Skeleton Tracking Solutions for a Low-cost Stroke Rehabilitation Support System

Ana Rita Cóias[1], Min Hun Lee[2], Alexandre Bernardino[1], and Asim Smailagic[3]

***Abstract*— Computer systems based on motion assessment are promising solutions to support stroke survivors' autonomous rehabilitation exercises. In this regard, researchers keep trying to achieve engaging and low-cost solutions suitable mainly for home use. Aiming to achieve a system with a minimal technical setup, we compare Microsoft Kinect, OpenPose, and MediaPipe skeleton tracking approaches for upper extremity quality of movement assessment after stroke. We determine if classification models assess accurately exercise performance with OpenPose and MediaPipe data against Kinect, using a dataset of 15 stroke survivors. We compute Root Mean Squared Error to determine the alignment of trajectories and kinematic variables. MediaPipe World Landmarks revealed high alignment with Kinect, revealing to be a potential alternative method.**

I. INTRODUCTION

After a stroke, rehabilitation therapy is crucial to diminish impairments, promote recovery, and prevent stroke recurrence [1]. Therapy requires a lot of time investment and an enormous allocation of human and financial resources [2]. Therapists encounter challenges in addressing the diversified needs of a growing number of patients [2]. Thus, task-oriented training [3] without supervision, or even at home, is often recommended, raising survivors' chances of recovery and easing chronicle stage management [4]. However, when exercising autonomously, survivors struggle to keep their engagement in therapy due to the lack of therapists' constant feedback and encouragement [4]. This absence of therapists' intervention may decrease survivors' compliance with treatment, leading to throwbacks in the recovery process and treatment withdrawals [4].

Based on skeleton tracking, computer systems have emerged as promising solutions to support autonomous stroke rehabilitation at home or healthcare facilities [5] [6] [7]. These aid rehabilitation training execution by providing exercise instructions, feedback on performance, and encouragement, keeping engagement and promoting movement pattern correction [5] [6]. Plus, they can provide therapists with objective information about patients' progress [8]. These systems require specific equipment and software to acquire body joints' pose data (e.g., Microsoft Kinect, OpenPose, and MediaPipe). They generate feedback on performance through kinematic analysis [6]. The accuracy and relevance of such analysis depend on precise skeleton tracking approaches.

Kinematic analysis has been done to characterize impaired motion patterns after stroke. Researchers assessed arm joints' speed profiles, motion angles, and displacements to distinguish impairment levels and healthy controls from impaired subjects, using Kinect for motion tracking [9]. Further, research teams generated performance scores based on kinematic assessment with Kinect, revealing a high agreement with therapist's evaluation [8].

When developing systems to support rehabilitation, a critical requirement is the simplicity and easy usability of solutions, promoting autonomous use while providing accurate feedback [10]. Solutions involving multiple devices (e.g., laptops, tablets, cameras, and objects) [5] may be considered as complex, of complicated use, and less adaptable to diverse settings (e.g., home). Multiple and specialized device usage, such as high-precision marker-based optical systems, implies extra complexity and cost inherent to the final solution.

Microsoft Kinect appeared as a low-cost and ease-of-use sensor and has been widely used in proposed rehabilitation solutions, enabling motion capture without in-body markers or sensors [7] [11]. With Kinect v2 discontinuation, the sensor is no longer officially distributed, its availability in the market decreased significantly, and support is no longer provided [12]. This fact poses a problem for the maintenance of previously developed solutions. At last, those may even become obsolete. This problem applies to every solution operating upon specialized devices with proper software, which lifetime is uncertain.

Solutions for pose estimation based on 2D RGB images have been proposed. OpenPose is presented as an open-source system for real-time multi-person 2D pose estimation [13]. MediaPipe BlazePose [14] is a single-person 2D pose estimator. It is presented as a fast and lightweight processing solution, alternative to Kinect and OpenPose methods. Besides providing 2D pose data, it provides a z coordinate representing depth movements.

In this paper, we determine whether novel motion capture solutions (OpenPose and MediaPipe) are suitable to assess stroke rehabilitation exercise performance against the widely used Microsoft Kinect v2. We envisage the development of a rehabilitation support solution based on a low-cost technical setup. We hypothesise that with the information directly provided by such methods, algorithms developed to assess performance will achieve desirable results likewise.

Kinect is validated for rehabilitation solutions against high precision marker-based optical systems [11] [15]. Faity

[1]Ana Rita Cóias and Alexandre Bernardino, Institute for Systems and Robotics, Instituto Superior Técnico, University of Lisbon, Lisbon, Portugal ana.coias@tecnico.ulisboa.pt alex@isr.tecnico.ulisboa.pt

[2]Min Hun Lee, Singapore Management University, Singapore, Singpore mhlee@smu.edu.sg

[3]Asim Amailagic, Carnegie Mellon University, Pittsburgh, PA, USA asim@cs.cmu.edu

et al. [15] validated Kinect against VICON for kinematic assessment in reaching tasks. Twenty-six healthy participants performed movements holding a dumbbell to induce movement patterns similar to stroke survivors (e.g., trembling and compensation). Researchers assessed several performance factors: elbow, shoulder, and trunk angles, movement efficiency, speed profiles, and limb and trunk displacements. They compared the measures with intra-class coefficient correlation, coefficient of determination, and root mean square error. Despite the study showing that Kinect does not assess some dimensions with satisfactory reliability, the authors affirm that the high potential of markerless motion capture solutions for rehabilitation applications for in-home and clinic use should encourage more validation studies.

Cóias *et al.* work [6] proposed a simpler and low-cost technical setup for compensatory movement analysis operating only on a laptop, with a built-in webcam, to support upper extremity exercise in real-time, using the OpenPose library for motion tracking. OpenPose was also validated for stroke rehabilitation support applications [16] [17]. Li *et al.* [16] evaluated OpenPose performance in assessing patients' balance to diminish falling risks against OptiTrack, with three healthy participants. MediaPipe also appears in proposed applications for upper limb stroke rehabilitation, revealing promising results enhancing the potential contribution of such solutions [18]. However, its evaluation against validated and widely used motion capture approaches is missing.

Previous studies [16] [11] [15] lack deep evaluation concerning the metrics and algorithms used to assess stroke-relevant exercise performance components. Most works performed simple evaluations with healthy participants, mainly correlating joint trajectories. The validation of motion capture solutions, markerless, low-cost, based on a single RGB camera, for stroke rehabilitation applications, is still needed.

In this work, seeking to achieve a system to support stroke rehabilitation composed only of a laptop with a built-in webcam, we determine an appropriate motion capture approach by comparing models used to assess upper extremity exercise performance. Microsoft Kinect v2 is a validated reference as a low-cost ease-of-use sensor widely used in assistive solutions. We explore OpenPose and MediaPipe BlazePose motion capture libraries against Kinect v2. We compare binary classifiers' performance in assessing three exercise performance components among the different motion-tracking approaches, using a dataset of 15 stroke survivors performing three exercises. Additionally, we compare movement trajectories and kinematic variables characterizing upper extremity impairments after stroke.

II. METHODOLOGY

In this section, we describe the methodology followed to compare the outcomes provided by Kinect, OpenPose, and MediaPipe BlazePose in upper extremity rehabilitation videos with stroke survivors. We perform three types of comparison. First, We compare the body keypoints' trajectories along the x and y axis. We also determine the correspondence between the z coordinate of Kinect and MediaPipe BlazePose. We describe *skeleton extraction* and data *normalization* steps. Second, based on the provided data, we compare relevant kinematic variables describing exercise performance. We present *performance components* significant for movement quality assessment and the *kinematic variables* describing them. Finally, we introduce the algorithms used to assess each performance component and compare their performance among skeleton tracking approaches taking the outcomes obtained with Kinect as our baseline.

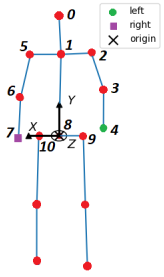
A. Body Skeleton Extraction

Kinect provides the 3D coordinates of 16 body joints (Table I), with the sensor as the origin of the coordinate system. We use OpenPose Demo[1] and MediaPipe[2] Python library to extract the body skeletons directly from the acquired images. OpenPose provides a set of 2D coordinates of 25 body keypoints in the image coordinate system, in pixels, and a confidence score associated with each keypoint. MediaPipe BlazePose provides two lists of 33 pose keypoints. One list denoted *Pose Landmarks* corresponds to body keypoints with x and y in the image coordinate system, normalized to the image width and height, respectively. It also provides a z coordinate representing a keypoint depth with origin at the midpoint between skeleton hips. Smaller the value of z, the closer the landmark is to the camera. Another list of keypoints denoted *Pose World Landmarks* provides pose data with x, y, and z in the world coordinate system, in meters, with origin at the midpoint between skeleton hips.

Table I presents the selected body keypoints for upper extremity motion analysis from the skeleton tracking methods and how we relate them. We apply a moving average filter with a window size of five frames to smooth trajectories.

TABLE I: Relation between Kinect, OpenPose, and MediaPipe body keypoints indices [19]. The skeleton shows the body keypoints and coordinate systems used in this work.

Body Joint	Abbr.	Kinect Joint Index	OpenPose Joint Index	MediaPipe Joint Index
head	hd	0	0	0
spine shoulder	ss	1	1	(11+12)/2
left shoulder	sh^l	2	2	12
left elbow	eb^l	3	3	14
left wrist	wr^l	4	4	16
right shoulder	sh^r	5	5	11
right elbow	eb^r	6	6	13
right wrist	wr^r	7	7	15
spine base	sb	8	8	(23+24)/2
left hip	hp^l	9	9	24
right hip	hp^r	10	12	23



In this work, we denote keypoint's position as ${}^{s}p_t(j,c)$, where s specifies the source tracking method in the set $S = \{\text{Kinect } (K), \text{OpenPose } (O), \text{MediaPipe Landmarks } (M), \text{MediaPipe World Landmarks } (W)\}$, j denotes a body joint in the set $J = \{hd, ss, sh^a, eb^a, wr^a, sb, hp^a\}$ (Table I), a specifies the body side in the set $A = \{\text{left } (l), \text{right } (r)\}$, c denotes a coordinate in the set $C \in \{x, y, z\}$, and t is the frame number of a total of T frames.

[1]https://github.com/CMU-Perceptual-Computing-Lab/openpose
[2]https://google.github.io/mediapipe/solutions/pose

B. Coordinate System Transformation & Normalization

Skeleton tracking methods provide keypoints in different coordinate systems with distinct origins: image coordinate system (O and M) and world coordinate system, with the centre at the camera (K) or a midpoint between skeleton hips (W). We apply a rigid body transformation to each keypoint from all methods to have them in the same coordinate system, with the *Spine Base* joint as the origin (Table I).

As the coordinates of the different methods have different scales and offsets, we compensate for this fact and perform a data normalization to compare joints' trajectories

$$ {}^{s}p'_t(j,c) = \frac{{}^{s}p_t(j,c) - {}^{s}\mu(j,c)}{{}^{s}\sigma(j,c)} \tag{1} $$

where μ is the mean position across a movement trial and σ is the standard deviation. Keypoints with reduced variation, i.e., less active joints, have low standard deviation values. A low standard deviation leads to unstable normalization, which may add a bias to our comparative analysis. We will focus essentially on joints actively moving.

C. Performance Components & Kinematic Variables

Lee *et al.* [8] defined three performance components to describe upper extremity exercise performance: *Range-of-Motion (ROM)*, *Smoothness*[3], and *Compensation*[4]. Table II presents the kinematic variables (detailed in [8]) that describe each component computed at each timestamp. [8]. We compute kinematic variable statistics at each timestamp (i.e., max, min, range, average, and standard deviation).

TABLE II: Kinematic variables (features) describing performance components [8].

Component	Kinematic Variables	Notation
ROM	· sh^a and eb^a angles	· $ja_t(hp^a, sh^a, eb^a)$, $ja_t(sh^a, eb^a, wr^a)$
	· eb^a and wr^a normalized relative trajectory	· $nrt_t(hd, eb^a)$, $nrt_t(hd, wr^a)$
	· eb^a and wr^a projected trajectory at each coordinate	· $npt_t(hd, wr^a, c)$, $npt_t(sh^a, wr^a, c)$ for $c \in C$
Smooth.	· eb^a and wr^a speed, acceleration, and jerk	· $sp_t(j)$,$ac_t(j)$,$jk_t(j)$
	· Normalized speed and jerk	· $nsp_t(j)$, $njk_t(j)$
	· sp and jk Mean Arrest Period Ratio	· $mapr_t(sp, j)$, $mapr_t(jk, j)$
	· ac and jk zero-crossing ratios	· $zc_t(ac, j)$, $zc_t(jk, j)$, $j \in \{eb^a, wr^a\}$
Comp.	· Spine angle	· $ja_t(ss_{init}, sb_{init}, ss)$
	· sh^a elevation and abduction angles	· $ja_t(sh^a_{init}, ss_{init}, sh^a)$, $ja_t(hp^a, sh^a, eb^a)$
	· Displacement between initial and current positions of the hd, ss, and sh^a at each coordinate	· $dpt_t(hd_{init}, hd, c)$, $dpt_t(ss_{init}, ss, c)$ $dpt_t(sh_{init}, sh, c)$ for $c \in C$

Most variables are normalized to overcome physical variabilities. As with keypoints trajectories, we need to acknowledge unstable normalized data to avoid biases.

D. Upper Extremity Exercises Performance Assessment

Aligned with therapists' assessment procedures, we use machine learning binary classification models (e.g. with output *0 - abnormal* or *1 - normal*) to assess performance components for entire movement trials. We use a Neural Network (NN), a non-sequential model, using as input features statistics summarizing the entire motion. Additionally, we explore a Long-Short Term Memory (LSTM), a sequential model, using kinematic variables at each timestamp as input features. In [8], NN and LSTM revealed increased performance on assessing the three performance components.

E. Metrics Under Evaluation and Statistical Analysis

We perform three analyses. First, We compare trajectories along the x and y axis provided by the skeleton tracking methods. We inspect the z coordinate from Kinect and MediaPipe. Second, We compare kinematic variables computed from different methods data. At last, we determine the performance achieved, with both classification approaches, on exercise assessment when using data from by OpenPose and MediaPipe against assessment based on Kinect data. We determine the alignment of OpenPose and MediaPipe against Kinect using the Root Mean Squared Error (RMSE)

$$ {}^{s}rmse(j,c) = \sqrt{\frac{1}{N}\sum_{t=1}^{T}\left({}^{K}v_t(j,c) - {}^{s}v_t(j,c)\right)^2} \tag{2} $$

where v denotes the variable subject to comparison, a joint position or a kinematic variable. RMSE enables the detection of systematic errors (offsets). We evaluate classifiers' performance with f_1 score and the mean squared error. The f_1 score is the harmonic mean between model precision and recall being suitable when the dataset is class-imbalanced.

III. EXPERIMENTS & RESULTS

A. The Upper Extremity Dataset

We utilize the dataset from [8] of 15 stroke survivors performing three task-oriented upper extremity rehabilitation exercises. Exercise 1 (E1) is *'Bring a Cup to the Mouth'*. Exercise 2 (E2) is *'Switch a Light On'*. Exercise 3 (E3) is *'Move a Cane Forward'*. Stroke survivors performed, on average, 10 movement trials for each exercise. The data was collected using the Microsoft Kinect v2, with a frame rate of 30 fps. Stroke survivors characterization is detailed in [8].

B. Body Skeleton Extraction

We extracted OpenPose keypoints by running the demo, from a command line tool to process sequences of frames[5]. Detected skeletons evaluation and data cleansing is described in our previous work [6]. We extracted MediaPipe landmarks using its Python library[6]. We did landmark extraction with a minimum detection confidence of 50%. However, MediaPipe could not process all frames and even entire video trials. For the study, we include the same subjects, videos, and frames among all skeleton tracking methods.

C. Keypoints Trajectories Comparison

In Figure 1, we can see a great alignment of the wrist joint (most active joint) among approaches, essentially along x and y axis. Along the z axis, MediaPipe data reveal increased displacement from the Kinect curve and a noisier pattern.

[3]Represents the level of joints' jittery and irregular motion patterns.

[4]Compensatory movements and postural patterns adopted by the stroke survivor to achieve the task target (e.g., shoulder elevation).

[5]i7-Intel(R) Core(TM) i7-8700 CPU @ 3.20GHz, 32 GB RAM, NVIDIA GeForce GTX 1070 GPU

[6]AMD Ryzen 5 3500U @ 2.1GHz, 12 GB RAM, AMD Radeon Vega GPU

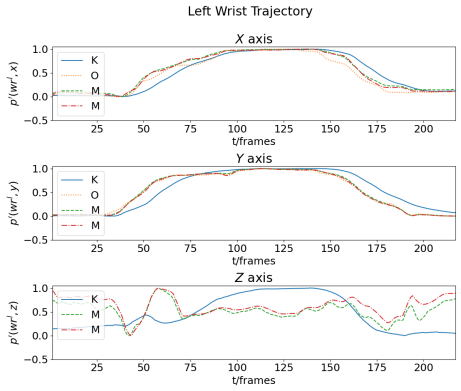


Fig. 1: Example of left wrist trajectory of a stroke survivor for exercise E1.

Table III presents the mean RMSE over stroke survivors and the three exercises for relevant body joints along each axis. From the table, we can infer that MediaPipe methods align with the Kinect data with a lower error when compared with OpenPose along x and y. Regarding MediaPipe approaches, there is no significant difference. Additionally, it is notable that the RMSE is much higher when we compare trajectory along the z axis.

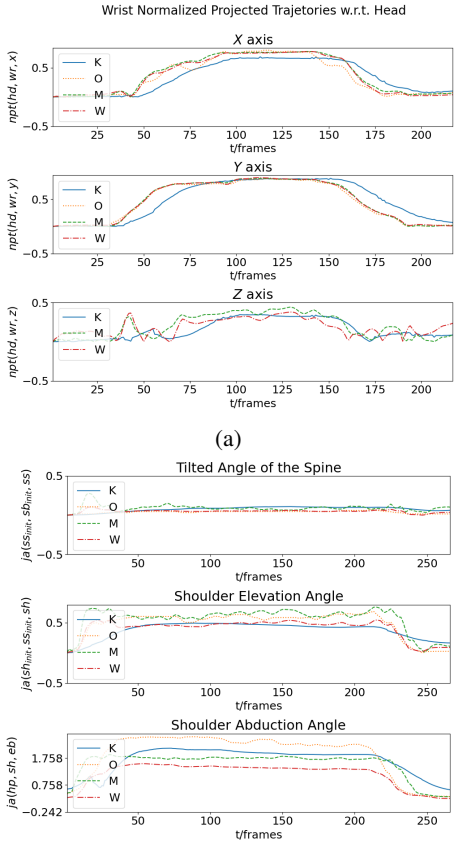
D. Kinematic Variables Comparison

Figure 2 shows a set of kinematic variables that describe performance components. We can observe that skeleton tracking methods' variables are nearly aligned with Kinect data. In Figure 2a, the main difference is between Kinect and MediaPipe normalized trajectories on the z axis, as expected. In Figure 2b, the MediaPipe Landmarks present significant irregularities and lower alignment with Kinect. Additionally, we may infer that the visible differences between joint angles are due to the lack of z component for the OpenPose method and low precision z component for MediaPipe approaches, mainly for the shoulder angle described between elbow and hip, which can represent a flexion or extension.

Table IV shows the mean RMSE and standard deviation, over exercises and stroke survivors, of kinematic variables. In this analysis, we only included the normalized variables. We excluded cases in which a low normalization factor could induce a bias in the analysis. From the Table, we can observe that variables from all methods align with Kinect variables at some level. We highlight the low error for MediaPipe World Landmarks (W), which is significant for some variables.

E. Quality of Movement Assessment Models

We utilize the *'Scikit-learn'* [20] and *'Pytorch'* [21] libraries for model implementation. For the NN, we explore architectures with one to three layers with 14, 16, 32, 48, 64, 96, 128, 256, and 512 hidden units with adaptive learning rate. We applied 500 iterations. For the LSTM approach, we explored a many-to-one architecture with one to three LSTM layers with 16, 32, and 64 hidden units with 0.5 dropout. We applied two fully connected layers with the same hidden units to produce an output. A Sigmoid function was applied to produce class probability at the last fully connected layer.



(a)

(b)

Fig. 2: Examples of kinematic variables describing performance: (a) wrist projected trajectory along each coordinate that described ROM; (b) shoulder and spine angles to detail compensatory movements.

The model converged in one epoch. We utilized the *'Adam'* optimizer, ReLu activation function, and Cross Entropy Loss for both approaches. We tested various initial learning rates (e.g., 0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1).

We conduct *Leave-One-Subject-Out* cross-validation method to evaluate our models. In this method, we take one subject for validation and train the model on the rest.

F. Performance Assessment Comparison

We test two binary classification approaches to assess three exercise performance components. Table V shows model parameters and classification performance results for the three exercises and performance components. As we can observe, with the data from OpenPose and MediaPipe methods, models have equal or increased accuracy in assessing performance in most scenarios compared with classification upon Kinect data. For the compensation component in exercise E3, the NN revealed a poor f_1 score, which could be associated with the lack of precise in-depth information. In this case, LSTM had a superior performance with OpenPose data.

IV. DISCUSSION

We performed three comparison steps to determine if novel skeleton tracking approaches are suitable alternatives

TABLE III: Alignment of the OpenPose and MediaPipe body keypoints (Table I) trajectories against Kinect v2 determined through mean RMSE over three exercises, stroke survivors movement trials, and mean standard deviation.

c		hd	ss	sh^l	eb^l	wr^l	sh^r	eb^r	wr^r	sb
x	***O vs. K***	0.69 ± 0.24	0.82 ± 0.30	0.82 ± 0.33	0.73 ± 0.19	0.67 ± 0.25	0.75 ± 0.16	0.76 ± 0.20	0.78 ± 0.23	1.29 ± 0.27
	M vs. K	**0.59 ± 0.22**	**0.68 ± 0.27**	**0.67 ± 0.33**	**0.63 ± 0.17**	**0.61 ± 0.23**	**0.66 ± 0.22**	**0.64 ± 0.20**	**0.75 ± 0.31**	**1.23 ± 0.29**
	W vs. K	**0.59 ± 0.22**	**0.68 ± 0.27**	**0.67 ± 0.33**	**0.63 ± 0.17**	**0.61 ± 0.23**	**0.66 ± 0.22**	**0.64 ± 0.20**	**0.75 ± 0.31**	**1.23 ± 0.29**
y	***O vs. K***	1.05 ± 0.29	1.00 ± 0.30	0.92 ±0.35	0.66 ± 0.28	0.75 ± 0.25	0.83 ± 0.25	0.61 ± 0.20	0.64 ±0.16	1.29 ± 0.25
	M vs. K	0.95 ±0.36	**0.96 ± 0.39**	**0.84 ± 0.39**	**0.65 ± 0.28**	**0.71 ± 0.23**	**0.72 ± 0.28**	**0.55 ± 0.18**	**0.62 ± 0.19**	**1.20 ± 0.32**
	W vs. K	**0.94 ± 0.35**	0.98 ± 0.39	0.87 ± 0.38	**0.65 ± 0.28**	0.73 ± 0.24	**0.72 ± 0.28**	**0.55 ± 0.18**	**0.62 ± 0.19**	1.22 ± 0.32
z	***M vs. K***	1.07 ± 0.33	1.19 ± 0.33	**1.16 ± 0.33**	**1.11 ± 0.32**	1.19 ± 0.34	1.31 ± 0.35	1.06 ± 0.39	1.20 ± 0.22	1.35 ± 0.22
	W vs. K	1.07 ± 0.33	1.19 ± 0.33	1.18 ± 0.33	1.14 ± 0.32	1.19 ± 0.34	1.31 ± 0.35	1.06 ± 0.39	1.20 ± 0.22	1.35 ± 0.22

TABLE IV: Alignment of the OpenPose (O) and MediaPipe Landmarks (M) and World Landmarks (W) kinematic variables with Kinect (K) represented by mean RMSE and standard deviation over three exercises, stroke survivors movement trials.

		$ja(hp^a, sh^a, eb^a)$	$ja(sh^a, eb^a, wr^a)$	$nrt(hd, eb^a)$	$nrt(hd, wr^a)$		
ROM	***O vs. K***	0.42 ± 0.17	0.68 ± 0.19	**0.10 ± 0.05**	0.14 ± 0.06		
	M vs. K	0.50 ± 0.19	0.88 ± 0.28	0.17 ± 0.08	0.21 ± 0.12		
	W vs. K	**0.34 ± 0.15**	**0.52 ± 0.17**	0.11 ± 0.05	**0.12 ± 0.05**		
		$npt(hd, wr^a, x)$	$npt(hd, wr^a, y)$	$npt(hd, wr^a, z)$	$npt(sh, wr^a, x)$	$npt(sh, wr^a, y)$	$npt(sh, wr^a, z)$
	O vs. K	0.88 ± 0.99	**0.13 ± 0.06**		1.31 ± 1.16	0.19 ± 0.09	
	M vs. K	0.95 ± 1.13	0.14 ± 0.06	1.53 ± 1.24	1.22 ± 1.13	0.19 ± 0.09	0.96 ± 1.33
	W vs. K	**0.79 ± 0.80**	**0.13 ± 0.06**	**1.25 ± 0.91**	**1.07 ± 0.96**	**0.18 ± 0.09**	**0.80 ± 0.74**
		$nsp(eb^a)$	$njk(eb^a)$	$mapr(sp, eb^a)$	$mapr(jk, eb^a)$	$zc(ac, eb^a)$	$zc(jk, eb^a)$
Smoothness	***O vs. K***	0.12 ± 0.02	0.05 ± 0.01	0.25 ± 0.05	0.15 ± 0.03	**0.13 ± 0.03**	**0.10 ± 0.02**
	M vs. K	**0.11 ± 0.02**	**0.04 ± 0.01**	0.25 ± 0.05	**0.13 ± 0.02**	0.17 ± 0.02	0.12 ± 0.02
	W vs. K	**0.11 ± 0.02**	0.04 ± 0.02	**0.22 ± 0.05**	**0.13 ± 0.02**	0.15 ± 0.03	0.15 ± 0.03
		$nsp(wr^a)$	$njk(wr^a)$	$mapr(sp, wr^a)$	$mapr(jk, wr^a)$	$zc(ac, wr^a)$	$zc(jk, wr^a)$
	O vs. K	**0.11 ± 0.02**	0.05 ± 0.02	**0.21 ± 0.04**	0.13 ± 0.02	**0.14 ± 0.03**	**0.10 ± 0.03**
	M vs. K	**0.11 ± 0.02**	**0.04 ± 0.01**	0.24 ± 0.06	**0.12 ± 0.02**	0.15 ± 0.03	0.11 ± 0.02
	W vs. K	**0.11 ± 0.02**	**0.04 ± 0.01**	0.22 ± 0.05	**0.12 ± 0.02**	0.25 ± 0.04	0.15 ± 0.03
		$ja(ss_{init}, sb_{init}, ss)$	$ja(sh^a_{init}, ss_{init}, sh^a)$	$ja(hp^a, sh^a, eb^a)$			
Compensation	***O vs. K***	**0.04 ± 0.03**	0.42 ± 0.24	0.42 ± 0.17			
	M vs. K	0.18 ± 0.13	0.39 ± 0.24	0.48 ± 0.17			
	W vs. K	0.06 ± 0.03	**0.19 ± 0.10**	**0.33 ± 0.16**			

TABLE V: Classifiers parameters and performance from *Leave-One-Subject-Out* Cross-Validation

Components			**ROM**			**Smoothness**			**Compensation**		
Source	Algorithms	Exercise	Param.	$f1$	mse	Param.	$f1$	mse	Param.	$f1$	mse
Kinect	NN	E1	(32,32) | 0.1	0.782 ± 0.409	0.171 ± 0.347	(16,16) | 0.005	**0.792 ± 0.305**	**0.244 ± 0.279**	(512) | 0.1	0.767 ± 0.381	0.257 ± 0.394
		E2	(16) | 0.1	**0.843 ± 0.326**	0.127 ± 0.277	(64,64,64) | 0.001	0.642 ± 0.407	**0.297 ± 0.281**	(16) | 0.001	**0.768 ± 0.398**	**0.247 ± 0.405**
		E3	(64,64,64) | 0.001	0.720 ± 0.390	0.316 ± 0.403	(32) | 0.05	0.803 ± 0.338	0.228 ± 0.349	(14,14,14) | 0.05	**0.624 ± 0.468**	**0.325 ± 0.437**
	LSTM	E1	(16) | 0.1	0.786 ± 0.410	0.214 ± 0.410	(16) | 0.1	0.857 ± 0.350	0.143 ± 0.350	(32,32) | 0.0005	0.643 ± 0.479	0.214 ± 0.410
		E2	(16,16) | 0.0005	0.800 ± 0.400	0.200 ± 0.400	(64) | 0.001	**0.933 ± 0.249**	**0.067 ± 0.249**	(16) | 0.0001	0.733 ± 0.442	0.267 ± 0.442
		E3	(16,16) | 0.0001	**0.800 ± 0.400**	**0.200 ± 0.400**	(16) | 0.1	0.800 ± 0.400	0.200 ± 0.400	(16) | 0.0001	**0.800 ± 0.400**	**0.200 ± 0.400**
OpenPose	NN	E1	(14) | 0.01	0.846 ± 0.346	0.148 ± 0.308	(14) | 0.1	0.767 ± 0.381	0.257 ± 0.394	(128) | 0.1	0.711 ± 0.450	0.285 ± 0.436
		E2	(32) | 0.05	0.812 ± 0.378	0.133 ± 0.322	(14) | 0.1	0.592 ± 0.484	0.415 ± 0.481	(64) | 0.05	0.681 ± 0.401	0.341 ± 0.378
		E3	(14,14) | 0.005	**0.774 ± 0.394**	**0.130 ± 0.218**	(256) | 0.05	**0.804 ± 0.340**	**0.227 ± 0.355**	(32,32,32) | 0.005	0.478 ± 0.410	0.437 ± 0.335
	LSTM	E1	(16) | 0.0005	**0.857 ± 0.350**	**0.143 ± 0.350**	(16) | 0.1	0.857 ± 0.350	0.143 ± 0.350	(32) | 0.0005	**0.929 ± 0.258**	**0.071 ± 0.258**
		E2	(32) | 0.0005	0.800 ± 0.400	0.200 ± 0.400	(16) | 0.0001	0.667 ± 0.471	0.333 ± 0.471	(16) | 0.0001	0.800 ± 0.400	**0.200 ± 0.400**
		E3	(16) | 0.001	0.733 ± 0.442	0.267 ± 0.442	(16) | 0.1	**0.800 ± 0.400**	0.200 ± 0.400	(16) | 0.0001	0.733 ± 0.442	0.267 ± 0.442
MediaPipe Landmarks	NN	E1	(14) | 0.001	0.839 ± 0.349	0.164 ± 0.336	(16) | 0.1	0.767 ± 0.381	0.257 ± 0.394	(64) | 0.1	**0.853 ± 0.349**	**0.142 ± 0.329**
		E2	(14,14) | 0.05	0.780 ± 0.395	0.212 ± 0.357	(14,14,14) | 0.005	0.623 ± 0.453	0.332 ± 0.393	(128) | 0.05	0.648 ± 0.459	0.356 ± 0.399
		E3	(64,64) | 0.05	0.635 ± 0.419	0.195 ± 0.272	(14) | 0.1	0.800 ± 0.339	0.235 ± 0.353	(16,16) | 0.01	0.490 ± 0.437	0.363 ± 0.316
	LSTM	E1	(16) | 0.001	**0.857 ± 0.350**	**0.143 ± 0.350**	(16) | 0.1	0.857 ± 0.350	0.143 ± 0.350	(16) | 0.005	**0.929 ± 0.258**	**0.071 ± 0.258**
		E2	(32,32) | 0.0005	**0.867 ± 0.340**	**0.133 ± 0.340**	(32) | 0.001	0.733 ± 0.442	0.267 ± 0.442	(16) | 0.0005	**0.800 ± 0.400**	**0.200 ± 0.400**
		E3	(32,32) | 0.0001	**0.800 ± 0.400**	**0.200 ± 0.400**	(16) | 0.1	0.800 ± 0.400	0.200 ± 0.400	(32) | 0.001	0.733 ± 0.442	0.267 ± 0.442
MediaPipe World Landmarks	NN	E1	(16) | 0.01	**0.914 ± 0.254**	**0.036 ± 0.048**	(16) | 0.1	0.767 ± 0.381	0.275 ± 0.394	(64,64) | 0.01	**0.853 ± 0.349**	**0.142 ± 0.329**
		E2	(16,16) | 0.05	0.841 ± 0.340	**0.113 ± 0.268**	(16,16) | 0.05	**0.658 ± 0.467**	0.348 ± 0.464	(32) | 0.005	0.567 ± 0.478	0.298 ± 0.395
		E3	(64) | 0.05	0.737 ± 0.381	0.284 ± 0.389	(16) | 0.1	0.800 ± 0.339	0.235 ± 0.353	(64,64,64) | 0.05	0.490 ± 0.457	0.427 ± 0.425
	LSTM	E1	(16) | 0.1	0.786 ± 0.410	0.214 ± 0.410	(16) | 0.1	0.857 ± 0.350	0.143 ± 0.350	(16) | 0.05	0.643 ± 0.479	0.357 ± 0.479
		E2	(32) | 0.0001	0.667 ± 0.471	0.333 ± 0.471	(64) | 0.001	0.867 ± 0.340	0.133 ± 0.340	(16) | 0.0001	0.733 ± 0.442	0.267 ± 0.442
		E3	(32) | 0.0001	**0.800 ± 0.400**	**0.200 ± 0.400**	(16) | 0.1	0.800 ± 0.400	0.200 ± 0.400	(16,16) | 0.0001	0.733 ± 0.442	0.267 ± 0.442

to the discontinued Microsoft Kinect: OpenPose, MediaPipe Landmarks, and MediaPipe World Landmarks.

First, we compare keypoint trajectories provided by the different approaches. Table III presents the RMSE for each method against our baseline, the Kinect v2. Due to the applied normalization, normalized trajectories for less active body joints during movement performance are unstable, which is revealed by the higher RMSE for body joints such as the spine base. Active body joints, such as the elbow and the wrist, with stable normalization, lead to a more precise comparison. MediaPipe Landmarks and World Landmarks revealed lower RMSE for trajectories along the x and y axis for all joints. Looking at the z component for MediaPipe methods, we observe that the RMSE is higher for this coordinate but not specially higher than for x and y components. Between Landmarks and World Landmarks, there is no meaningful difference.

Second, we compare the kinematic variables describing exercise performance components calculated directly from each skeleton tracking approach data after coordinate system transformation. Since most variables are already normalized to overcome physical variabilities, we compare these directly through RMSE. From Table IV, all variables are at some level aligned with Kinect variables. MediaPipe World landmarks reveal lower RMSE, mainly for shoulder elevation and abduction/flexion angles. This result may indicate that the z component from MediaPipe is relevant to determine joint extensions and flexions in each exercise.

Third, we compare classifiers' performance. Table V shows that based on the 2D skeleton data, for all exercises and performance components, classifiers equal or even surpass performance based on Kinect data in most scenarios. One exception is visible for the NN classifier when assessing the compensation component for exercise E3. The lower performance could be explained by the lack of z component in the OpenPose case or the lower precision in

depth information provided by MediaPipe. The lack of a z component, could lead to inaccurate measures of shoulder angles and trunk displacements that characterize compensation. However, LSTM outperformed NN in this scenario, which makes us discard this hypothesis.

Previous works do not compare classification models performance on assessing rehabilitation exercises [11] [15]. This analysis is essential to inspect alternatives to Kinect v2. NN and LSTM were explored in [6] and [8]. In [8], binary classification output was used to generate global performance scores to determine the agreement between computer-based and therapists' assessments. In [6], the NN was adapted for real-time compensation assessment, providing feedback during exercise performance in a user interface to support rehabilitation. Binary classification might act as a trigger mechanism for feedback engines, which notify patients which motion patterns they should correct.

V. CONCLUSION

In this work, we aim to determine a suitable novel motion capture approach for the development of a low-cost system to support stroke rehabilitation. We compare OpenPose and MediaPipe against the widely used and already validated in this field, Microsoft Kinect.

Keypoint trajectory comparison revealed that MediaPipe Landmarks and World Landmarks have higher alignment with Kinect and less irregular patterns. The z component provided by these methods shows less alignment than x and y components. Comparison between kinematic variables describing exercise performance reveals that all methods can follow Kinect. MediaPipe World Landmarks show higher alignment with Kinect, mainly in compensatory angles.

As we hypothesize, classifiers assessed exercise performance with equal or higher accuracy based on data provided by OpenPose and MediaPipe, making us conclude that for the specified set of exercises, the lack of in depth information is not a limitation for quality of movement assessment.

Allied with the advantages given by MediaPipe as a faster and lightweight library for skeleton tracking, the provided World Landmarks revealed significant alignment with Kinect keypoints trajectories and kinematic variables.

ACKNOWLEDGMENT

This work was supported by FCT with the LARSyS - FCT Project UIDB/50009/2020, partially supported by the Portuguese Foundation for Science and Technology - FCT with a Ph.D. grant (2021.05239.BD) and by the Singapore Ministry of Education (MOE) Academic Research Fund (AcRF) Tier 1 grant.

REFERENCES

[1] B. Semenko, L. Thalman, E. Ewert, R. Delorme, S. Hui, H. Flett, and N. Lavoie, "An evidence based occupational therapy toolkit for assessment and treatment of the upper extremity post stroke," *Screening*, vol. 4, pp. 4–1, 2015.

[2] K. L. Meadmore, E. Hallewell, C. Freeman, and A.-M. Hughes, "Factors affecting rehabilitation and use of upper limb after stroke: views from healthcare professionals and stroke survivors," *Topics in stroke rehabilitation*, vol. 26, no. 2, pp. 94–100, 2019.

[3] M. Rensink, M. Schuurmans, E. Lindeman, and T. Hafsteinsdottir, "Task-oriented training in rehabilitation after stroke: systematic review," *Journal of advanced nursing*, vol. 65, no. 4, pp. 737–754, 2009.

[4] S. A. Billinger, R. Arena, J. Bernhardt, J. J. Eng, B. A. Franklin, C. M. Johnson, M. MacKay-Lyons, R. F. Macko, G. E. Mead, E. J. Roth, *et al.*, "Physical activity and exercise recommendations for stroke survivors: a statement for healthcare professionals from the american heart association/american stroke association," *Stroke*, vol. 45, no. 8, pp. 2532–2553, 2014.

[5] T. Rikakis, A. Kelliher, J. Choi, J.-B. Huang, K. Kitani, S. Zilevu, and S. L. Wolf, "Semi-automated home-based therapy for the upper extremity of stroke survivors," in *PErvasive Technologies Related to Assistive Environments Conference*, 2018, pp. 249–256.

[6] A. R. Cóias, M. H. Lee, and A. Bernardino, "A low-cost virtual coach for 2d video-based compensation assessment of upper extremity rehabilitation exercises," *Journal of NeuroEngineering and Rehabilitation*, vol. 19, no. 1, pp. 1–16, 2022.

[7] D. Webster and O. Celik, "Systematic review of kinect applications in elderly care and stroke rehabilitation," *Journal of NeuroEngineering and Rehabilitation*, vol. 11, no. 1, pp. 1–24, 2014.

[8] M. H. Lee, D. P. Siewiorek, A. Smailagic, A. Bernardino, and S. B. i. Badia, "Learning to assess the quality of stroke rehabilitation exercises," in *International Conference on Intelligent User Interfaces*, 2019, pp. 218–228.

[9] 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.

[10] E. B. Brokaw, E. Eckel, and B. R. Brewer, "Usability evaluation of a kinematics focused kinect therapy program for individuals with stroke," *Technology and Health Care*, vol. 23, no. 2, pp. 143–151, 2015.

[11] D. Webster and O. Celik, "Experimental evaluation of microsoft kinect's accuracy and capture rate for stroke rehabilitation applications," in *IEEE Haptics Symposium (HAPTICS)*, 2014, pp. 455–460.

[12] M. Tölgyessy, M. Dekan, L. Chovanec, and P. Hubinskỳ, "Evaluation of the azure kinect and its comparison to kinect v1 and kinect v2," *Sensors*, vol. 21, no. 2, p. 413, 2021.

[13] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "Openpose: Realtime multi-person 2d pose estimation using part affinity fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 1, pp. 172–186, 2021.

[14] V. Bazarevsky, I. Grishchenko, K. Raveendran, T. Zhu, F. Zhang, and M. Grundmann, "Blazepose: On-device real-time body pose tracking," *CVPR Workshop on Computer Vision for Augmented and Virtual Reality*, 2020.

[15] G. Faity, D. Mottet, and J. Froger, "Validity and reliability of kinect v2 for quantifying upper body kinematics during seated reaching," *Sensors*, vol. 22, no. 7, p. 2735, 2022.

[16] B. Li, J. Williamson, N. Kelp, T. Dick, and A. P. Bo, "Towards balance assessment using openpose," in *International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2021, pp. 7605–7608.

[17] S. Zhao, Y. Tong, G. Chen, and Y. Zhang, "Testing the feasibility of quantizing the progress of stroke patients' rehabilitation with a computer vision method," in *International Conference on Computer Vision, Image and Deep Learning & International Conference on Computer Engineering and Applications*. IEEE, 2022, pp. 1–4.

[18] R. Baluz, A. Teles, J. E. Fontenele, R. Moreira, R. Fialho, P. Azevedo, D. Sousa, F. Santos, V. H. Bastos, and S. Teixeira, "Motor rehabilitation of upper limbs using a gesture-based serious game: Evaluation of usability and user experience," *Games for Health Journal*, vol. 11, no. 3, pp. 177–185, 2022.

[19] T. B. de Gusmao Lafayette, A. d. Q. Burle, A. d. A. Almeida, V. L. Ventura, V. M. Carvalho, A. E. Fontes da Gama, J. M. Xavier Natario Teixeira, and V. Teichrieb, "The virtual kinect," in *Symposium on Virtual and Augmented Reality*, 2021, pp. 111–119.

[20] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, *et al.*, "Scikit-learn: Machine learning in python," *Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.

[21] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in pytorch," in *Conference on Neural Information Processing Systems (NIPS)*, 2017.