#N/A	0.844	0.285	0.725
DA	0.677	0.484	0.844	0.467	0.758
NB	0.708	N/A	0.866	0.459	0.798
SVM	0.708	N/A	0.825	0.378	0.768
KNN	0.692	N/A	0.786	0.308	0.673
EM	0.754	0.520	0.867	0.486	0.820

*Same abbreviation as Table A2, #N/A = Not Applicable.

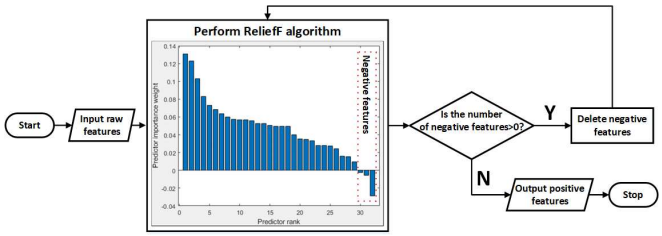

Fig. A1. The flowchart of the multi-ReliefF method.

ETHICS STATEMENT

The ethical approval has been approved by the Ethics Advisory Committee of Fujian University of Traditional Chinese Medicine (No. 2018KY-019-02). Informed consent was obtained from all participants.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CREDIT AUTHORSHIP CONTRIBUTION STATEMENT

Bo Sheng: Conceptualization, Formal analysis, Methodology, Software, Validation, Writing – original draft. **Xiaohui Chen:** Software, Formal analysis, Writing– revised draft. **Jian Cheng:** Investigation, Data curation, Writing – original draft. **Yanxin Zhang:** Supervision, Writing – review & editing. **Shane (Sheng Quan) Xie:** Supervision, Writing – review & editing. **Jing Tao:** Methodology, Software, Validation – review & editing. **Chaoqun Duan:** Supervision, Writing – review & editing.

ACKNOWLEDGMENT

The authors would like to thank Jennifer Qiao and Rylea Hart for the English improvement and Zhenhui Li for the clinical data collection. This work was supported by the National Natural Science Foundation of China under Grants 62103252, 12002177, 51875358, and 62033001, and the Shanghai Pujiang Program under Grant 21PJ1404000.

REFERENCES

[1] A. de los Reyes-Guzmán, I. Dimbwadyo-Terrer, F. Trincado-Alonso, F. Monasterio-Huelin, D. Torricelli, and A. Gil-Agudo, "Quantitative assessment based on kinematic measures of functional impairments during upper extremity movements: A review," *Clin. Biomech.*, vol. 29, no. 7, pp. 719–727, 2014, doi: 10.1016/j.clinbiomech.2014.06.013.

[2] L. Yu, D. Xiong, L. Guo, and J. Wang, "A remote quantitative Fugl-Meyer assessment framework for stroke patients based on wearable sensor networks," *Comput. Methods Programs Biomed.*, vol. 128, pp. 100–110, 2016.

[3] D. J. Gladstone, C. J. Danells, and S. E. Black, "The Fugl-Meyer assessment of motor recovery after stroke: a critical review of its measurement properties," *Neurorehabil. Neural Repair*, vol. 16, no. 3, pp. 232–240, 2002, doi: 10.1177/154596802401105171.

[4] S. V Duff *et al.*, "Interrater reliability of the Wolf Motor Function Test–Functional Ability Scale: Why it matters," *Neurorehabil. Neural Repair*, vol. 29, no. 5, pp. 436–443, 2015, doi: 10.1177/1545968314553030.

[5] B. Sheng, X. Wang, M. Hou, J. Huang, S. Xiong, and Y. Zhang, "An automated system for motor function assessment in stroke patients using motion sensing technology: A pilot study," *Measurement*, vol. 161, p. 107896, 2020.

[6] H. Yang, J. Wan, Y. Jin, X. Yu, and Y. Fang, "EEG and EMG Driven Post-Stroke Rehabilitation: A Review," *IEEE Sens. J.*, 2022.

[7] Á. Costa, M. Itkonen, H. Yamasaki, F. S. Alnajjar, and S. Shimoda, "Importance of muscle selection for EMG signal analysis during upper limb rehabilitation of stroke patients," in *2017 39th Annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, 2017, pp. 2510–2513.

[8] M. Munoz-Novoa, M. B. Kristoffersen, K. S. Sunnerhagen, A. Naber, M. Alt Murphy, and M. Ortiz-Catalan, "Upper Limb Stroke Rehabilitation

Using Surface Electromyography: A Systematic Review and Meta-Analysis," *Front. Hum. Neurosci.*, vol. 16, p. 897870, 2022.

[9] N. Riahi, V. A. Vakorin, and C. Menon, "Estimating Fugl-Meyer upper extremity motor score from functional-connectivity measures," *IEEE Trans. neural Syst. Rehabil. Eng.*, vol. 28, no. 4, pp. 860–868, 2020.

[10] X. Zhang, R. D'Arcy, and C. Menon, "Scoring upper-extremity motor function from EEG with artificial neural networks: a preliminary study," *J. Neural Eng.*, vol. 16, no. 3, p. 36013, 2019.

[11] X. Zhang, R. D'Arcy, L. Chen, M. Xu, D. Ming, and C. Menon, "The Feasibility of Longitudinal Upper Extremity Motor Function Assessment Using EEG," *Sensors*, vol. 20, no. 19, p. 5487, 2020.

[12] T. Kawano *et al.*, "Electroencephalographic phase synchrony index as a biomarker of poststroke motor impairment and recovery," *Neurorehabil. Neural Repair*, vol. 34, no. 8, pp. 711–722, 2020.

[13] X. Zhang, X. Yong, and C. Menon, "Evaluating the versatility of EEG models generated from motor imagery tasks: An exploratory investigation on upper-limb elbow-centered motor imagery tasks," *PLoS One*, vol. 12, no. 11, p. e0188293, 2017.

[14] C. W. Wu *et al.*, "Synchrony between default-mode and sensorimotor networks facilitates motor function in stroke rehabilitation: a pilot fMRI study," *Front. Neurosci.*, vol. 14, p. 548, 2020.

[15] M. Antico *et al.*, "Postural control assessment via Microsoft Azure Kinect DK: An evaluation study," *Comput. Methods Programs Biomed.*, vol. 209, p. 106324, 2021.

[16] X. Song, S. Chen, J. Jia, and P. B. Shull, "Cellphone-Based Automated Fugl-Meyer Assessment to Evaluate Upper Extremity Motor Function After Stroke," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 10, pp. 2186–2195, 2019.

[17] E. Martini *et al.*, "Enabling Gait Analysis in the Telemedicine Practice through Portable and Accurate 3D Human Pose Estimation," *Comput. Methods Programs Biomed.*, vol. 225, p. 107016, 2022.

[18] P. Otten, J. Kim, and S. H. Son, "A framework to automate assessment of upper-limb motor function impairment: A feasibility study," *Sensors*, vol. 15, no. 8, pp. 20097–20114, 2015.

[19] S. Lee, Y.-S. Lee, and J. Kim, "Automated evaluation of upper-limb motor function impairment using Fugl-Meyer assessment," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 1, pp. 125–134, 2017.

[20] H. Yan, B. Hu, G. Chen, and E. Zhengyuan, "Real-time continuous human rehabilitation action recognition using OpenPose and FCN," in *2020 3rd International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE)*, 2020, pp. 239–242.

[21] V. F. Bento, V. T. Cruz, D. D. Ribeiro, and J. P. S. Cunha, "Towards a movement quantification system capable of automatic evaluation of upper limb motor function after neurological injury," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2011, pp. 5456–5460.

[22] V. T. Cruz, V. F. Bento, D. D. Ribeiro, I. Araújo, C. A. Branco, and P. Coutinho, "A novel system for automatic classification of upper limb motor function after stroke: an exploratory study," *Med. Eng. Phys.*, vol. 36, no. 12, pp. 1704–1710, 2014.

[23] E. Wade, A. R. Parnandi, and M. J. Mataric, "Automated administration of the wolf motor function test for post-stroke assessment," in *2010 4th International Conference on Pervasive Computing Technologies for Healthcare*, 2010, pp. 1–7.

[24] A. Parnandi, E. Wade, and M. Matarić, "Motor function assessment using wearable inertial sensors," in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, 2010, pp. 86–89.

[25] S. Patel *et al.*, "A novel approach to monitor rehabilitation outcomes in stroke survivors using wearable technology," *Proc. IEEE*, vol. 98, no. 3, pp. 450–461, 2010, doi: 10.1109/JPROC.2009.2038727.

[26] S. Patel *et al.*, "Tracking motor recovery in stroke survivors undergoing rehabilitation using wearable technology," in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, 2010, pp. 6858–6861.

[27] K. Bogard, S. Wolf, Q. Zhang, P. Thompson, D. Morris, and D. Nichols-Larsen, "Can the wolf motor function test be streamlined?," *Neurorehabil. Neural Repair*, vol. 23, no. 5, pp. 422–428, 2009, doi: 10.1177/1545968308331141.

[28] L. Cai, Y. Ma, S. Xiong, and Y. Zhang, "Validity and Reliability of Upper Limb Functional Assessment Using the Microsoft Kinect V2 Sensor," *Appl. bionics Biomech.*, vol. 2019, 2019, doi: 10.1155/2019/7175240.

[29] D. H. Schoellhamer, "Singular spectrum analysis for time series with missing data," *Geophys. Res. Lett.*, vol. 28, no. 16, pp. 3187–3190, 2001.

[30] F. J. Alonso, J. M. Del Castillo, and P. Pintado, "Application of singular spectrum analysis to the smoothing of raw kinematic signals," *J. Biomech.*, vol. 38, no. 5, pp. 1085–1092, 2005.

[31] H. Hassani, "Singular spectrum analysis: methodology and comparison," 2007.

[32] A. Ozturk, A. Tartar, B. E. Huseyinsinoglu, and A. H. Ertas, "A clinically feasible kinematic assessment method of upper extremity motor function impairment after stroke," *Measurement*, vol. 80, pp. 207–216, 2016.

[33] W. D. Schot, E. Brenner, and J. B. J. Smeets, "Robust movement segmentation by combining multiple sources of information," *J. Neurosci. Methods*, vol. 187, no. 2, pp. 147–155, 2010.

[34] D. G. Liebermann, S. Berman, P. L. Weiss, and M. F. Levin, "Kinematics of reaching movements in a 2-D virtual environment in adults with and without stroke," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 20, no. 6, pp. 778–787, 2012.

[35] S. Balasubramanian, A. Melendez-Calderon, and E. Burdet, "A robust and sensitive metric for quantifying

movement smoothness," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 8, pp. 2126–2136, 2011.

[36] E. Vergaro, M. Casadio, V. Squeri, P. Giannoni, P. Morasso, and V. Sanguineti, "Self-adaptive robot training of stroke survivors for continuous tracking movements," *J. Neuroeng. Rehabil.*, vol. 7, no. 1, pp. 1–12, 2010.

[37] B. Rohrer *et al.*, "Movement smoothness changes during stroke recovery," *J. Neurosci.*, vol. 22, no. 18, pp. 8297–8304, 2002.

[38] N. Hogan and D. Sternad, "Sensitivity of smoothness measures to movement duration, amplitude, and arrests," *J. Mot. Behav.*, vol. 41, no. 6, pp. 529–534, 2009.

[39] D. A. Winter, *Biomechanics and motor control of human movement*. John wiley & sons, 2009.

[40] B. Sheng, S. Xiong, X. Wang, M. Hou, and Y. Zhang, "Kinematic Metrics for Upper-limb Functional Assessment of Stroke Patients," in *the International Conference on Intelligent Informatics and BioMedical Sciences*, 2019. doi: 10.1109/ICIIBMS46890.2019.8991507.

[41] M. Alt Murphy, C. Willén, and K. S. Sunnerhagen, "Responsiveness of upper extremity kinematic measures and clinical improvement during the first three months after stroke," *Neurorehabil. Neural Repair*, vol. 27, no. 9, pp. 844–853, 2013.

[42] J. C. Bezdek and N. R. Pal, "Some new indexes of cluster validity," *IEEE Trans. Syst. Man, Cybern. Part B*, vol. 28, no. 3, pp. 301–315, 1998.

[43] A. Stief, J. R. Ottewill, and J. Baranowski, "Relief F-based feature ranking and feature selection for monitoring induction motors," in *2018 23rd International Conference on Methods {\&} Models in Automation {\&} Robotics (MMAR)*, 2018, pp. 171–176.

[44] M. Robnik-Šikonja and I. Kononenko, "Theoretical and empirical analysis of ReliefF and RReliefF," *Mach. Learn.*, vol. 53, no. 1, pp. 23–69, 2003.

[45] M. Robnik-Šikonja and I. Kononenko, "An adaptation of Relief for attribute estimation in regression," in *Machine Learning: Proceedings of the Fourteenth International Conference (ICML'97)*, 1997, vol. 5, pp. 296–304.

[46] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, 2015, doi: 10.1016/j.neunet.2014.09.003.

[47] E. I. Georga, D. I. Fotiadis, and S. K. Tigas, *Personalized Predictive Modeling in Type 1 Diabetes*, 1st editio. Massachusetts: Academic Press, 2017.

[48] D. Svozil, V. Kvasnicka, and J. Pospichal, "Introduction to multi-layer feed-forward neural networks," *Chemom. Intell. Lab. Syst.*, vol. 39, no. 1, pp. 43–62, 1997.

[49] Z. Zhang, L. Liparulo, M. Panella, X. Gu, and Q. Fang, "A fuzzy kernel motion classifier for autonomous stroke rehabilitation," *IEEE J. Biomed. Heal. informatics*, vol. 20, no. 3, pp. 893–901, 2015.

[50] Y. Jiang *et al.*, "Seizure classification from EEG signals using transfer learning, semi-supervised learning and TSK fuzzy system," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 12, pp. 2270–2284, 2017.

[51] M. Z. Alom, C. Yakopcic, M. S. Nasrin, T. M. Taha, and V. K. Asari, "Breast cancer classification from histopathological images with inception recurrent residual convolutional neural network," *J. Digit. Imaging*, vol. 32, pp. 605–617, 2019.

[52] Y. Zhang and Y. Ma, "Application of supervised machine learning algorithms in the classification of sagittal gait patterns of cerebral palsy children with spastic diplegia," *Comput. Biol. Med.*, vol. 106, pp. 33–39, 2019, doi: 10.1016/j.compbiomed.2019.01.009.

[53] The MathWorks, "Assess Classifier Performance in Classification Learner," 2021.

[54] A. Aggarwal and E. Kean, "Comparison of the Folstein Mini Mental State Examination (MMSE) to the Montreal Cognitive Assessment (MoCA) as a cognitive screening tool in an inpatient rehabilitation setting," *Neurosci. Med.*, vol. 1, no. 02, pp. 39–42, 2010.

[55] N. Carson, L. Leach, and K. J. Murphy, "A re-examination of Montreal Cognitive Assessment (MoCA) cutoff scores," *Int. J. Geriatr. Psychiatry*, vol. 33, no. 2, pp. 379–388, 2018.

[56] Z. Zhang, Q. Fang, and X. Gu, "Objective assessment of upper-limb mobility for poststroke rehabilitation," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 4, pp. 859–868, 2015, doi: 10.1109/TBME.2015.2477095.

[57] Z. Zhang, Q. Fang, and X. Gu, "Fuzzy inference system based automatic Brunnstrom stage classification for upper-extremity rehabilitation," *Expert Syst. Appl.*, vol. 41, no. 4, pp. 1973–1980, 2014, doi: 10.1016/j.eswa.2013.08.094.

[58] J. D. Mejia-Trujillo *et al.*, "Kinect™ and Intel RealSense™ D435 comparison: a preliminary study for motion analysis," in *2019 IEEE International Conference on E-health Networking, Application {\&} Services (HealthCom)*, 2019, pp. 1–4.

[59] P. Subedi, "Machine Learning — The different ways to evaluate your Classification models and choose the best one!" 2018. Accessed: Apr. 21, 2021. [Online]. Available: https://medium.com/kharpann/machine-learning-the-different-ways-to-evaluate-your-classification-models-and-choose-the-best-1281542432c

[60] M. Grandini, E. Bagli, and G. Visani, "Metrics for multi-class classification: an overview," *arXiv Prepr. arXiv2008.05756*, 2020.

[61] E. D. Ona, A. Jardón, E. Monge, F. Molina, R. Cano, and C. Balaguer, "Towards automated assessment of upper limbs motor function based on fugl-meyer test and virtual environment," in *International Conference on NeuroRehabilitation*, 2018, pp. 297–301.

[62] E. D. Oña, P. Sánchez-Herrera, A. Cuesta-Gómez, S. Martinez, A. Jardón, and C. Balaguer, "Automatic outcome in manual dexterity assessment using colour segmentation and nearest neighbour classifier," *Sensors*, vol. 18, no. 9, p. 2876, 2018, doi: 10.3390/s18092876.

[63] A. Scano, A. Chiavenna, M. Malosio, L. M. Tosatti, and F. Molteni, “Kinect V2 implementation and testing of the reaching performance scale for motor evaluation of patients with neurological impairment,” *Med. Eng. Phys.*, vol. 56, pp. 54–58, 2018, doi: 10.1016/j.medengphy.2018.04.005.

[64] E. D. O. Simbaña, P. S.-H. Baeza, A. J. Huete, and C. Balaguer, “Review of automated systems for upper limbs functional assessment in neurorehabilitation,” *IEEE Access*, vol. 7, pp. 32352–32367, 2019.

[65] J. Sarsfield *et al.*, “Clinical assessment of depth sensor based pose estimation algorithms for technology supervised rehabilitation applications,” *Int. J. Med. Inform.*, vol. 121, pp. 30–38, 2019, doi: 10.1016/j.ijmedinf.2018.11.001.

[66] M. Tölgyessy, M. Dekan, Ľ. Chovanec, and P. Hubinský, “Evaluation of the azure Kinect and its comparison to Kinect V1 and Kinect V2,” *Sensors*, vol. 21, no. 2, p. 413, 2021.

[67] K. M. Steele, C. Papazian, and H. A. Feldner, “Muscle activity after stroke: perspectives on deploying surface electromyography in acute care,” *Front. Neurol.*, p. 1076, 2020.

[68] J. Tang, S. Alelyani, and H. Liu, “Feature selection for classification: A review,” *Data Classif. Algorithms Appl.*, p. 37, 2014.