Article

BiomacVR: A Virtual Reality-Based System for Precise Human Posture and Motion Analysis in Rehabilitation Exercises Using Depth Sensors

Rytis Maskeliūnas [1], **Robertas Damaševičius** [1,*], **Tomas Blažauskas** [1], **Cenker Canbulut** [1], **Aušra Adomavičienė** [2] **and Julius Griškevičius** [3]

[1] Faculty of Informatics, Kaunas University of Technology, LT-44029 Kaunas, Lithuania
[2] Center of Rehabilitation, Physical and Sports Medicine, Vilnius University Hospital Santaros Clinics, LT-08661 Vilnius, Lithuania
[3] Department of Biomechanical Engineering, Vilnius Gediminas Technical University, LT-10223 Vilnius, Lithuania
[*] Correspondence: robertas.damasevicius@ktu.lt

Abstract: Remote patient monitoring is one of the most reliable choices for the availability of health care services for the elderly and/or chronically ill. Rehabilitation requires the exact and medically correct completion of physiotherapy activities. This paper presents BiomacVR, a virtual reality (VR)-based rehabilitation system that combines a VR physical training monitoring environment with upper limb rehabilitation technology for accurate interaction and increasing patients' engagement in rehabilitation training. The system utilises a deep learning motion identification model called Convolutional Pose Machine (CPM) that uses a stacked hourglass network. The model is trained to precisely locate critical places in the human body using image sequences collected by depth sensors to identify correct and wrong human motions and to assess the effectiveness of physical training based on the scenarios presented. This paper presents the findings of the eight most-frequently used physical training exercise situations from post-stroke rehabilitation methodology. Depth sensors were able to accurately identify key parameters of the posture of a person performing different rehabilitation exercises. The average response time was 23 ms, which allows the system to be used in real-time applications. Furthermore, the skeleton features obtained by the system are useful for discriminating between healthy (normal) subjects and subjects suffering from lower back pain. Our results confirm that the proposed system with motion recognition methodology can be used to evaluate the quality of the physiotherapy exercises of the patient and monitor the progress of rehabilitation and assess its effectiveness.

Keywords: posture analysis; pose recognition; motion analysis; action recognition; depth sensors; rehabilitation exercises; virtual reality; telehealth

1. Introduction

For elderly and/or chronically ill individuals, remote patient monitoring is considered one of the most-trusted options for healthcare solutions [1]. Furthermore, observing people's interactions is crucial in diagnosing and treating those who are unwell. The preservation of activities of daily living (ADLs) in seniors [2] is vital to keep medical expenditures low, especially as the elderly population grows. Physiotherapy activities such as active range-of-motion (ROM) exercise (extension, flexion, and rotation), muscle strength, and endurance training are essential for patients recovering from stroke (PSR) [3]. A physiotherapist uses a variety of approaches to help people regain their daily mobility, including task training, muscular strengthening, and the use of assistive devices. However, guiding a patient through physiotherapy exercises is a time-consuming, tiresome, and costly endeavour [4].

Numerous studies have been reported to evaluate the feasibility and usefulness of new information technology tools and their design with the goal of helping rehabilitation at home after stroke or trauma [5,6] as well as rehabilitation of muskoskeletal disorders [7] and for hospitalised patients with COVID-19 [8]. Many studies have been conducted to investigate the effectiveness of computer-assisted treatment or virtual reality (VR) in rehabilitation and recovery of upper limb motor skills, balance control and gait, lower limbs, posture, and walking [9]. Additionally, researchers have researched the therapeutic benefits of telerehabilitation, which allows patients to conduct therapy with therapists using telecommunication technology in their home setting and has been widely used for motor and cognitive recovery [10]. Simultaneously, there has been an increase in studies that use information technology to help patients with rehabilitation at home. Home-based recovery [11] in the context of telehealth [12] is increasingly being utilised to cut health-care expenditures; yet there is a danger of worsening clinical results due to the patient's lack of desire and the difficulty of performing a rigorous medical control. This is especially true for patients who must strictly adhere to the efficient rehabilitation methods of physicians. The creation of methods for monitoring the functional recovery of the subject, for example, after a stroke [13], is thus thought critical both for the patient's motivation to complete the appropriate number of exercises and for the patient's motivation to perform the appropriate number of workouts. Monitoring the patient's posture in real time to assess posture in his range of motion is one of the acknowledged ways to analyse the accuracy of a series of recovery exercises [14]. The overall quality of the exercises is determined by assessing thoracic orientation, hip and knee joint rotations, and leg length. One of the key advantages of this technology for human motion monitoring is its inexpensive cost, which, together with its compact dimensions, allows widespread use in rehabilitation clinics, gyms, and even at home. Moreover, observation of posture is important for healthy subjects as well, who are at risk due to unhealthy work practices and poor ergonomics such as for office employees with excessive sitting work at an office [15,16]. A rehabilitation system that combines a virtual gamified rehabilitation environment with upper limb rehabilitation technology is interactive and engaging, which increases the enthusiasm and initiative of patients for rehabilitation training as well as the efficiency and effectiveness of rehabilitation training and treatment [17].

To gather raw sensor data to monitor human activities, a variety of approaches are used. Wearable sensors are studied in [18,19]. Sensors from the Inertial Measurement Unit (IMU) are used in combination with clinical tests and outcome measurements. IMUs are devices that combine linear acceleration from accelerometers with angular turning rates from gyroscopes to create an integrated motion unit. IMUs were chosen not only because of their mobility and inexpensive cost, but also because they provide precise motion modelling of the participants. Smartphones with built-in IMUs and optical cameras are also among the most-frequently used approaches [20,21]. Because smartphones have practically become a must-have companion for people's daily lives, much human body and human activity-related research is conducted using cellphones. An optical camera, as a commonly utilised sensing device, is a common approach for recognising human activity. When compared to standard optical cameras, a depth camera offers a distinct advantage due to the additional dimension of depth information.

Currently, Microsoft Kinect (Microsoft Corp., Redmond, WA, USA) is a standard low-cost motion sensor that can be used for the measurement of posture and balance during human exercise [22]. Several studies have found the Kinect RGB-D sensor to be a good and usable choice [23] for motion analysis in rehabilitative and industrial settings when professional medical-grade markers-based systems cannot be employed. How to use a single camera to accurately solve the problem of multi-joint spatial motion evaluation of human lower limbs and how to avoid camera occlusion [24] in the rehabilitation physician's operation process are all important foundations for accurate input of lower limb motion information. Another approach is to use multiple Kinect cameras connected to a single computer and synchronised to provide a multiview of a human pose [25].

Traditional algorithms and deep learning-based recognition algorithms are the two primary approaches to solving the challenge of human action recognition. Recognition algorithms based on deep learning learn object characteristics using neural networks and immediately output the final recognition results, whereas traditional algorithms employ the approach of "feature extraction and expression + feature matching" to detect human behaviour. Traditional algorithms analyse intrinsic properties of human behaviour, such as motion information features, points of interest in temporal and spatial time, and geometric details, to detect human behaviour or to analyse human shape for the identification of biometric information such as sex [26] or health status. Deep learning methods have emerged in recent years, and neural network characteristics have more abstract and comprehensive descriptions of behaviour characteristics [27]. Machine learning and deep learning techniques provide outstanding performance tasks that previously required a lot of knowledge and time to model and to handle and process data collected from sensors, allowing a more accurate and faster assessment of human health condition [28].

The contributions and novelty of this study are as follows. This study presents an action identification system for upper limb home rehabilitation, which is a physical exercise programme for particular joints and muscle groups with benefits such as cost effectiveness, suitability for rehabilitation, and ease of use. The study's main contributions are: (1) the suggested system is a physical movement training programme for specific muscle groups; (2) the upper limb rehabilitation system's hardware incorporates a personal computer and a Kinect depth camera; (3) patients may complete therapeutic activities while in a VR environment; and (4) the suggested upper extremity rehabilitation system is real-time, efficient in vision-based action recognition, and uses hardware and software that are inexpensive.

Section 2 presents an overview of the related work. Section 3 presents a description of the implementation of the BiomacVR system and the methods used for the evaluation of the subject's condition. Section 4 provides a description of the physical rehabilitation training implemented and details each posture required by the system. Section 5 describes the experiments and the results. The evaluation of results and discussion are given in Section 6. Finally, the conclusions are given in Section 7.

2. Related Work

Many human position estimation methods collect skeletons or skeleton points from depth sensors these days [29,30]. However, for rehabilitation purposes, identification of motions (including directions and angles) and joint centres is required. Although various posture estimation methods and exergames have been created and presented, most existing systems focus on extracting skeletons or skeleton points from a depth sensor and primarily focus on human pose estimate rather than joint movement determination (including directions and angles) [31]. For example, Lee et al. [32], for the evaluation of upper extremity motor function in stroke patients, proposed a sensor-based system with a depth sensor (Kinect V2) and an algorithm to classify the continuous Fugl–Meyer (FM) scale based on fuzzy inference. The values obtained from the FM scale had a strong correlation with the scores evaluated by the clinician, indicating the prospect of a more sensitive evaluation of motor function in the upper extremities. Ayed et al. [33] suggested a method to assess the Functional Reach Test (FRT) using Kinect v2; FRT is one of the most widely used balancing clinical instruments for predicting falls. These findings indicated that the Kinect v2 unit is capable of calculating the conventional FRT. Using a bidirectional long- and short-term memory neural network (BLSTM-NN), Saini et al. [34] propose a Kinect sensor-based interaction monitoring system between two people. This methodology is used to help people get back on their feet by assessing their actions. Capecci et al. [35] present a data set of rehabilitation activities for low back pain (LBP) acquired by an RGB-D sensor. The RGB, depth videos, and joint locations of the skeleton are included in the data set. These characteristics are utilised to calculate a score for the subject's performance, which may be used in rehabilitation to measure human mobility. Wang et al. [36] presented

an experimental platform of a virtual rehabilitation training system to analyse upper limb rehabilitation exercises for subjects with stroke caused by hemiplegic dyskinesia. Hand motion tracking is realised by the Kinect's bone tracking based on the Kinect depth image and colour space model. Sarsfield et al. [37] conduct a clinical qualitative and quantitative analysis of the pose estimation algorithms of the Xbox One Kinect in order to determine their suitability for technology-assisted rehabilitation and to help develop pose recognition methods for rehabilitation scenarios. In the upper-body stroke rehabilitation scenario, the researchers discovered difficulty with occluded depth data for shoulder, elbow, and wrist joint tracking. They determined that in order to infer joint locations with complete or partial occlusion, pose estimation algorithms should consider leveraging temporal information and extrapolating from prior frames. This should also decrease the potential of mistakenly inferring joint position by excluding quick and dramatic changes in position.

For patients with mobility impairments, Xiao et al. [38] built a VR rehabilitation system based on Kinect, a vision capture sensor. The technology gathers real-time motion data from the user and detects compensation. The system analyses the patients based on their training performance once they have completed the programme. Bijalwan et al. [39] suggested using an RGB-Depth camera to detect and recognise upper limb activities, allowing patients to complete real-time physiotherapy exercises without the need for human assistance. A deep convolutional neural network (CNN) identifies the physiotherapy activity by extracting characteristics from pre-processed data. Recurrent neural networks (RNNs) are used to extract and use temporal relationships. CNN-GRU is a hybrid deep learning model that uses a unique focussed loss criterion to overcome the limitations of ordinary cross-entropy loss. The RGB-D data received from Kinect v2 sensors are used to evaluate a dataset of ten distinct physiotherapy exercises with extremely high accuracy. Junata et al. [40] proposed a Kinect-based Rapid Movement Training (RMT) system to evaluate the overall balance of chronic stroke sufferers and the responsiveness of balance recovery. He et al. [41] offer a novel Kinect-based posture identification approach in a physical sports training system based on urban data. The spatial coordinates of human body joints are first obtained using Kinect. The two-point approach is then used to determine the angle, and the body posture library is created. Finally, to assess posture identification, angle matching is used using a posture library. Wang et al. [42] developed a new data collection method for patients before they begin rehabilitation training to guarantee that the robot used for rehabilitation training should not overextend any of the joints of stroke sufferers. In the RGB-D camera picture, the ranges of motion of the hip and joint, knee joint in the sagittal plane, and hip joint in the coronal plane are modelled using least square analysis as a mapping between the camera coordinate system and the pixel coordinate system. The Kinect V2.0 colour and depth sensors were used in the HemoKinect system [43] to obtain 3D joint locations. HemoKinect can evaluate the following workouts utilising angle calculations and centre-of-mass (COM) estimations based on these joint positions: elbow flexion/extension, knee flexion/extension (squat), step climb (ankle exercise), and multidirectional balancing based on COM. The programme creates reports and progress graphs and can email the data directly to the physician. The exercises were tested on ten healthy people and eight sick. The HemoKinect system effectively recorded elbow and knee activities, with real-time joint angle measurements shown at an accuracy of up to 78%. Leightley et al. [44] proposed a system that decomposes the skeletal stream into a collection of unique joint-group properties in order to obtain motion capture (MoCap) using a single Kinect depth sensor. Analysis techniques are used to offer joint group input that highlights the condition of mobility, providing doctors with specific information. Walking, sitting, standing, and balancing are used for the evaluation of the system. Patalas-Maliszewska et al. [45] offer a unique architecture for automatic workplace instruction production and real-time identification of industrial worker activities. To detect and validate completed job activities, the suggested technique includes CNN, CNN with Support Vector Machine (SVM), and Region-Based CNN (Yolov3 Tiny). To begin, video records of the work process are evaluated, and reference video frames corresponding to different stages of the job

activity are identified. Subsequently, depending on the properties of the reference frames, the work-related characteristics and objects are identified using CNN with SVM (reaching 94% accuracy) and the Yolov3 Tiny network.

To summarise, several attempts have been made to construct home-based diagnostic and rehabilitation monitoring systems. These devices have been validated and have the potential to aid in home mobility monitoring; however, they do not assess mobility constraints, which is what our work seeks to address. Furthermore, other existing systems only provide a single health indicator, but a more thorough descriptive signal, as in the BIOMAC approach, might be more valuable to a medical expert for evaluating rehabilitation efficiency.

We offer a novel deep learning-based approach for remote rehabilitative analysis of non-invasive (no RGB data utilised) monitoring of human motions during rehabilitation exercises. Our method can analyse inverse kinematics to precisely recreate human skeletal body components and joint centres, with the end objective of producing a comprehensive 3D model. Our model differs from others in that it is activity-independent while maintaining anthropometric regularity and good joint mapping accuracy and motion analysis with smooth motion frames. The suggested method extracts the entire humanoid figure motion curve, which can then be connected to a 3D model for near-real-time preview. Furthermore, because the full video feed is considered as a single entity rather than being processed frame-by-frame, smooth interpolation between postures is possible, with the interpolation accuracy controlled by the video stream sampling rate. The sample rate can be reduced for quicker video preprocessing in return for accuracy or vice versa, reaching a very rapid processing speed of 23 ms.

3. Materials and Methods

3.1. Biomechanical Model

Our technique reproduces human skeletal postures, deforms surface geometry, and is independent of camera position at each time step of capturing the depth video. In contrast to conventional human activity capture algorithms, the algorithm we created works well in processing frequent unsupervised indoor situations in which a potential patient films himself/herself performing rehabilitation activities using a sensor set. Figure 1 shows the 14 points recorded during the exercise in real time. When the recording is stopped, the system attempts to automatically categorise the beginning and end of the training movement by evaluating the motion pattern. The beginning and end of the movement exercise can also be adjusted by the supervisor (e.g., a rehab nurse or a medical doctor).

Figure 1. Location of 14 points in the human body that are recorded during exercise: 0. Head; 1. Chin; 2. Right shoulder joint; 3. Right arm elbow joint; 4. Wrist joint of right hand; 5. Left shoulder joint; 6. Left arm elbow joint; 7. Left wrist joint; 8. Right side hip joint; 9. Knee joint of the right leg; 10. Ankle joint of right leg; 11. Left side hip joint; 12. Left leg knee joint; and 13. Left ankle joint.

We are utilising the person's height, and arm and leg length. Using this information, the "skeleton" processing framework estimates the centre of each monitored joint. Clothing has no effect on the accuracy of the computation because the tracking system is based on depth (interpolated from HTC Vive sensors on the body) and does not employ an RGB camera.

A semantic connection is formed between the keypoints of the body and declares the order in which the recorded points must connect; that is, point "0" of the head cannot connect to point "11" on the left side of the hip joint.

The length of body parts (BPs), the ratio of BP sizes (based on predicted joint positions), and the angles of body parts connected by joints are the three key static properties based on skeletons. The first two are predicated on the idea that the subject's joints are kept in their relative placements (proportions) and lengths and, as a result, should retain their values over time. Since each BP should maintain its length, the subject's body dimensions and BPs spread, which are generated from depth images, serve as useful body joint properties to identify various human positions. Based on the estimated lengths of the joints, we include the total lengths of the BPs. The characteristic of the length is the following:

$$L^m = \sum_{i,i' \in I} D\left(\hat{J}_i^m - \hat{J}_{i'}^m\right) \tag{1}$$

where D is the Euclidean distance metric, J is the set of joint indices, and $D(\hat{J}_i^m, \hat{J}_{i'}^m)$ is the length of the BP between joints i' and i represented as $BP_{i,i'}$.

Another area of skeletal features can be based on the connection between joint locations. It may be calculated using the length ratio of the BPs. The ratio can be used to distinguish between patients according to their BP measurements. At time instance m, the ratio feature may be described as a subset of ratios between a collection of BPs.

The ratio between $BP_{i,i'}$ and $BP_{l,l'}$ for a subset of two BPs is defined as:

$$R^m = \frac{D\left(\hat{J}_i^m - \hat{J}_{i'}^m\right)}{D\left(\hat{J}_l^m - \hat{J}_{l'}^m\right)} \tag{2}$$

On the basis of the angles between two BPs and a common joint, the angular position (vertex) is determined. This might be an absolute angle relative to a reference system or a relative angle, such as the angle formed by the intersection of two BPs and their common joint. The following definition describes the degree-based relationship between the bodily components $BP_{i,i'}$ and $BP_{k,i'}$:

$$\theta_{i-i',i'-k}^m = \tan^{-1}\left(\left\|\frac{\left(\hat{J}_i^m - \hat{J}_{i'}^m\right) \times \left(\hat{J}_k^m - \hat{J}_{i'}^m\right)}{\left(\hat{J}_i^m - \hat{J}_{i'}^m\right) \cdot \left(\hat{J}_k^m - \hat{J}_{i'}^m\right)}\right\|\right) \tag{3}$$

where the operators $\times$, and $\cdot$ are the cross and dot products, respectively, and $\|\|$ is the Euclidean norm.

There is often no obvious preservation of bone lengths in human movement, and axial rotations of the limbs are not measured. Our method uses keypoint recognition across several camera frames as input for robust kinematic optimization (e.g., [46,47]) to achieve an accurate global 3D posture in real time that corresponds quite well to the overall system performance reported. When the subject is in the initiation position, the 3D coordinates of markers tied to anatomical landmarks are used to determine person-specific joint centres and axes.

During the start-up, the inertial characteristics of all body segments are determined using a regression model based on segment lengths and total body mass. We calculate the total body mass calculation following the method suggested in [48], which approximates the subject using the elliptical cylinder method and calculates his/her area. The density values are based on cadaver data from the literature and taken from [48].

3.2. Motion Recognition Model

The motion recognition model is based on Convolutional Pose Machines (CPM), which are based on deep neural networks. Neural networks are trained to detect keypoints in the human body, namely the joints, scalp, and chin. Figure 2 shows the architecture of a motion recognition system based on deep neural networks. The system architecture uses 16 convolutional layers (indicated by the letter "C" in Figure 2) and 6 pull-out layers (indicated by the letter "P" in Figure 2). Localisation of key points is performed using regions of different sizes in different convolutional layers. Seven differently sized regions are used to process an image with sizes of 400×400 pixels, 9×9, 26×26, 60×60, 96×96, 160×160, 240×240, and 320×320, respectively. The use of regions of different sizes ensures an accurate search for keypoints. The model processes the images received in real time from the video camera and outputs the changes in the x and y coordinates of the keypoints recorded over time.
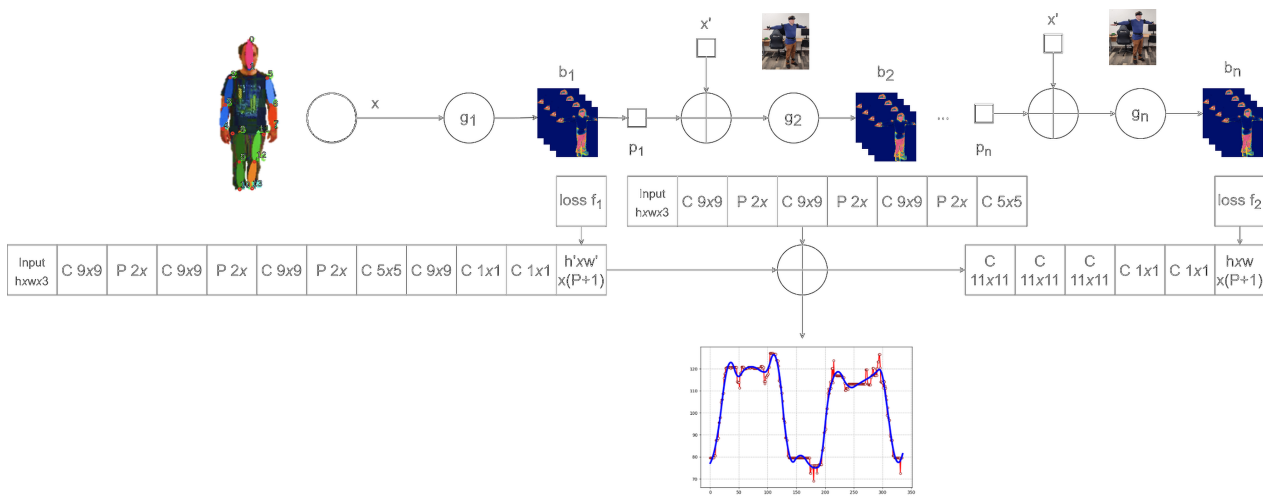

Figure 2. Deep neural network architecture of motion recognition system. Network supplements the IK model with motion curves in its output (in blue) and motion deviations (in red).

3.3. Architecture of BiomacVR System

The developed software system, named BiomacVR, for recording patient movements is described in this section. The system was developed using the Unreal Engine (Epic Games, Cary, NC, USA). The VR sensors required to record patient movement use HTC Vive VR equipment (HTC Corporation, New Taipei, Taiwan). The system requires a Microsoft Windows operating system (Microsoft Corporation, Redmond, WA, USA) and Steam software package (Valve Corporation, Bellevue, WA, USA). The programme uses the Steam VR subsystem to configure the VR environment used for sensor tracking.

The system consists of four software packages. The first package consists of a VR programme for a personal computer and VR glasses, the purpose of which is to record, edit and export data in .csv format for patient exercises. This part of the system, the VR session recorder, is used by both the doctor and the patient. The doctor controls this programme by selecting exercises and records the movements performed by the patient.

The calibration package links the patient and the virtual avatar. The calibration data are used to display the virtual avatar and track sensor data. The functionality of this package is used in the initial stage. The calibration package contains two functions used to calibrate the vectors showing the position of the bones and joints of the three-dimensional (3D) character (avatar), representing the appropriate parts of the body. Each segment is described by three vectors. These vectors may need to be calibrated in the original and other frames of the animation data. The same is true for the reference bones. The reference bones and joints are statically determined vectors located at specific body locations. Each reference bone and joint is compared to an analogous moving bone to calculate the angle between them.

The processing package allows viewing and editing of recorded data. When a patient's exercise is recorded, some of the data are redundant (what happens before the exercise and what happens after the exercise is stopped). This package allows us to view the entire session and to crop the entry by marking the beginning and end of the session. This information is then used in the export package. This package has three auxiliary functions that are used to calculate the required information. The frame calculator takes the node information for each frame of the session (from the marked start to the end), calculates the required vectors (vector calculator) and the angles between them (angle calculator), and either displays or exports the calculated information for each frame.

The export and import packages are for exporting and importing session records. Data are exported in .csv format. They can then be imported into a recording programme for further editing or can be used for data analysis.

The scenario package performs the analysis of the data to be recorded. Different nodes and vectors showing the positions of the bones need to be recorded for each exercise. When selecting an exercise to be recorded, one of the scripts is activated, which registers the data required for that exercise.

3.4. Keypoint Tracking Program

The graphical interface window of the keypoint tracking programme is shown in Figure 3. The programme is implemented in the Python (Python Software Foundation, Wilmington, DE, USA) programming environment using the following libraries:

- TensorFlow—an in-depth machine learning tool and ecosystem for the use of decision-making models in automated recognition processes;
- PySimpleGUI—a tool for creating a graphical user interface, also used to create Web applications;
- OpenCV (Open Source Computer Vision Library)—image processing library, which is used in the programme as communication between the hardware system part (colour cameras) and the software part (image processing);
- Scipy—a software tool for filtering time series signals.

The programme is divided into three main windows:

- "Real View" window is for displaying "raw" data. In this window, the user has the option to analyse the video (in .avi and .mp4 formats) or to perform the analysis using a video input device (standard webcam).
- "Keypoints" window visualizes the results of the keypoint recognition algorithm. The detected points are drawn on the analysed frame in a certain order. The user can assess how successfully and how accurately the keypoints of the human body are detected.
- "Dynamics" window presents the recorded values in graphical form. The user can export the selected data. These can be the coordinates of all keypoints or the estimated dynamics of the change in distances between selected points.
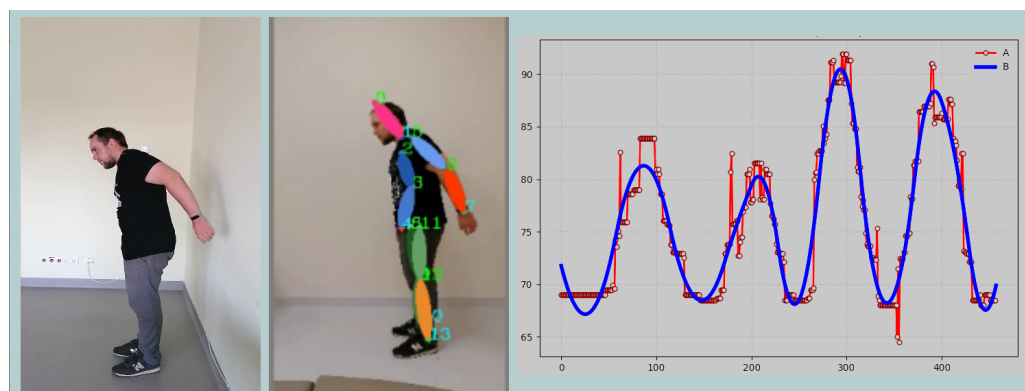

Figure 3. A snapshot of the application window showing live view (**left**), skeleton keypoints (**middle**), and dynamic data view (**right**).

3.5. Movement Tracking System Configuration

This section introduces the configuration of the human exercise tracking system that was used to record exercises. The developed exercise tracking system uses the HTC Vive system and at least eight second-generation HTC Vive sensors. These sensors allow you to track the position in space and angles of rotation. The system requires at least two HTC Vive units (called base stations), but it is recommended to use four stations for easier use and more-accurate tracking. The layout of the sensors is shown in Figure 4.
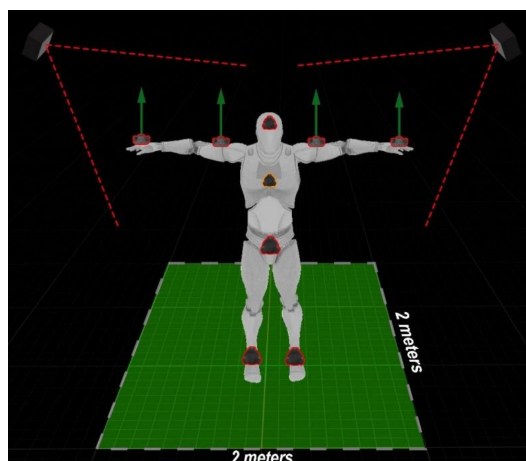

Figure 4. System and sensor layout view: At least 2 m should be made available for unrestricted movements of a subject.

The sensors are layout as follows:

- two sensors are placed on the hands, pointing them upwards;
- two sensors are placed on the arm, pointing them upwards;
- two sensors are placed on the legs, pointing them forward;
- one sensor is placed on the hips, pointing it forward;
- one sensor is placed on the head, pointing it forward;
- an additional 1 sensor can be placed on the chest, pointing it forward.

At the beginning of the exercise tracking session, the sensors must be placed first on the patient, as shown in Figure 5. Attention must be paid to both the position and the orientation of the sensors. Sensors must be tightened so that they do not become distorted during exercise. After the sensors have been installed on the subject, the sensors must be matched to the virtual avatar during system calibration.
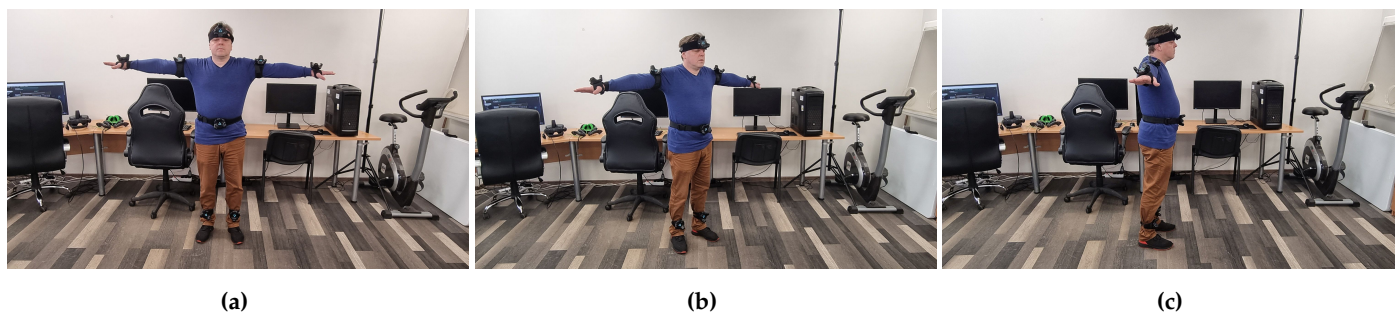

(**a**)　　　　　　　　　　　　　　(**b**)　　　　　　　　　　　　　　(**c**)

Figure 5. Example of sensor mounting: (**a**) front view, (**b**) 45-degree view, and (**c**) side view.

The subject is first asked to stand in the so-called 'T' position (standing upright with arms outstretched) then extend both hands forward after waiting 10 s. Naturally, standing in the "T" posture for 10 s might be challenging for post-stroke patients due to hemiparesis or loss of muscle tone, strength, or coordination. In this instance, the aid of a nurse or home caregiver is required to hold both affected arms in the correct position during system

calibration (Figure 6). Because the assistant (e.g., a nurse or home caregiver) is not wearing sensors, she is not detected by our system and has no bearing on the skeleton reference registration even if her hands cover the actual sensors on some body part(s) of the patient. The calibration of VR sensors aims to link VR trackers with a human and a virtual avatar. During calibration, we indicate which sensor is attached to a certain part of the body, and after a couple of movements, the person is associated with his virtual avatar (Figure 7). The calibration data are then recorded and the exercise can be performed.

Figure 6. Assignment of sensors to body parts during calibration of the BiomacVR system. A person is assisted by a nurse.
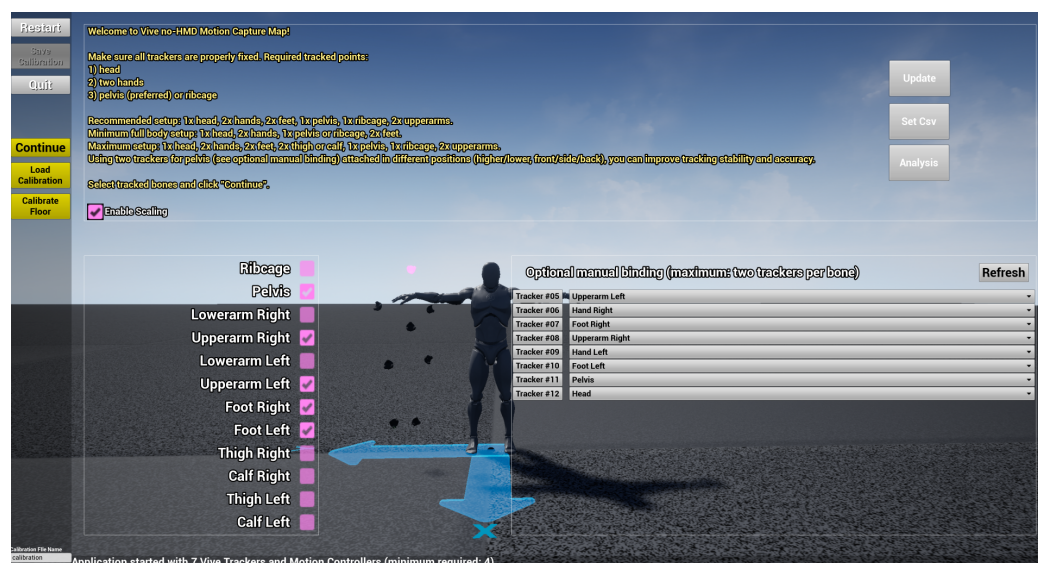

Figure 7. Assignment of sensors to body parts during calibration of the BiomacVR system. An assistant person is not visible in the system and has no effect on the representation of a patient.

3.6. Classification Methods

Various classification and prediction methods can be used for health assessment and can identify possible pathologies from digital images, biological or motion signals, survey data, etc. [49,50]. Machine learning (ML) involves the use of advanced statistical and probabilistic methods in the construction of systems that can automatically learn from the data provided. Because ML algorithms perform fast and high-quality analysis of complex data, they are extremely popular in the study of various health disorders to improve the patient's condition and to increase the understanding of the patient's physiological condition and its control [51]. Depending on the amount of data or the information available

on the data sample itself, an algorithm category or several algorithms shall be selected for the study. After testing, the model that best describes the data is selected.

3.6.1. Random Forest

A random forest (RF) consists of individual decision trees (DTs) that can be trained sequentially or simultaneously with a random sample of data [52]. In each tree, all nodes have defined conditions that specify the properties by which the data are broken down into two separate sets. Examining the recorded signals (angles of motion) of a healthy person can show significant differences compared to a person with a movement disorder.

RF has many parameters whose values need to be defined in advance. There is no single general rule that specifies which parameter set is most appropriate for the data being analysed. Their setting can take a very long time, so we use a random grid search, that is, a fixed number of parameter value combinations randomly selected 1000 times. The following hyperparameters of the RF model are set:

- Learning rate (*'learning_rate'*)—Determines the effect of each newly added tree on the final result. The values are from the range $[0.2, 0.15, 0.1, 0.05, 0.01, 0.005]$;
- Number of trees (*'n_estimators'*)—The number of trees in the RF model. The values are from the range $[4, 6, 8, 10, 15, 20, 25, 30, 35, 40, 45, 50]$;
- Maximum depth (*'max_depth'*)—The maximum height of a single tree. The values are from the range $[2, 3, 4, 5, 6, 7]$;
- Minimum number of elements (*'min_samples_split'*)—Nodes with fewer elements are not split. The values are from the range $[2, 4, 6, 8, 10, 12, 18, 20, 40]$;
- Minimum number of items per sheet (*'min_samples_leaf'*)—The smallest possible number of elements on a sheet. The values are from the range $[1, 3, 5, 7, 8, 9]$;
- Maximum number of features (*'max_features'*)—Defines the maximum number of characteristics that describe each breakdown. An array of values is used in the grid $[2, 3, 4]$.

A total of 70% of the data sample (randomly) is used for training and validation of the RF model, and 30% is for testing.

3.6.2. Convolutional Neural Network

A Convolutional Neural Network (CNN) is a multilayer neural network using at least one convolutional layer. A convolutional layer (conv) is a layer of artificial neurones in which mathematical cross-correlation calculations are performed by combining two different data samples. This operation replaces the data description function by reducing the dimension of the input data. This layer is required in any CNN model in order to reduce the number of parameters describing the data and, at the same time, to shorten the learning time. Pooling works in a similar way to convolutional layers. It reduces the amount of data by leaving only the most important numerical values in the data segment. These are usually average pooling values or maximum pooling values. One way to protect an emerging CNN model from overfitting is by introducing a dropout layer. This layer makes the learning process noisier by introducing randomness. This makes the model more-or-less dependent on the input data [53]. After the input data pass through all the layers listed so far, a flatten procedure is performed, during which the data from the matrix form are transformed into a vector. They are then used as input data in the artificial neural network (ANN). A "dense" operation is already performed in the ANN network, where each neurone in the described layer of the model receives output data from each previous layer neurones. Moving to the next layer with a smaller number of neurones, a matrix product of vectors is performed. During the study, convolutional neural networks of different sizes and different layers were analysed. Finally, a CNN model consisting of two convolutional layers, one screening layer, one junction layer, and smoothing and two compression procedures in the neural network was chosen to address the problem of classifying people with and without mobility impairment. Sequence, input and output data, and visual representation are provided in Figures 8 and 9.
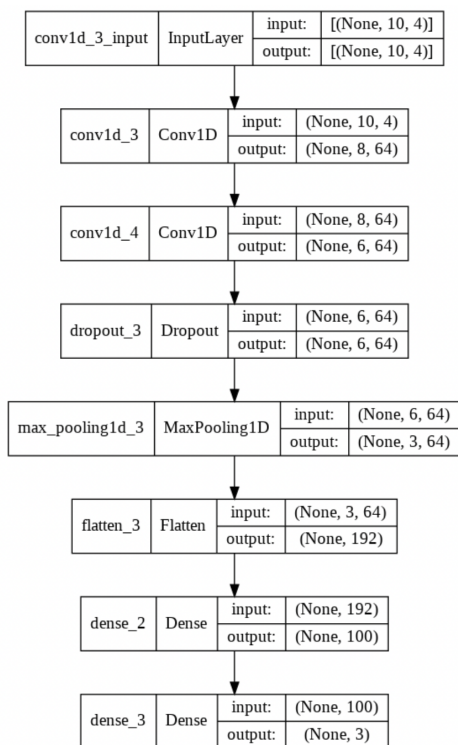
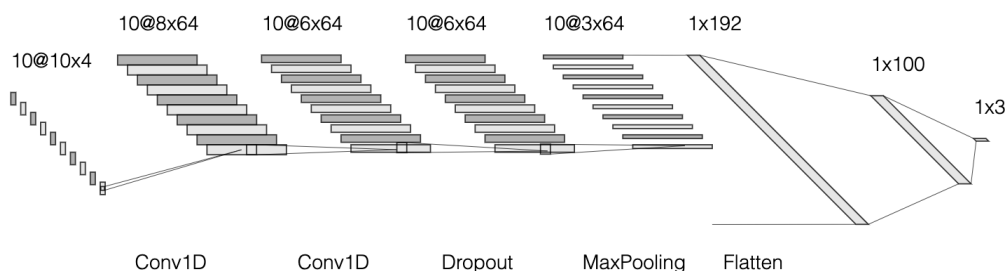**Figure 8.** Sequence of layers of a convolutional neural network.

Figure 9. Layout diagram of a convolutional neural network.

Our network employs a two-stage technique, first estimating the 2D joints from the input images and then estimating the 3D posture from the predicted 2D joints. As the 2D joint estimation module, we employ the cutting-edge stacked hourglass technique [54], and to produce various 3D pose hypotheses, we use our own processor, which comprises a feature extractor and a hypotheses generator [55]. Each hourglass is comprised of an encoder and a decoder that perform downsampling using convolution and pooling and upsampling with bilinear interpolation, respectively. Because it layers these hourglasses to repeat bottom-up and top-down processes, the model is known as a stacked hourglass network [56]. The model collects data at various input sizes. In addition, interim supervision is applied to the heatmaps produced by each stack.

The stacking process is described by

$$X_{i+1} = X_i + (1 + C(x))T \tag{4}$$

where X_i is the input of the ith-level hourglass network, and T is the output of the main network; $C(x) = \sigma\left(W_f^2 \delta\left(W_f^1 g\right)\right)$ where δ and σ are, respectively, the ReLU and sigmoid

functions, and $W_f = \left\{ W_f^1, W_f^2 \right\}$ are the parameters of the fully-connected (dense) layers; and $g = \frac{1}{H \times W} \sum\limits_{i=1}^{H} \sum\limits_{j=1}^{W} x(i,j)$ are weights of input image x with height H and width W.

After upsampling and the insertion of residual connections, the network employs two layers of linear modules with 33 convolutions to create output heat maps that predict the existence of a joint at each pixel. This helps to get the maximum features with convolutions and keeps the original data with residual skip connections. The residual block processes convolutions and the max-pooling feature down to a very low resolution, and the network achieves this low resolution by permitting smaller spatial filters as follows:

$$F(x_n, W_r) = conv_{1 \times 1} \left(conv_{3 \times 3} \left(conv_{1 \times 1} \left(x_n, W_r^1 \right), W_r^2 \right), W_r^3 \right) \tag{5}$$

where $W_r = \left\{ W_r^1, W_r^2, W_r^3 \right\}$ are the parameters of the residual branch, and x_n is the input feature map.

Using the skip connection, we mix the two distinct resolutions by closest neighbour upsampling the one with low resolution and doing an element-wise addition. Note that just the depth is modified here, while the size remains constant. The projected position of numerous joints is depicted in these heatmaps, and the loss function $\mathcal{L}$ is used to train the network:

$$\mathcal{L} = \frac{1}{N} \sum\limits_{n=1}^{N} \sum\limits_{ij} \left\| H_n(i,j) - \hat{H}_n(i,j) \right\|^2 \tag{6}$$

which calculates the loss between the ground truth heat map $\hat{H}$ and the heat map H predicted by the network using the mean squared error (MSE).

The intermediate supervision given to the predictions of each hourglass output is the important principle that underpins the stacking of hourglass modules presented by the authors in the original study [57]. This means that not only are the predictions of the previous hourglass monitored, but each hourglass is supervised as well.

In addition to the CNN network structure (layers), it is still necessary to define the parameters required for training, such as the number of epochs, the sample size, and the number of sample data required for validation. As with the RF model, a random search grid is created to automatically generate a CNN model for each movement under study. After 100 iterations, the model is selected that most accurately classifies the data. The following hyperparameters of the CNN model are considered:

- Number of epochs (*'epochs'*)—Specifies the number of times learning will be performed using all input data. The values are from the range $[6, 8, 10, 12, 14, 16, 18, 20, 21, 22, 50]$;
- Sample size (*'batch_size'*)—Shows the sample size of the input data used (the sample is used in the learning process before updating the model and moving on to the next era). The values are from the range $[20, 28, 34, 40, 48, 55, 68, 74, 80]$;
- Validation sample size (*'validation_split'*)—Shows what part of the sample data is used in the learning process and what is used during validation. An array of values is used in the grid $[0.05, 0.1, 0.15, 0.2, 0.25, 0.3]$.

3.7. Performance Analysis

The performance analysis algorithm is presented in the form of Algorithm 1. The programme code can process the image directly from the webcam 'WEBCAM ID' (if there is more than one camera, its identifier must be specified, indexed from 0 up, in whole numbers) or from a video file with the path 'PathToVideo'. In Step 1, the programme creates an object cap that connects to the specified video sources. In Step 2, the programme performs iterative calculations until all video frames are analysed or the programme is terminated. In Step 3, the cap object scans the image frame and saves it as a variable frame. The colour frame frame is processed in Step 5, where the frame size is changed as needed. The modified-sized image frame is analysed, and the coordinate points of the keypoints

and their hierarchy are derived (see Step 6). The value of all keypoints is displayed in the image in Step 10.

Algorithm 1 Image processing and recognition

```
# Reading the video source
1: cap <- cv2.VideoCapture ('Path to video or WEBCAM ID')
2: while (True):
# Scans a frame from an image source
3: ret, frame <- cap.read()
4: try:
5: img <- cv2.resize(frame, (width, height))
# A point detection model is applied
6: points, hierarchy <- model.predict(img)
# Displays all found points in a circle
# using a radius of 10 pixels
7: radius = 10
# color—The color of key points
8: color = (150, 150, 150)
9: for k in range (0, len(points)):
10: cv2.circle(img, (point[k, 0], point[k, 1]), radius, color, −1)
11: endfor
12: endwhile
```

3.8. Evaluation of Classifier Performance

To solve the problem of two classes of supervised learning, each element in the validation (or testing) sample is assigned a positive or negative class (usually 0 or 1). In this study, most exercises have two negative classes, denoted 1 and 2, and both indicate incorrect movement. The machine learning algorithm teaches the model to separate these two or three classes according to the provided data. In the end, a prediction is made for each item of the test data sample. The algorithm then assigns the elements to one of the categories provided based on the predictions obtained (Table 1).

Table 1. Categories of classification items obtained from the forecast.

Name	Abbreviation	Description
True negative	TN	The item is classified as incorrect in the prediction and is in fact incorrect
True positive	TP	The element is classified as correct in the prediction and is actually correct
False positive	FP	The item is assigned as predictive in the forecast and is in fact incorrect
False positive	FN	The item is classified as incorrect in the prediction and is actually correct

If many elements from the TP or TN categories are obtained during testing/validation of the ML algorithm, it means that the algorithm is able to correctly classify as positive elements that were actually positive in the validation data sample (TP) or as negative those that had a negative value in the validation data sample (TN). The table for all categories is called the confusion matrix. To understand how well the resulting algorithm performs in the general case, the overall accuracy of the model is calculated.

3.9. Statistical Analysis

The exercises performed in this study are analysed separately. The nodes to be analysed are selected for each exercise and describe how the angles of the nodes should change during the exercise correctly and incorrectly. The data obtained from the experiment are then analysed. The following tools are used for the analysis:

- Statistical confidence intervals indicate that 95% of the measured values are in these ranges, i.e., there is only a 5% probability that a specific value will not be in this range.
- Student's *t*-test evaluates the equality of the means to zero. The null hypothesis is that the means of the two samples are equal. If the value of p obtained is greater than the significance level of 0.05, then the null hypothesis cannot be rejected. Otherwise, the null hypothesis is rejected.
- One-way analysis of variance (ANOVA) tests the hypothesis that the samples in the y-columns are composed of populations with the same mean compared to the alternative hypothesis that the population values are not all the same, and the hypothesis that all measures in a group are equal to an alternative hypothesis that at least one group is different from the others.

4. Exercises Used in Experimental Studies

4.1. Background and Motivation

Postural examination is often the initial component of any patient's tests and measurements for any musculoskeletal problem. The therapist looks at the patient from the front, rear, and sides. Postural assessment is a critical component of objective evaluation, and ideal static postural alignments have been proposed. However, both static and dynamic postures must be assessed to determine the patient's functional mobility and capacity to self-correct a static habitus. Scoliosis, postural decompensation, anatomic short leg, previous trauma or surgery, trunk control (after stroke), or particular segmental somatic dysfunctions in areas of the body where asymmetry is present can all be caused by postural misalignment or asymmetries. There are some items that are critical to examine at the at the beginning of the rehabilitation programme [58]. Active range of motion (ROM) in all peripheral joints, particularly the shoulder and upper limbs, should be monitored and recorded during a postural exam and especially during training due to observing exercises performed correctly and accurately. Pain, weakness, muscle shortening, and oedema can all be causes of joints mobility restriction. Shoulder ROM, muscular performance, and strength deficiencies can all influence postural alterations or inappropriate posture and incorrectly performed exercises during training. That is why during evaluation we employed full active ROM of the shoulder (extension/flexion; adduction/abduction; horizontal adduction/abduction; and rotation) as factors that most impact postural alterations or occurring compensatory mechanisms (for example, complete shoulder flexion affects trunk control and leads to back hyperextension) [59,60].

4.2. Setup

The main physical exercises used in stroke rehabilitation were selected to recover motor function from the upper limb and shoulder. Subjects without health disorders simulated correct and incorrect rehabilitation movements (exercises) 5 times. Each data set consists of one correct and one two incorrect logged signals. Physical movements were recorded by measuring the angles between the respective sensors. Each exercise has two or three possible scenarios:

- Correct movements (the subject performs the exercise exactly)—the angles registered during the movement are assigned to Class 0;
- Incorrect movement 1 (the subject performs the exercise incorrectly)—the angles recorded during the movement are assigned to Class 1;
- (Optional) incorrect movement 2 (the experimenter performs the exercise incorrectly but in a different way than described above)—the angles recorded during the movement are assigned to Class 2.

The following exercises were selected for the experimental study:

1. Reaching the nose with the index finger of the left hand (shoulder adduction 0–90 degrees, frontal plane, sagittal axis through the centre of humeral head/elbow flexion 0–145 degrees, sagittal plane, transverse axis through the centre of lateral epicondyles) (see Figure 10);

2. Reaching the nose with the index finger of the left hand performing motion with compensation through the torso and neck (shoulder adduction 0–90 degrees, frontal plane, sagittal axis through the centre of humeral head/elbow flexion 0–145 degrees, sagittal plane, transverse axis through the centre of lateral epicondyles) (see Figure 11);
3. Bending of the arm up to 180 degrees (shoulder flexion 0–180 degree, sagittal plane, transverse axis through the centre of humeral head) (see Figure 12);
4. Bending of the arm up to 90 degrees (shoulder flexion 0–90 degree, sagittal plane, transverse axis through the centre of humeral head) (see Figure 13);
5. Bending of the arm up to 90 degrees with compensation through the torso (shoulder flexion 0–90 degree, sagittal plane, transverse axis through the centre of humeral head) (see Figure 14);
6. Bending of the arm up to 90 degrees with compensation through the shoulder/neck (shoulder flexion 0–90 degree, sagittal plane, transverse axis through the centre of humeral head) (see Figure 15);
7. Lifting back the arm with compensation through the torso and neck/head (shoulder extension 0–45 degrees, sagittal plane, transverse axis through the centre of humeral head) (see Figure 16);

All exercises were performed in the range of 60 to 120 s. Statistical analysis was applied to the coordinates of the change to the registered keypoints.

4.3. Finger–Nose

This exercise is carried out while standing. The participant must reach the tip of the nose with the index finger of the left hand during the exercise. Initially, the left arm should be raised to the side and at shoulder level bend over the elbow joint to reach the nose. The tip of the nose must be reached without the use of any other additional movements, i.e., without involving movements of the shoulder, head, and/or torso.
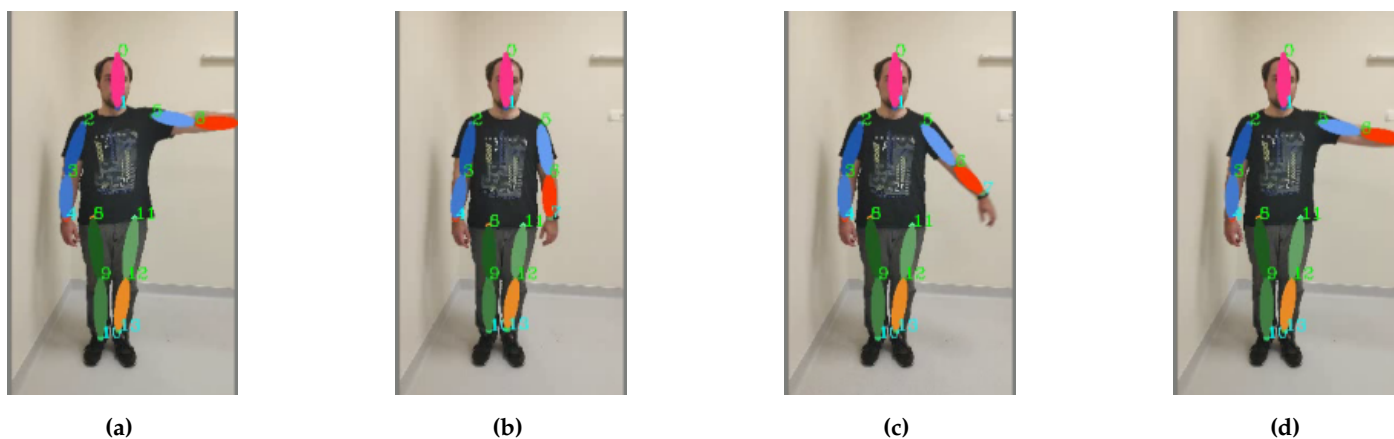

(a)　　　　　　　　　　(b)　　　　　　　　　　(c)　　　　　　　　　　(d)

Figure 10. Footage of the Finger–Nose exercise: the motion sequence (**a**–**d**) starts on the left and ends on the right.

4.4. Finger–Nose with Compensation through the Torso and Neck

Exercise is carried out while standing. The task of the exercise is to reach the tip of the nose with the index finger of the left hand by including motion compensation. Motion compensation must be done using a vest. This exercise simulates a disorder of coordination and control of the body.
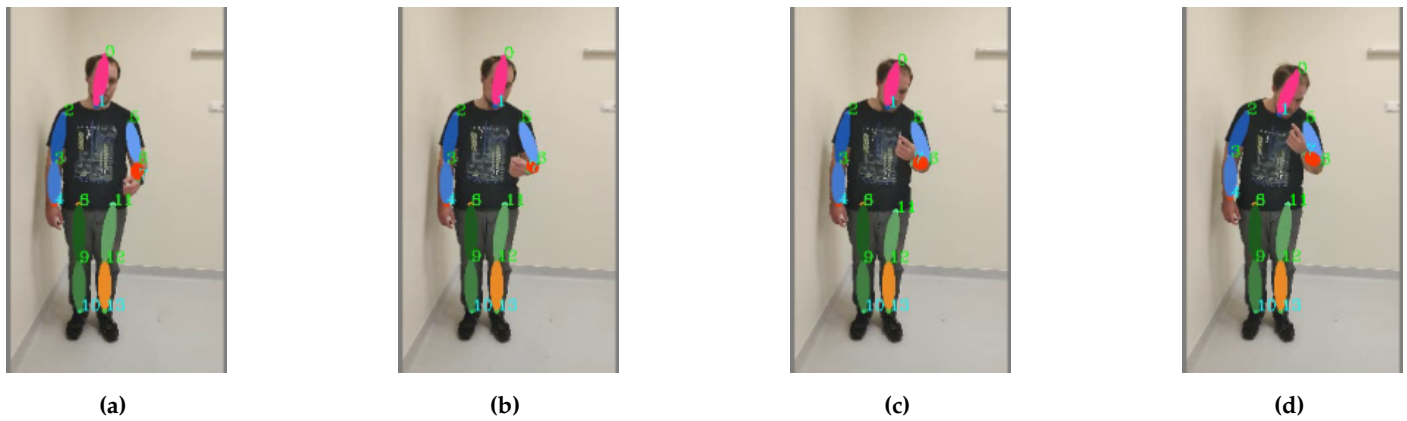
(a) (b) (c) (d)

Figure 11. Exercise "Finger–Nose" is performed with compression of the torso: the motion sequence (**a**–**d**) starts on the left and ends on the right.

4.5. Arm Bending to 180 Degrees

The exercise is performed while standing and aims to raise the left arm above your head using only the muscles associated with the arm. The left hand must be raised above the head from the lowered position. The hand must be extended and raised in front of you.
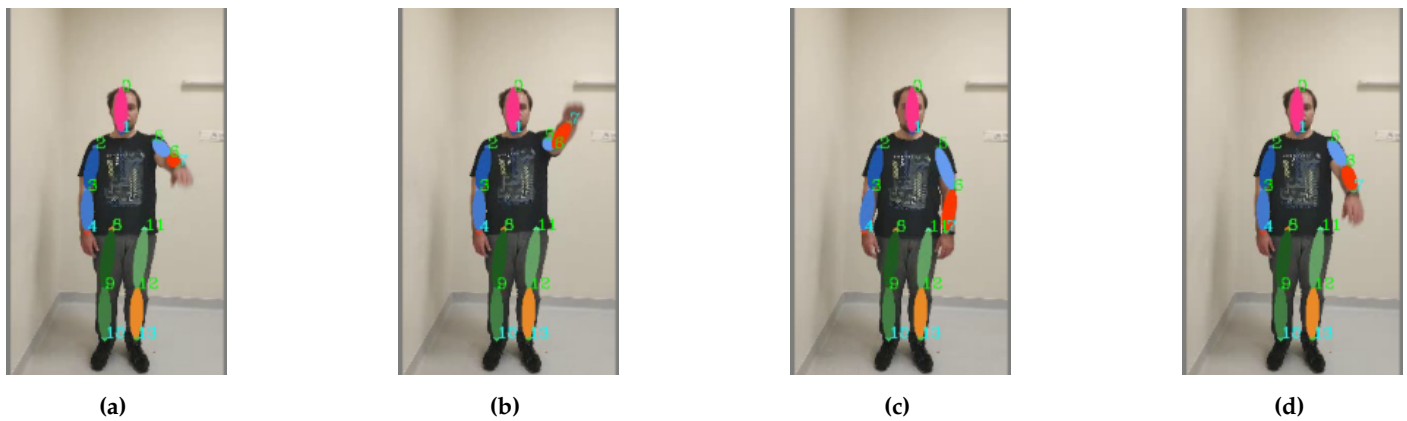

(a) (b) (c) (d)

Figure 12. Footage of the exercise "Bend the Arm": the motion sequence (**a**–**d**) starts on the left and ends on the right.

4.6. Arm Bend to 90 Degrees

The exercise is performed while standing and aims to raise the left arm to shoulder level. It is requested that no compensatory movements be used during the exercise. Exercise should be done for up to 120 s at a constant speed.
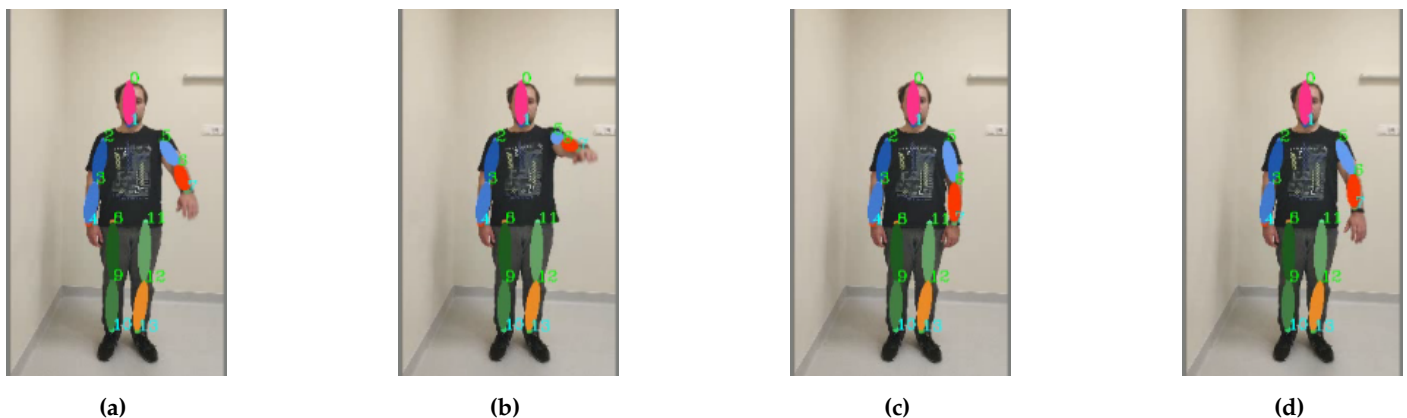

(a) (b) (c) (d)

Figure 13. Arm bending up to 90 degrees freestyle exercise footage: the motion sequence (**a**–**d**) starts on the left and ends on the right.

4.7. Arm Bending with Compensation through the Torso

The exercise is performed while standing and aims to raise the outstretched left arm above the head using additional torso movements. The exercise aims to simulate movement coordination disorders. The left hand should be extended in front of you during the bending of the arm and raised. Torso movement must be connected to exercise halfway through the movement.
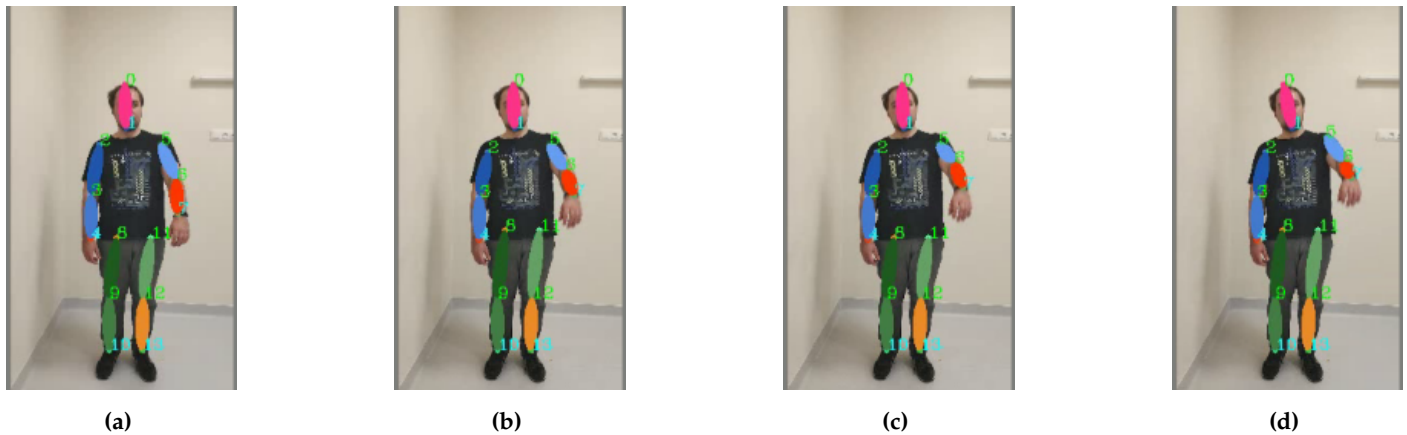

| (a) | (b) | (c) | (d) |

Figure 14. Arm bending with compensation from the torso: the motion sequence (**a**–**d**) starts on the left and ends on the right.

4.8. Arm Bending with Compensation through the Shoulder

The exercise is performed while standing and aims to raise the outstretched left arm above the head using additional shoulder movements. The exercise aims to simulate movement coordination disorders. The left hand should be extended in front of you during the bending of the arm and raised. The shoulder movement must be connected to the exercise halfway through the arm's trajectory. The movement of the shoulders must compensate for the movements of the arm.
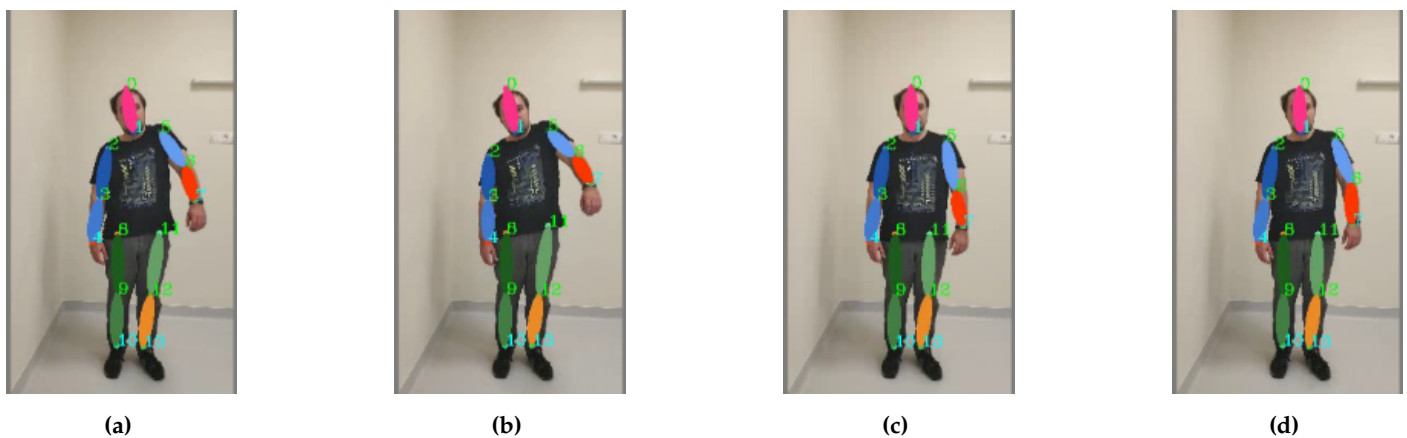

| (a) | (b) | (c) | (d) |

Figure 15. The arm bends with compensatory shoulder support: the motion sequence (**a**–**d**) starts on the left and ends on the right.

4.9. Lifting Back the Arm with Compensation through the Torso and Neck/Head

The exercise is performed while standing, and the left arm is lifted back as far possible. The exercise aims to mimic movement coordination disorders when arm extension is compensated by other movements across the torso or shoulder. Movement of the shoulder or torso must be connected to the exercise halfway through the ascent of the arm. The movement of the shoulder or torso must compensate for movements of the arm.
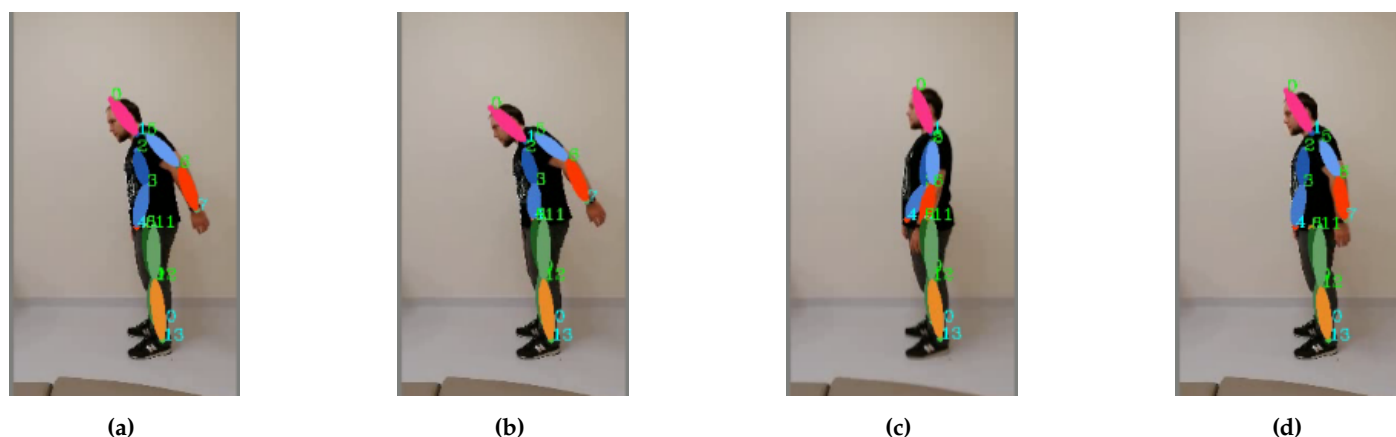
(a)   **(b)**   **(c)**   **(d)**

Figure 16. Arm raising with compensation: the motion sequence (**a**–**d**) starts on the left and ends on the right.

5. Experimental Evaluation

5.1. Dataset Collection

The experiment used a database of various exercises compiled by Vilnius University Hospital Santaros clinics (Vilnius, Lithuania). The recordings used in the study were obtained using two different depth sensors; an example of a data set is shown in the figure. The first depth sensor, the Intel Realsense L515, was placed in front of the subject, and the second depth sensor, the Intel Realsense D435i, was located 90 degrees to the right of the subject. Both sensors were installed at a height of 1.4 m above the ground and a distance of 1.8 m from the subject.

A total of 16 healthy subjects (mean (SD) = 43 (S.D. 11) years) volunteered from our institutions throughout Stage 1 of our study. We began our study with these healthy volunteers to confirm that our system was functioning properly and to better understand the usability of our technology. This was followed by Stage 2, in which 10 post-stroke (mean (SD) = 57 (S.D. 13) years) patients took part. The goal was to determine how our method may help with upper limb rehabilitation. The stroke patients enlisted after being referred by a physiotherapist.

The criteria for inclusion were:

1. A post-stroke timespan of >6 months;
2. The capacity to follow directions;
3. The capacity to observe a computer or TV screen from a 1.5-meter distance (safe distance to avoid hitting the device during exercises);
4. A clinician-performed MMST—Mini-Mental test following a minimum 24 points rule. This is commonly used after a stroke to assess cognitive function to check if motor skills are good and a person can understand or follow commands. Then, muscle tone evaluation was applied using a modified Ashford scale to avoid the condition of hypertonic limbs, because such persons would not be able to perform the full movement required and forced attempts to do this would "lock" the limbs, leading to a spasmic limb. Finally, active range-of-motion (ROM) capabilities were checked.

Individuals after recent surgery (within the last 5 months) and individuals with pacemakers were omitted from the trial. Persons that were not vaccinated against heinous Covid virus were excluded from this study. The results in the paper contain the results only from the second subject group.

The database consisted of 10 subjects, each of whom performed the following steps:

1. Shoulder flexion (shoulder flexion 0–180 degree, sagittal plane, transverse axis through the centre of humeral head);
2. Shoulder flexion and internal rotation (shoulder flexion 0–90 degree, sagittal plane, transverse axis through the centre of humeral head/internal rotation 0–70 degrees, transverse plane, vertical axis through the centre humeral head);

3. Shoulder flexion and internal rotation, elbow flexion (shoulder flexion 0–90 degree, sagittal plane, transverse axis through the centre of humeral head/Internal rotation 0–70 degrees, transverse plane, vertical axis through the centre humeral head/elbow flexion 0–145 degrees, sagittal plane, transverse axis through the centre of lateral epicondyles);

4. Shoulder extension and internal rotation (shoulder extension 0–45 degrees, sagittal plane, transverse axis through the centre of humeral head/internal rotation 0–70 degrees, transverse plane, vertical axis through the centre humeral head);

5. Shoulder flexion and external rotation, elbow flexion (shoulder flexion 0–180 degree, sagittal plane, transverse axis through the centre of humeral head/internal rotation 0–90 degrees, transverse plane, vertical axis through the centre humeral head/elbow flexion 0–145 degrees, sagittal plane, transverse axis through the centre of lateral epicondyles);

6. Shoulder horizontal adduction/abduction (shoulder adduction 0–120 degrees/abduction 0–30 degrees, transverse plane, vertical axis);

7. Shoulder adduction (shoulder adduction 0–180 degrees, frontal plane, sagittal axis through the centre of humeral head).

Each video in both data sets was pre-processed for use in neural network training. This was done by extracting frames every 0.5 s to minimise similar frames in the video channel. The given frame was then passed to an optional deep learning network to extract its connections, and each joint was then calculated to determine orientations to teach the IC neural network.

5.2. Data

Human skeletal movement was observed using information visible from depth cameras. A graphical representation (joint coordinates) of the variables analysed is presented in Figure 17 and Table 2. The coordinates of the joints are indicated by J_i, where i is given in Figure 17.
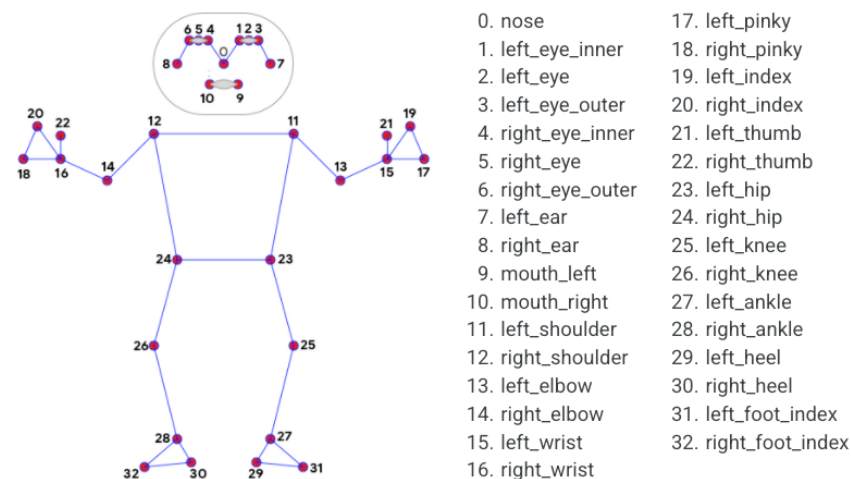

Figure 17. Position of skeletal joints: individual joints are numbered.

5.3. Movement Analysis

Exercises that tested the motion detection system are described above. The dynamic change of the three distances during the above-mentioned exercises was evaluated to determine the accuracy and speed of the system. We analysed the distances between:

- The point of the yoke and the elbow joint of the left arm;
- The point of the yoke and the wrist joint of the left arm;
- The distance between the point of the yoke and the centre of the face.

The coordinates of the yoke point are calculated taking into account the locations of the keypoints shown in Figure 17.

Table 2. Features of the analysed data. Here J_i is the position of the i-th joint, x is the horizontal coordinate, and y is the vertical coordinate.

Variable	Explanation
$|J_{11}(y) - J_{12}(y)|$	Shoulder Line—Absolute difference between the vertical coordinates of the 11th and 12th points
$|J_2(y) - J_5(y)|$	Eye Line—Absolute difference between the vertical coordinates of the 2nd and 5th points
$\left|\frac{(J_{11}(x) + J_{12}(x))}{2} - \frac{(J_{23}(x) + J_{24}(x))}{2}\right|$	Spine line—Absolute difference between the horizontal coordinates of the 0th and 12th points between the 11th and 12th points

The coordinates of the keypoints and the geometric distances between the keypoints are expressed in pixels of the digital image as follows:

$$x_J = \frac{(x_2 + x_5)}{2}, y_J = \frac{(y_2 + y_5)}{2}, z_J = \frac{(z_2 + z_5)}{2} \tag{7}$$

The central coordinates of the face are calculated as follows:

$$x_V = \frac{(x_0 + x_1)}{2}, y_V = \frac{(y_0 + y_1)}{2}, z_V = \frac{(z_0 + z_1)}{2} \tag{8}$$

The distance between the point of the yoke and the elbow of the left arm is calculated as follows:

$$D_{JA} = \sqrt{(x_J - x_6)^2 + (y_J - y_6)^2 + (z_J - z_6)^2} \tag{9}$$

The distance D_{JR} between the Jung point and the wrist joint of the left arm is calculated according to:

$$D_{JR} = \sqrt{(x_J - x_7)^2 + (y_J - y_7)^2 + +(z_J - z_7)^2} \tag{10}$$

The distance between the Jung point and the centre of the face is calculated as follows.

$$D_{JV} = \sqrt{(x_J - x_V)^2 + (y_J - y_V)^2 + (z_J - z_V)^2} \tag{11}$$

where x, y, z are the spatial coordinates of the respective joints.

5.4. Case Study: Analysis of Spine Line for Health Diagnostics

The most prevalent condition among all occupational-related disorders is "low back pain" (LBP) [61]. We aim to assess spinal functioning traits in sedentary workers. The results of spine line measurments are presented in Figure 18. Statistical analysis of the average, minimum, maximum, and median horizontal deviation shows a statistically significant difference (see Table 3, $p < 0.001$) between normal and diseased subjects with LBP.

Table 3. Results of the ANOVA test for spine line data using different statistical functions. Rejection of hypothesis means that the means of data are significantly different.

Function	p	F	Equality Hypothesis
Min	4.2670×10^{-35}	328.5703	Rejected
Median	7.0028×10^{-57}	965.2931	Rejected
Max	7.0954×10^{-45}	546.0185	Rejected
Mean	7.9203×10^{-57}	962.9296	Rejected

(a) Mean horizontal deviation.

(b) Minimum horizontal deviation.

(c) Maximum horizontal deviation.
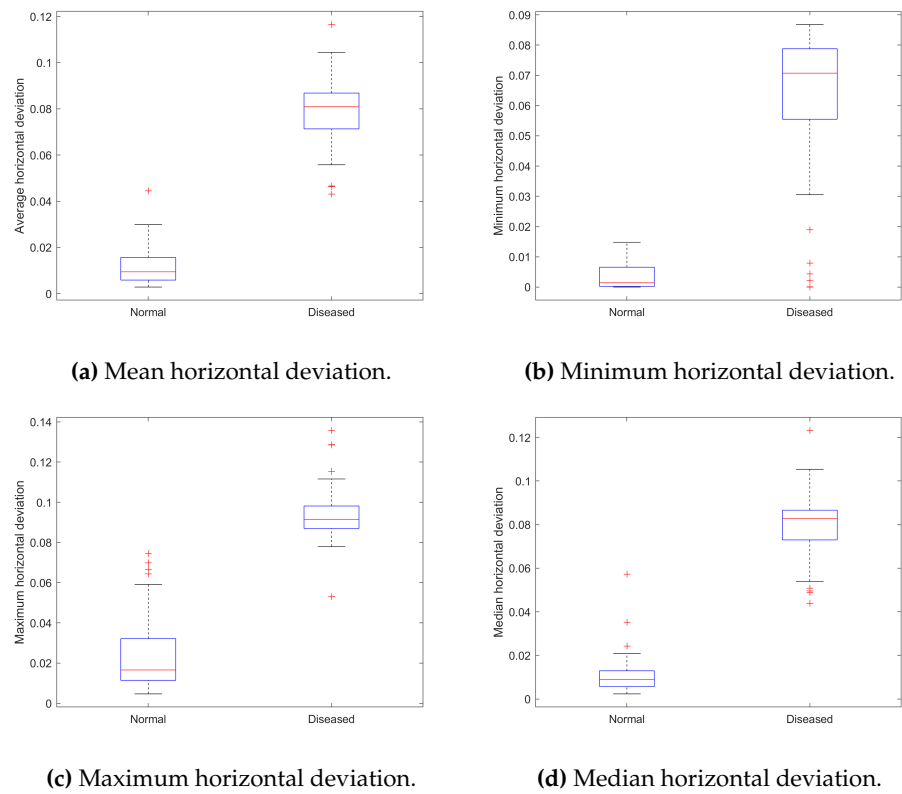
(d) Median horizontal deviation.

Figure 18. Summary of statistical characteristics of data for spine line: (**a**) Mean horizontal deviation, (**b**) Minimum horizontal deviation, (**c**) Maximum horizontal deviation, and (**d**) Median horizontal deviation.

The importance of features (spine line, eye line, or shoulder line) was evaluated using feature ranking by class separability criterion (absolute value of two-sample *t*-test with pooled variance estimate). The results presented in Figure 19 show that the spine line feature is the most important in discriminating between healthy subjects and subjects suffering from LBP.
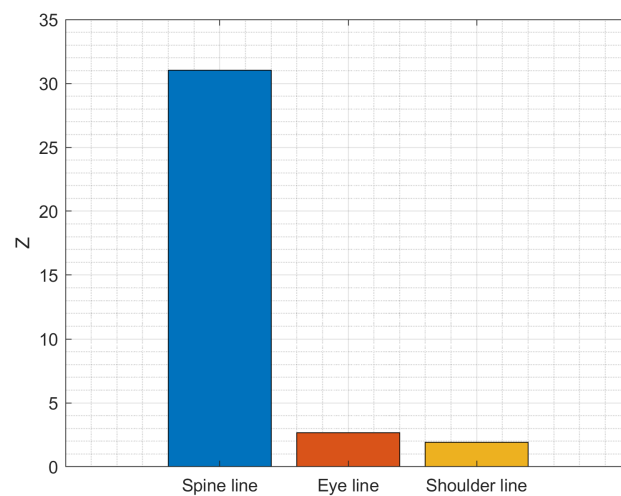

Figure 19. Importance of symptoms in predicting the condition of the subject (Healthy vs. Patient).

Using the horizontal deviation of the spine line versus the vertical deviation of the shoulder line, a good separation can be observed between normal (healthy) subjects and diseased subjects (Figure 20).
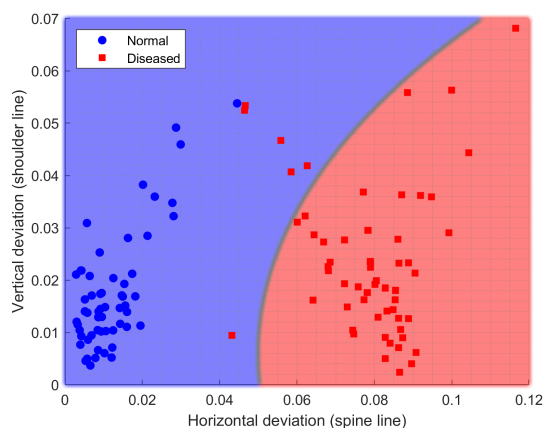
Figure 20. Separation of subjects by correctness of physical exercise performance using spine and shoulder line data.

5.5. Classification Results

5.5.1. Random Forest

In this case, the three statistics of the two angles 'angle 1' and 'angle 3' and the average of 'angle 4' (A4m) are used to build a random forest model. The parameters of the RF model obtained during the search for random parameters in the grid are presented in Table 4.

Table 4. Hyperparameters of the RF classifier and their values used for classification.

Parameters	Meaning
number of iterations	1000
learning grade ('learning_rate')	0.15
number of trees ('n_estimators')	50
maximum depth ('max_depth')	7
minimum number of items ('min_samples_split')	8
minimum number of items per sheet ('min_samples_leaf')	1
max number of features ('max_features')	3

5.5.2. Convolutional Neural Network

A 100-n random grid search of possible hyper parameter values found a CNN model with 14 epochs, a sample size of 48, and a validation sample size of 0.3 (Table 5). The classification results for the eight exercises are summarised in Table 6. The estimated accuracy is 63.4%. The CNN classification process uses full signals recorded during motion (angles are measured every 16.6 ms).

Table 5. Hyperparameter values for the CNN classifier.

Parameters	Meaning
number of iterations	100
number of epochs	14
sample size ('batch_size')	48
validation sample size ('validation_split')	0.3

5.6. Results of Performance Analysis

The performance of the proposed keypoint detection software code was tested using a fragment of video material composed of 500 frames. Three frame sizes were selected, namely: 400×400 pixels, 700×700 pixels, and Full HD 1920×960 pixels. The prototype of the keypoints programme was tested on a computer with a GTX 3080 12GB video card and 16GB system RAM.

Table 6. Classification results of data (angles) recorded during motion using RF and CNN methods.

Exercise	RF Accuracy	CNN Accuracy
Arm bending	100%	66%
Arm raising	100%	66%
Horizontal retraction/retraction of the arm	100%	89.5%
External arm rotation and bending	94%	60.7%
Arm rotations with supination/pronation	100%	92.7%
Lifting and placing a heavy object high	100%	93.8%
Finger-nose	100%	90.3%
Forearm supination/pronation	86%	63.4%

Figure 21 shows the performance of a keypoint recognition programme prototype, where the video frame number is plotted on the horizontal axis and the processing time per frame is expressed in seconds on the vertical axis. The blue curve corresponds to the processing speed of video material with a frame size of 400×400 pixels, the red curve corresponds to 700×700 pixels, and the green curve corresponds to 1280×960 pixels.
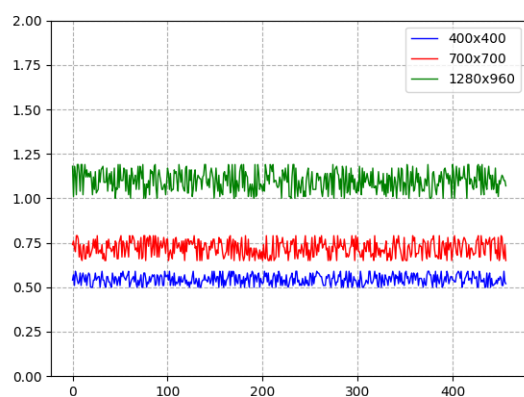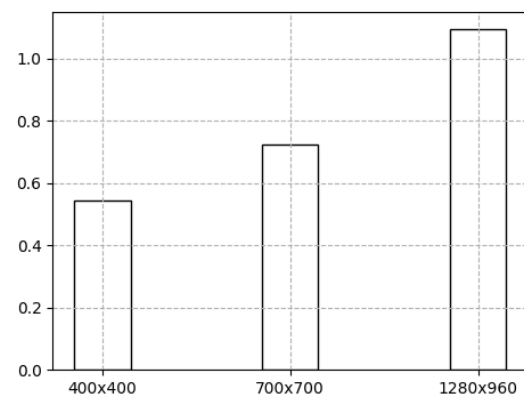

Figure 21. Keypoint recognition performance: frame processing time (s) vs. frame number in video sequence.

Figure 22 shows the average image processing performance values, where the horizontal axis contains the used digital image size values, and the vertical axis shows the delayed performance in seconds. Digital images of 400×400 pixels are likely to be processed. The processing of the prototype of the programme took 0.543 ± 0.028 s to process such images and to obtain keypoints and 0.722 ± 0.043 s to analyse images measuring 700×700 pixels. Digital images with a size of 1280×960 pixels were analysed the longest; it took 1.093 ± 0.058 s to detect keypoints.

Figure 22. Average image processing time (s) vs. frame image size in pixels.

The signal processing speed using an Intel (R) Core (TM) i5-4570 CPU with 8GB RAM, Windows 10 OS computer on average was 23.1 ms. The CNN classifier process uses full motion signals measured every 16.6 ms, and about 6.5 ms was needed to process these data. Measurements are summarised in Figure 23.
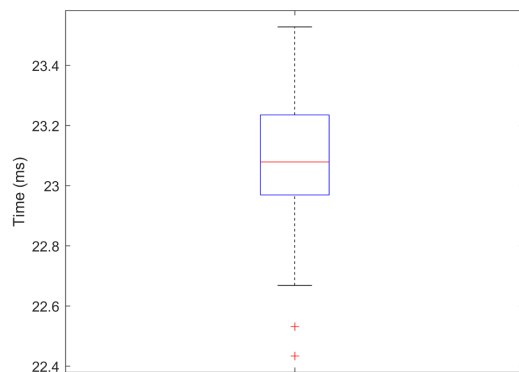

Figure 23. Time value distribution for each frame analysis (signal processing speed).

6. Discussion

In recent years, technology has revolutionized [62] all aspects of medical rehabilitation, from developments in the provision of cutting edge treatments to the actual delivery of the specific interventions. It is becoming more and more popular to use mobile devices such as phones, tablets, cameras, and smartphones in medicine, public health and telerehabilitation [63,64]. The use of information and communication technologies allows providing of rehabilitation services to people remotely in their home environments [18,65]. New technologies are seen as an enabler of change worldwide due to being high-reach and low-cost solutions that are acceptable and easy for the user, especially for individuals who require constant monitoring of progress, consultation, and training [24,27]. Home-based training system for post-stroke patients highlight the usability needs and issues concerned with the appropriateness and acceptability of the equipment in domestic settings, are convenient for users, and are especially important to provide good feedback—the clear interpretation of the screen presentation creates understandability or acceptability of information to patients and physiotherapist [9,12]. It is clear that the success and good results of home-training system are determined by a system that is properly and comfortably adapted to the patient: prototypes based on advanced movement sensors that are friendly for users (patients, their carers, and physiotherapists) [66,67], have simple methods of attachment and use [40,41], and do not cause discomfort when performing prescribed exercises [34,38,68]. Sensor data from sensors are sent to a computer that displays the patient movements, body pose, and postures can identify improper posture or incorrect motions during exercising in real time. Good feedback is very important to adapt telerehabilitation techniques and approaches [11] to post-stroke patients' needs, which would subsequently help to improve the quality of rehabilitation and to reach a more effective continuous treatment in the home environment [58]. Virtual environments are another technological method introduced to healthcare or rehabilitation. These allow users to interact with computer-generated environments in real time and begin with real-world scenes that are then virtualized, thus mimicking real-world environments. This enables rehabilitation professionals to design environments that can be used in areas such as physical rehabilitation, training, education, and other activities promoting patient independence in daily life [40].

BiomacVR system was applied for post-stroke patients independently to perform a home-based physical training. The main contribution of this article is twofold: methodological and practical. The developed methodology is of practical use as the developed BiomacVR system is applied to real life to address the training/exercising protocol in the context of post-stroke rehabilitation. The methodological contribution is based on new technological implementation into system development. The developed BiomacVR system

is privacy-oriented, as it uses no camera and works only using a sensor set (14 keypoints) recorded during the exercise in real time. The usage of the sensors follows established practices to not require much effort and to not cause discomfort for the patient during exercising [69,70]. The model was able to accurately identify key parameters of the posture and body movements and correct or incorrect human motions during training to assess the effectiveness of performing a different exercise and can be considered as a potential tool for implementation of a rehabilitation monitoring system.

Our approach makes it possible for rehabilitation specialists to monitor patients training progress and to evaluate the effectiveness. At the end of a patient rehabilitation program, a physiotherapist can make a further plan for an individual patient that targets his/her specific problems [71,72]. The exercise schedule can be modified with repetitions, sets, and/or additional strengthening/ endurance components [73] to be performed by the patient in his/her home environment. Continuation of training at home ensures that a patient will begin to improve and that their progress can be monitored throughout time [74]. The continued rehabilitation must be tailored from each session by the physiotherapist and must be complemented by the patient's continued determination and trust in his/her shifting program. The understanding and commitment between both the patient and physiotherapist to achieve both short-term and long-term goals indicates a better outcome for the patient once he/she is back to (or potentially better than) his/her baseline [9,10]. The physiotherapist can monitor the patient's progress during time (week and months), constantly analyse the functional changes, identify mistakes during exercising or incorrect movements, and monitor training effectiveness to accordingly increase/decrease physical load during training [75]. Therefore, the feedback is also convenient for post-stroke patients, allowing monitoring of exercises performed correctly/incorrectly, providing a feeling of control during exercising, and increasing self-confidence and self-efficacy in their own progress and recovery [25,33,76]. Thus, home-rehabilitation-based technologies have the advantage of providing flexibility in location, time, and costs, are friendly for users, and can provide remote feedback from therapist. However, they cause a significant challenge for engineers/developers in expanding the applicability of technologies for different health disorders and disability needs [77].

7. Conclusions

Sensors are found to be able to identify key parameters of posture of a person performing a rehabilitation exercise. The resulting average response time is 23 ms, which allows the system to be used in real-time physical exercise monitoring systems for teleoperated rehabilitation. The results confirm that the proposed system can be used to determine the quality of patient movements, monitor movements, monitor the progress of training, control the physical load and the complexity of the exercises, and evaluate the effectiveness of the rehabilitation programme.

The development of new technology systems that allow patients who have had a stroke to independently perform rehabilitation activities at home is an important aspect of today's health research and a significant problem. It is critical that precalculated movement patterns are connected and matched with the motions of patients. A system that is properly and comfortably adapted to the patient determines the success and good results of a home-based physical training and rehabilitation monitoring system. The developed BiomacVR system is based on advanced depth camera sensors that subjects wear when performing prescribed exercises or activities. Sensor data are sent to a computer, which displays the user's motions, postures, improper posture, and training progress.

Our developed system demonstrated very promising results, having an advantage of determining accurate measurements not only for these demonstrated posture movements but also for the complex evaluation of all body movements (upper and lower limbs, incorrect posture, balance, and gait parameters), with complex movement measurements being very accurate and correct. Furthermore, our framework is designed to modify serious movement in the context of rehabilitation using 3D motion tracking and virtual

reality environment to create a personalised and adaptive movement tracking system that allows patients to correctly perform the selected physical actions as prescribed by their physiotherapists.

When analysing performance on real data, i.e., methods that use deep learning networks to directly infer a 3D pose from a real person filmed in real time, the network does not infer the dependencies of the human joints and needs to make predictions, which affects the performance of the solution under real-life conditions. To overcome this problem while keeping the computational cost low during inference, a deep learning regression framework for structured 3D human posture prediction from monocular photos might be included. This method would use a redundant automated encoder to learn a high-dimensional latent pose representation by projecting the locations of body joints. We could train a CNN-based network and map it to the resulting high-dimensional posture mapping after the automated encoder is learnt. This enables us to detect implicit limits on human posture, maintain body statistics, and improve prediction accuracy and processing speed. Finally, we should go over the connections of the autoencoder's decoding layers to the CNN network and fine-tune the entire model. We may try a structure-matching regression to boost the method's speed even more by using the joint dependencies. This method aims to identify both 2D and 3D poses, and therefore balances data optimisation well.

To address detection-based constraints in real image data, we could consider the pose estimation task as a problem of determining the location of a keypoint in discrete 3D space. A specific model could help predict the probability of each joint for each voxel. The resulting volumetric representation would be more sensitive to the 3D nature of this task. To further improve the initial estimation of the data, we could use a coarse- and fine-grained prediction scheme that progressively increases the resolution of the 3D observation volume. This step would address the problem of the large number of data dimensions observed in this study and the increasing number of data dimensions observed in the 3D domain as the number of observations grows, ultimately allowing iterative refinement of the data. It would also allow us to investigate the joint location, obtained as an integration of all locations in the heatmap, weighted by their probabilities. This combines the advantages of both approaches while solving the limitations of data processing.

In the future, it is also worth investigating weakly supervised methods, as they may use unlabelled 2D poses or multiview images, which require less supervision.

The analysis of the large amount of data revealed the following limiting conditions for the accurate detection of keypoints and the trajectory of movements:

- The critical point detection and tracking system can process up to 15 frames per second if the size of the digital image being analysed does not exceed 400×400 pixels. A digital image of this size is sufficient to provide real-time tracking of key points, approximately 70 ms per frame.
- A model based on deep neural networks uses a digital image of 400×400 pixels in which a person must be recorded at full height. The human legs and arms outstretched about 50 cm above the head must be visible. The environment surrounding the person must be contrasting, homogeneous, and as uniform as possible.
- Due to the size of the digital image being processed, the accuracy of the coordinates of the keypoints depends on the resolution of the image. In order to detect various movement disorders, it is recommended to record human movements as close as possible to the video camera without violating the second point of the recommendation.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and was ethically approved by the Institutional Review Board of VilniusTech University Faculty Committee 64-2221.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data created in this study cannot be shared due to privacy requirements.

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

Abbreviations

The following abbreviations are used in this manuscript:

ROM	Range of Motion
ADL	Activities of Daily Living
PSR	Post-stroke Rehabilitation Patients
VR	Virtual Reality
IMU	Inertial Measurement Unit
FM	Fugl–Meyer scale
LBP	Low Back Pain
BLSTM	Bidirectional Long Short-Term Memory Neural Network
RGB	Red Green Blue
RGBD	Red Green Blue + Depth
CNN	Cconvolutional Neural Network
RNN	Recurrent Neural Networks
RMT	Rapid Movement Training
COM	Centre-of-Mass
SVM	Support Vector Machine
BP	Body Part
CPM	Convolutional Pose Machines
HTC Vive	VR Equipment
RF	Random Forest
TN	True Negative
TP	True Positive
FP	False Postive
FN	False Negative
OS	Operating System

References

1. Kumra, S.; Monika, S. A Survey of Acceptability and Use of IoT for Patient Monitoring. In Proceedings of the 2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON), Faridabad, India, 26–27 May 2022; Volume 1, pp. 30–36. [CrossRef]
2. Chen, L.H. *Activities of Daily Living*; WW Norton: New York, NY, USA, 2022.
3. Reis, A.; Lains, J.; Paredes, H.; Filipe, V.; Abrantes, C.; Ferreira, F.; Mendes, R.; Amorim, P.; Barroso, J. Developing a System for Post-Stroke Rehabilitation: An Exergames Approach. In *Universal Access in Human-Computer Interaction. Users and Context Diversity*; Springer International Publishing: Cham, Switzerland, 2016; pp. 403–413. [CrossRef]
4. Goeldner, M.; Herstatt, C.; Canhão, H.; Oliveira, P. User Entrepreneurs for Social Innovation: The Case of Patients and Caregivers as Developers of Tangible Medical Devices. *SSRN Electron. J.* **2019**. [CrossRef]
5. Schröder, J.; van Criekinge, T.; Embrechts, E.; Celis, X.; Schuppen, J.V.; Truijen, S.; Saeys, W. Combining the benefits of telerehabilitation and virtual reality-based balance training: A systematic review on feasibility and effectiveness. *Disabil. Rehabil. Assist. Technol.* **2018**, *14*, 2–11. [CrossRef] [PubMed]
6. Kikuchi, A.; Taniguchi, T.; Nakamoto, K.; Sera, F.; Ohtani, T.; Yamada, T.; Sakata, Y. Feasibility of home-based cardiac rehabilitation using an integrated telerehabilitation platform in elderly patients with heart failure: A pilot study. *J. Cardiol.* **2021**, *78*, 66–71. [CrossRef] [PubMed]

7. Pournajaf, S.; Goffredo, M.; Pellicciari, L.; Piscitelli, D.; Criscuolo, S.; Pera, D.L.; Damiani, C.; Franceschini, M. Effect of balance training using virtual reality-based serious games in individuals with total knee replacement: A randomized controlled trial. *Ann. Phys. Rehabil. Med.* **2022**, *65*, 101609. [CrossRef] [PubMed]

8. Rodrigues, I.M.; Lima, A.G.; dos Santos, A.E.; Santos, A.C.A.; do Nascimento, L.S.; Serra, M.V.C.L.; de Jesus Santos Pereira, T.; Barbosa, F.D.S.; Seixas, V.M.; Monte-Silva, K.; et al. A Single Session of Virtual Reality Improved Tiredness, Shortness of Breath, Anxiety, Depression and Well-Being in Hospitalized Individuals with COVID-19: A Randomized Clinical Trial. *J. Pers. Med.* **2022**, *12*, 829. [CrossRef]

9. Jung, H.; Jeong, J.G.; Cheong, Y.S.; Nam, T.W.; Kim, J.H.; Park, C.H.; Park, E.; Jung, T.D. The Effectiveness of Computer-Assisted Cognitive Rehabilitation and the Degree of Recovery in Patients with Traumatic Brain Injury and Stroke. *J. Clin. Med.* **2021**, *10*, 5728. [CrossRef] [PubMed]

10. Peretti, A.; Amenta, F.; Tayebati, S.K.; Nittari, G.; Mahdi, S.S. Telerehabilitation: Review of the State-of-the-Art and Areas of Application. *JMIR Rehabil. Assist. Technol.* **2017**, *4*, e7. [CrossRef]

11. Uccheddu, F.; Governi, L.; Furferi, R.; Carfagni, M. Home physiotherapy rehabilitation based on RGB-D sensors: A hybrid approach to the joints angular range of motion estimation. *Int. J. Interact. Des. Manuf.* **2021**, *15*, 99–102. [CrossRef]

12. Vanagas, G.; Engelbrecht, R.; Damaševičius, R.; Suomi, R.; Solanas, A. eHealth Solutions for the Integrated Healthcare. *J. Healthc. Eng.* **2018**, *2018*, 3846892. [CrossRef]

13. Okuyama, K.; Kawakami, M.; Tsuchimoto, S.; Ogura, M.; Okada, K.; Mizuno, K.; Ushiba, J.; Liu, M. Depth sensor-based assessment of reachable work space for visualizing and quantifying paretic upper extremity motor function in people with stroke. *Phys. Ther.* **2021**, *100*, 870–879. [CrossRef]

14. Milosevic, B.; Leardini, A.; Farella, E. Kinect and wearable inertial sensors for motor rehabilitation programs at home: State of the art and an experimental comparison. *BioMed. Eng. OnLine* **2020**, *19*, 25. [CrossRef] [PubMed]

15. Kulikajevas, A.; Maskeliunas, R.; Damaševičius, R. Detection of sitting posture using hierarchical image composition and deep learning. *PeerJ Comput. Sci.* **2021**, *7*, 1–20. [CrossRef] [PubMed]

16. Li, M.; Jiang, Z.; Liu, Y.; Chen, S.; Wozniak, M.; Scherer, R.; Damasevicius, R.; Wei, W.; Li, Z.; Li, Z. Sitsen: Passive sitting posture sensing based on wireless devices. *Int. J. Distrib. Sens. Netw.* **2021**, *17*, 15501477211024846. [CrossRef]

17. Palestra, G.; Rebiai, M.; Courtial, E.; Koutsouris, D. Evaluation of a rehabilitation system for the elderly in a day care center. *Information* **2019**, *10*, 3. [CrossRef]

18. Boukhennoufa, I.; Zhai, X.; Utti, V.; Jackson, J.; McDonald-Maier, K.D. Wearable sensors and machine learning in post-stroke rehabilitation assessment: A systematic review. *Biomed. Signal Process. Control.* **2022**, *71*, 103197. [CrossRef]

19. Kristoffersson, A.; Lindén, M. A Systematic Review of Wearable Sensors for Monitoring Physical Activity. *Sensors* **2022**, *22*, 573. [CrossRef]

20. Meegahapola, L.; Gatica-Perez, D. Smartphone Sensing for the Well-Being of Young Adults: A Review. *IEEE Access* **2021**, *9*, 3374–3399. [CrossRef]

21. Straczkiewicz, M.; James, P.; Onnela, J. A systematic review of smartphone-based human activity recognition methods for health research. *npj Digit. Med.* **2021**, *4*, 148. [CrossRef]

22. Clark, R.A.; Vernon, S.; Mentiplay, B.F.; Miller, K.J.; McGinley, J.L.; Pua, Y.H.; Paterson, K.; Bower, K.J. Instrumenting gait assessment using the Kinect in people living with stroke: Reliability and association with balance tests. *J. Neuroeng. Rehabil.* **2015**, *12*, 15. [CrossRef]

23. Scano, A.; Mira, R.M.; Cerveri, P.; Tosatti, L.M.; Sacco, M. Analysis of upper-limb and trunk kinematic variability: Accuracy and reliability of an RGB-D sensor. *Multimodal Technol. Interact.* **2020**, *4*, 14. [CrossRef]

24. Kulikajevas, A.; Maskeliunas, R.; Damasevicius, R.; Scherer, R. Humannet-a two-tiered deep neural network architecture for self-occluding humanoid pose reconstruction. *Sensors* **2021**, *21*, 3945. [CrossRef] [PubMed]

25. Ryselis, K.; Petkus, T.; Blažauskas, T.; Maskeliūnas, R.; Damaševičius, R. Multiple Kinect based system to monitor and analyze key performance indicators of physical training. *Hum.-Centric Comput. Inf. Sci.* **2020**, *10*, 51. [CrossRef]

26. Camalan, S.; Sengul, G.; Misra, S.; Maskeliūnas, R.; Damaševičius, R. Gender detection using 3d anthropometric measurements by kinect. *Metrol. Meas. Syst.* **2018**, *25*, 253–267.

27. Su, M.; Tai, P.; Chen, J.; Hsieh, Y.; Lee, S.; Yeh, Z. A Projection-Based Human Motion Recognition Algorithm Based on Depth Sensors. *IEEE Sens. J.* **2021**, *21*, 16990–16996. [CrossRef]

28. Meng, Z.; Zhang, M.; Guo, C.; Fan, Q.; Zhang, H.; Gao, N.; Zhang, Z. Recent progress in sensing and computing techniques for human activity recognition and motion analysis. *Electronics* **2020**, *9*, 1357. [CrossRef]

29. Sharif, M.I.; Khan, M.A.; Alqahtani, A.; Nazir, M.; Alsubai, S.; Binbusayyis, A.; Damaševičius, R. Deep Learning and Kurtosis-Controlled, Entropy-Based Framework for Human Gait Recognition Using Video Sequences. *Electronics* **2022**, *11*, 334. [CrossRef]

30. Ali, S.F.; Aslam, A.S.; Awan, M.J.; Yasin, A.; Damaševičius, R. Pose estimation of driver's head panning based on interpolation and motion vectors under a boosting framework. *Appl. Sci.* **2021**, *11*, 11600. [CrossRef]

31. Chen, Y.; Liu, C.; Yu, C.; Lee, P.; Kuo, Y. An upper extremity rehabilitation system using efficient vision-based action identification techniques. *Appl. Sci.* **2018**, *8*, 1161. [CrossRef]

32. Lee, S.; Hwang, Y.; Lee, H.; Kim, Y.; Ogrinc, M.; Burdet, E.; Kim, J. Proof-of-concept of a sensor-based evaluation method for better sensitivity of upper-extremity motor function assessment. *Sensors* **2021**, *21*, 5926. [CrossRef]

33. Ayed, I.; Jaume-I-capó, A.; Martínez-Bueso, P.; Mir, A.; Moyà-Alcover, G. Balance measurement using microsoft kinect v2: Towards remote evaluation of patient with the functional reach test. *Appl. Sci.* **2021**, *11*, 6073. [CrossRef]

34. Saini, R.; Kumar, P.; Kaur, B.; Roy, P.P.; Dogra, D.P.; Santosh, K.C. Kinect sensor-based interaction monitoring system using the BLSTM neural network in healthcare. *Int. J. Mach. Learn. Cybern.* **2018**, *10*, 2529–2540. [CrossRef]

35. Capecci, M.; Ceravolo, M.G.; Ferracuti, F.; Iarlori, S.; Monteriu, A.; Romeo, L.; Verdini, F. The KIMORE Dataset: KInematic Assessment of MOvement and Clinical Scores for Remote Monitoring of Physical REhabilitation. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2019**, *27*, 1436–1448. [CrossRef] [PubMed]

36. Wang, L.; Liu, J.; Lan, J. Feature Evaluation of Upper Limb Exercise Rehabilitation Interactive System Based on Kinect. *IEEE Access* **2019**, *7*, 165985–165996. [CrossRef]

37. Sarsfield, J.; Brown, D.; Sherkat, N.; Langensiepen, C.; Lewis, J.; Taheri, M.; McCollin, C.; Barnett, C.; Selwood, L.; Standen, P.; et al. Clinical assessment of depth sensor based pose estimation algorithms for technology supervised rehabilitation applications. *Int. J. Med. Inform.* **2019**, *121*, 30–38. [CrossRef] [PubMed]

38. Xiao, B.; Chen, L.; Zhang, X.; Li, Z.; Liu, X.; Wu, X.; Hou, W. Design of a virtual reality rehabilitation system for upper limbs that inhibits compensatory movement. *Med. Nov. Technol. Devices* **2022**, *13*, 100110. [CrossRef]

39. Bijalwan, V.; Semwal, V.B.; Singh, G.; Mandal, T.K. HDL-PSR: Modelling Spatio-Temporal Features Using Hybrid Deep Learning Approach for Post-Stroke Rehabilitation. *Neural Process. Lett.* **2022**. [CrossRef]

40. Junata, M.; Cheng, K.C.; Man, H.S.; Lai, C.W.; Soo, Y.O.; Tong, R.K. Kinect-based rapid movement training to improve balance recovery for stroke fall prevention: A randomized controlled trial. *J. Neuroeng. Rehabil.* **2021**, *18*, 150. [CrossRef]

41. He, D.; Li, L. A New Kinect-Based Posture Recognition Method in Physical Sports Training Based on Urban Data. *Wirel. Commun. Mob. Comput.* **2020**, *2020*, 8817419 . [CrossRef]

42. Wang, X.; Liu, G.; Feng, Y.; Li, W.; Niu, J.; Gan, Z. Measurement Method of Human Lower Limb Joint Range of Motion through Human-Machine Interaction Based on Machine Vision. *Front. Neurorobot.* **2021**, *15*, 753924. [CrossRef]

43. Mateo, F.; Soria-Olivas, E.; Carrasco, J.J.; Bonanad, S.; Querol, F.; Pérez-Alenda, S. Hemokinect: A microsoft kinect V2 based exergaming software to supervise physical exercise of patients with hemophilia. *Sensors* **2018**, *18*, 2439. [CrossRef]

44. Leightley, D.; McPhee, J.S.; Yap, M.H. Automated Analysis and Quantification of Human Mobility Using a Depth Sensor. *IEEE J. Biomed. Health Inform.* **2017**, *21*, 939–948. [CrossRef]

45. Patalas-maliszewska, J.; Halikowski, D.; Damaševičius, R. An automated recognition of work activity in industrial manufacturing using convolutional neural networks. *Electronics* **2021**, *10*, 2946. [CrossRef]

46. Valdez, S.I.; Gutierrez-Carmona, I.; Keshtkar, S.; Hernandez, E.E. Kinematic and dynamic design and optimization of a parallel rehabilitation robot. *Intell. Serv. Robot.* **2020**, *13*, 365–378. [CrossRef]

47. Georgakis, G.; Li, R.; Karanam, S.; Chen, T.; Košecká, J.; Wu, Z. Hierarchical Kinematic Human Mesh Recovery. In *Computer Vision – ECCV 2020*; Springer International Publishing: Cham, Switzerland, 2020; pp. 768–784. [CrossRef]

48. Smith, S.H.; Bull, A.M. Rapid calculation of bespoke body segment parameters using 3D infra-red scanning. *Med. Eng. Phys.* **2018**, *62*, 36–45. [CrossRef]

49. Ayar, M.; Sabamoniri, S. An ECG-based feature selection and heartbeat classification model using a hybrid heuristic algorithm. *Inform. Med. Unlocked* **2018**, *13*, 167–175. [CrossRef]

50. Shatte, A.B.R.; Hutchinson, D.M.; Teague, S.J. Machine learning in mental health: A scoping review of methods and applications. *Psychol. Med.* **2019**, *49*, 1426–1448. [CrossRef]

51. Luo, J.; Wu, M.; Gopukumar, D.; Zhao, Y. Big Data Application in Biomedical Research and Health Care: A Literature Review. *Biomed. Inform. Insights* **2016**, *8*, BII.S31559. [CrossRef]

52. Zhou, Y.; Qiu, G. Random forest for label ranking. *Expert Syst. Appl.* **2018**, *112*, 99–109. [CrossRef]

53. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.

54. Xu, T.; Takano, W. Graph Stacked Hourglass Networks for 3D Human Pose Estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 16105–16114.

55. Wu, Y.; Ma, S.; Zhang, D.; Huang, W.; Chen, Y. An Improved Mixture Density Network for 3D Human Pose Estimation with Ordinal Ranking. *Sensors* **2022**, *22*, 4987. [CrossRef]

56. Kim, S.T.; Lee, H.J. Lightweight Stacked Hourglass Network for Human Pose Estimation. *Appl. Sci.* **2020**, *10*, 6497. [CrossRef]

57. Newell, A.; Yang, K.; Deng, J. Stacked Hourglass Networks for Human Pose Estimation. In *Computer Vision–ECCV 2016*; Springer International Publishing: Cham, Switzerland, 2016; pp. 483–499. [CrossRef]

58. Norouzi-Gheidari, N.; Hernandez, A.; Archambault, P.S.; Higgins, J.; Poissant, L.; Kairy, D. Feasibility, Safety and Efficacy of a Virtual Reality Exergame System to Supplement Upper Extremity Rehabilitation Post-Stroke: A Pilot Randomized Clinical Trial and Proof of Principle. *Int. J. Environ. Res. Public Health* **2019**, *17*, 113. [CrossRef]

59. Stinear, C.M.; Lang, C.E.; Zeiler, S.; Byblow, W.D. Advances and challenges in stroke rehabilitation. *Lancet Neurol.* **2020**, *19*, 348–360. [CrossRef]

60. Schepers, V.P.M.; Ketelaar, M.; van de Port, I.G.L.; Visser-Meily, J.M.A.; Lindeman, E. Comparing contents of functional outcome measures in stroke rehabilitation using the International Classification of Functioning, Disability and Health. *Disabil. Rehabil.* **2007**, *29*, 221–230. [CrossRef]

61. Petit, A.; Roquelaure, Y. Low back pain, intervertebral disc and occupational diseases. *Int. J. Occup. Saf. Ergon.* **2015**, *21*, 15–19. [CrossRef]
62. Roth, S. The Great Reset. Restratification for lives, livelihoods, and the planet. *Technol. Forecast. Soc. Chang.* **2021**, *166*, 120636. [CrossRef]
63. Sarfo, F.S.; Ulasavets, U.; Opare-Sem, O.K.; Ovbiagele, B. Tele-Rehabilitation after Stroke: An Updated Systematic Review of the Literature. *J. Stroke Cerebrovasc. Dis.* **2018**, *27*, 2306–2318. [CrossRef]
64. Appleby, E.; Gill, S.T.; Hayes, L.K.; Walker, T.L.; Walsh, M.; Kumar, S. Effectiveness of telerehabilitation in the management of adults with stroke: A systematic review. *PLoS ONE* **2019**, *14*, e0225150. [CrossRef]
65. Lee, S.I.; Adans-Dester, C.P.; Grimaldi, M.; Dowling, A.V.; Horak, P.C.; Black-Schaffer, R.M.; Bonato, P.; Gwin, J.T. Enabling Stroke Rehabilitation in Home and Community Settings: A Wearable Sensor-Based Approach for Upper-Limb Motor Training. *IEEE J. Transl. Eng. Health Med.* **2018**, *6*, 1–11. [CrossRef]
66. Moonen, E.J.; Haakma, J.R.; Peri, E.; Pelssers, E.; Mischi, M.; den Toonder, J.M. Wearable sweat sensing for prolonged, semicontinuous, and nonobtrusive health monitoring. *View* **2020**, *1*, 20200077. [CrossRef]
67. Cranen, K.; Drossaert, C.H.C.; Brinkman, E.S.; Braakman-Jansen, A.L.M.; IJzerman, M.J.; Vollenbroek-Hutten, M.M.R. An exploration of chronic pain patients' perceptions of home telerehabilitation services. *Health Expect.* **2011**, *15*, 339–350. [CrossRef] [PubMed]
68. Smith, B.E.; Hendrick, P.; Bateman, M.; Holden, S.; Littlewood, C.; Smith, T.O.; Logan, P. Musculoskeletal pain and exercise—Challenging existing paradigms and introducing new. *Br. J. Sport. Med.* **2018**, *53*, 907–912. [CrossRef]
69. Bavan, L.; Surmacz, K.; Beard, D.; Mellon, S.; Rees, J. Adherence monitoring of rehabilitation exercise with inertial sensors: A clinical validation study. *Gait Posture* **2019**, *70*, 211–217. [CrossRef] [PubMed]
70. Tagliaferri, S.D.; Miller, C.T.; Owen, P.J.; Mitchell, U.H.; Brisby, H.; Fitzgibbon, B.; Masse-Alarie, H.; Oosterwijck, J.V.; Belavy, D.L. Domains of Chronic Low Back Pain and Assessing Treatment Effectiveness: A Clinical Perspective. *Pain Pract.* **2019**, *20*, 211–225. [CrossRef] [PubMed]
71. Melin, J.; Nordin, Å.; Feldthusen, C.; Danielsson, L. Goal-setting in physiotherapy: Exploring a person-centered perspective. *Physiother. Theory Pract.* **2019**, *37*, 863–880. [CrossRef] [PubMed]
72. Ritschl, V.; Stamm, T.A.; Aletaha, D.; Bijlsma, J.W.J.; Böhm, P.; Dragoi, R.G.; Dures, E.; Estévez-López, F.; Gossec, L.; Iagnocco, A.; et al. 2020 EULAR points to consider for the prevention, screening, assessment and management of non-adherence to treatment in people with rheumatic and musculoskeletal diseases for use in clinical practice. *Ann. Rheum. Dis.* **2020**, *80*, 707–713. [CrossRef]
73. Smith, S.S.; Osmotherly, P.G.; Rivett, D.A. What elements of the exercise prescription process should clinicians consider when prescribing exercise for musculoskeletal rehabilitation in a one on one setting? A review of the literature and primer for exercise prescription. *Phys. Ther. Rev.* **2022**, 1–11. [CrossRef]
74. Taylor, J.L.; Holland, D.J.; Spathis, J.G.; Beetham, K.S.; Wisløff, U.; Keating, S.E.; Coombes, J.S. Guidelines for the delivery and monitoring of high intensity interval training in clinical populations. *Prog. Cardiovasc. Dis.* **2019**, *62*, 140–146. [CrossRef]
75. Kiper, P.; Luque-Moreno, C.; Pernice, S.; Maistrello, L.; Agostini, M.; Turolla, A. Functional changes in the lower extremity after non-immersive virtual reality and physiotherapy following stroke. *J. Rehabil. Med.* **2020**, *52*, jrm00122. [CrossRef]
76. Nott, M.; Wiseman, L.; Seymour, T.; Pike, S.; Cuming, T.; Wall, G. Stroke self-management and the role of self-efficacy. *Disabil. Rehabil.* **2019**, *43*, 1410–1419. [CrossRef]
77. Lou, Z.; Wang, L.; Jiang, K.; Wei, Z.; Shen, G. Reviews of wearable healthcare systems: Materials, devices and system integration. *Mater. Sci. Eng. R Rep.* **2020**, *140*, 100523. [CrossRef]